MOLECULAR BIOLOGY

DUX4 double whammy: The transcription factor that causes a rare muscular dystrophy also kills the precursors of the human nose

Kaoru Inoue[1], Hamed Bostan[2], MaKenna R. Browne[1], Owen F. Bevis[1], Carl D. Bortner[3], Steven A. Moore[4], Aaron A. Stence[5], Negin P. Martin[6], Shih-Heng Chen[6], Adam B. Burkholder[2], Jian-Liang Li[2], Natalie D. Shaw[1]*

SMCHD1 **mutations cause congenital arhinia (absent nose) and a muscular dystrophy called FSHD2. In FSHD2, loss of SMCHD1 repressive activity causes expression of double homeobox 4 (DUX4) in muscle tissue, where it is toxic. Studies of arhinia patients suggest a primary defect in nasal placode cells (human nose progenitors). Here, we show that upon SMCHD1 ablation, DUX4 becomes derepressed in H9 human embryonic stem cells (hESCs) as they differentiate toward a placode cell fate, triggering cell death. Arhinia and FSHD2 patient-derived induced pluripotent stem cells (iPSCs) express DUX4 when converted to placode cells and demonstrate variable degrees of cell death, suggesting an environmental disease modifier. HSV-1 may be one such modifier as herpesvirus infection amplifies DUX4 expression in SMCHD1 KO hESC and patient iPSC. These studies suggest that arhinia, like FSHD2, is due to compromised SMCHD1 repressive activity in a cell-specific context and provide evidence for an environmental modifier.**

INTRODUCTION

Congenital arhinia (absent nose) is an extremely rare malformation with fewer than 100 cases reported in the past century (*1*, *2*). We assembled an international cohort of 40 arhinia patients and 55 family members and, through exome sequencing, determined that 86% of patients had an extremely rare, heterozygous missense mutation in the gene *SMCHD1* (*2*). These findings were replicated in a smaller, independent cohort (*1*). The heterogeneity of clinical phenotypes associated with *SMCHD1* mutations—ranging from no craniofacial phenotype, to anosmia (absent sense of smell), to nasal hypoplasia, to complete arhinia—that we observed among unrelated individuals and within multiplex families, however, indicates that there are likely to be genetic or environmental modifiers yet to be discovered. *SMCHD1* encodes a master epigenetic repressor that silences the inactive X-chromosome as well as autosomal transposable elements/repeat sequences, imprinted genes, and clustered genes (e.g., protocadherins and Hox) (*3*–*10*). The precise mechanisms used by SMCHD1 to exact these repressive functions are unknown, but it is believed to stabilize repressive DNA and histone marks and long noncoding RNAs (e.g., *XIST*) and to modulate three-dimensional chromosomal architecture (*3*, *8*, *11*–*14*).

The identification of SMCHD1 as the major genetic driver of arhinia was unexpected because mutations in *SMCHD1* were previously shown to cause facioscapulohumeral muscular dystrophy type 2 (FSHD2), a rare oligogenic disorder with no associated frontonasal dysmorphia or olfactory deficits (*15*). In arhinia, all described mutations have been missense and are restricted to exons 3 to 13 (of 48) (*1*, *2*). *SMCHD1* mutations identified in patients with FSHD2 have been either loss of function (nonsense, frameshift, and deletions) or missense and affect all known functional domains of the protein (*16*, *17*). In FSHD2, loss of SMCHD1 activity is associated with hypomethylation of a subtelomeric macrosatellite repeat array on chromosome 4q35 called D4Z4, which encodes the double homeobox 4 (DUX4) retrogene. DUX4 is normally silenced in somatic cells, but in the presence of hypomethylation, a permissive haplotype (4A, which contains a poly-adenylation signal distal to the array that is required to stabilize the DUX4 transcript), and a shortened or "contracted" array (defined as <28 repeats), DUX4 is ectopically expressed in skeletal muscle (*15*). There, it acts not only as a traditional transcription factor but also as a pioneer factor, recruiting p300 and CREB (adenosine 3′,5′-monophosphate response element–binding protein)–binding protein (CBP), two coactivators with histone acetyltransferase (HAT) activity, to gain access to promoters at H3K27Ac-depleted loci (*18*). DUX4 thereby induces a set of genes involved in oxidative stress, DNA damage, innate immune cell activation, and blockade of non–sense-mediated mRNA decay that together trigger muscle cell apoptosis [reviewed in (*19*)]. Intriguingly, DUX4 is also normally expressed at the four-cell human embryo stage and the ortholog, *Dux*, is expressed in the two-cell mouse embryo (*20*, *21*) where it induces a cleavage-stage transcriptional program, as well as in human embryonic stem cells (hESCs) and in control and FSHD induced pluripotent stem cells (iPSCs) (*22*).

We hypothesized that *SMCHD1* missense mutations in patients with arhinia, like mutations in FSHD2 patients, are loss of function based on several pieces of evidence spanning clinical, translational, and basic studies: (i) Arhinia patients demonstrate the same pattern of D4Z4 hypomethylation in blood cells as do FSHD2 patients (i.e., methylation rate of <25%) (*2*, *23*); (ii) arhinia patients who meet full

[1]Pediatric Neuroendocrinology Group, Clinical Research Branch, National Institute of Environmental Health Sciences (NIEHS), Research Triangle Park, NC, USA. [2]Integrative Bioinformatics, NIEHS, Research Triangle Park, NC, USA. [3]Signal Transduction Laboratory, NIEHS, Research Triangle Park, NC, USA. [4]Department of Pathology, University of Iowa Carver College of Medicine and Senator Paul D. Wellstone Muscular Dystrophy Specialized Research Center, Iowa City, IA, USA. [5]University of Iowa Hospitals and Clinics, Iowa City, IA, USA. [6]Viral Vector Core, NIEHS, Research Triangle Park, NC, USA.
*Corresponding author. Email: natalie.shaw@nih.gov

genetic criteria for FSHD2 (i.e., permissive haplotype and contracted repeat) express DUX4, albeit at low levels, in primary skin fibroblasts after transdifferentiation into a myogenic lineage (so-called "fibro-myoblasts") (*24*), consistent with loss of SMCHD1 repressive activity at the D4Z4 locus; (iii) there is mutational overlap between arhinia and FSHD2 patients at at least three amino acid positions (107, 137, and 242) (*25*); and (iv) *smchd1* morphant zebrafish embryos demonstrate dose-dependent aberrations in facial cartilage patterning and microphthalmia as well as a 45% reduction in the length of the terminal nerve that houses gonadotropin-releasing hormone (GnRH) type 3 neurons (*2*); however, findings have been inconsistent across transgenic animal models (*1*, *2*).

In addition to open questions surrounding the full spectrum of SMCHD1 biological functions, questions remain regarding (i) the role of genetic and/or environmental modifiers in the pathogenesis of arhinia, (ii) the directional effect of *SMCHD1* missense mutations (loss versus gain of function), and (iii) the specific identity of the embryonic cell population that is most vulnerable to SMCHD1 dysfunction and whose destruction ultimately leads to congenital arhinia. Arhinia is frequently accompanied by two comorbidities: ocular defects, such as an-/micro-ophthalmia, coloboma, and cataracts, and hypogonadotropic hypogonadism due to GnRH deficiency. Together, these defects define a clinical triad named Bosma arhinia microphthalmia syndrome (*26*). The unique Bosma phenotype plausibly reflects the failed morphogenesis of a subset of cranial placode cells and/or cranial neural crest cells (NCCs) (*27*–*29*). Studies in the mouse and Xenopus have demonstrated strong Smchd1 expression in these particular cell types (*1*). Placode cells and NCCs are transient embryonic populations that originate from the anterior neural plate and interact in a highly coordinated fashion as they assemble the face and skull. Placodes are focal ectodermal thickenings that give rise to the sensory organs of the head. The lens and olfactory placodes, in particular, form the lens vesicle (future anterior eye) and the olfactory pit, respectively. The olfactory pit induces the formation of the nasal cartilage and bone (*30*, *31*) and gives rise to olfactory neurons, their support

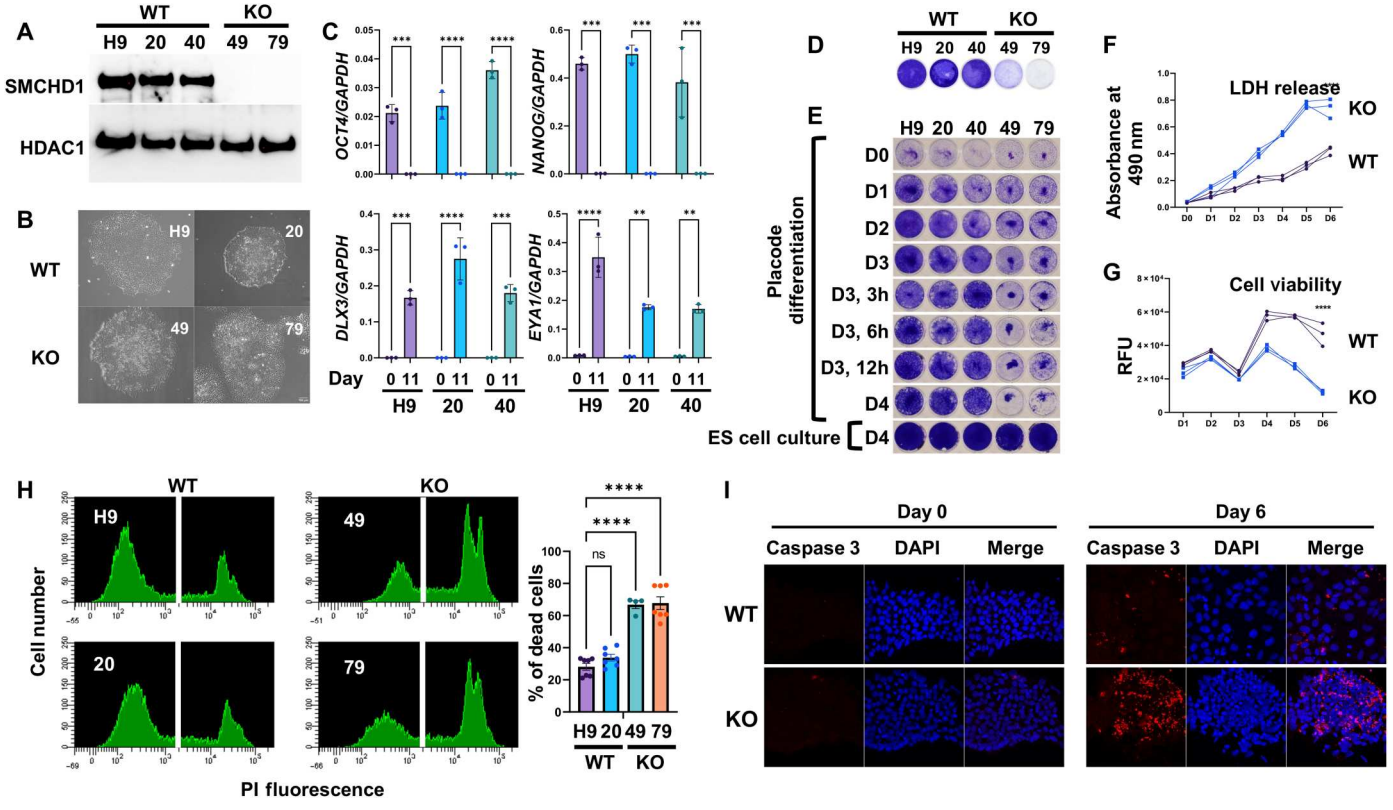

Fig. 1. SMCHD1 KO human ES cells die during differentiation toward a placodal lineage. (**A**) Western blots for the SMCHD1 protein in WT and SMCHD1 KO human ES cells before differentiation. H9 refers to WT human ES cells, clones 20 and 40 are failed CRISPR-Cas9 KO clones (super controls), and clones 49 and 79 are SMCHD1 KO human ES cells. HDAC1 is shown as a loading control. (**B**) Representative microscope images of WT human ES cells (clones H9 and 20) and SMCHD1 KO cells (clones 49 and 79) before differentiation. (**C**) Relative mRNA levels of pluripotency markers (OCT4 and NANOG) and placode cell markers (DLX3 and EYA1) normalized to GAPDH and using three biological replicates. Asterisks indicate statistically significant differences determined using two-way analysis of variance (ANOVA; $^{**}P < 0.01$; $^{***}P < 0.0002$; $^{****}P < 0.0001$). (**D**) Crystal violet–stained cells at day 11 of placode differentiation. (**E**) Crystal violet–stained cells at day zero (D0), D1, D2, D3, D3 and 3 hours, D3 and 6 hours, D3 and 12 hours, and D4 during placode cell differentiation and on D4 when ES cells were maintained in mTeSR medium. (**F** and **G**) Time course experiment for LDH release (F) and cell viability (G) during placode cell differentiation in WT and SMCHD1 KO ES (clone 79) cells using three biological replicates. Asterisks indicate statistically significant differences determined using two-way ANOVA ($^{****}P < 0.0001$). RFU, relative fluorescence units. (**H**) Propidium iodide (PI) fluorescence in two WT and two SMCHD1 KO at day 6 of placode differentiation. Asterisks indicate statistically significant differences determined using one-way ANOVA [not significant (ns) > 0.999; $^{****}P < 0.0001$]. (**I**) Immunofluorescent images showing nuclear staining of DAPI (blue) and activated caspase-3 (red) in WT and SMCHD1 KO cells (clone 79) at baseline (day 0) and during day 6 of placode differentiation.

cells, and GnRH neurons (*32*, *33*). Coincident with placode development, a stream of cranial NCCs migrates into the fronto-nasal region to lay the foundation (mesenchyme) for the craniofacial skeleton.

