p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models

Grow, E.J.; Weaver, B.D.; Smith, C.M.; Guo, J.T.; Stein, P.; Shadle, S.C.; ... ; Cairns, B.R.

nature genetics

ARTICLES

https://doi.org/10.1038/s41588-021-00893-0

p53 convergently activates *Dux/DUX4* in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models

Edward J. Grow[1], Bradley D. Weaver[1,7], Christina M. Smith[1,7], Jingtao Guo[1,2], Paula Stein[3], Sean C. Shadle[1], Peter G. Hendrickson[1], Nicholas E. Johnson[4], Russell J. Butterfield[4], Roberta Menafra[5], Susan L. Kloet[5], Silvère M. van der Maarel[6], Carmen J. Williams[3] and Bradley R. Cairns[1]✉

In mammalian embryos, proper zygotic genome activation (ZGA) underlies totipotent development. Double homeobox (DUX)-family factors participate in ZGA, and mouse *Dux* is required for forming cultured two-cell (2C)-like cells. Remarkably, in mouse embryonic stem cells, *Dux* is activated by the tumor suppressor p53, and *Dux* expression promotes differentiation into expanded-fate cell types. Long-read sequencing and assembly of the mouse *Dux* locus reveals its complex chromatin regulation including putative positive and negative feedback loops. We show that the p53–DUX/DUX4 regulatory axis is conserved in humans. Furthermore, we demonstrate that cells derived from patients with facioscapulohumeral muscular dystrophy (FSHD) activate human *DUX4* during p53 signaling via a p53-binding site in a primate-specific subtelomeric long terminal repeat (LTR)10C element. In summary, our work shows that p53 activation convergently evolved to couple p53 to *Dux/DUX4* activation in embryonic stem cells, embryos and cells from patients with FSHD, potentially uniting the developmental and disease regulation of DUX-family factors and identifying evidence-based therapeutic opportunities for FSHD.

Fertilization and early embryogenesis involve the transition from specialized unipotent gametes to totipotent embryos. After fertilization, mammalian embryos rely on maternally deposited RNA but subsequently initiate ZGA during which embryonic transcription begins[1]. Diverse mechanisms control the timing of ZGA, such as controlling RNA polymerase activity, the nuclear/cytoplasmic ratio, or translation of critical ZGA transcription factors (TFs) in *Caenorhabditis elegans*, *Drosophila melanogaster* or *Danio rerio*, respectively[1]. Currently, we have an incomplete understanding of how the transcriptional machinery (RNA polymerases) and/or sequence-specific TFs dictate the timing of ZGA in mammals and contribute to developmental potential.

Recent work identified the TF DUX (DUX4 in humans) as a key regulator of ZGA gene expression[2–5]. When ectopically expressed in cells, DUX and DUX4 activate many ZGA genes, including the earliest wave of ZGA genes in humans and mice[2,3]. However, the extent to which DUX is required for appropriate ZGA is unclear, as the reported effect of genetic loss of *Dux* ranges from minor molecular to major transcriptional defects and decreased development in mouse or human embryos[4–6].

To study ZGA using a cellular model, we and others have used 2C-embryo-like cells (2CLCs), which are an endogenously fluctuating subpopulation of mouse embryonic stem cells (mESCs) that recapitulate several key features of ZGA[7]. The 2CLCs activate transcripts characteristic of the 2C mouse embryo (including *Dux*, *Zscan4*, the endogenous retrovirus MERVL and *Kdm4dl*), lack chromocenters and downregulate pluripotency markers such as OCT4 (refs. [2,7,8]). Importantly, *Dux* is required for 2CLC formation, and, when expressed in mESCs, is sufficient to activate the 2CLC state at transcriptional and chromatin levels[2,3]. Mouse DUX is encoded by a retrogene array of >28 copies (unassembled in mm10), and a set of repressors are known to coordinate *Dux* array repression[8–10]. However, it is currently unclear which TF activators directly activate *Dux*. Notably, although DPPA2 and DPPA4 TFs and the pause release factor NELFA are reported to bind the *Dux* promoter, they lack clear DNA sequence-specific binding and are probably not gene-selectivity factors[11,12]. As *Dux* transcripts are not maternally inherited and as *Dux* is activated in early ZGA[2], we hypothesized that a maternally deposited (and previously unidentified) TF activates mouse *Dux* and serves as the upstream trigger for ZGA and the emergence of 2CLCs in mESC cultures.

Likewise, the activator for the human ortholog *DUX4* remains unknown, either during ZGA or in pathogenic contexts. *DUX4* reactivation from the *4qA* permissive haplotype containing a polyadenylation signal causes the human disease FSHD[13], characterized by progressive degeneration of affected muscle groups[14]. Normally, the *DUX4* locus exists as ~11–100 tandem repeats of the *D4Z4* repeat unit, but, in patients with FSHD1, *D4Z4* contraction to less than eight repeat units relieves epigenetic silencing of the *DUX4* locus and allows for stochastic *DUX4* activation[14,15].

[1]Howard Hughes Medical Institute, Department of Oncological Sciences and Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA. [2]Division of Urology, Department of Surgery, University of Utah School of Medicine, Salt Lake City, UT, USA. [3]Reproductive and Developmental Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA. [4]Department of Neurology, University of Utah School of Medicine, Salt Lake City, UT, USA. [5]Leiden Genome Technology Center, Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands. [6]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands. [7]These authors contributed equally: Bradley D. Weaver, Christina M. Smith. ✉e-mail: bradley.cairns@hci.utah.edu

FSHD2 is caused by loss-of-function mutations in the locus encoding heterochromatin protein SMCHD1 (*Smchd1* gene in mice[16]) and confers *DUX4* activation of the wild-type (WT) *D4Z4* repeat locus[17]. *DUX4* activation causes PKR- and MYC-dependent cell death in cultured FSHD myoblasts[18], and the *DUX4* locus is normally silenced by several heterochromatin proteins (such as SMCHD1 and CHD4)[17,19]. However, as with the mouse *Dux* locus, it is unclear which transcriptional activator(s) regulate the human *DUX4* locus and whether regulation of *DUX4* during embryonic genome activation is similar to the FSHD disease state.

Here, we use the 2CLC system to identify p53 as a key driver of *Dux* expression. First, we reveal that *Dux* activation in mESCs requires the DNA-damage response (DDR) pathway[20]. In contrast to a recent report[20], we demonstrate that p53 is required for DNA-damage-mediated DUX induction and 2CLC emergence. Critically, there are multiple sources of endogenous DNA damage present in the early embryo[21–24], and we find that p53 is activated soon after fertilization. Although not strictly required, p53 is important for full or proper *Dux* activation and DUX-target expression during ZGA. By sequencing and assembling the mouse *Dux* locus, we discover an unusual 'poised' chromatin signature and regulatory features including a p53-dependent *Dux* promoter, a DUX positive feedback loop and ZSCAN4 binding to a CA repeat embedded in each *Dux* repeat unit. Transient DUX expression alters the cellular differentiation of mESCs, biasing them to an expanded-fate potential. Importantly, we find that the regulatory relationship between p53 and *DUX4* was conserved in humans and that cells derived from patients with FSHD contain inducible *DUX4* alleles, are hypersensitive to DNA damage and use a primate-specific p53-bound LTR10C element to activate the locus. Surprisingly, our data show that the mouse and human *Dux* and *DUX4* loci likely convergently evolved p53 regulation. Previously, the signal initiating *DUX4* expression in FSHD was elusive, and our findings identify a promising disease mechanism for therapeutic intervention. Together, our results uncover a regulatory role for p53 in 2CLCs and *DUX4* expression in FSHD.

Results

DNA damage induces *Dux* expression and 2CLC emergence. First, we searched for conditions in which downstream DUX targets were activated. DNA-damaging agents can induce the expression of *Zscan4* (ref. [25]), which is a direct DUX target[2]. Here, chemicals that induce double-stranded DNA breaks directly or indirectly (doxorubicin[25–27] or hydroxyurea–aphidicolin, respectively) led to a much higher fraction of MERVL–GFP$^+$ (a reporter for DUX activity[2]) cells in mESC cultures (Fig. 1a and Extended Data Fig. 1a). Additionally, expression of the PpoI endonuclease (which makes double-stranded DNA breaks[26]) induced the MERVL–GFP reporter (Fig. 1b), indicating that DNA breaks induce 2C reporter activation. To test whether DUX itself was induced by DNA damage upstream of *Zscan4* and MERVL activation, we raised a highly specific antibody to mouse DUX (Extended Data Fig. 1b–e), which stained mESC nuclei treated with doxorubicin but not *Dux*-knockout (KO) mESCs (Fig. 1c), showing that DUX protein is induced by doxorubicin.

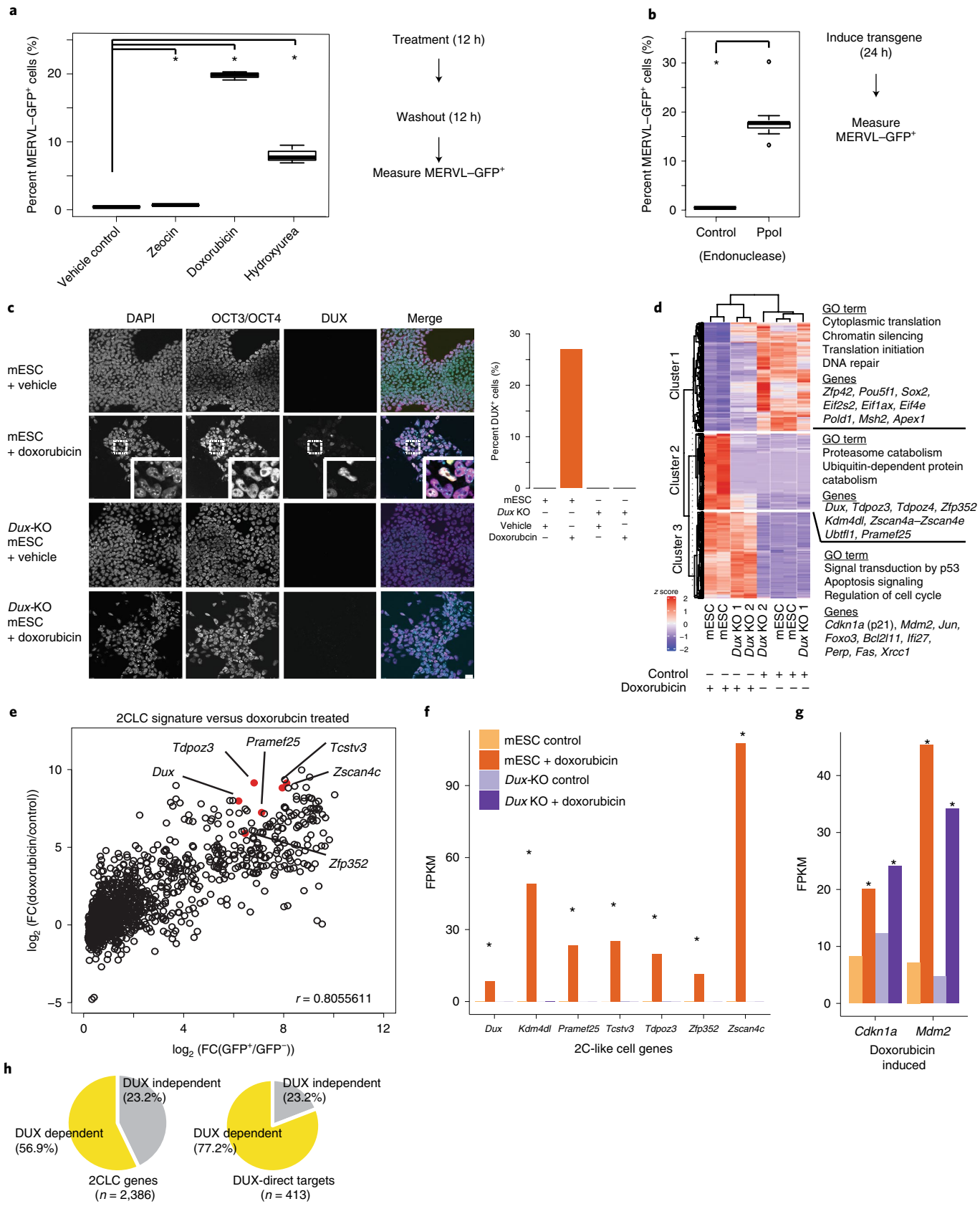
To test whether *Dux* is required for the DNA-damage induction of 2C genes, we performed RNA-seq with *Dux*-KO and control mESCs treated with doxorubicin. Hierarchical clustering of the differentially expressed transcripts revealed three main classes of genes: cluster 1 contained genes downregulated after doxorubicin treatment, and cluster 2 contained 2C-like genes that failed to activate in *Dux*-KO mESCs, while cluster 3 contained classical DDR genes[27] that were activated in both *Dux*-KO and control *Dux*-WT mESCs (Fig. 1d). Surprisingly, the 2C-like network is evidently a subset of a non-canonical DDR in mESCs. Naturally fluctuating MERVL–GFP$^+$ 2CLCs also exhibit a 2C gene expression profile similar to that of doxorubicin-treated mESCs[2], highlighting the similarity between the stochastic occurrence of 2CLCs and the DNA-damage-stimulated signature (correlation $r = 0.8055$, Fig. 1e). Doxorubicin treatment induced *Dux* RNA and 2C gene expression in *Dux*-WT control mESCs but not in *Dux*-KO mESCs, which exhibited negligible induction of classical 2C genes such as *Prame**fl16*, *Zscan4c*, *Zfp352*, *Tdpoz3*, *Kdm4dl* and *Tsvtv3* (Fig. 1f). Importantly, both *Dux*-KO and WT control mESCs activate classical DDR genes[27] such as *Cdkn1a* (p21) and *Mdm2* after doxorubicin treatment, indicating that DUX is not required for the canonical DDR typical of somatic cells (Fig. 1g). Most of the 2CLC gene network shows *Dux*-dependent activation after doxorubicin treatment; thus, DNA damage is coupled to 2CLC induction through DUX (Fig. 1h and Extended Data Fig. 1f,g).