Here, we use hESC and a cranial placode differentiation protocol to demonstrate that loss of SMCHD1 activity leads to placode cell death. Intriguingly, akin to muscle cell death in patients with FSHD2, placode cell death is mediated by DUX4 toxicity. We show further that arhinia and FSHD2 patient iPSC, but not control iPSC, express DUX4 upon differentiation to placode cells, and that herpesvirus infection amplifies DUX4 induction in SMCHD1 knockout (KO) hESC and arhinia and FSHD2 patient iPSC. Thus, we provide evidence that the combination of SMCHD1 dysfunction and environmental modifiers triggers DUX4-dependent placode cell death and may underlie human congenital arhinia.

RESULTS

SMCHD1 ablation induces cell death in human cranial placode cells but not in cranial NCC

On the basis of Smchd1 expression studies (*1*) and deep phenotyping studies in arhinia patients (*34*), we hypothesized that the primary defect in arhinia involves the anteriormost cranial placode cells. We therefore used CRISPR-Cas9 gene editing (fig. S1) to knock out SMCHD1 in the H9 human embryonic cell line (clones 49 and 79; Fig. 1A and table S1). SMCHD1KO H9 cells did not differ from wild-type (WT) cells in morphology (Fig. 1B) or proliferation (fig. S2). Cranial placode cells were derived from H9 using a placode induction protocol (PIP) by Dincer *et al.* (*35*) that has been validated in several hESC (including H9) and human iPSC lines. All experiments were performed in H9 WT cells, two SMCHD1 KO clones (clones 49 and 79), and two clones in which CRISPR-Cas9 editing had failed to excise SMCHD1 (clones 20 and 40, "super controls") unless noted otherwise. The PIP includes timed exposure to SB-431542, a transforming growth factor–β (TGF-β) inhibitor (days 1 to 11), and Noggin, a bone morphogenetic protein (BMP) inhibitor (for days 1 to 3) (*35*). It initially produces an anterior pre-placode cell population ($PAX6^+/SIX1^+/TFAP2A^+$ cells, which comprise the olfactory, lens, and anterior pituitary placodes), and after PIP culture day 7, the cells spontaneously acquire more posterior placode markers (e.g., PAX3). Thus, this protocol allowed us to specifically interrogate the effect of SMCHD1 ablation in placode precursors and early to mature placode cells.

SMCHD1 WT H9 cells subjected to the PIP recapitulated the original findings of Dincer *et al.* (*35*). That is, pluripotency markers, such as *NANOG* and *OCT4*, were rapidly down-regulated and placode marker genes were up-regulated, as measured on day 11 (Fig. 1C). However, only a very small proportion of KO cells were viable on day 11 (Fig. 1D). A major defect in cell viability unfolded in the KO cells on days 3 and 4 of the PIP (Fig. 1, E to G), suggesting either cell death or proliferation arrest. The increase in lactate dehydrogenase (LDH) release in KO cells suggested the former (Fig. 1F), and cell death was confirmed by propidium iodide (PI) staining (Fig. 1H). While the decrease in cell viability coincided with the time of Noggin withdrawal on day 3, maintaining Noggin for the full 11 days did not rescue this cellular phenotype. Cells were viable at culture day 4 if maintained in hESC medium (mTeSR), implicating the process of cell differentiation toward a placodal lineage in the pathogenesis of the decrease in cell viability (Fig. 1E). Placode cells, by their (neuro-)epithelial nature, tend to organize into stacks or clumps. However, KO cells were unable to adhere to one another and dead cells were observed floating in the medium (fig. S3A). Immunohistochemical staining of the intracellular adhesion markers E-cadherin, vinculin, and actin demonstrated alterations in cell morphology and cell-cell adhesion in KO cells (fig. S3B). Positive cleaved caspace-3 immunofluorescence (*36*) along with chromatin condensation and nuclear fragmentation suggested that the observed changes in placode cell viability are due to placode cell death via apoptosis (Fig. 1I and fig. S4).

To determine whether the change in cell viability was specific to SMCHD1 KO cranial placode cells, we also converted WT and SMCHD1 KO H9 to cranial NCC using an established protocol

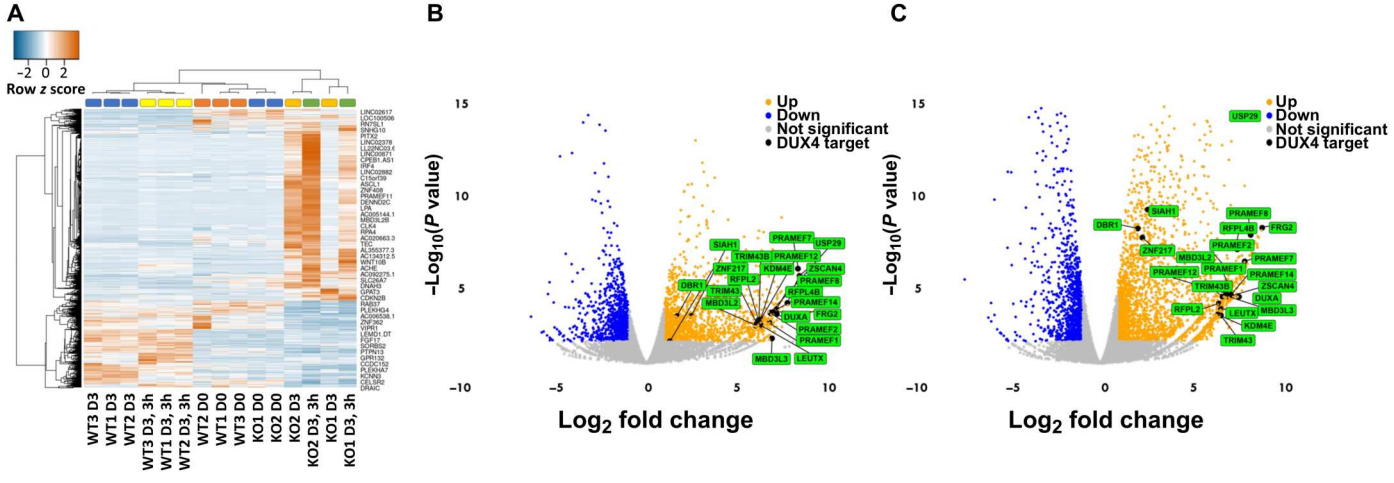

Fig. 2. Expression profile of DEGs. Heatmap of gene expression across all samples (DESeq2 size factor normalized) reordered by hierarchical clustering of samples and DEGs [KO versus WT at days 3 and 3, 3h (day 3 and 3 hours)] using average linkage and Pearson correlation (**A**) and volcano plot of genes differentially expressed between KO and WT at day 3 (**B**) and day 3, 3 hours (**C**) considering ($|\log_2 FC| \geq 1$ and FDR < 0.05) and illustrating up-regulated, down-regulated, nonsignificant, and DUX4 target (Choi '16) genes. RNA-seq was performed in three WT and two SMCHD1 KO clones.

(*37*). With this protocol, hESCs are first differentiated in suspension to neuroectodermal spheres ("rosettes"). Upon rosette attachment, NCCs spontaneously migrate away from the rosette clusters (*37*, *38*). We observed no gross defects in SMCHD1 KO rosette formation, and SMCHD1 KO NCC showed normal proliferation and migration (fig. S5).

SMCHD1 KO cranial placode cells demonstrate a DUX4 transcriptional signature

We initially took an unbiased approach in investigating the underlying etiology of placode cell death by performing a PIP time course RNA sequencing (RNA-seq) experiment. On day 3 and day 3 and 3 hours, the time when we began to observe phenotypic differences in KO versus WT cells (Fig. 1E), there were 1061 and 2676 differentially expressed genes (DEGs) between WT and KO cells, respectively [absolute fold change (FC) ≥ 2, $P_{adj} < 0.05$; Fig. 2 and tables S2 to S5]. Ingenuity Pathway Analyses (using data from day 3 and 3 hours) indicated that the top network functions involved embryonic, organismal, and nervous system development, and intriguingly, DUX4 was among the top activated upstream regulators (activation *z* score 4.586). We therefore chose to pursue DUX4 as our primary candidate for placode cell death. We found that the H9 cell line is particularly well suited for this investigation because we determined that it contains a permissive 4q35 haplotype (4qA161-L) and a contracted (25-repeat) array. We found that mRNA and protein levels of DUX4 were increased on days 2 to 4 of the PIP relative to day 0, and mRNA levels of many of its known downstream targets (e.g., TRIM43 and ZSCAN4) were increased on days 3 and 4 of the PIP relative to day 0 (Fig. 3 and fig. S6). Furthermore, HOMER motif analysis [day 3, using DEGs with false discovery rate (FDR) < 0.5 and $\log_2$FC ≥ 1] identified DUX4 as a potential driver of these transcriptomic changes (fig. S7), and there was significant overlap between genes up-regulated in SMCHD1 KO placode cells on day 3 and day 3 and 3 hours and genes previously shown to be up-regulated after DUX4 overexpression in immortalized human myoblasts ($P < 0.0001$ by hypergeometric test) (*18*). The increase in DUX4 expression was accompanied by D4Z4 hypomethylation in KO clones (35% in KO versus 95% in WT clones on day 2 and 37% in KO versus 96% in WT clones on day 3). However, hypomethylation cannot explain DUX4 induction in placode cells, as D4Z4 hypomethylation without DUX4 or DUX4 target gene expression was present at baseline in the SMCHD1 KO H9 cells (methylation 25% in KO clone 79 versus 100% in WT; table S1). SMCHD1 KO NCC cells were also hypomethylated (methylation 26% in KO clone 79 NCC versus 95% in WT NCC) and did not express DUX4 (fig. S5).

We next performed DUX4 chromatin immunoprecipitation sequencing (ChIP-seq) studies to confirm that DUX4 target genes were directly activated by DUX4 and to potentially identify targets unique to placode cells. We identified 5869 peaks (fig. S8 and tables S6 to S8). We compared these data with that of others who used a more artificial system wherein DUX4 ChIP-seq was performed in human iPSC (~17,000 peaks) (*21*) or in human myoblasts (~14,000 peaks) 24 hours after treatment with doxycycline (*39*) to promote DUX4 overexpression. We observed overlap among the three independent DUX4 ChIP-seq datasets (fig. S9). Enrichment analyses revealed "aplasia/hypoplasia of the central nervous system," "neurodegeneration," and "regulation of expression of SLITs and ROBOs" as terms unique to placode cell DUX4 targets (table S9). SLIT-ROBO signaling is of interest because it modulates N-cadherin protein expression and placodal adhesion (*40*).