DDR and p53 are required for stimulus-induced 2CLCs. Next, we identified factors downstream of DNA damage and upstream of *Dux* induction. DDR involves the kinases ATR and ATM that sense DNA damage and the kinase transducers CHK1 and CHK2 (ref. [28]). Treatment of cells with all four small-molecule inhibitors of ATR, ATM, CHK1 and CHK2 together nearly eliminated MERVL–GFP$^+$ cells after doxorubicin treatment (Fig. 2a). Interestingly, small interfering (si)RNA-mediated depletion of p53, the main transcriptional effector of DDR, likewise led to an almost complete loss of MERVL–GFP$^+$ mESCs (Fig. 2b). We confirmed that p53 was required for full 2C gene and *Cdkn1a* and *Mdm2* expression induction after DNA

Fig. 1 | DNA damage induces a 2CLC signature in mESCs and requires *Dux*. a, Small molecules that cause DNA damage induce MERVL-GFP expression. mESCs were treated with the indicated DNA-damage-inducing agents; MERVL-GFP$^+$ cells were quantified by flow cytometry. **False discovery rate (FDR) < 0.05, one-way ANOVA, one-sided *t*-test, $n = 3$ biological replicates. **b**, Direct formation of DNA double-stranded breaks induces expression of the MERVL-GFP reporter. mESCs were induced to express PpoI endonuclease or a control; MERVL-GFP$^+$ cells were quantified 24 h later. *$P < 0.05$, one-sided *t*-test, $P = 4.361 \times 10^{-6}$, $n = 8$ biological replicates. **c**, Immunostaining of control or *Dux*-KO mESCs treated with vehicle control or doxorubicin for 6 h and washed out for 6 h. Merge: 4′,6-diamidino-2-phenylindole (DAPI), cyan; OCT4, magenta; DUX, yellow. Mean quantification is on the right; representative images are shown of three fields of view. Representative example from five independent experiments. Scale bar, 25 µm. **d**, Hierarchical clustering of RNA-seq data from control or *Dux*-KO mESCs treated with vehicle (control) or doxorubicin for 6 h and allowed a 12-h washout recovery, using DESeq2 with FDR < 0.01. $n = 2$ biological replicates for each genotype and condition. **e**, Doxorubicin treatment in mESCs induces a 2CLC transcriptional signature. Endogenously fluctuating 2CLC genes[2] (*x* axis) were compared to the effect of doxorubicin versus control (*y* axis). $n = 2$ biological replicates. Doxorubicin, 6 h; washout, 12 h. FC, fold change. **f**, Key 2CLC genes require *Dux* for their induction after doxorubicin treatment. RNA-seq, mean of $n = 2$ biological replicates. *FDR < 0.05, DESeq2. Doxorubicin, 6 h; washout, 12 h. FPKM, fragments per kb per million reads. **g**, *Dux*-KO mESCs induce DNA-damage-responsive genes. Mean of $n = 2$ biological replicates, *FDR < 0.05, DESeq2. Doxorubicin, 6 h; washout, 12 h. **h**, Most 2CLC transcripts show DUX-dependent induction after doxorubicin treatment (left), and 77.2% of direct DUX targets (those bound by DUX via ChIP-seq[2] and transcriptionally induced by *Dux* transgene expression) require *Dux* for their induction. $n = 2$ biological replicates of *Dux*-KO and WT mESCs for each treatment (vehicle versus doxorubicin), FDR < 0.05, DESeq2. Doxorubicin, 6 h; washout, 12 h. For box plots in **a**,**b**, the median is shown as a line in the box, and the outline of the box is depicted at the 25th and 75th percentiles. Extended whiskers depict quartile (Q)1 − 1.5 × interquartile range (IQR) and Q3 + 1.5 × IQR. Outliers are depicted as dots.

damage through RNA-seq analysis of two independent *Trp53*-KO mESC clones treated with doxorubicin (Fig. 2c and Extended Data Fig. 2a). Thus, p53 is evidently required for *Dux* and 2C gene expression after DNA damage (Fig. 2c).

Loss of p53 decreases endogenously fluctuating 2CLCs. We next tested whether endogenously fluctuating 2CLCs require p53 activation for their generation. Staining of mESCs indicated that ~4% were phospho-p53^{+} (Fig. 2d), and mitotic cells with phosphorylated



histone H2AX (γH2AX)$^+$ chromosomes (not shown) or phospho-p53 were often observed, consistent with G2/M enrichment of 2CLCs[29] (Fig. 2d). The loss of *Dux* or *Trp53* decreased the proportion of MERVL–GFP$^+$ cells, indicating that DUX and p53 are required for the normal fluctuation of mESCs into the 2CLC state (Fig. 2e). This suggests that DNA-damage-induced 2CLCs and endogenously fluctuating 2CLCs share a similar mechanistic trigger.

p53 expression rescues *Dux* activation in *Dppa2*- and *Dppa4*-double KO mESCs. We next determined the epistatic relationship between *Trp53* and other genes that influence 2CLC emergence and encode putative *Dux* activators, such as *Dppa2* and *Dppa4*. We confirmed that double KO (dKO) of *Dppa2* and *Dppa4* led to the loss of DNA-damage-dependent *Dux* and *Zscan4* induction (Extended Data Fig. 2b,c). However, overexpression (OE) of p53 strongly reactivated *Dux* and *Zscan4* after doxorubicin treatment in *Dppa2* and *Dppa4*-dKO mESCs, indicating that p53 can rescue the loss of *Dppa2* and *Dppa4* and is likely downstream of *Dppa2* and *Dppa4* (Fig. 2f). By contrast, OE of DPPA2 and DPPA4 was not sufficient to reactivate *Dux*, even in *Trp53*-WT mESCs (Extended Data Fig. 2e). These observations favor a model in which *Dppa2* and *Dppa4* help preserve the inducibility of the *Dux* locus[11,30], but loss of *Dppa2* and *Dppa4* can be bypassed through OE of an activator of *Dux* such as p53.

***Dux* locus assembly reveals new regulatory features.** To understand the regulatory landscape of the mouse *Dux* locus, which is currently in an assembly gap in the mm10 mouse genome build, we performed long-read sequencing (via PacBio) and assembled the locus (Extended Data Fig. 3a). Consistent with prior Southern blotting results[31], the mouse ~350-kb *Dux* locus exists as two tandem repeats of 18 or ten units (which we call 18×*Dux* or 10×*Dux*, respectively) separated by ~77-kb 'linker' sequences (Extended Data Fig. 3a).

We next remapped epigenomic datasets to understand chromatin regulation of the *Dux* locus. In mESCs, we detected an unusual chromatin signature at the repeated *Dux* gene promoters: high levels of monomethyl histone 3 lysine 4 (H3K4me1) and trimethyl histone 3 lysine 27 (H3K27me3) and low levels of trimethyl histone 3 lysine 4 (H3K4me3) (Fig. 3a). After doxorubicin treatment, the *Dux* arrays gained acetyl histone 3 lysine 27 (H3K27ac), open chromatin, and were marked by H3K4me3 in *Zscan4*-reporter$^+$ mESCs (Fig. 3a–c and Extended Data Fig. 1f). This signature is similar to 'poised' enhancers and is characterized by H3K4me1 and H3K27me3 until activation, when these poised enhancers gain H3K27ac[32]. We also detected broad p53 chromatin immunoprecipitation followed by sequencing (ChIP–seq) signal across the *Dux* arrays, and MACS2 peak calling identified p53 ChIP–seq peaks at the promoters of the *Dux* repeat array (Fig. 3a,c), although we could not rule out an additional indirect effect of p53 on the *Dux* locus. In reporter assays, a 1.3-kb fragment of the *Dux* promoter had strong doxorubicin-stimulated activity, and this effect was only observed in *Trp53*-WT mESCs (Fig. 3d). CRISPR activation targeting to the *Dux* promoter, but not four other regions surrounding the locus, strongly induced *Dux* expression in *Trp53*-KO mESCs (Extended Data Fig. 3b). Although we could not rule out additional distal regulatory elements for the *Dux* locus, targeting a strong activator to the *Dux* promoter was sufficient to activate expression and bypass loss of p53. This indicates that de-repression is not required per se for *Dux* activation in mESCs and underscores the potent inducibility of the locus consistent with its 'poised' chromatin status.

Next, we analyzed other TFs and DNA-binding proteins implicated in the 2CLC state. We detected strong enrichment of DPPA2 and DPPA4 binding near the transcriptional start site (TSS) of the *Dux* repeats, consistent with several recent reports[11] (Fig. 3a,e). Surprisingly, we also detected strong enrichment of DUX protein binding to the *Dux* gene through a perfect match to the DUX motif ~500 bp upstream of the TSS (Fig. 3e). A codon-altered *Dux* transgene activated the endogenous *Dux* locus ~100-fold (Fig. 3f), identifying a possible positive feedback loop.