Reintroduction of SMCHD1 or repression of DUX4 prevents cell death during placode cell differentiation in SMCHD1 KO cells

To understand the relationship between SMCHD1, DUX4, and decreased viability in placode cells, we first reexpressed SMCHD1 in SMCHD1 KO H9 cells. Reexpression of SMCHD1 completely abolished DUX4 induction (Fig. 4A and fig. S10) at day 4 of differentiation and restored cell viability (Fig. 4B and fig. S10). DUX4 target genes (TRIM43 and ZSCAN4) were also repressed by reexpression of SMCHD1 (Fig. 4, C and D, and fig. S10). Next, we inhibited DUX4 induction to confirm its role in the decrease in cell viability. We tested four short hairpin RNAs (shRNAs) targeting different

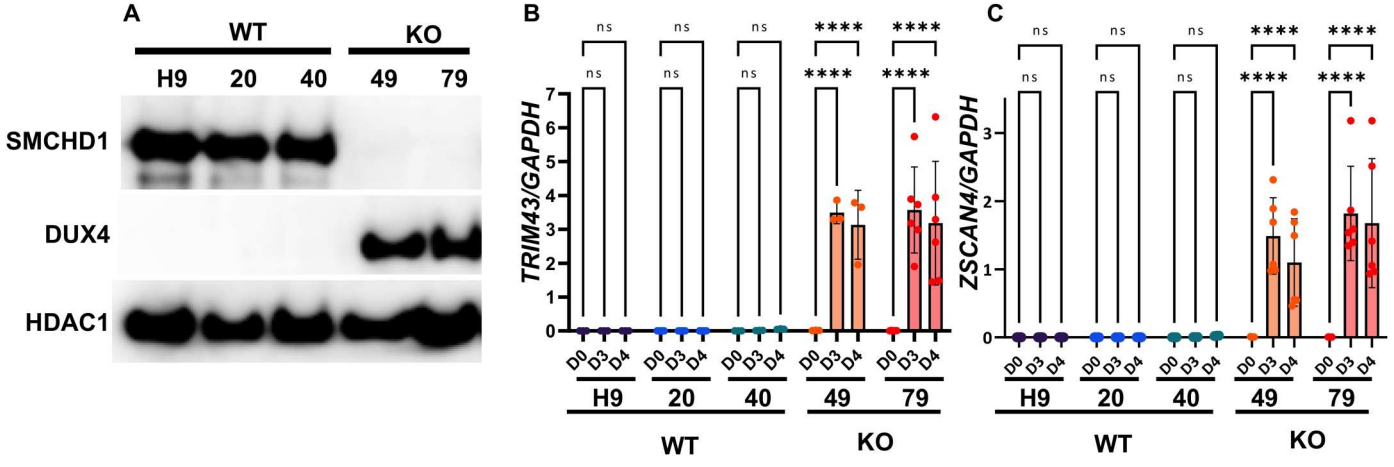

Fig. 3. DUX4 is induced during placode differentiation in SMCHD1 KO cells. (**A**) Western blots for SMCHD1 and DUX4 in WT human ES cells and SMCHD1 KO cells at day 4 of placode differentiation with HDAC1 as a loading control. (**B** and **C**) Relative mRNA levels of DUX4 target genes (TRIM43 and ZSCAN4) at baseline (day 0) and during days 3 and 4 of placode differentiation. Normalized to GAPDH. All experiments included at least three biological replicates. Asterisks indicate statistically significant differences determined using two-way ANOVA (ns > 0.999; ****$P < 0.0001$).

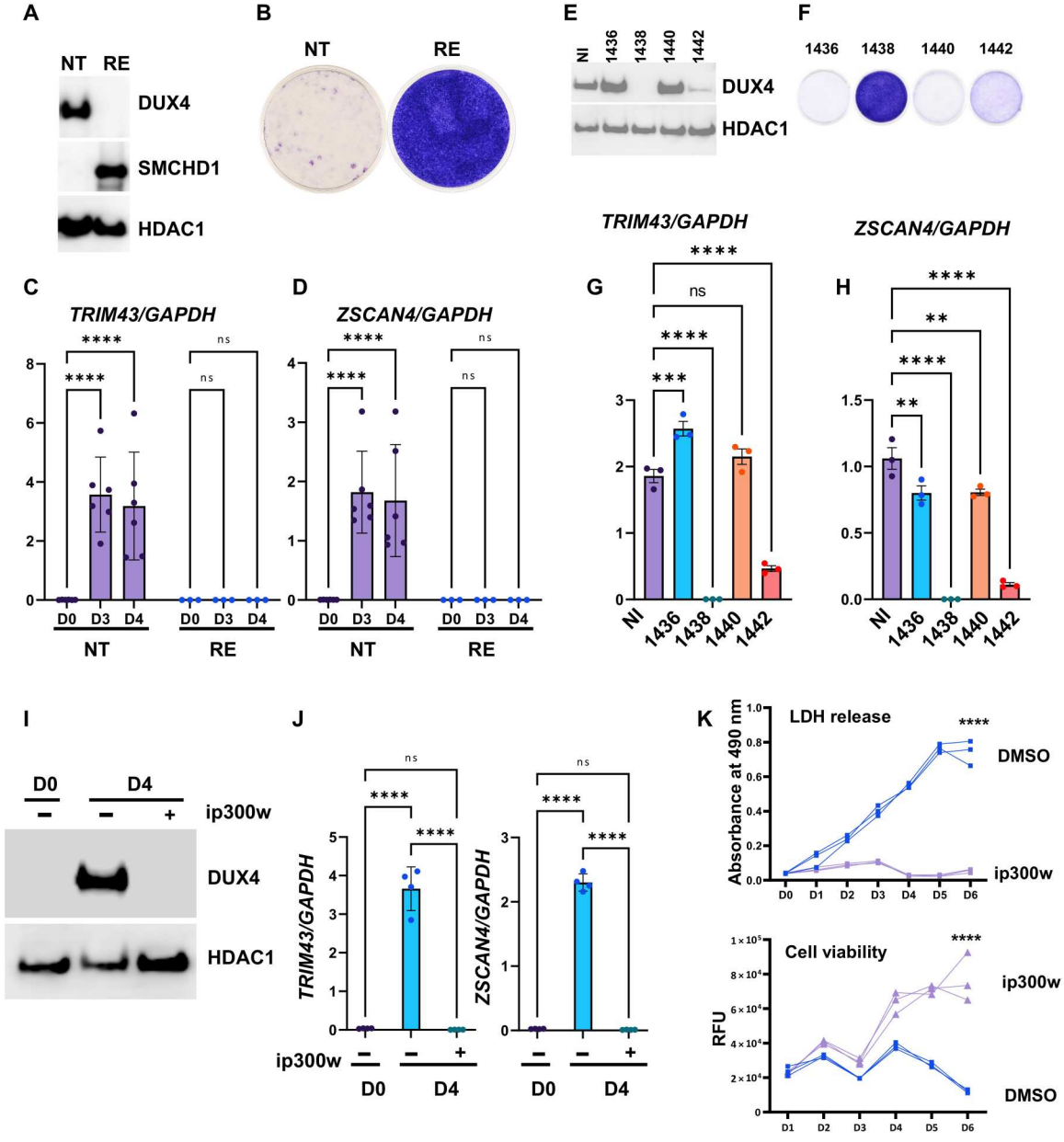

Fig. 4. Reexpression of SMCHD1 in SMCHD1 KO cells and DUX4 knockdown with shRNA or the ip300w inhibitor repress DUX4 induction, averting cell death. (**A**) Western blots for SMCHD1 and DUX4 in SMCHD1 KO (clone 79) cells at day 4 of placode cell differentiation with (RE, reexpression) or without (NT, no treatment) SMCHD1 reexpression. HDAC1 is shown as a loading control. (**B**) Crystal violet–stained cells on day 11 of placode cell differentiation with (RE) and without (NT) reexpression of SMCHD1 in SMCHD1 KO cells. (**C** and **D**) Relative mRNA levels of DUX4 target genes (TRIM43 and ZSCAN4) at baseline (D0) and at days 3 and 4 of placode cell differentiation. Asterisks indicate statistically significant differences determined using two-way ANOVA (ns > 0.999; ****$P < 0.0001$). (**E**) Western blots for DUX4 in SMCHD1 KO cells at day 4 of placode differentiation with four different short hairpin RNAs (shRNAs) targeting. HDAC1 is shown as a loading control. (**F**) Crystal violet–stained cells on day 11 of placode differentiation in the presence of shRNAs targeting DUX4. (**G** and **H**) Relative mRNA levels of DUX4 target genes (TRIM43 and ZSCAN4) at day 4 of placode differentiation in the presence of an shRNA against DUX4. Asterisks indicate statistically significant differences determined using one-way ANOVA (ns > 0.999; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$). NI, no infection. mRNA normalized to GAPDH. $n \geq 3$ biological replicates per experiment. (**I**) Western blots for DUX4 in SMCHD1 KO (clone 79) cells at baseline (day 0) and day 4 of placode differentiation in the presence and absence [dimethyl sulfoxide (DMSO)] of ip300w. HDAC1 is shown as a loading control. (**J**) Relative mRNA levels of DUX4 target genes (TRIM43 and ZSCAN4) at baseline (day 0) and day 4 of placode differentiation. mRNA normalized to GAPDH. Asterisks indicate statistically significant differences determined using one-way ANOVA (ns > 0.9; ****$P < 0.0001$). (**K**) Time course experiment for LDH release and cell viability during placode cell differentiation with ip300w (labeled ip300w) and without ip300w (labeled DMSO). Asterisks indicate statistically significant differences determined using two-way ANOVA (****$P < 0.0001$). $n \geq 3$ biological replicates for all experiments.

sequences located at either the 3′ untranslated region or within the DUX4 gene body (exon 1) to knock down DUX4. One of the shRNAs (sh1438) completely blocked DUX4 induction, whereas DUX4 protein was greatly diminished with a second shRNA (sh1442; Fig. 4E and fig. S10). Both cell viability and DUX4 target gene expression (Fig. 4, F to H, and fig. S10) were correlated with the degree of DUX4 knockdown.

We next tested the ability of a pharmacological DUX4 inhibitor to rescue placode cell viability. DUX4 recruits p300 and CBP, two transcriptional coactivators with HAT activity, to gain access to a set of target genes at H3K27Ac-depleted regions of chromatin (*18*). Bosnakovski *et al.* (*41*) recently discovered ip300w, a selective HAT inhibitor, that blocks DUX4-mediated target gene transcription and cytotoxicity in a number of different cell types including immortalized human and mouse myoblasts harboring a dox-inducible DUX4 transgene and primary FSHD myoblasts and in tissue from an FSHD mouse model. Treatment of SMCHD1 KO placode cells with the HAT inhibitor ip300w likewise diminished DUX4 and DUX4 target gene expression and reversed the DUX4-mediated decrease in cell viability (Fig. 4, I to K, and fig. S10). In summary, these results demonstrate that DUX4 expression occurs specifically because of the loss of SMCHD1 repressive activity in placode cells and that the decrease in placode cell viability is caused by DUX4 expression.

Arhinia and FSHD2 patient-derived cranial placode cells, but not healthy control cells, express DUX4

To extend these observations in SMCHD1 KO placode cells to patient-derived cells, we converted four FSHD2, four arhinia, and two control iPSCs to cranial placode cells using the PIP (table S1). FSHD2 and arhinia patient-derived placode cells demonstrated DUX4 protein and/or DUX4 target gene expression (TRIM43), while control-derived placode cells did not (Fig. 5, A and B, and fig. S11). Among the eight patient cell lines, HT558, HT647, and HT661 showed the greatest amount of DUX4 protein (Fig. 5B and fig. S11) and the highest TRIM43 expression (Fig. 5A and fig. S11), and HT661 showed the greatest decrease in cell viability (Fig. 5A). The amount of DUX4 protein, however, was much lower in patient cells than in placode cells derived from SMCHD1 KO H9 cells (Fig. 5C). TRIM43 gene expression was also approximately 10 times less in patient cells than in SMCHD1 KO placode cells normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH; 5 versus 0.5; Figs. 3 and 5A and fig. S11). Notably, these data in patient-derived placode cells demonstrating variability in DUX4 expression and cell viability are reminiscent of the incomplete penetrance of SMCHD1 mutations where associated phenotypes range from congenital arhinia or severe FSHD2 to the complete absence of craniofacial and/or neuromuscular abnormalities. Together, these data demonstrate that in human placode cells, loss of SMCHD1 activity unleashes the potent DUX4 toxin that triggers apoptotic cell death and this may ultimately lead to complete or partial agenesis of nasal and ocular sensory organs in the developing embryo.