NELFA was previously reported to drive the 2CLC state by directly activating *Dux*[12], but our reanalysis of published ChIP–seq data showed low enrichment of NELFA at the *Dux* locus (Fig. 3a and Extended Data Fig. 3c). Instead, we found that *Nelfa* is a direct target of DUX through a tandem MERVL inserted near the 3′ end of the *Nelfa* gene (Extended Data Fig. 3c). Importantly, *Dux* transgene expression activated *Nelfa* ~900-fold, *Nelfa* transcription was induced by doxorubicin, and this effect required p53 and DUX (Extended Data Fig. 3c,d). In sum, our findings indicate that NEFLA expression is a marker rather than a driver of the 2CLC state, but *Nelfa*-KO mESCs or maternal–zygotic (MZ) *Nelfa*-KO embryos will help clarify the role of NELFA in 2CLCs or ZGA embryos.

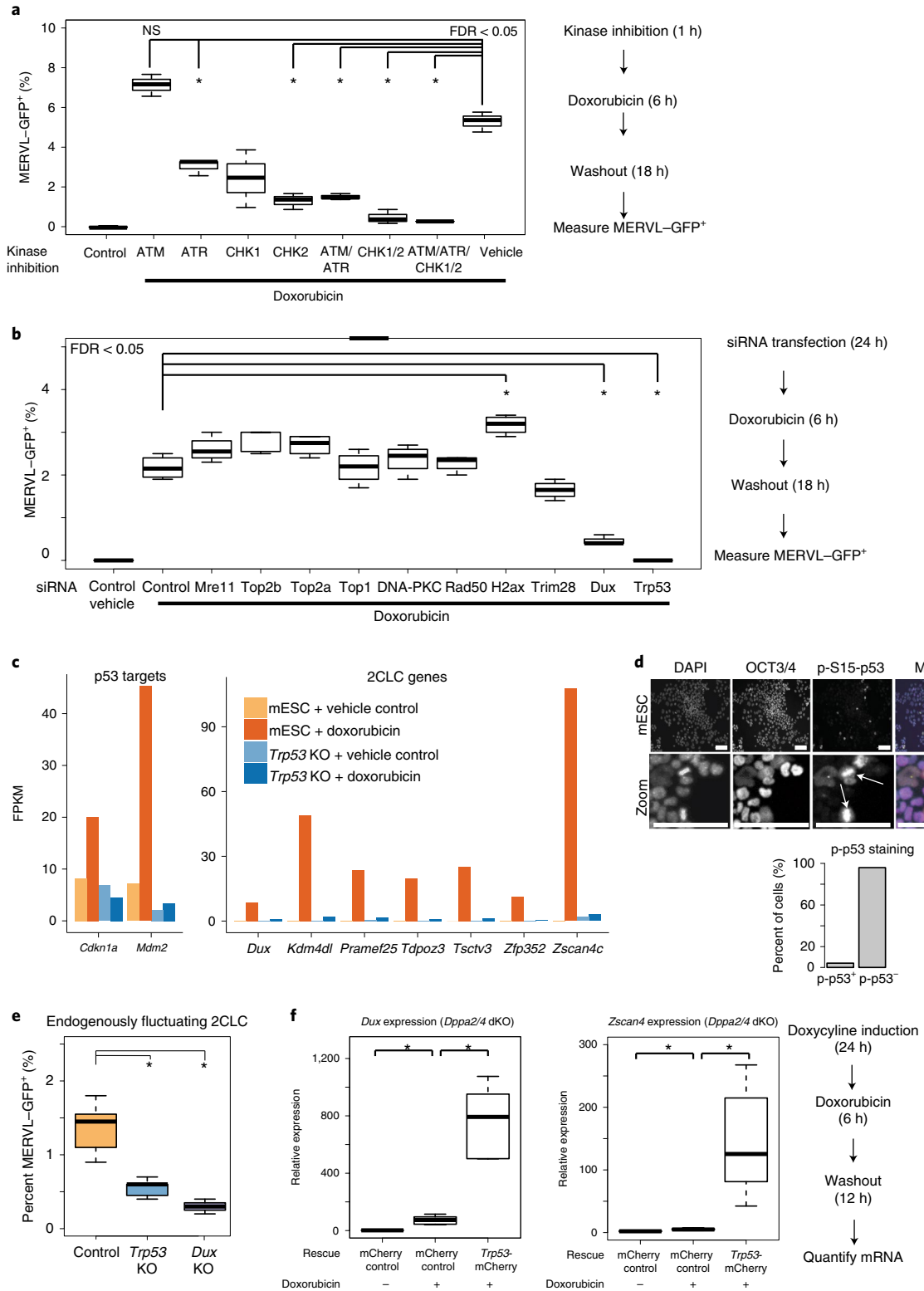
Additionally, we detect strong ZSCAN4 protein binding to a CA repeat ~800 bp upstream of the TSS (Fig. 3a,g). As *Zscan4* is one of the strongest direct targets of DUX[2], it is possible that ZSCAN4 transcriptionally regulates the *Dux* locus, thus providing a feedback loop affecting *Dux* expression. In clonal ZSCAN4-OE mESCs, *Dux* and other direct p53 targets such as *Mdm2* and *Krt5* were significantly downregulated after doxorubicin treatment compared to those of control doxorubicin-treated mESCs (Fig. 3h and Extended Data Fig. 3e). Although we cannot exclude a direct transcriptional repression effect or a competitive-inhibition effect of ZSCAN4 on *Dux* expression, we propose that ZSCAN4 OE prevents DNA damage and/or decreases p53 signaling, consistent with its function in protecting 2C-stage mouse embryos from transcription-induced DNA damage[24] (Extended Data Fig. 3f). Consistent with this, ZSCAN4-OE mESCs exhibited decreased phospho-p53 levels, lower p53-target gene induction and less cell death after doxorubicin treatment (Fig. 3h and Extended Data Fig. 3e–g).

MZ *Trp53*-KO embryos exhibit lower DUX-target gene activation. Next, we asked whether maternally deposited p53 impacts ZGA in vivo. Although some MZ *Trp53*-KO embryos are viable[33],

Fig. 2 | 2CLC emergence and *Dux* expression after doxorubicin treatment require canonical DDR signaling and p53. a, DDR kinase inhibition in mESCs shows that ATR, CHK1 and CHK2 activities are required for MERVL-GFP induction in mESCs after doxorubicin treatment. MERVL-GFP$^+$ cells were quantified by flow cytometry. $n=3$ biological replicates, *FDR < 0.05, one-way ANOVA. NS, not significant. **b**, siRNA-mediated depletion in mESCs demonstrates that *Trp53* is required for MERVL-GFP induction after doxorubicin treatment. MERVL-GFP$^+$ cells were quantified by flow cytometry. $n=3$ biological replicates, *FDR < 0.05, one-way ANOVA. **c**, RNA-seq analysis of *Trp53*-KO mESCs after doxorubicin treatment shows that p53 is required for 2CLC gene induction. WT control or *Trp53*-KO mESC mean RNA-seq data. $n=2$ biological replicates. *FDR < 0.05, DESeq2. **d**, Phospho-S15-p53^{hi} mESCs. WT mESCs were stained with the indicated antibodies, showing mitotic phospho-p53$^+$ cells, indicated by arrows. Mean quantification is shown below, $n=365$ cells. Merge: DAPI, cyan; OCT4, magenta; phospho-p53, yellow. Scale bar, 125 µm. Results are representative from three independent experiments. p, phospho. **e**, Quantification of endogenously fluctuating MERVL-GFP$^+$ 2CLCs in WT, *Dux*-KO or *Trp53*-KO mESCs. $n=8$ biological replicates, *FDR < 0.05, one-way ANOVA. **f**, *Dppa2*- and *Dppa4*-dKO mESCs lose high levels of *Dux* and *Zscan4c* inducibility after doxorubicin treatment, which can be rescued by p53 OE. Quantitative PCR with reverse transcription (RT-qPCR) quantification of $n=3$ biological replicates, normalized to values from *Actb* RNA. *FDR < 0.05, one-way ANOVA. For box plots in **a**,**b**,**e**,**f**, the median is shown as a line in the box, and the outline of the box is depicted at the 25th and 75th percentiles. Extended whiskers depict Q1 − 1.5 × IQR and Q3 + 1.5 × IQR . Outliers are depicted as dots.

there is a neural tube defect resulting in female-specific midgestational loss of *Trp53*-KO embryos[34,35], and it is unclear whether loss of maternal p53 confers defects at earlier stages. Notably, we found enriched phospho-p53 in the paternal pronucleus of zygotes and early 2C-stage mouse embryos (Fig. 4a and Extended Data Fig. 4a), consistent with known high γH2AX levels in mouse zygotes[21,23]. MZ *Trp53*-KO embryos had fewer *Dux* transcripts at the pronuclear 5 (PN5) zygote stage and lower DUX protein expression at the early 2C stage, although MZ *Trp53*-KO embryos still stained with the DUX antibody, indicating a potential compensatory or partly redundant mechanism for *Dux* activation (Fig. 4b and Extended Data Fig. 4b).

We then tested whether MZ *Trp53*-KO embryos exhibited normal ZGA gene expression. Hierarchical clustering revealed a class of transcripts significantly downregulated in MZ *Trp53*-KO embryos ($n = 1{,}056$ decreased of $n = 1{,}971$ differentially expressed