HSV-1 infection of SMCHD1 KO hESC and arhinia and FSHD2 patient iPSC induces DUX4

Having demonstrated that cranial placode cells derived from patient iPSC harboring an *SMCHD1* mutation show variable degrees of upregulation of DUX4 and its downstream targets as well as cell death, we were interested in identifying potential environmental modifiers of DUX4 expression in these cells. It has recently been discovered that herpesviruses induce a germline-specific transcriptional response that is driven by DUX4 in human host cells (*42*). We therefore infected WT and SMCHD1 KO H9 cells with herpes simplex virus 1 (HSV-1) and measured DUX4 expression. DUX4 protein was induced by HSV-1 infection [multiplicity of infection (MOI) = 5 at 18 hours], and the amount was greater in SMCHD1 KO cells than in WT cells (Fig. 6A and fig. S12). TRIM43 mRNA levels were also higher in SMCHD1 KO cells than in WT cells, and levels correlated with DUX4 protein expression (Fig. 6B and fig. S12). We next investigated whether HSV-1 infection might induce DUX4 expression in arhinia and FSHD2 iPSC, as observed in SMCHD1 KO H9 cells. Infection with HSV-1 induced DUX4 in arhinia and FSHD2 patient-derived iPSC >> control iPSC (Fig. 6, C and D, and fig. S12). Thus, intrauterine HSV-1 infection may be an environmental agent that augments DUX4 expression in patient cells.

DISCUSSION

We (*2*) and others (*1*) discovered that missense mutations in *SMCHD1* are the primary genetic driver of congenital arhinia and related nasal phenotypes. This finding was a complete surprise because mutations in *SMCHD1*, some identical to those harbored by individuals with arhinia, also cause a late-onset neuromuscular disorder, FSHD2. In arhinia patients, mutations are entirely missense and are restricted to exons 3 to 13, which encode the extended adenosine triphosphatase (ATPase) domain. In FSHD2 patients, loss-of-function variants span the entire *SMCHD1* coding region, yet as in arhinia, missense variants are significantly enriched in the exons encoding the extended ATPase domain (*23*). To date, despite conducting extensive phenotyping studies in individuals with arhinia (*24*, *34*) and in individuals with FSHD2 (*25*), we have not identified any individuals with pathognomonic features of both conditions, suggesting the presence of distinct genetic and/or environmental modifiers that act during early development. In this work, we show that in arhinia, *SMCHD1* missense mutations confer an overall loss of protein function in cranial placode cells, which leads to derepression of DUX4 and placode cell death. We also implicate herpesviruses as a potential environmental modifier that may exacerbate DUX4 toxicity in the cranial placode cells of the developing fetus.

Cranial placode cells are an evolutionarily ancient cell population that give rise to the sensory organs of the head (*43*). The morphogenesis of the olfactory placode, which gives rise to the nasal skeleton, olfactory neurons, and GnRH neurons, is universally impaired in patients with arhinia, whereas that of the lens placode (anterior eye) is variably affected. The olfactory and lens placode initially develop in close proximity in the anteriormost portion of the neural folds, and both show strong Smchd1 expression in animal models (*1*). While placode cell death has been implicated in other congenital malformations due to genetic (e.g., otic placode in Branchio-Oto-Renal (BOR) syndrome due to mutations in *EYA1*) (*44*) and/or teratogenic mechanisms (e.g., hypoxia, alcohol, and hyperthermia), to our knowledge, arhinia represents the first example of a craniofacial malformation caused by epigenetic derepression of a toxic protein.

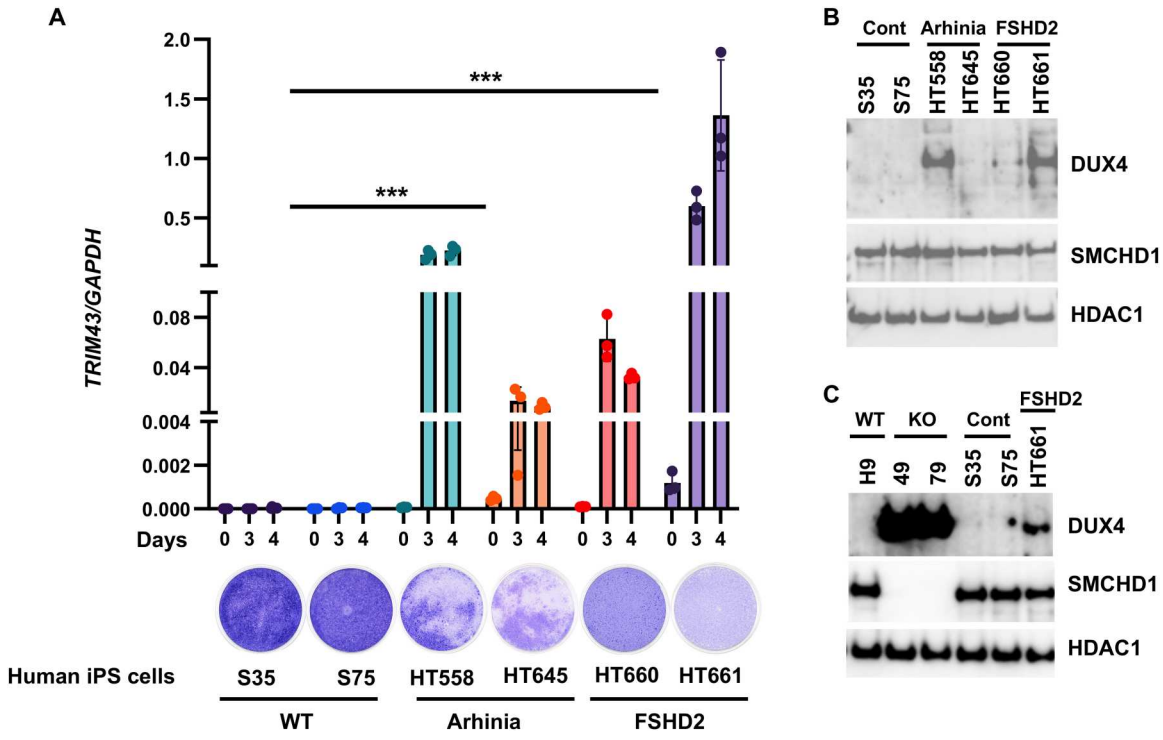

Fig. 5. DUX4 is induced in arhinia and FSHD2 patient-derived cells upon placode cell differentiation. (**A**) Top: Relative mRNA levels of DUX4-inducible gene TRIM43 at baseline (D0) and at days 3 and 4 of placode cell differentiation. Asterisks indicate statistically significant differences between groups determined using two-way repeated-measures ANOVA (***$P < 0.001$). mRNA normalized to GAPDH. $n \geq 3$ biological replicates per experiment. Bottom: Crystal violet–stained cells on day 11 of placode cell differentiation for control, arhinia, and FSHD2 patient-derived cells. (**B** and **C**) Western blots for SMCHD1 and DUX4 in control, arhinia, and FSHD2 patient-derived placode cells, WT human ES cells (H9), and SMCHD1 KO ES cells (clones 49 and 79) at day 4 of differentiation. HDAC1 is shown as a loading control. See table S1 for phenotypic and genotypic data for controls and patients.

Cranial placodes closely interact with another cell population—cranial NCCs—during development of the head (*43*). The ontogeny of these two cell types is highly interdependent due to a "chase and run"-type behavior of attractive and repulsive signaling (*45*). For example, as the olfactory and lens placodes invaginate, they repel nearby NCC, in effect dictating the path of the NCC migratory stream into the fronto-nasal region. Conversely, sensory neurons that develop from the placodes follow the "channels" plowed by NCC through the head mesenchyme. We observed DUX4-mediated cell death that was specific to SMCHD1 KO cranial placode cells; arhinia and FSHD2 placode cells also demonstrated increased expression of DUX4 and its downstream targets. In contrast, and despite D4Z4 hypomethylation in SMCHD1 KO NCC, DUX4 was not induced in SMCHD1 KO NCC. Notably, a recent study of NCC derived from one arhinia patient iPSC and one control iPSC also failed to demonstrate DUX4 induction in arhinia NCC but suggested that there may be a very mild defect in arhinia NCC migration related to diminished AKT signaling (*46*). Thus, while the primary defect in arhinia appears to be placodal, placode cell death may impair the proliferation and/or migration of nearby NCC such that a combined defect prevents the development of the human nose.

Despite carrying the same deleterious *SMCHD1* mutation, not all carriers in multiplex families manifest arhinia; the phenotypic spectrum also includes a normal nose, hemi-arhinia, uni- or bilateral nasal hypoplasia, hyposmia, and anosmia. This variable expressivity points to the presence of genetic and/or environmental modifiers. Our model system, for example, suggests that genetic modifiers may contribute to DUX4 expression and/or toxicity in placode cells or may control an individual's overall complement of placode cells in utero. Recent work in mouse ESC, FSHD2 iPSC, and FSHD2 myoblasts suggests that endogenous DNA damage stimulates p53 activity, which drives DUX4 expression in these cells (*47*); thus, interindividual differences in DNA damage or p53 responsiveness may dictate DUX4 expression and ultimately determine the fate of cranial placode cells in *SMCHD1* mutation carriers. We also suggest that herpesvirus is a relevant environmental modifier as we find that HSV-1 augments DUX4 production in SMCHD1 KO hESC and arhinia and FSHD2 iPSC more so than in WT hESC and control iPSC, respectively. HSV-1 may also induce DUX4 via a DNA damage response and p53 activity (*48*), although recent studies suggest a direct role for the viral immediate early proteins ICP0 and ICP4 in DUX4 up-regulation (*49*). HSV-1 is a common sexually transmitted disease among women of reproductive age, and it can, in rare circumstances, be transmitted to the fetus during pregnancy with severe consequences (e.g., microcephaly and fetal demise) (*50*). It is therefore conceivable that a low-grade infection in the first trimester (when the face is forming) in the context of a genetic background of SMCHD1 dysfunction could specifically derail olfactory and lens placodogenesis.

While we have made major strides in understanding the pathogenesis of arhinia by identifying the relevant cell type (cranial

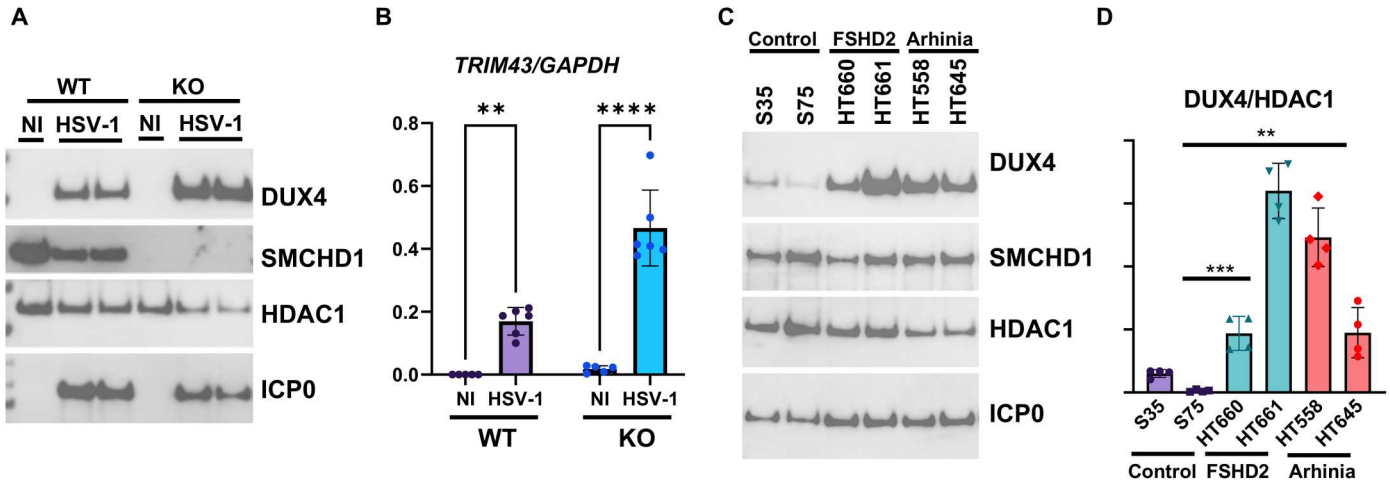

Fig. 6. HSV-1 infection induces greater DUX4 expression in SMCHD1 KO ES cells than in WT ES cells and in arhinia and FSHD2 patient-derived iPSC than in control iPSC. (**A**) Western blots for SMCHD1 and DUX4 in WT and SMCHD1 KO ES (clone 79) cells with or without HSV-1 infection (MOI 5, 18 hours). HDAC1 is shown as a loading control. Infected cell protein 0 (ICP0) is an HSV-1 protein confirming viral infection. (**B**) Relative mRNA levels of the DUX4 target gene TRIM43 with or without HSV-1 infection normalized to GAPDH. Asterisks indicate statistically significant differences determined using two-way ANOVA (**$P < 0.01$; ****$P < 0.0001$). (**C**) Western blots for SMCHD1 and DUX4 in control and patient iPSCs infected with HSV-1. (**D**) Graphical representation of DUX4 protein amount in control and patient iPSC after HSV-1 infection. The DUX4 protein level was quantified by measuring the intensity of protein bands and expressed relative to HDAC1. Asterisks indicate statistically significant differences determined using one-way ANOVA (***$P < 0.001$, **$P < 0.005$). $n \geq 3$ biological replicates per experiment.

placode) and mechanism of cell death (DUX4), our in vitro model is limited in that both arhinia and FSHD2 patient-derived iPSC demonstrate variable degrees of cell death upon placode cell differentiation; hence, our data do not answer the question of why arhinia patients, but not FSHD2 patients, demonstrate craniofacial abnormalities. We suspect that there are genetic modifiers that have important effects in utero, perhaps in fine-tuning the temporospatial expression of DUX4 or in determining the total number of placode cells that form, but these effects, as well as those of unknown environmental modifiers, are not captured by this model system. Intriguingly, a permissive 4q35 haplotype does not appear to be required for DUX4 expression in placode cells, as demonstrated with the arhinia cell line HT647.

In summary, we propose that in patients with arhinia and related nasal phenotypes (e.g., anosmia and nasal hypoplasia), nasal morphogenesis is completely or partially arrested when *SMCHD1* missense mutations unleash DUX4 toxicity in cranial placode cells, leading to cell death. Additional genetic variants and/or environmental factors, such as viral infections, appear to modulate this process, resulting in incomplete penetrance and variable expressivity among mutation carriers. Further studies are necessary to determine the unique repertoire of transcriptional activators and repressors that control DUX4 expression in human placode cells.

MATERIALS AND METHODS

Study design

Statistical analyses

Statistical details, including the number of biological replicates, methods used, and *P* values, are reported in each figure legend, in the text, or in the supplementary tables. All bar graphs show individual data points as well as mean ± 1 SD.

Sample size—A sample size of two to three per group (three WT clones, two KO clones, four FSHD2 iPSCs, and four arhinia iPSCs) was chosen to achieve appropriate statistical power based on previous experience working with these cell types.

Data exclusions—None.

Replication—There were three biological replicates for the control group (clones H9, 20, and 40) and two for the SMCHD1 KO ESC (clones 49 and 79).

Randomization—N/A.

Blinding—N/A.

Enrollment of research subjects

Written informed consent was obtained from all participants. Ethical approval was obtained from the institutional review board at the National Institute of Child Health and Development/National Institutes of Health (NIH; protocols 2012-CH-0050 and 2012-CH-0049). Under this protocol, dermal fibroblasts were collected by skin punch biopsy from subjects with arhinia. Fibroblasts from patients with FSHD2 were gifts from R. Tawil (Richard Fields Center for FSHD Research, University of Rochester Medical School, Rochester, NY) and C. Emerson Jr. (University of Massachusetts Medical School Wellstone Center for FSHD, Worcester, MA). iPSC lines derived from primary fibroblasts from two deidentified female controls were a gift from the laboratory of T. Archer [National Institute of Environmental Health Sciences (NIEHS)/NIH]. Sample information is listed in table S1.

Derivation and characterization of iPSC

Arhinia and FSHD2 iPSC were reprogrammed from primary fibroblasts or myoblasts using Sendai virus, as previously described (*51*). The presence of pluripotency markers (NANOG, TRA-1-60) was confirmed by fluorescence-activated cell sorting, and cells were confirmed to be vector-free using anti-SeV antibodies. Control iPSCs were reprogrammed from primary fibroblasts via lentiviral transduction, as previously described (*52*, *53*). The control iPSC demonstrated the expected expression pattern of genes associated with pluripotency, ectoderm, mesoderm, and endoderm lineage commitment (*52*, *54*).

Human pluripotent cell culture

Human iPSCs or H9 hESCs (WiCell, WA09) were maintained on Matrigel (Corning, 356235)–coated plates using mTeSR1 (STEMCELL Technologies, 85850) and passaged using ReLeSR (STEMCELL Technologies, 05872). After passaging, rock inhibitor (Y-27632) was added to the medium and removed 24 hours after plating. mTeSR1 was changed every day, and cells were passaged at about 60% confluency.

Genotyping of SMCHD1 KO cell lines

For genotyping of SMCHD1 KO cell lines (49 and 79), genomic DNA was prepared from each cell line using the DNeasy Blood and Tissue Kit (Qiagen, 69504) following the instructions of the manufacturer. Regular polymerase chain reaction (PCR) was performed using primer sets 735 and 736 with 100 ng of genomic DNA as a template. PCR product was run onto 1% agarose gel, and 497 base pairs of PCR product was extracted using a QIAquick Gel Extraction kit and cloned into pCR 2.1-TOPO TA vector using the TOPO TA Cloning Kit (Thermo Fisher Scientific, K450001). Picked colonies were cultured in LB with ampicillin, and plasmid was prepared using the QIAprep Spin Miniprep Kit (Qiagen, 28704) and sequenced with primer 735. The clones containing proper region of genomic DNA were aligned and compared with WT sequence.

Derivation of cranial NCC and cranial placode cells

Cranial NCC and placode cells were derived from H9 cells as previously described (*35*, *37*). NCCs migrating away from spheres were detached from dishes by Accutase and replated in six-well plates at 1×10^6 cells per well. When cells are confluent in a well (generally, 24 to 48 after plating), a wound is made with a pipette tip and migration was observed by taking pictures every 2 to 3 hours.

Scratch wound assay

NCCs were harvested by Accutase and passed through a cell strainer (mesh size, 40 μm) to remove spheres, and 1×10^5 cells were plated in one well of 96-well plate. Twelve hours after plating, wound was made using Incucyte 96-Well Woundmaker Tool (Sartorius, catalog no. 4563), the well was washed once with phosphate-buffered saline (PBS), and 100 μl of medium was added into each well. Cell migration was monitored by IncucyteS3 (Sartorius), and data were analyzed using Incucyte Scratch Wound Analysis Software Module (Sartorius, catalog no. 9600-0012).

Analysis of cellular proliferation by flow cytometry

Human ES cells (WT and SMCHD1KO) were plated with 2×10^6 density in 10-cm dish, and cells are collected 3 days after plating, fixed with 80% EtOH, and kept at −20°C until analysis. Cell proliferation was measured using anti-Ki67–PE (phycoerythrin) (BD Biosciences) according to the manufacturer's instructions. Ten thousand cells were analyzed using a BD Fortessa flow cytometer (San Jose, CA) equipped with FACSDiva software. Cells were excited with a 561-nm laser, and PE fluorescence was detected at 585 nm. Changes in cellular proliferation were determined by analyzing single cells on an anti-Ki67–PE histogram.

Flow cytometry for PI staining to detect necrotic cells

Cells were plated with 3×10^5 density in a 6-cm dish and differentiated into placode cells 24 hours after plating. At day 6 of differentiation, both cells in medium (detouched) and attached cells harvested by Accutase were collected in one tube by centrifugation (500*g* × 5 min). Plasma membrane integrity was measured by adding PI (R&D Systems) to live cells at a final concentration of 5 μg/ml immediately before flow cytometric analysis. Ten thousand cells were analyzed using a BD Fortessa flow cytometer (San Jose, CA) equipped with FACSDiva software. Cells were excited with a 561-nm laser, and PI fluorescence was detected at 585 nm. A gate was drawn on a PI histogram for the control sample to determine the percent of viable and dead experimental cells.

Cytotoxicity assay

hESCs were plated on white 96-well plates coated with Matrigel (1×10^4 per well), and 24 hours after plating, placode cell differentiation was started. Samples of the cell culture medium were taken every day up to 6 days of differentiation and frozen for assay. LDH release was measured using CytoTox96 Non-Radio Cytotoxicity Assay (Promega, G1781) following the product's instructions. Cell viability was measured each day using the CellTiter-Fluor Cell Viability Assay, as described above.

Crystal violet staining to determine cell viability

hESCs (both WT and SMCHD1 KO) were differentiated into placode cells, and at each time point, medium was removed from the dishes or wells, and crystal violet solution (2% in 10% ethanol) was added to cover the well. The plates or dishes sat at room temperature for at least 1 hour. Plates or dishes were washed four times in a stream of tap water. After washing, plates or dishes were air-dried without their lids overnight and pictures were taken from the bottom angle.

PCR and primer extension

For regular PCR and primer extensions, PrimeSTAR Max DNA Polymerase (Takara, R045B) and PfuTurbo DNA Polymerase (Agilent, 600250) were used following the manufacturer's instructions, respectively.

Plasmid construction

pCDH-EF1α-MCS-T2A-SMCHD1 and pCDH-EF1α-MCS-T2A-Hygro were made from pCDH-EF1α-MCS-T2A-Puro (System Bioscience, CD527A-1). To remove the puromycin-resistant gene from CD527A-1, primer extension was performed with primer set 971/972 and transformed after Dpn I treatment. Xho I site was inserted by performing primer extension using primer set 973/974. Hygromycin-resistant gene was amplified from pCDH-UbC-MCS-EF1α-Hygro (System Bioscience, CD615B-1) using primer set 975/976 and inserted into Xho I–digested pCDH-EF1α-MCS-T2A-Hygro by SLIC (*55*). Mutations in *SMCHD1* around guide RNA (gRNA) without changing the amino acid sequence to avoid Cas9 digestion were introduced by primer extension with two primers (828/829) using pcDNA3.1HASMCHD1 vector as a template. gRNA-mutated SMCHD1 was amplified with primer set 965/966 and inserted into pCDH-EF1α-MCS-T2A-Hygro–digested with Xba I.

pLKO.1 hygro-DUX4 shRNAs. Sets of oligos with 1436/1437, 1438/1439, 1440/1441, and 1442/1443 were annealed and ligated into pLKO.1 hygro digested with Age I and Eco RI by T4 ligase and transformed.

pNL2.1-DUX4RE-tk. Two oligos were ordered as an Ultramer Duplex from IDT and inserted into pGL3B-tk (*56*) digested with Xho I. DUX4-responsive element conjugated with tk promoter was amplified using primer set 1454/1459 and inserted into pNL2.1 digested with Xho I. The plasmid was sequenced, and the one that had the insertion with the correct direction was selected. The final product has 21 repeats of the DUX4-responsive element (TAATCTAATCA) upstream of the tk promoter. All construct sequences were confirmed by Sanger sequencing (Genewiz).

Lentivirus production

All lentiviruses, pLentiCRISPR v2-SMCHD1guideRNA3, pCDH-EF1α-MCS-T2A-Hygro-SMCHD1, and pLKO.1 hygro with DUX4 shRNA, were packaged in human embryonic kidney (HEK) 293T/17 cells according to protocol (*57*). Briefly, HEK293T cells were transiently transfected with pMD2G, psPAX2, and transfer vector containing the desired gene using Lipofectamine 2000 (Invitrogen). Culture medium was collected 48 hours after transfection and cleared from debris before storage. All titers were determined by performing quantitative or digital droplet PCR to measure the number of lentiviral particles that integrated into the host genome.