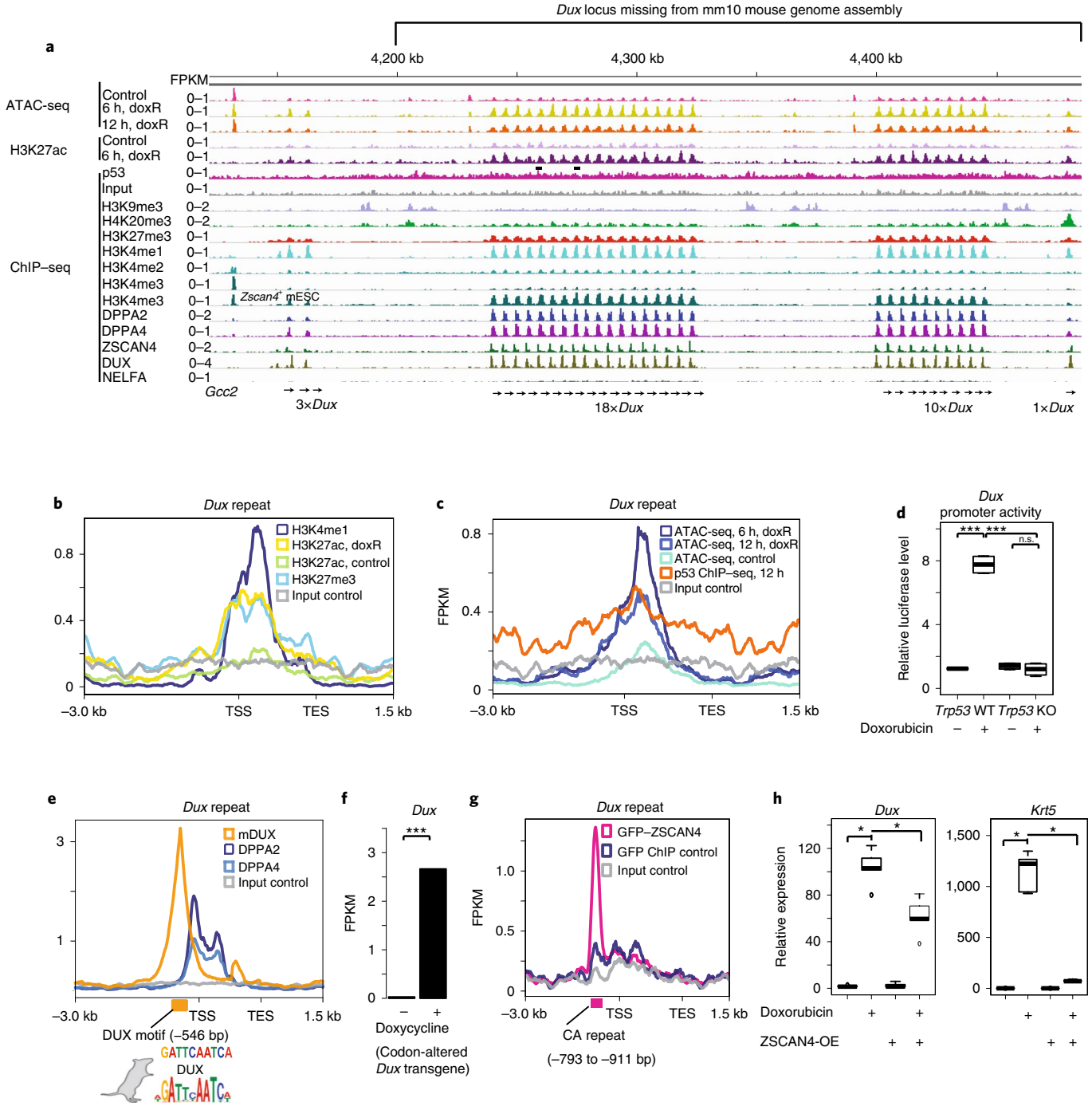


Fig. 3 | Assembly of the mouse *Dux* locus reveals chromatin and transcriptional regulatory features. a, Genome browser snapshots of the newly assembled PacBio-sequenced mouse *Dux* locus with the indicated epigenomic datasets (see Methods for all Gene Expression Omnibus accessions). doxR, doxorubicin. **b**, Metagene plot of the *Dux* repeat unit with the indicated histone ChIP–seq datasets; TES, transcription end site. **c**, Metagene plot of the *Dux* repeat unit with ATAC-seq or p53 ChIP–seq datasets. **d**, Luciferase assay testing a 1.3-kb fragment of the *Dux* promoter in WT or *Trp53*-KO mESCs, with vehicle or doxorubicin treatment. Data were normalized to values from samples cotransfected to express *Renilla* luciferase. $n = 4$ biological replicates, ***FDR < 0.001, one-way ANOVA. **e**, Metagene plot of the *Dux* repeat showing enrichment of DPPA2 or DPPA4 or DUX with the DUX TF motif shown below[2], mDUX, mouse DUX. **f**, Doxycycline induction of the codon-altered *Dux* transgene activates the endogenous *Dux* locus; RNA-seq data were reprocessed from ref. [2]. Mean of $n = 2$ biological replicates, FDR < 0.01, DESeq2. **g**, Metagene plot of the *Dux* array showing enrichment of the ZSCAN4 ChIP–seq signal at the CA repeat unit; data were reprocessed from ref. [24]. **h**, Clonal ZSCAN4-OE mESCs show lower levels of *Dux* and *Krt5* expression after doxorubicin treatment than those of the control. $n = 5$ biological replicates, *FDR < 0.05, one-way ANOVA. For box plots in **d**,**h**, the median is shown as a line in the box, and the outline of the box is depicted at the 25th and 75th percentiles. Extended whiskers depict Q1 − 1.5 × IQR and Q3 + 1.5 × IQR . Outliers are depicted as dots.

ZGA-specific genes) characterized by gene ontology (GO) terms such as p53 signaling, cell cycle checkpoints and transcription (Fig. 4c). Notably, MZ *Trp53*-KO embryos had significantly lower expression levels of direct DUX targets *Zscan4c*, *Kdm4dl*, *Eif1a*-like retrogenes and MERVL[2] (Fig. 4d,e). Thus, p53 loss correlates with transcriptional defects of DUX targets alongside indirect maternal

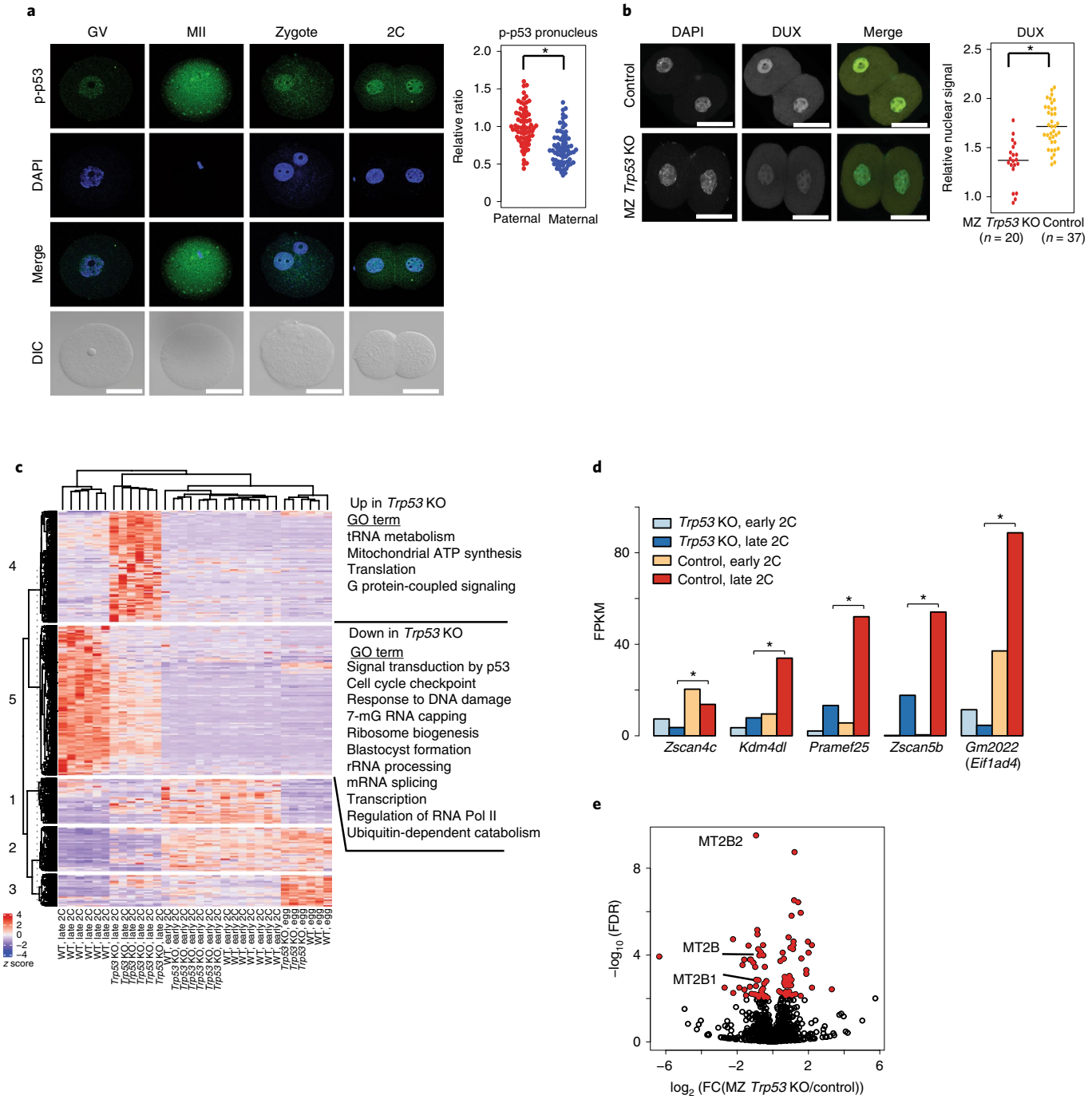


Fig. 4 | p53 accumulates in mouse zygotes after fertilization, and MZ *Trp53*-KO embryos have lower DUX and DUX-target expression. a, Immunostaining of mouse oocytes, eggs and embryos using a phospho-S15 (S18 in mice) p53-specific antibody. Quantification of the significantly higher phospho-p53 signal in the paternal pronucleus is shown on the right, $n = 69$ zygotes, $^*P < 0.05$, one-sided *t*-test, $P = 2.9 \times 10^{-10}$. Scale bar, 40 µm. GV, germinal vesicle stage oocyte; MII, meiosis II egg; DIC, differential interference contrast. Data were combined from three independent experiments. Scale bar, 40 µm. **b**, Immunostaining of MZ *Trp53*-KO early 2C-stage embryos with anti-DUX antibody. Merge: DAPI, cyan; DUX, yellow. Mean quantification is shown on the right, $^*P < 0.05$, one-sided *t*-test, $P = 5.543 \times 10^{-7}$. Scale bar, 40 µm. **c**, RNA-seq analysis of MZ *Trp53*-KO eggs and embryos compared to WT controls. Hierarchical clustering of transcripts from DESeq2 with FDR < 0.01. $n = 6$ embryos, $n = 3$ eggs per genotype per stage. **d**, RNA-seq data from MZ *Trp53*-KO embryos versus WT control embryos; key ZGA genes are shown, *FDR < 0.05, DESeq2. $n = 6$ embryos per genotype per stage. **e**, Repetitive element expression, RNA-seq data from MZ *Trp53*-KO embryos versus WT control embryos at the late 2C stage; red indicates FDR < 0.05, TEtranscripts. $n = 6$ biological replicates per genotype.

mRNA-clearance defects (Fig. 4c), consistent with the effects of siRNA-mediated DUX4 depletion in human embryos[5].

DUX biases mESCs toward extra-embryonic differentiation. We next tested the developmental outcome of 2CLCs that transiently express DUX. Notably, 2CLCs contribute to both embryonic (post-implantation epiblast-derived) and extra-embryonic (primitive endoderm and trophectoderm-derived) tissues in chimeras[7], which is often referred to as 'expanded potential', and single-cell injection of 2CLCs into embryos demonstrates their totipotency[36].

To characterize downstream cellular fates of DUX-expressing mESCs, we pulsed a low dose of DUX in mESCs cultured in 2iLIF (CHIR99021, PD0325901, leukemia inhibitory factor) medium and then withdrew 2iLIF and formed embryoid bodies (EBs), which allows cells to exit pluripotency and differentiate in a non-directed fashion, enabling evaluation of intrinsic developmental potential in an unbiased manner[37,38]. Single-cell RNA-seq (scRNA-seq) of day 9 EBs formed from DUX-pulsed mESCs and control mESCs revealed multiple populations of cells representing endoderm, ectoderm and mesoderm, with high concordance between biological replicates (Fig. 5a and Extended Data Fig. 5a,b).

We then identified cell types that were generated preferentially in DUX-pulsed EBs by quantifying the proportion of cells in each cluster (Fig. 5a,b). Interestingly, cells from DUX-pulsed EBs were enriched in clusters expressing markers from embryonic day (E)8.25 mouse embryos[39]: amnion (cluster 3), cardiac, blood and endothelial (clusters 4, 8 and 9, respectively) and non-post-implantation epiblast derivatives (extra-embryonic ectoderm (cluster 6) and extra-embryonic endoderm (cluster 7)) (Fig. 5b and Extended Data Fig. 5b,c). We further characterized cluster 7, which expressed higher levels of classical extra-embryonic endoderm (ExEnd) genes such as *Trap1a* and *Rhox5*, distinguishing these cells from embryonic endoderm[40] (Fig. 5c,d). We confirmed scRNA-seq data with flow cytometry from EBs using a *Rhox5*-mCherry-knock-in mESC clone (Fig. 5e). We note that p53 and/or p73 activation occurs upon withdrawal of leukemia inhibitory factor during differentiation[41], which may contribute to the low percentage of RHOX5$^+$ cells in control non-DUX-pulsed EBs. Using co-staining with the pan-endoderm marker SOX17, we detected RHOX5$^+$SOX17$^+$ ExEnd cells at low frequencies on the surface of EBs but found that the frequency of RHOX5$^+$SOX17$^+$ cells increased after treatment with fibroblast growth factor (FGF), which promotes ExEnd differentiation in the inner cell mass in vivo[42] (Fig. 5f). Strikingly, we found that DUX-pulsed EBs were covered in an outer layer of RHOX5$^+$ cells and RHOX5$^+$ outgrowths (Fig. 5f), even in the absence of exogenous FGF.

Under normal conditions, mESCs cultured in 2iLIF rarely generate ExEnd, but mESCs treated with retinoic acid and activin can generate 'converted ExEnd endoderm stem cells' (cXEN)[43]. In cXEN differentiation, we found that DUX-pulsed mESCs generated a higher percentage of *Rhox5*-mCherry$^+$ cells, most of which were SOX17$^+$ and GATA6$^+$, markers for endoderm and ExEnd, respectively (Fig. 5g,h). In summary, we use two different assays (EB formation and cXEN differentiation) to demonstrate that DUX-pulsed mESCs exhibit 'expanded-fate' potential, and we speculate that this capacity might impact cell fate after ZGA (Discussion).

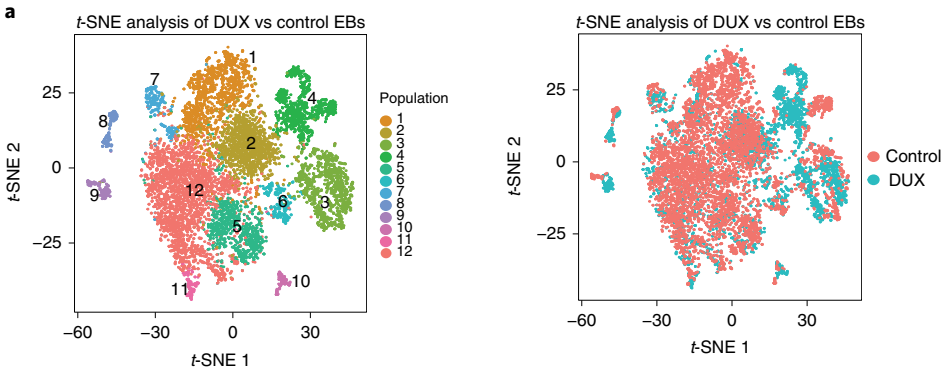
FSHD induced pluripotent stem cells preferentially activate *DUX4* via p53 signaling. Next, we tested whether the regulatory relationships that we found between p53 and mouse *Dux* were preserved at the human *DUX4* locus. Importantly, the mouse *Dux* and human *DUX4* loci are not in syntenic positions and are thought to have been generated through independent retrogene formation[31,44], raising the possibility that their regulation has also evolved independently. Because *D4Z4* contractions of the *4qA* permissive haplotype cause FSHD1 (ref. [15]), we derived independent induced pluripotent stem cell (iPSC) lines from three unrelated patients with FSHD1 to model the unusual repeat-contraction involvement in this disease (Extended Data Fig. 6a,b). Surprisingly, doxorubicin treatment elicited *DUX4* transcriptional induction in FSHD1 iPSCs but not in control non-FSHD iPSCs and less potently in an independent non-FSHD human embryonic stem cell (hESC) line (Fig. 6a and Extended Data Fig. 6c), and this induction required p53 (Fig. 6a,b). Furthermore, the isogenic FSHD1 male iPSC line 1, in which we knocked out *SMCHD1* (which encodes a chromatin protein required for efficient silencing of the *DUX4* locus[17]) in three independent clones, was likewise hyper-responsive to doxorubicin treatment compared to the isogenic *SMCHD1*-WT parental FSHD1 iPSC line (note $\log_{10}$ scale; Fig. 6c and Extended Data Fig. 6d). These data indicate that loss of epigenetic silencing, through *DUX4* repeat contraction or combined *SMCHD1* loss, potentiates the inducibility of the locus by p53 signaling. In sum, these data suggest that the human *DUX4* locus, similar to the mouse *Dux* locus, is induced by DNA damage through p53 and epigenetic de-repression of the *DUX4* locus to create a hyperinducible state.

An LTR10C element is required for full p53 responsiveness of *DUX4*. To test whether the *DUX4* locus is directly regulated by p53, we searched for nearby p53-binding sites. By reanalyzing published data from etoposide-treated human lung fibroblasts[45], nutlin-3-treated human foreskin fibroblasts[46] or doxorubicin-treated hESCs[47], we found a strong p53-binding site between the *DUX4* locus and the *4qA* telomere[45] (Fig. 6d and Extended Data Fig. 6e). This p53-binding site also gained open chromatin in FSHD1 iPSCs after doxorubicin treatment in a p53-dependent manner (Fig. 6e). Remarkably, the p53-binding site resides within an LTR10C element, which is a primate-specific LTR known to be responsive to p53 (Fig. 6e)[48]. Notably, LTR10C generates RNA transcripts at higher levels in human cleavage-stage embryos than in eggs or blastocysts (Fig. 6f) and shows an open chromatin signal in active transposition into active chromatin sequencing (ATAC-seq) datasets from human cleavage-stage embryos (Extended Data Fig. 6e), suggesting contemporaneous activation of LTR10c and *DUX4* at embryonic genome activation. To test whether the LTR10C distal to *DUX4* is p53 responsive, we cloned a 3.1-kb fragment that includes LTR10C to test its activity in luciferase reporter assays. After doxorubicin treatment, the LTR10C directed high levels of expression, comparable to those of the relatively strong p53-binding site near *CDKN1A* (p21) (Fig. 7a and Extended Data Fig. 6f)[27]. Importantly, LTR10C element activity required p53, as *TP53* KO eliminated

Fig. 5 | Transient DUX expression in mESCs confers increased expanded-fate potential. a, scRNA-seq identifies cell populations (clusters 1–12) from differentiated EBs (left), and scRNA-seq data are colored by treatment (right). $n=2$ biological replicates per condition. *t*-SNE, *t*-distributed stochastic neighbor embedding. **b**, Table of scRNA-seq clusters, with putative cell identity using E8.25 mouse embryo scRNA-seq data[34]. $n=2$ biological replicates per condition. N/A, not applicable. **c**, Marker expression for cluster 7 showing canonical ExEnd genes. **d**, Median quantification of expression of *Trap1a* and *Rhox5* (ExEnd) markers in cluster 7, showing higher levels in DUX-pulsed cells. *FDR < 0.05, Seurat. **e**, Quantification of percent RHOX5$^+$ cells in day 8 EBs from a *Rhox5*-mCherry knock-in (KI) mESC clone. $n=15$ biological replicates per condition, $^*P<0.05$, $P=0.04517$, one-sided *t*-test. **f**, Immunofluorescence of EBs stained with anti-SOX17 and anti-RHOX5 antibodies. SOX17$^+$RHOX5$^+$ cells can be induced in WT control EBs treated with FGF. DUX-pulsed EBs show RHOX5$^+$ cells on the outside layer. Representative image from $n=10$ imaged EBs from each experimental condition. Merge: DAPI, cyan; RHOX5, green; SOX17, red. Scale bar, 150 µm. **g**, ExEnd cell (XEN) differentiation of DUX-pulsed or control mESCs and quantification of *Rhox5*-mCherry-knock-in cells by flow cytometry. $^*P<0.05$, $P=0.001301$, one-sided *t*-test, $n=3$ biological replicates per condition. **h**, XEN differentiation of DUX-pulsed mESCs using a *Rhox5*-mCherry-knock-in clone stained with antibodies against the ExEnd marker GATA6 ($n=103$ cells) or the endoderm marker SOX17 ($n=120$ cells). Immunofluorescent images are on the bottom, with quantification of RHOX5$^+$ cells on top. Merge: DAPI, cyan; GATA6 or SOX17, magenta; *Rhox5*-mCherry, yellow. Scale bar, 50 µm. For box plots in **e**,**g**, the median is shown as a line in the box, and the outline of the box is depicted at the 25th and 75th percentiles. The extended whiskers depict Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. Outliers are depicted as dots.

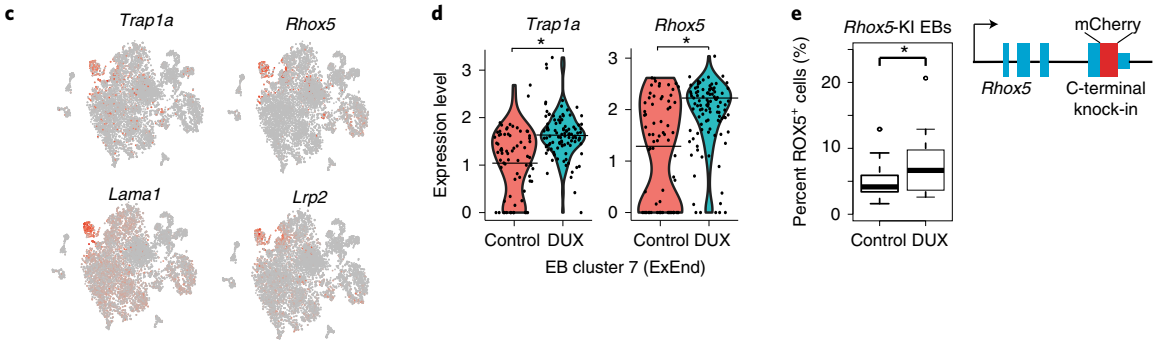
activity (Fig. 7a). We further mapped the minimal region within the 3.1-kb fragment to the ~500-bp LTR10C itself, which contains the p53 motif (Fig. 7a and Extended Data Fig. 6f)[49].