Lentivirus infection

Human ES cells (H9 cells or SMCHD1 KO H9 cells) were plated at a density of either 1×10^5 or 5×10^5 in a six-well plate. Cells were infected with 5 or 10 MOIs for each cell density, respectively. Medium was changed every day, and cells were split with 10× serial dilution in 10-cm dishes. Cells were treated with puromycin (0.25 to 0.5 μg/ml) or hygromycin (10 to 50 μg/ml) for drug selection. For SMCHD1 KO and SMCHD1 reexpression cells, surviving colonies were picked and expanded individually. Each clone was examined for genomic DNA sequences after preparing genomic DNA with DNAzol (Invitrogen, 10503027) or DNeasy (Qiagen, 69504) and performing PCR using primer set 735/736 and/or Western blotting with SMCHD1 antibody. For DUX4 knockdown cells, all surviving cells following hygromycin treatment were mixed and used for further experiments.

Anti-SMCHD1 antibody production

We used bacterially expressed and purified SMCHD1 24-580AA (*58*) for rabbit injections using GenScript. GenScript affinity purified the antibody using an antigen column.

Human herpes virus (HSV-1) production and infection

Human ES cells or iPSCs were plated 5×10^5 per well in a six-well plate and infected with HSV-1 at an MOI of 5 24 hours after plating. Cells were collected 18 hours after infection, and RNA and protein samples were prepared.

Western blotting

Cell pellets were suspended with Buffer A [10 mM Hepes (pH 7.6), 10 mM KCl, and 1.5 mM $MgCl_2$ containing 0.3% NP-40] supplemented with cOmplete Mini, EDTA-free (Roche, 11836170001) protease inhibitor and kept on ice for 10 min. Suspension was centrifuged at 3000*g* for 5 min, and supernatant was stored as cytoplasmic fraction. The pellet was washed with Buffer A again, suspended with cell lysis buffer [10 mM Hepes (pH 7.6), 0.5 mM EDTA, 100 mM NaCl, 1% Triton X-100, and 10% glycerol supplemented with cOmplete Mini], and sonicated for 20 s with a Q500 Qsonica sonicator using a microtip probe and a 30% amplitude setting. Protein concentration was quantified with Pierce BCA protein assay kit (Pierce, 23225), and protein (5 to 10 μg) was loaded onto SDS–polyacrylamide gel electrophoresis gel (NuPAGE 4 to 12%, bis-tris, 1.0 mM, Mini Protein Gel, 10- or 15-well; Invitrogen, NP0321BOX or NP0321BOX). After transferring proteins onto polyvinylidene difluoride membranes (Immobilon-P, Millipore, IPVH85R), membranes were blocked with TBST (tris-buffered saline with 0.05% Tween 20) containing 5% skim milk and incubated with primary antibodies overnight at 4°C. Membranes were then incubated with horseradish peroxidase–labeled secondary antibodies, and signals were detected by WesternBright Sirius or WesternBright ECL (Advansta, K-12043 or K-12045) using C-DiGit (LI-COR) or ChemiDoc Imaging system (Bio-Rad).

Quantitative PCR with reverse transcription

RNA was isolated using TRIzol extraction. All RNA samples were treated with deoxyribonuclease (DNase) on RNeasy columns and eluted. Reverse transcription was performed with High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, 4368814), and quantitative PCR (qPCR; see table S7 for qPCR primers and probes) was performed with Universal PCR Master Mix (Thermo Fisher Scientific, 4304437) using a Bio-Rad qPCR CFX. Experiments were performed in biological replicates as indicated in figure legends. GAPDH or 18*S* ribosomal RNA (rRNA) was used to normalize data.

Immunofluorescence of cultured cells

Immunofluorescence was performed according to the protocol described by Kuragano *et al.* (*59*). Briefly, cells were cultured on Matrigel-coated cover glass and fixed in 3.7% formaldehyde in PBS at room temperature for 10 min and washed with PBS. Cells were permeabilized using 0.1% Triton X-100 in PBS at room temperature for 10 min and blocked using 3% bovine serum albumin (BSA) in PBS at room temperature for 30 min. Primary antibodies (see table S7 for dilution range) were diluted in PBS with 3% BSA and incubated at room temperature for 1 hour or overnight at 4°C, washed three times in PBS and incubated with secondary antibodies (table S7) diluted in PBS with 3% BSA, and incubated for 1 hour at room temperature. Samples were washed three times with PBS, briefly rinsed in water, and mounted with ProLong Gold Anti-fade reagent and DAPI (4′,6-diamidino-2-phenylindole) mounting medium (Fisher Scientific, P-36935).

Definition of biological replicate

Biological replicates are samples analyzed from distinct samples and, when possible, are different independent clones (for example, KO clones isolated separately to control for clonal-derivation technical variation).

Mycoplasma testing

Cells were routinely tested for mycoplasma using the method described previously (*60*), and the results were negative.

FSHD diagnostic testing

All testing was performed at the University of Iowa Molecular Pathology Laboratory using standard methods for the clinical diagnosis of FSHD (*61*). (i) Determination of haplotype, 4q35, and 10q26 D4Z4 repeat sizes (*62*): Cultured stem cells were extracted using the Specimen Preparation Fresh Cells DNA Isolation Protocol (Bionano Genomics, San Diego, CA; catalog no. 80030). Genomic DNA was quantitated by using a Qubit dsDNA BR assay kit in a Qubit 4.0 fluorometer (Thermo Fisher Scientific, Waltham, MA). Genomic DNA was fluorescently tagged by using the Direct Label and Stain DLS DNA Kit (Bionano Genomics, catalog no. 80005). Labeled and stained DNA was quantitated using the Qubit dsDNA HS assay kit in a Qubit 4.0 fluorometer. Labeled DNA molecules were electrophoresed through low-voltage nanochannel arrays on a Saphyr chip (Bionano Genomics, catalog no. 20366) to linearize the DNA. High-throughput sequential imaging of the nanochannels was performed by using the Saphyr Genome Imaging Instrument to produce thousands of high-resolution images from which the molecule maps could be derived. Data were processed with Bionano Solve software version 3.5 to first align labeled molecules against the reference sequence predicted label pattern; the hg38 reference carries both the 4qA and 4qB

D4Z4 haplotypes. Molecules aligned to the reference chromosome 4q35 or 10q26 regions were further collected to generate representative allelic profiles of structural variation to interpret FSHD genotypes by using the custom EnFocus FSHD analysis (version 1.0; Bionano Genomics). (ii) Determination of D4Z4 methylation status: DNA was isolated from cultured cells and digested with the restriction enzymes Eco RI and Bgl II, followed by a second digestion with the methylation-sensitive enzyme Fse I. After linear gel electrophoresis, the blots were hybridized with the probe p13E-11. The relative amounts of methylated and hypomethylated DNA were quantitated using Image Lab software.

Next generation sequencing and analysis

Samples and library preparation

DNA sequencing. Rare sequence variants in *SMCHD1* were identified by exome sequencing and/or targeted (Sanger) sequencing, as previously described (*2*).

RNA sequencing. For RNA-seq of hESCs, RNA was extracted using TRIzol and prepared using DNeasy with deoxyribonuclease treatment, and concentration was measured with Qubit using Qubit RNA HS Assay Kit. Libraries were generated using TruSeq RNA Library Prep Kit v2 (Illumina, RS-122-2001) according to the manufacturer's instructions. Purified libraries were quantified on the Agilent Technologies 2100 Bioanalyzer with the Agilent High Sensitivity DNA Kit.

DUX4 ChIP-seq. SMCHD1 KO human ES (H9) cells (clones 79 and 49) were differentiated into placode cells in 10-cm dishes. On day 3, cells were cross-linked following the instructions from Active Motif (www.activemotif.com/documents/1848.pdf). Briefly, 1 ml of freshly prepared formaldehyde solution [11% formaldehyde (Sigma-Aldrich, F-8775), 0.1 M NaCl, 1 mM EDTA (pH 8.0), and 50 mM Hepes (H7.9)] was added into 10 ml of cell culture medium in a 10-cm dish. The dish was agitated for 15 min at room temperature. Fixation was stopped by adding 0.55 ml of 2.5 M glycine. After incubating at room temperature for 5 min, cells were collected and washed with chilled PBS containing 0.5% IGEPAL CA-630 (Sigma-Aldrich, I-8896) two times. After collecting the cells by centrifugation, the supernatant was completely removed by aspiration. Cells were then frozen on dry ice and shipped to Active Motif for further analysis. Abcam monoclonal anti-DUX4 antibody [E5-5] (Abcam, ab124699) was used for ChIP.

Reads processing. All next generation sequencing (NGS) reads were subjected to FASTQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) and FASTQ-SCREEN (www.bioinformatics.babraham.ac.uk/projects/fastq_screen/) for sequence quality verification and contamination screening. Sequences were then fed into TrimGalore (www.bioinformatics.babraham.ac.uk/projects/trim_galore/) for adapter sequence trimming and low-quality reads filtering (--cores 7 --fastqc). Pair-end reads were fed to TrimGalore selecting the pair-end mode (--paired), and only proper pair reads were retained. Cleaned high-quality reads from each dataset were used for the downstream analysis.

All the alignments are based on the reference human genome hg38 and the GENCODE gene model prediction. STAR (*63*) was used for RNA-seq read alignment to the reference genome considering the following parameters (--runMode alignReads --outFilterType BySJout --outSAMtype BAM SortedByCoordinate; table S8). BOWTIE2 (*64*, *65*) was used for aligning the ChIP-seq reads to the reference genome using single-end mode and --phred33 -p 21 parameters (table S9). GATK v4.1.9.0 was used to add appropriate read groups to each sample, remove PCR duplicates, sort and index, and write them in BAM format as analysis read ready for downstream analysis.

RNA-seq DEG analysis. featureCounts (*66*) was used to quantify reads uniquely intersecting exons of hg38 GENCODE gene models in each RNA-seq sample. The raw read count matrix was then loaded into R where size factors were determined from the complete dataset and then applied to each sample during each separately analyzed pairwise comparison (see tables S2 to S4; $|\log_2 FC| \geq 1$ and FDR < 0.05) using the DESeq2 package available on R Bioconductor (*67*). The size factor normalized read counts and DEGs from each pairwise comparison were then generated using the standard DESeq2 workflow.

ChIP-seq peak calling, annotation, and motif identification. Callpeak module from MACS2 (*68*) was run to call peaks (table S5). Peaks were then annotated using annotatePeaks module implemented in HOMER (*69*) cross checking the peaks with the GENCODE v32 reference human gene prediction and annotation. Module findMotif from HOMER was then used on the annotated peaks to identify known and de novo active motifs for each set (table S10).

Functional enrichment analysis. Enrichment analysis was carried out using gprofiler2 (*70*), an R package for gene list functional enrichment analysis and namespace conversion toolset g: Profiler (table S6).