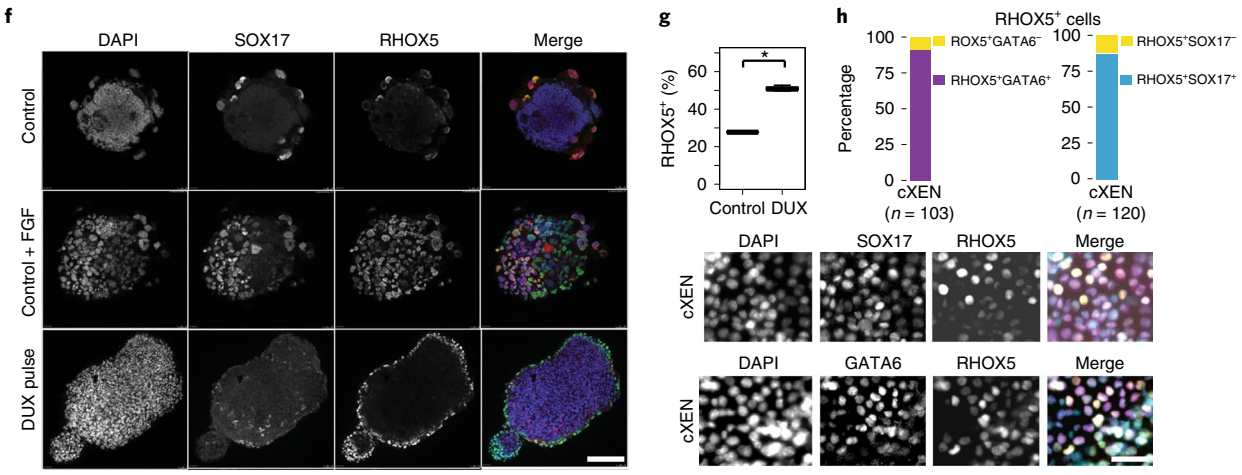
To test whether the p53-binding site within the LTR10C distal to *DUX4* is required for locus inducibility, we used CRIPSR interference to repress the LTR10C. By using dCas9 (nuclease-dead Cas9) fused to a strong transcriptional repressor (KRAB–MeCP2)[50] expressed in FSHD1 iPSCs, we found that repressing the LTR10C significantly reduced *DUX4* inducibility after doxorubicin treatment (Fig. 7b). Importantly, *CDKN1A* activation was largely unaffected, indicating that both experimental groups experienced similar DNA damage and/or p53 activation (Fig. 7b)[45–47]. Thus, we conclude that the



b

Cluster	% control	% DUX	DUX enrichment	Putative cell type	Markers
1	0.201	0.049	0.246	ExEnd	*Spink1, Slc7a7, Fgb, Fxyd2, Gldc*
2	0.147	0.127	0.866	N/A	
3	0.060	0.215	3.580	Amnion	*Pmp22, Postn, Col5a2, Cdh11, Lum, Nnat*
4	0.044	0.196	4.478	Cardiac	*Ttn, Actc1, Myl4, Myh7, Myl3, Tnnt2, Myh6*
5	0.113	0.061	0.537	N/A	
6	0.023	0.043	1.911	ExEnd, placode	*Gjb3, Rhox9, Tfap2c, Gata2, Wnt6, Pramef12*
7	0.023	0.040	1.712	ExEndoderm	*Nostrin, Trap1a, Rhox5, Lama1, Lrp2*
8	0.009	0.036	4.119	Blood, endothelial	*Pf4, Clec1b, Fcer1g, Itga2b, Rgs10*
9	0.010	0.032	3.170	Endothelial	*Icam2, Plvap, Pecam1, Cdh5*
10	0.008	0.026	3.420	Blood	*Hba-x, Hba-a1, Hbb-bh1, Alas2*
11	0.012	0.018	1.556	ExEnd	*Itih2, Paqr9, Ambp*
12	0.352	0.158	0.448	Notochord	*Ly6h, Tff3*





p53-bound LTR10C distal to the *DUX4* locus is required for its full transcriptional activation by p53. As the LTR10C becomes occupied by p53 following doxorubicin treatment even in non-FSHD cells (Fig. 6d and Extended Data Fig. 6e), p53 binding is evidently insufficient to strongly activate the *DUX4* locus; rather, it likely must be combined with the permissive *4qA* haplotype[15] and *D4Z4* contractions or loss of repressors such as SMCHD1 (ref. [16]), which relieve chromatin repression and/or may permit looping between LTR10C and *DUX4* promoters.

Developmental or disease DUX4 regulation in FSHD myoblasts. It is unknown whether the regulatory sequences that activate *Dux* or *DUX4* loci in the early embryo are the same ones that activate the *DUX4* locus in muscle cells during FSHD pathogenesis. Remarkably, we found that myoblasts from five different patients with FSHD1 or FSHD2 (ref. [19]) induced *DUX4* expression after doxorubicin treatment, indicating that the DDR hypersensitivity of the *DUX4* locus is not only present in iPSCs but also in myoblasts, a cell type relevant to the muscle pathology observed in patients with FSHD (Fig. 7c).

Discussion

Previous work identified DUX as a driver of the 2CLC state and of sets of genes and retrogenes during ZGA[2–4], but how *Dux* itself was regulated at the transcriptional level was unknown. In this work, we show a coupling of p53 to *Dux* or *DUX4* activation, potentially uniting the developmental regulation of the *Dux* locus at ZGA or in 2CLCs and the regulation of the *DUX4* locus in patients with FSHD (Fig. 7d).

Our work is in apparent contrast with a recent report[20] claiming that p53 is not required for ATR-dependent replication stress-induced *Dux* activation. However, that work did not use a clonally derived and validated *Trp53*-KO mESC line[20]. Here, our data provide several independent lines of evidence that the majority of the effect on *Dux* induction and DUX-target activation is mediated by p53. Loss-of-function assays using Trp53 siRNA, two independent early-passage *Trp53*-KO clones and *Trp53*-KO mESCs derived from *Trp53*$^{fl/fl}$-KO blastocysts indicate that p53 is required for full *Dux* activation (Fig. 2b,c and Extended Data Fig. 2d). Our work also establishes *Nelfa* as a DUX target and therefore as a marker for the 2CLC state rather than a driver. Additionally, p53 OE strongly rescues the failure of *Dppa2*- and *Dppa4*-dKO cells to induce *Dux* and DUX targets (Fig. 2f and Extended Data Fig. 2b). Although we demonstrate that p53 is a clear *Dux* activator, we cannot rule out a minor p53-independent role in residual *Dux* expression in mESCs and emphasize that our data support both p53-dependent and p53-independent mechanisms for regulating DUX levels in embryos.

What is the benefit of coupling p53 activation to DUX expression? As DNA damage (from paternal DNA demethylation[23], alternative lengthening of telomere elongation[22], replication stress[23] and transcription-mediated DNA damage[24]) and subsequent p53 activation is an intrinsic process of preimplantation development, we speculate that p53 activation could allow the embryo to ensure that a robust DDR is coupled to ZGA and thus development. Critically, early cleavage-stage embryos are not competent to undergo apoptosis[51], which we suggest provides a unique developmental time window to use active p53 signaling without triggering cell death. However, our analysis of MZ *Trp53*-KO embryos indicates that p53 impacts but is not absolutely required for DUX-target expression, highlighting a key difference between the mouse ZGA embryo and 2CLCs. Interestingly, DNA damage was recently shown to control the onset of ZGA in *C. elegans* primordial germ cells[52].

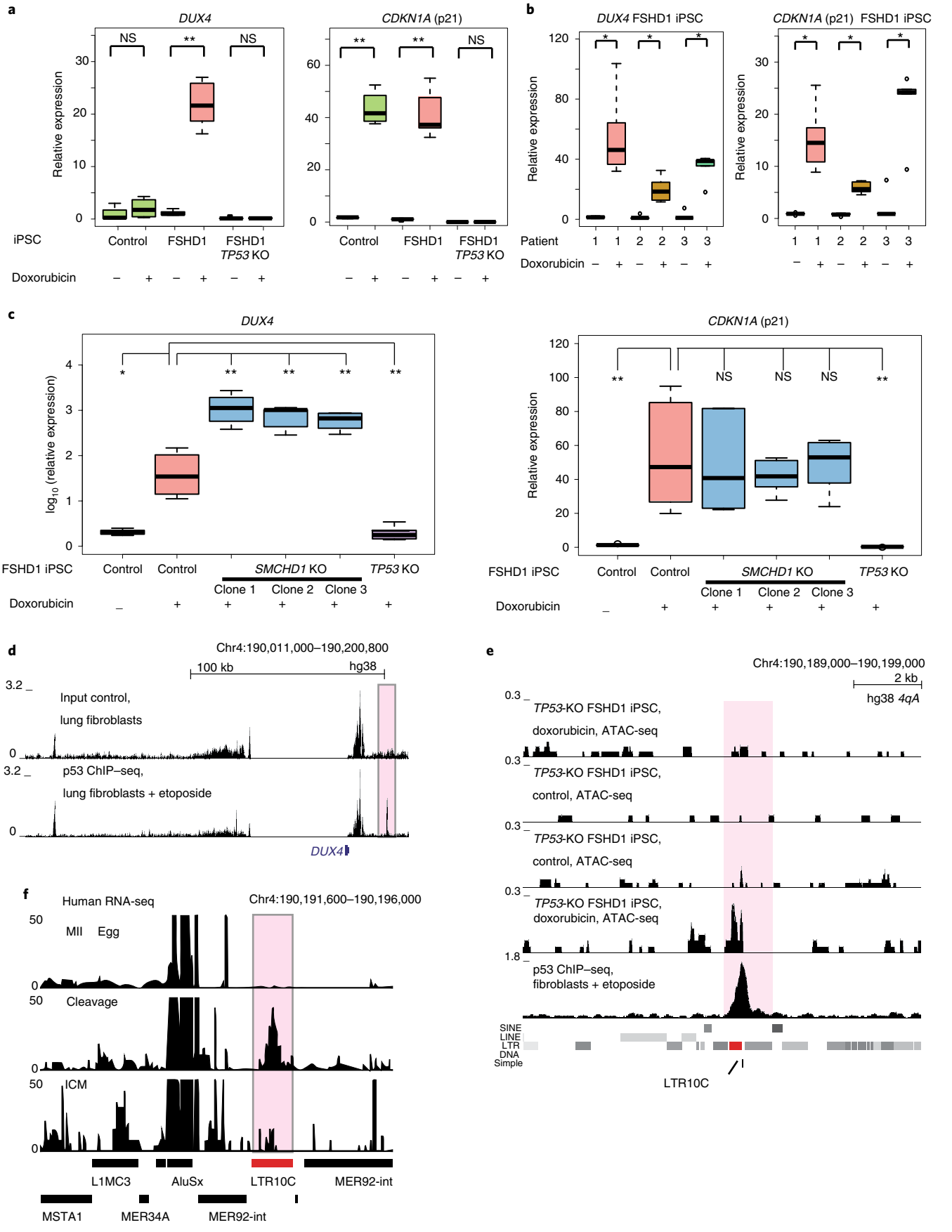
Additionally, DUX is a potent transcriptional activator of *Zscan4*, which prevents excessive transcription-induced DNA damage during ZGA[22,24]. We show that ZSCAN4 binds to the mouse *Dux* array at a CA repeat present in each repeat unit (Fig. 3a,g), and ZSCAN4-OE mESCs have lower *Dux* expression levels after doxorubicin-induced DNA damage (Fig. 3h). However, because ZSCAN4-OE mESCs also have lower phospho-p53 levels, p53 target activation and less cell death after doxorubicin treatment, the decrease in *Dux* expression that we observed in ZSCAN4-OE mESCs is likely an indirect effect on p53 signaling strength (Fig. 3h and Extended Data Fig. 3e–g). Instead, we speculate that ZSCAN4 might bind to the CA repeat at each *Dux* repeat unit and promote locus and/or genomic stability, perhaps by protecting Z-DNA-prone regions (known to mark sites of chromosomal translocations[53]), consistent with ZSCAN4 preventing sister-chromatid exchange[54]. Recombination-mediated expansion underlies gene 'accordions'[55], prompting future work to determine whether and how ZSCAN4 regulates *Dux* via this process.

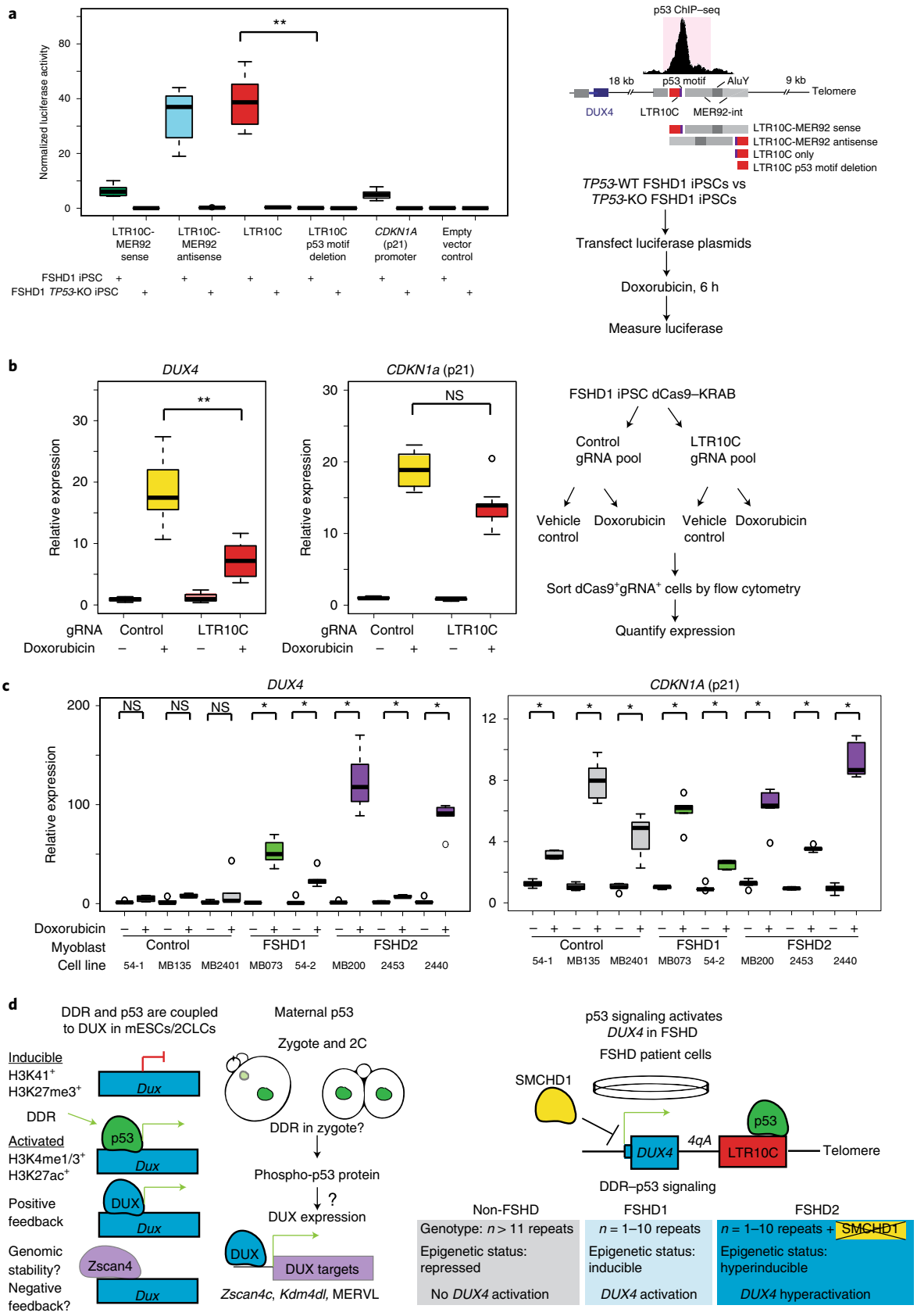
A major implication of our work is that embryonic cells with excessive and/or unrepaired DNA damage may have higher and prolonged DUX expression and therefore adopt a non-post-implantation epiblast fate, which may serve to protect the soma or the germline from damaged cells. Consistent with a 'quality control' function for DUX-conferred expanded fate is the finding that *Trp53*-KO tetraploid cells contribute to the embryo, while almost all tetraploid cells from *Trp53*-WT embryos are excluded from the embryo proper and instead contribute to extra-embryonic tissues[56]. Whether *DUX* reactivation in vivo occurs outside of ZGA and confers a similar 'expanded' potential fate will be the focus of future work. As expanded fate of mESCs is observed after poly(ADP-ribose) polymerase (PARP)1 inhibition[57] or loss of H2AX[58], and both conditions

Fig. 6 | p53 is required for *DUX4* activation after DNA damage in FSHD cells. a, In FSHD1 iPSCs, *DUX4* is hypersensitive to doxorubicin treatment, and this effect requires p53. Non-FSHD control iPSCs (male WT33), FSHD1 iPSCs (male patient 1) or isogenic FSHD1 cells with *TP53* KO from patient 1 (FSHD1) were treated with doxorubicin or the vehicle control; *DUX4* expression and *CDKN1A* (p21) expression were quantified by RT–qPCR and normalized to values from 18S rRNA. $n = 5$ biological replicates, *FDR < 0.05, one-sided *t*-test. **b**, Additional FSHD1 iPSC lines also show *DUX4* induction after doxorubicin treatment. Line 1 is an additional clone of patient 1 (shown in Fig. 5a), and lines 2 and 3 are from unrelated female patients with FSHD1, normalized to values from 18S rRNA. Line 1, six biological replicates; lines 2 and 3, five biological replicates. *FDR < 0.05, one-sided *t*-test. **c**, Loss of *SMCHD1* in male FSHD1 iPSCs leads to non-additive activation of *DUX4* expression. FSHD1 iPSC patient 1 line or isogenic FSHD1 and *SMCDH1*-KO iPSCs from the same patient 1 with FSHD1 (three independent *SMCHD1*-KO clones) or *TP53*-KO FSHD iPSCs from the same patient 1 with FSHD1 were treated with doxorubicin or the vehicle control, and *DUX4* or *CDKN1a* (p21) expression was quantified by RT–qPCR and normalized to values from 18S rRNA. $n = 6$ biological replicates, *$P < 0.05$, one-sided *t*-test. **d**, A strong, subtelomeric p53-binding site 18 kb from the *DUX4* locus. Public data from human lung fibroblasts treated with etoposide were used for p53 ChIP–seq[39]. **e**, Comparison of the subtelomeric p53-binding site that overlaps an LTR10C element shows that it gains open chromatin in a doxorubicin-treatment- and p53-dependent manner in iPSCs from patient 1 (FSHD1) (top four tracks of ATAC-seq data). $n = 2$ biological replicates per condition per genotype. LINE, long interspersed nuclear element; SINE, short interspersed nuclear element. **f**, RNA-seq coverage from human MII-stage eggs, cleavage-stage embryos or isolated inner cell mass (ICM) shows that LTR10C generates RNA transcripts. Data were reanalyzed from ref. [2]. For **a**–**c**, the median is shown as a line in the box, and the outline of the box is depicted at the 25th and 75th percentiles. Extended whiskers depict Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. Outliers are depicted as dots.

inhibit DNA-break repair and prolong DDR or p53 signaling, the coupling of DDR or p53 activation to DUX activation may be a unifying theme for conferring expanded-fate potential to mESCs.

We note three other reported links of p53 to 2CLC induction. First, *Mir34a*-KO mESCs have a higher percentage of stochastically fluctuating 2CLCs[36]. As *Mir34a* is a direct target of p53 (ref. [59]),





loss of the tumor suppressor *Mir34a* may allow p53^{hi} mESCs to survive, thus allowing 2CLC accumulation in culture. Second, depletion of UBC9, an E2 enzyme required for sumoylation, was also shown to activate 2CLCs[60]. Interestingly, p53 is also sumoylated by the enzyme UBC9 (ref. [61]), and this sumoylation can decrease p53 activity[62]. Therefore, loss of UBC9 could indirectly activate *Dux* through p53 sumoylation or activity, in addition to the reported bona fide chromatin targets of sumoylation[60]. Third, loss of the

Fig. 7 | The subtelomeric p53-binding site at the LTR10C is required for full activation of the *DUX4* locus, and myoblasts from patients with FSHD are similarly hypersensitive to DNA damage. a, Luciferase reporter assay on deletion analysis of the ~3-kb LTR10C-MER92 fragment near the *DUX4* locus identifies the LTR10C sequence as sufficient for doxorubicin-treatment-responsive activity in a p53-dependent manner, and deletion of the p53-binding site (motif) within LTR10C abolishes the activity of this element. $n = 8$ biological replicates of control FSHD1 iPSCs or isogenic *TP53*-KO FSHD1 iPSCs both from patient 1 (FSHD), *FDR < 0.05, one-sided *t*-test. Values were normalized to those from samples cotransfected to express the *Renilla* luciferase plasmid. **b**, CRISPR interference analysis using dCas9–KRAB–MeCP2-mediated repression of the LTR10C in iPSCs from patient 1 (FSHD1) indicates that it is required for full activation of *DUX4* expression. RT–qPCR quantification of the indicated genes after doxorubicin or vehicle treatment of FSHD iPSCs, normalized to values from 18S rRNA. $n = 9$ biological replicates; **$P < 0.005$, *DUX4* expression, $P = 4.958 \times 10^{-5}$; p21 expression, $P = 0.7343$; one-sided *t*-test. **c**, FSHD1 and FSHD2 myoblasts are hypersensitive to doxorubicin treatment compared to non-FSHD control myoblasts. RT–qPCR analysis of non-FSHD control (54-1, 135, 2401), FSHD1 (073, 54-2) or FSHD2 (MB200, 2453, 2440) myoblasts treated with doxorubicin or vehicle, normalized to values from 18S rRNA. $n = 5$ biological replicates, *FDR < 0.02, one-sided *t*-test. **d**, Model figure integrating p53 activation of *Dux* or *DUX4* in mouse embryos, 2CLCs and cells from patients with FSHD. While control cell lines with non-contracted *D4Z4* arrays are refractory to activation by p53, *D4Z4* contractions in patients with FSHD1 confer p53 inducibility on the *DUX4* alleles in FSHD1 cells, and loss-of-function mutations in *SMCHD1* confer p53 hyperinducibility on *DUX4* alleles in FSHD1 cells. For **a**–**c**, the median is shown as a line in the box, and the outline of the box is depicted at the 25th and 75th percentiles. Extended whiskers depict Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. Outliers are depicted as dots.

N^6-methyladenosine reader protein YTHDC1 results in *Dux* activation and increased numbers of 2CLCs, which was attributed to long interspersed nuclear element 1-mediated *Dux* induction[63]. However, our reanalysis of *Ythdc1*-KO mESC RNA-seq[63] data reveals strong induction of direct p53 targets (*Mdm2*, *Cdkn1a* (p21) and *Krt5*) (Extended Data Fig. 6g), suggesting p53 activation of *Dux* as a likely unifying theme in the biology of 2CLCs.

Although p53 activates both mouse *Dux* and human *DUX4*, the molecular mechanism in mouse and human cells is nevertheless quite different. While mouse *Dux* has a p53-reponsive promoter (Fig. 3d), the human *DUX4* locus is regulated by a p53-bound LTR10C distal enhancer (Figs. 6d,e and 7b). Because the LTR10C was acquired in the primate lineage and is absent in rodents, this implies that the mouse *Dux* and the human *DUX4* loci evolved p53 regulation convergently. Characterization of the intron-containing progenitor *DUXC* array could help clarify whether p53 regulation is an ancestral property that was supplanted by the LTR10C enhancer in primates or whether rodent and primate lineages independently evolved this regulatory feature.