Supplementary Materials

This PDF file includes:
Figs. S1 to S12
Legends for tables S1 to S10

Other Supplementary Material for this manuscript includes the following:
Tables S1 to S10

REFERENCES AND NOTES

1. C. T. Gordon, S. Xue, G. Yigit, H. Filali, K. Chen, N. Rosin, K.-I. Yoshiura, M. Oufadem, T. J. Beck, R. McGowan, A. C. Magee, J. Altmüller, C. Dion, H. Thiele, A. D. Gurzau, P. Nürnberg, D. Meschede, W. Mühlbauer, N. Okamoto, V. Varghese, R. Irving, S. Sigaudy, D. Williams, S. F. Ahmed, C. Bonnard, M. K. Kong, I. Ratbi, N. Fejjal, M. Fikri, S. C. Elalaoui, H. Reigstad, C. Bole-Feysot, P. Nitschké, N. Ragge, N. Lévy, G. Tunçbilek, A. S. M. Teo, M. L. Cunningham, A. Sefiani, H. Kayserili, J. M. Murphy, C. Chatdokmaiprai, A. M. Hillmer, D. Wattanasirichaigoon, S. Lyonnet, F. Magdinier, A. Javed, M. E. Blewitt, J. Amiel, B. Wollnik, B. Reversade, De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. *Nat. Genet.* **49**, 249–255 (2017).
2. N. D. Shaw, H. Brand, Z. A. Kupchinsky, H. Bengani, L. Plummer, T. I. Jones, S. Erdin, K. A. Williamson, J. Rainger, A. Stortchevoi, K. Samocha, B. B. Currall, D. S. Dunican, R. L. Collins, J. R. Willer, A. Lek, M. Lek, M. Nassan, S. Pereira, T. Kammin, D. Lucente, A. Silva, C. M. Seabra, C. Chiang, Y. An, M. Ansari, J. K. Rainger, S. Joss, J. C. Smith, M. F. Lippincott, S. S. Singh, N. Patel, J. W. Jing, J. R. Law, N. Ferraro, A. Verloes, A. Rauch, K. Steindl, M. Zweier, I. Scheer, D. Sato, N. Okamoto, C. Jacobsen, J. Tryggestad, S. Chernausek, L. A. Schimmenti, B. Brasseur, C. Cesaretti, J. E. Garcia-Ortiz, T. P. Buitrago, O. P. Silva, J. D. Hoffman, W. Muhlbauer, K. W. Ruprecht, B. L. Loeys, M. Shino, A. M. Kaindl, C. H. Cho, C. C. Morton, R. R. Meehan, V. van Heyningen, E. C. Liao, R. Balasubramanian, J. E. Hall, S. B. Seminara, D. Macarthur, S. A. Moore, K. I. Yoshiura, J. F. Gusella, J. A. Marsh, J. M. Graham Jr., A. E. Lin, N. Katsanis, P. L. Jones, W. F. Crowley Jr., E. E. Davis, D. R. FitzPatrick, M. E. Talkowski, SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. *Nat. Genet.* **49**, 238–248 (2017).
3. K. Chen, J. Hu, D. L. Moore, R. Liu, S. A. Kessans, K. Breslin, I. S. Lucet, A. Keniry, H. S. Leong, C. L. Parish, D. J. Hilton, R. J. L. F. Lemmers, S. M. van der Maarel, P. E. Czabotar, R. C. J. Dobson, M. E. Ritchie, G. F. Kay, J. M. Murphy, M. E. Blewitt, Genome-wide binding

and mechanistic analyses of Smchd1-mediated epigenetic regulation. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E3535–E3544 (2015).

4. M. E. Blewitt, A. V. Gendrel, Z. Pang, D. B. Sparrow, N. Whitelaw, J. M. Craig, A. Apedaile, D. J. Hilton, S. L. Dunwoodie, N. Brockdorff, G. F. Kay, E. Whitelaw, SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* **40**, 663–669 (2008).
5. M. E. Blewitt, N. K. Vickaryous, S. J. Hemley, A. Ashe, T. J. Bruxner, J. I. Preis, R. Arkell, E. Whitelaw, An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7629–7634 (2005).
6. A. V. Gendrel, A. Apedaile, H. Coker, A. Termanis, I. Zvetkova, J. Godwin, Y. A. Tang, D. Huntley, G. Montana, S. Taylor, E. Giannoulatou, E. Heard, I. Stancheva, N. Brockdorff, Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev. Cell* **23**, 265–279 (2012).
7. A. V. Gendrel, Y. A. Tang, M. Suzuki, J. Godwin, T. B. Nesterova, J. M. Greally, E. Heard, N. Brockdorff, Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. *Mol. Cell. Biol.* **33**, 3150–3165 (2013).
8. N. Jansz, A. Keniry, M. Trussart, H. Bildsoe, T. Beck, I. D. Tonks, A. W. Mould, P. Hickey, K. Breslin, M. Iminitoff, M. E. Ritchie, E. McGlinn, G. F. Kay, J. M. Murphy, M. E. Blewitt, Smchd1 regulates long-range chromatin interactions on the inactive X chromosome and at Hox clusters. *Nat. Struct. Mol. Biol.* **25**, 766–777 (2018).
9. H. S. Leong, K. Chen, Y. Hu, S. Lee, J. Corbin, M. Pakusch, J. M. Murphy, I. J. Majewski, G. K. Smyth, W. S. Alexander, D. J. Hilton, M. E. Blewitt, Epigenetic regulator Smchd1 functions as a tumor suppressor. *Cancer Res.* **73**, 1591–1599 (2013).
10. A. W. Mould, Z. Pang, M. Pakusch, I. D. Tonks, M. Stark, D. Carrie, P. Mukhopadhyay, A. Seidel, J. J. Ellis, J. Deakin, M. J. Wakefield, L. Krause, M. E. Blewitt, G. F. Kay, Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. *Epigenetics Chromatin* **6**, 19 (2013).
11. N. J. Brideau, H. Coker, A. V. Gendrel, C. A. Siebert, K. Bezstarosti, J. Demmers, R. A. Poot, T. B. Nesterova, N. Brockdorff, Independent mechanisms target SMCHD1 to trimethylated histone H3 lysine 9-modified chromatin and the inactive X chromosome. *Mol. Cell. Biol.* **35**, 4053–4068 (2015).
12. M. R. Gdula, T. B. Nesterova, G. Pintacuda, J. Godwin, Y. Zhan, H. Ozadam, M. McClellan, D. Moralli, F. Krueger, C. M. Green, W. Reik, S. Kriaucionis, E. Heard, J. Dekker, N. Brockdorff, The non-canonical SMC protein SmcHD1 antagonises TAD formation and compartmentalisation on the inactive X chromosome. *Nat. Commun.* **10**, 30 (2019).
13. C. Y. Wang, H. Brand, N. D. Shaw, M. E. Talkowski, J. T. Lee, Role of the chromosome architectural factor SMCHD1 in X-chromosome inactivation, gene regulation, and disease in humans. *Genetics* **213**, 685–703 (2019).
14. C. Y. Wang, T. Jegu, H. P. Chu, H. J. Oh, J. T. Lee, SMCHD1 merges chromosome compartments and assists formation of super-structures on the inactive X. *Cell* **174**, 406–421.e25 (2018).
15. R. J. L. F. Lemmers, R. Tawil, L. M. Petek, J. Balog, G. J. Block, G. W. E. Santen, A. M. Amell, P. J. van der Vliet, R. Almomani, K. R. Straasheijm, Y. D. Krom, R. Klooster, Y. Sun, J. T. den Dunnen, Q. Helmer, C. M. Donlin-Smith, G. W. Padberg, B. G. M. van Engelen, J. C. de Greef, A. M. Aartsma-Rus, R. R. Frants, M. de Visser, C. Desnuelle, S. Sacconi, G. N. Filippova, B. Bakker, M. J. Bamshad, S. J. Tapscott, D. G. Miller, S. M. van der Maarel, Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* **44**, 1370–1374 (2012).
16. A. D. Gurzau, M. E. Blewitt, P. E. Czabotar, J. M. Murphy, R. W. Birkinshaw, Relating SMCHD1 structure to its function in epigenetic silencing. *Biochem. Soc. Trans.* **48**, 1751–1763 (2020).
17. M. Larsen, S. Rost, N. El Hajj, A. Ferbert, M. Deschauer, M. C. Walter, B. Schoser, P. Tacik, W. Kress, C. R. Muller, Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. *Eur. J. Hum. Genet.* **23**, 808–816 (2015).
18. S. H. Choi, M. D. Gearhart, Z. Cui, D. Bosnakovski, M. Kim, N. Schennum, M. Kyba, DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res.* **44**, 5161–5173 (2016).
19. A. E. Campbell, S. C. Shadle, S. Jagannathan, J. W. Lim, R. Resnick, R. Tawil, S. M. van der Maarel, S. J. Tapscott, NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. *eLife* **7**, e31023 (2018).
20. A. De Iaco, E. Planet, A. Coluccio, S. Verp, J. Duc, D. Trono, DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **49**, 941–945 (2017).
21. P. G. Hendrickson, J. A. Dorais, E. J. Grow, J. L. Whiddon, J. W. Lim, C. L. Wike, B. D. Weaver, C. Pflueger, B. R. Emery, A. L. Wilcox, D. A. Nix, C. M. Peterson, S. J. Tapscott, D. T. Carrell, B. R. Cairns, Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
22. L. Snider, L. N. Geng, R. J. L. F. Lemmers, M. Kyba, C. B. Ware, A. M. Nelson, R. Tawil, G. N. Filippova, S. M. van der Maarel, S. J. Tapscott, D. G. Miller, Facioscapulohumeral dystrophy: Incomplete suppression of a retrotransposed gene. *PLOS Genet.* **6**, e1001181 (2010).
23. R. J. L. F. Lemmers, N. van der Stoep, P. J. V. Vliet, S. A. Moore, D. San Leon Granado, K. Johnson, A. Topf, V. Straub, T. Evangelista, T. Mozaffar, V. Kimonis, N. D. Shaw, R. Selvatici, A. Ferlini, N. Voermans, B. van Engelen, S. Sacconi, R. Tawil, M. Lamers, S. M. van der Maarel, SMCHD1 mutation spectrum for facioscapulohumeral muscular dystrophy type 2 (FSHD2) and Bosma arhinia microphthalmia syndrome (BAMS) reveals disease-specific localisation of variants in the ATPase domain. *J. Med. Genet.* **56**, 693–700 (2019).
24. P. Mohassel, N. Chang, K. Inoue, A. Delaney, Y. Hu, S. Donkervoort, D. Saade, B. J. Billioux, B. Meader, R. Volochayev, C. G. Konersman, A. M. Kaindl, C. H. Cho, B. Russell, A. Rodriguez, K. W. Foster, A. R. Foley, S. A. Moore, P. L. Jones, C. G. Bonnemann, T. Jones, N. D. Shaw, Cross-sectional neuromuscular phenotyping study of patients with arhinia with SMCHD1 variants. *Neurology* **98**, e1384–e1396 (2022).
25. K. Mul, R. J. L. F. Lemmers, M. Kriek, P. J. van der Vliet, M. L. van den Boogaard, U. A. Badrising, J. M. Graham Jr., A. E. Lin, H. Brand, S. A. Moore, K. Johnson, T. Evangelista, A. Töpf, V. Straub, S. K. Garcia, S. Sacconi, R. Tawil, S. J. Tapscott, N. C. Voermans, B. G. M. van Engelen, C. G. C. Horlings, N. D. Shaw, S. M. van der Maarel, FSHD type 2 and Bosma arhinia microphthalmia syndrome: Two faces of the same mutation. *Neurology* **91**, e562–e570 (2018).
26. J. M. Graham Jr., J. Lee, Bosma arhinia microphthalmia syndrome. *Am. J. Med. Genet. A* **140**, 189–193 (2006).
27. P. Barraud, A. A. Seferiadis, L. D. Tyson, M. F. Zwart, H. L. Szabo-Rogers, C. Ruhrberg, K. J. Liu, C. V. H. Baker, Neural crest origin of olfactory ensheathing glia. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21040–21045 (2010).
28. P. E. Forni, C. Taylor-Burds, V. S. Melvin, T. Williams, S. Wray, Neural crest and ectodermal cells intermix in the nasal placode to give rise to GnRH-1 neurons, sensory neurons, and olfactory ensheathing cells. *J. Neurosci.* **31**, 6915–6927 (2011).
29. M. Simoes-Costa, M. E. Bronner, Insights into neural crest development and evolution from genomic analysis. *Genome Res.* **23**, 1069–1080 (2013).
30. J. Corsin, Influence des placodes olfactives et des ebauches optiques sur la morphagenese du squelette cranien chez Pleurodeles waltii michah. *Annales d'Embryologie et de Morphogenese* **1**, 41–48 (1971).
31. H. L. Szabo-Rogers, P. Geetha-Loganathan, C. J. Whiting, S. Nimmagadda, K. Fu, J. M. Richman, Novel skeletogenic patterning roles for the olfactory pit. *Development* **136**, 219–229 (2009).
32. T. Humphrey, The development of the olfactory and the accessory olfactory formations in human embryos and fetuses. *J. Comp. Neurol.* **73**, 431–468 (1940).
33. M. Schwanzel-Fukuda, D. Bick, D. W. Pfaff, Luteinizing hormone-releasing hormone (LHRH)-expressing cells do not migrate normally in an inherited hypogonadal (Kallmann) syndrome. *Brain Res. Mol. Brain Res.* **6**, 311–326 (1989).
34. A. Delaney, R. Volochayev, B. Meader, J. Lee, K. Almpani, G. Y. Noukelak, J. Henkind, L. Chalmers, J. R. Law, K. A. Williamson, C. M. Jacobsen, T. P. Buitrago, O. Perez, C.-H. Cho, A. Kaindl, A. Rauch, K. Steindl, J. E. Garcia, B. E. Russell, R. Prasad, U. K. Mondal, H. M. Reigstad, S. Clements, S. Kim, K. Inoue, G. Arora, K. B. Salnikov, N. P. DiOrio, R. Prada, Y. Capri, K. Morioka, M. Mizota, R. M. Zechi-Ceide, N. M. Kokitsu-Nakata, C. Tonello, S. Vendramini-Pittoli, G. da Silva Dalben, R. Balasubramanian, A. A. Dwyer, S. B. Seminara, W. F. Crowley, L. Plummer, J. E. Hall, J. M. Graham, A. E. Lin, N. D. Shaw, Insight into the ontogeny of GnRH neurons from patients born without a nose. *J. Clin. Endocrinol. Metab.* **105**, 1538–1551 (2020).
35. Z. Dincer, J. Piao, L. Niu, Y. Ganat, S. Kriks, B. Zimmer, S. H. Shi, V. Tabar, L. Studer, Specification of functional cranial placode derivatives from human pluripotent stem cells. *Cell Rep.* **5**, 1387–1402 (2013).
36. L. A. Rojas, E. Valentine, A. Accorsi, J. Maglio, N. Shen, A. Robertson, S. Kazmirski, P. Rahl, R. Tawil, D. Cadavid, L. A. Thompson, L. Ronco, A. N. Chang, A. M. Cacace, O. Wallace, p38α regulates expression of DUX4 in a model of facioscapulohumeral muscular dystrophy. *J. Pharmacol. Exp. Ther.* **374**, 489–498 (2020).
37. R. Bajpai, D. A. Chen, A. Rada-Iglesias, J. Zhang, Y. Xiong, J. Helms, C. P. Chang, Y. Zhao, T. Swigut, J. Wysocka, CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature* **463**, 958–962 (2010).
38. R. Bajpai, G. Coppola, M. Kaul, M. Talantova, F. Cimadamore, M. Nilbratt, D. H. Geschwind, S. A. Lipton, A. V. Terskikh, Molecular stages of rapid and uniform neuralization of human embryonic stem cells. *Cell Death Differ.* **16**, 807–825 (2009).
39. L. N. Geng, Z. Yao, L. Snider, A. P. Fong, J. N. Cech, J. M. Young, S. M. van der Maarel, W. L. Ruzzo, R. C. Gentleman, R. Tawil, S. J. Tapscott, DUX4 activates germline genes, retroelements, and immune mediators: Implications for facioscapulohumeral dystrophy. *Dev. Cell* **22**, 38–51 (2012).
40. C. E. Shiau, M. Bronner-Fraser, N-cadherin acts in concert with Slit1-Robo2 signaling in regulating aggregation of placode-derived cranial sensory neurons. *Development* **136**, 4155–4164 (2009).
41. D. Bosnakovski, M. T. da Silva, S. T. Sunny, E. T. Ener, E. A. Toso, C. Yuan, Z. Cui, M. A. Walters, A. Jadhav, M. Kyba, A novel P300 inhibitor reverses DUX4-mediated global histone H3 hyperacetylation, target gene expression, and cell death. *Sci. Adv.* **5**, eaaw7781 (2019).

42. F. Full, M. van Gent, K. M. J. Sparrer, C. Chiang, M. A. Zurenski, M. Scherer, N. H. Brockmeyer, L. Heinzerling, M. Sturzl, K. Korn, T. Stamminger, A. Ensser, M. U. Gack, Centrosomal protein TRIM43 restricts herpesvirus infection by regulating nuclear lamina integrity. *Nat. Microbiol.* **4**, 164–176 (2019).
43. T. Yuan, J. R. York, D. W. McCauley, Neural crest and placode roles in formation and patterning of cranial sensory ganglia in lamprey. *Genesis* **58**, e23356 (2020).
44. S. Abdelhak, V. Kalatzis, R. Heilig, S. Compain, D. Samson, C. Vincent, D. Weil, C. Cruaud, I. Sahly, M. Leibovici, M. Bitner-Glindzicz, M. Francis, D. Lacombe, J. Vigneron, R. Charachon, K. Boven, P. Bedbeder, N. Van Regemorter, J. Weissenbach, C. Petit, A human homologue of the Drosophila eyes absent gene underlies branchio-oto-renal (BOR) syndrome and identifies a novel gene family. *Nat. Genet.* **15**, 157–164 (1997).
45. E. Theveneau, B. Steventon, E. Scarpa, S. Garcia, X. Trepat, A. Streit, R. Mayor, Chase-and-run between adjacent cell populations promotes directional collective migration. *Nat. Cell Biol.* **15**, 763–772 (2013).
46. C. Laberthonniere, E. M. Novoa-Del-Toro, R. Chevalier, N. Broucqsault, V. V. Rao, J. P. Trani, K. Nguyen, S. Xue, B. Reversade, J. D. Robin, A. Baudot, F. Magdinier, AKT signaling modifies the balance between cell proliferation and migration in neural crest cells from patients affected with bosma arhinia and microphthalmia syndrome. *Biomedicine* **9**, 751 (2021).
47. E. J. Grow, B. D. Weaver, C. M. Smith, J. Guo, P. Stein, S. C. Shadle, P. G. Hendrickson, N. E. Johnson, R. J. Butterfield, R. Menafra, S. L. Kloet, S. M. van der Maarel, C. J. Williams, B. R. Cairns, p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models. *Nat. Genet.* **53**, 1207–1220 (2021).
48. C. E. Lilley, C. T. Carson, A. R. Muotri, F. H. Gage, M. D. Weitzman, DNA repair proteins affect the lifecycle of herpes simplex virus 1. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5844–5849 (2005).
49. S. Walter, V. Franke, N. Drayman, E. Wyler, S. Tay, M. Landthaler, A. Akalin, A. Ensser, F. Full, Herpesviral induction of germline transcription factor DUX4 is critical for viral gene expression. bioRxiv 2021.03.24.436599 [**Preprint**]. 24 March 2021. https://doi.org/10.1101/2021.03.24.436599.
50. D. M. Money, M. Steben; INFECTIOUS DISEASES COMMITTEE, Guidelines for the management of herpes simplex virus in pregnancy. *J. Obstet. Gynaecol. Can.* **30**, 514–519 (2008).
51. J. Beers, K. L. Linask, J. A. Chen, L. I. Siniscalchi, Y. Lin, W. Zheng, M. Rao, G. Chen, A cost-effective and efficient reprogramming platform for large-scale production of integration-free human induced pluripotent stem cells in chemically defined culture. *Sci. Rep.* **5**, 11319 (2015).
52. L. C. Mackey, L. A. Annab, J. Yang, B. Rao, G. E. Kissling, S. H. Schurman, D. Dixon, T. K. Archer, Epigenetic enzymes, age, and ancestry regulate the efficiency of human iPSC reprogramming. *Stem Cells* **36**, 1697–1708 (2018).
53. J. Yu, K. Hu, K. Smuga-Otto, S. Tian, R. Stewart, I. I. Slukvin, J. A. Thomson, Human induced pluripotent stem cells free of vector and transgene sequences. *Science* **324**, 797–801 (2009).
54. L. S. Bisogno, J. Yang, B. D. Bennett, J. M. Ward, L. C. Mackey, L. A. Annab, P. R. Bushel, S. Singhal, S. H. Schurman, J. S. Byun, A. M. Napoles, E. J. Perez-Stable, D. C. Fargo, K. Gardner, T. K. Archer, Ancestry-dependent gene expression correlates with reprogramming to pluripotency and multiple dynamic biological processes. *Sci. Adv.* **6**, eabc3851 (2020).
55. M. Z. Li, S. J. Elledge, Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251–256 (2007).
56. T. Sueyoshi, T. Kawamoto, I. Zelko, P. Honkakoski, M. Negishi, The repressed nuclear receptor CAR responds to phenobarbital in activating the human CYP2B6 gene. *J. Biol. Chem.* **274**, 6043–6046 (1999).
57. I. Barde, P. Salmon, D. Trono, Production and titration of lentiviral vectors. *Curr. Protoc. Neurosci.* **Chapter 4**, Unit 4.21 (2010).
58. L. C. Pedersen, K. Inoue, S. Kim, L. Perera, N. D. Shaw, A ubiquitin-like domain is required for stabilizing the N-terminal ATPase module of human SMCHD1. *Commun. Biol.* **2**, 255 (2019).
59. M. Kuragano, T. Q. P. Uyeda, K. Kamijo, Y. Murakami, M. Takahashi, Different contributions of nonmuscle myosin IIA and IIB to the organization of stress fiber subtypes in fibroblasts. *Mol. Biol. Cell* **29**, 911–922 (2018).
60. C. C. Uphoff, H. G. Drexler, Detecting mycoplasma contamination in cell cultures by polymerase chain reaction. *Methods Mol. Biol.* **731**, 93–103 (2011).
61 . M. K. Preston, R. Tawil, L. H. Wang, Facioscapulohumeral muscular dystrophy, in *GeneReviews((R))*, M. P. Adam, D. B. Everman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. W. Gripp, A. Amemiya, H. H. Ardinger, Eds. (University of Washington, Seattle,1993).
62. A. A. Stence, J. G. Thomason, J. A. Pruessner, R. R. Sompallae, A. N. Snow, D. Ma, S. A. Moore, A. D. Bossler, Validation of optical genome mapping for the molecular diagnosis of facioscapulohumeral muscular dystrophy. *J. Mol. Diagn.* **23**, 1506–1514 (2021).
63. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
64. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
65. B. Langmead, C. Wilks, V. Antonescu, R. Charles, Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).
66. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
67. R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. Yang, J. Zhang, Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
68. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
69. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
70. L. Kolberg, U. Raudvere, I. Kuzmin, J. Vilo, H. Peterson, gprofiler2—An R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res.* **9**, 709 (2020).

Acknowledgments: We acknowledge the following Core Facilities for their scientific and technical assistance: Epigenomics and DNA Sequencing Core Facility, Viral Vector Core Facility, Microscopy Core (in particular, C. "Jeff" Tucker and E. Scappini), the Flow Cytometry Core, and the Structural Biology Core Facility at NIEHS and the iPSC Core Facility at NHLBI. In particular, we wish to thank J. Zou and J. Beers for their expertise in preparing and characterizing patient-derived iPSC. We thank C. Emerson Jr., R. Tawil, and T. Archer for providing fibroblasts or iPSC; G. Hu for providing H9 cells; T. Sueyoshi for providing the pGL3B-tk plasmid; and R. Bajpai and L. Annab for sharing their expertise in neural crest cell and iPSC derivation, respectively. We also thank S. McCaw for taking high-resolution photographs of iPSC and ESC. **Funding:** This work was supported, in part, by the Intramural Research Program of the NIH, NIEHS (1ZIAES103327-05, N.D.S.). N.D.S. is also supported as a Lasker Clinical Research Scholar (1SI2ES025429-01). S.A.M. is funded by P50 NS053672 to the Iowa Wellstone Muscular Dystrophy Specialized Research Center. **Author contributions:** K.I. designed and performed experiments, analyzed and interpreted data, and wrote and finalized the manuscript. H.B. and A.B.B. designed and performed all bioinformatic analyses and revised the manuscript. J.-L.L. supervised bioinformatic analyses. M.R.B. and O.F.B. performed experiments and analyzed data. N.P.M. and S.-H.C. conducted viral studies. C.D.B. performed flow cytometry studies. N.D.S. conceived and conceptualized the study design, supervised the experiments, and wrote and finalized the manuscript. A.A.S. performed D4Z4 haplotyping and methylation analyses, and S.A.M. supervised these experiments. All authors reviewed and approved the final manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All sequencing data is publicly available at SRA (BioProject ID: PRJNA818884).

Submitted 29 April 2022
Accepted 12 January 2023
Published 17 February 2023
10.1126/sciadv.abq7744