Our work also has implications for the mechanistic understanding of FSHD etiology. Although there is a model of how DUX4 expression leads to downstream cell death[14,18], it was previously unclear which transcriptional activators control *DUX4* expression. Here, we discovered that, in FSHD1 and FSHD2 cells, *DUX4* is hypersensitive to p53 activation, while *DUX4* from non-FSHD cells is not, likely due to the relief of chromatin or epigenetic silencing only in FSHD genotypes (Fig. 7d), potentially explaining stochastic *DUX4* expression in muscle cells from patients with FSHD, which are known to exhibit DNA damage[64]. Importantly, p53 is necessary for regulated myogenesis, as p53 activation during muscle differentiation is required for balance between quiescence and differentiation[65]. Future work will be needed to identify the contexts that lead to p53 activation in patients with FSHD and whether effective interventions (such as CHK1 inhibitors[66]) can be pursued to mitigate *DUX4* activation and FSHD pathogenesis in humans.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00893-0.

Received: 1 May 2020; Accepted: 1 June 2021;
Published online: 15 July 2021

References

1. Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic genome activation in vertebrates. *Dev. Cell* **42**, 316–332 (2017).
2. Hendrickson, P. G. et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
3. De Iaco, A. et al. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **49**, 941–945 (2017).
4. Iaco, A. D., Verp, S., Offner, S., Grun, D. & Trono, D. DUX is a non-essential synchronizer of zygotic genome activation. *Development* **147**, dev177725 (2020).
5. Vuoristo, S. et al. DUX4 regulates oocyte to embryo transition in human. Preprint at *bioRxiv* https://doi.org/10.1101/732289 (2019).
6. Chen, Z. & Zhang, Y. Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nat. Genet.* **51**, 947–951 (2019).
7. Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
8. Ishiuchi, T. et al. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* **22**, 662–671 (2015).
9. Percharde, M. et al. A LINE1–Nucleolin partnership regulates early development and ESC identity. *Cell* **174**, 391–405 (2018).
10. Rodriguez-Terrones, D. et al. A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat. Genet.* **50**, 106–119 (2018).
11. Eckersley-Maslin, M. et al. Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes Dev.* **33**, 194–208 (2019).
12. Hu, Z. et al. Maternal factor NELFA drives a 2C-like state in mouse embryonic stem cells. *Nat. Cell Biol.* **22**, 175–186 (2020).
13. Dixit, M. et al. *DUX4*, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of *PITX1*. *Proc. Natl Acad. Sci. USA* **104**, 18157–18162 (2007).
14. Himeda, C. L. & Jones, P. L. The genetics and epigenetics of facioscapulohumeral muscular dystrophy. *Annu. Rev. Genomics Hum. Genet.* **20**, 265–291 (2019).
15. Lemmers, R. J. L. F. et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* **329**, 1650–1653 (2010).
16. Blewitt, M. E. et al. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* **40**, 663–669 (2008).
17. Lemmers, R. J. L. F. et al. Digenic inheritance of an *SMCHD1* mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* **44**, 1370–1374 (2012).
18. Shadle, S. C. et al. DUX4-induced dsRNA and *MYC* mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy. *PLoS Genet.* **13**, e1006658 (2017).
19. Campbell, A. E. et al. NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. *eLife* **7**, e31023 (2018).
20. Atashpaz, S. et al. ATR expands embryonic stem cell fate potential in response to replication stress. *eLife* **9**, e54756 (2020).
21. Ziegler-Birling, C., Helmrich, A., Tora, L. & Torres-Padilla, M.-E. Distribution of p53 binding protein 1 (53BP1) and phosphorylated H2A.X during mouse preimplantation development in the absence of DNA damage. *Int. J. Dev. Biol.* **53**, 1003–1011 (2009).
22. Liu, L. et al. Telomere lengthening early in development. *Nat. Cell Biol.* **9**, 1436–1441 (2007).

23. Wossidlo, M. et al. Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *EMBO J.* **29**, 1877–1888 (2010).
24. Srinivasan, R. et al. Zscan4 binds nucleosomal microsatellite DNA and protects mouse two-cell embryos from DNA damage. *Sci. Adv.* **6**, eaaz9115 (2020).
25. Storm, M. P. et al. Zscan4 is regulated by PI3-kinase and DNA-damaging agents and directly interacts with the transcriptional repressors LSD1 and CtBP2 in mouse embryonic stem cells. *PLoS ONE* **9**, e89821 (2014).
26. Britton, S., Coates, J. & Jackson, S. P. A new method for high-resolution imaging of Ku foci to decipher mechanisms of DNA double-strand break repair. *J. Cell Biol.* **202**, 579–595 (2013).
27. Fischer, M. Census and evaluation of p53 target genes. *Oncogene* **36**, 3943–3956 (2017).
28. Blackford, A. N. & Jackson, S. P. ATM, ATR, and DNA-PK: the trinity at the heart of the DNA damage response. *Mol. Cell* **66**, 801–817 (2017).
29. Eckersley-Maslin, M. A. et al. MERVL/Zscan4 network activation results in transient genome-wide DNA demethylation of mESCs. *Cell Rep.* **17**, 179–192 (2016).
30. Eckersley-Maslin, M. A. et al. Epigenetic priming by Dppa2 and 4 in pluripotency facilitates multi-lineage commitment. *Nat. Struct. Mol. Biol.* **27**, 696–705 (2020).
31. Clapp, J. et al. Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **81**, 264–279 (2007).
32. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
33. Hu, W., Feng, Z., Teresky, A. K. & Levine, A. J. p53 regulates maternal reproduction through LIF. *Nature* **450**, 721–724 (2007).
34. Sah, V. P. et al. A subset of p53-deficient embryos exhibit exencephaly. *Nat. Genet.* **10**, 175–180 (1995).
35. Delbridge, A. R. D. et al. Loss of p53 causes stochastic aberrant X-chromosome inactivation and female-specific neural tube defects. *Cell Rep.* **27**, 442–454 (2019).
36. Hoi, Y. J. et al. Deficiency of microRNA *miR-34a* expands cell fate potential in pluripotent stem cells. *Science* **355**, aag1927 (2017).
37. De Los Angeles, A. et al. Hallmarks of pluripotency. *Nature* **525**, 469–478 (2015).
38. Behringer, R., Gertsenstein, M., Nagy, K. V. & Nagy, A. Differentiating mouse embryonic stem cells into embryoid bodies by hanging-drop cultures. *Cold Spring Harb. Protoc.* https://doi.org/10.1101/pdb.prot092429 (2016).
39. Ibarra-Soria, X. et al. Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. *Nat. Cell Biol.* **20**, 127–134 (2018).
40. Chan, M. M. et al. Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019).
41. He, H. et al. p53 and p73 regulate apoptosis but not cell-cycle progression in mouse embryonic stem cells upon DNA damage and differentiation. *Stem Cell Rep.* **7**, 1087–1098 (2016).
42. Yamanaka, Y., Lanner, F. & Rossant, J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**, 715–724 (2010).
43. Niakan, K. K., Schrode, N., Cho, L. T. Y. & Hadjantonakis, A.-K. Derivation of extraembryonic endoderm stem (XEN) cells from mouse embryos and embryonic stem cells. *Nat. Protoc.* **8**, 1028–1041 (2013).
44. Leidenroth, A. et al. Evolution of *DUX* gene macrosatellites in placental mammals. *Chromosoma* **121**, 489–497 (2012).
45. Catizone, A. N., Good, C. R., Alexander, K. A., Berger, S. L. & Sammons, M. A. Comparison of genotoxic versus nongenotoxic stabilization of p53 provides insight into parallel stress-responsive transcriptional networks. *Cell Cycle* **18**, 809–823 (2019).
46. Karsli Uzunbas, G., Ahmed, F. & Sammons, M. A. Control of p53-dependent transcription and enhancer activity by the p53 family member p63. *J. Biol. Chem.* **294**, 10720–10736 (2019).
47. Akdemir, K. C. et al. Genome-wide profiling reveals stimulus-specific functions of p53 during differentiation and DNA damage of human embryonic stem cells. *Nucleic Acids Res.* **42**, 205–223 (2014).
48. Wang, T. et al. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc. Natl Acad. Sci. USA* **104**, 18613–18618 (2007).
49. Tutton, S. et al. Subtelomeric p53 binding prevents accumulation of DNA damage at human telomeres. *EMBO J.* **35**, 193–207 (2016).
50. Yeo, N. C. et al. An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat. Methods* **15**, 611–616 (2018).
51. Hardy, K. Apoptosis in the human embryo. *Rev. Reprod.* **4**, 125–134 (1999).
52. Butuči, M., Williams, A. B., Wong, M. M., Kramer, B. & Michael, W. M. Zygotic genome activation triggers chromosome damage and checkpoint signaling in *C. elegans* primordial germ cells. *Dev. Cell* **34**, 85–95 (2015).
53. Wang, G., Christensen, L. A. & Vasquez, K. M. Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc. Natl Acad. Sci. USA* **103**, 2677–2682 (2006).
54. Zalzman, M. et al. Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature* **464**, 858–863 (2010).
55. Elde, N. C. et al. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. *Cell* **150**, 831–841 (2012).
56. Horii, T. et al. p53 suppresses tetraploid development in mice. *Sci. Rep.* **5**, 8907 (2015).
57. Yang, Y. et al. Derivation of pluripotent stem cells with in vivo embryonic and extraembryonic potency. *Cell* **169**, 243–257 (2017).
58. Wu, T. et al. Histone variant H2A.X deposition pattern serves as a functional epigenetic mark for distinguishing the developmental potentials of iPSCs. *Cell Stem Cell* **15**, 281–294 (2014).
59. Rokavec, M., Li, H., Jiang, L. & Hermeking, H. The p53/miR-34 axis in development and disease. *J. Mol. Cell Biol.* **6**, 214–230 (2014).
60. Cossec, J.-C. et al. SUMO safeguards somatic and pluripotent cell identities by enforcing distinct chromatin states. *Cell Stem Cell* **23**, 742–757 (2018).
61. Kahyo, T., Nishida, T. & Yasuda, H. Involvement of PIAS1 in the sumoylation of tumor suppressor p53. *Mol. Cell* **8**, 713–718 (2001).
62. Association of Ubc9, an E2 ligase for SUMO conjugation, with p53 is regulated by phosphorylation of p53. *FEBS Lett.* **573**, 15–18 (2004).
63. Liu, J. et al. The RNA m6A reader YTHDC1 silences retrotransposons and guards ES cell identity. *Nature* **591**, 322–326 (2021).
64. Sasaki-Honda, M. et al. A patient-derived iPSC model revealed oxidative stress increases facioscapulohumeral muscular dystrophy-causative DUX4. *Hum. Mol. Genet.* **27**, 4024–4035 (2018).
65. Le Roux, I., Konge, J., Le Cam, L., Flamant, P. & Tajbakhsh, S. Numb is required to prevent p53-dependent senescence following skeletal muscle injury. *Nat. Commun.* **6**, 8528 (2015).
66. Dent, P. et al. CHK1 inhibitors in combination chemotherapy. *Mol. Interv.* **11**, 133–140 (2011).

Methods

This research complies with all relevant ethical regulations. Animal experiments performed at the University of Utah were first granted IACUC and Institutional approval under protocol 18-07014. Informed consent was obtained from all human participants and overseen by the University of Utah Institutional Review Board (IRB)-approved protocol 40092. No compensation was provided to participants. Written informed consent was obtained from all participating individuals under a University of Utah IRB-approved protocol (IRB registration 30923).

Mouse embryonic stem cell culture. Mycoplasma-free E14 mESCs were cultured on gelatin in 2i+LIF medium containing Gibco KO-DMEM with nonessential amino acids, 2-mercaptoethanol and dipeptide glutamine and were supplemented with 15% ESC-grade FBS, leukemia inhibitory factor (Thermo Fisher), 1 mM PD0325901 (Sigma-Aldrich) and 3 mM CHIR99021 (Sigma-Aldrich). Stable cell lines were selected using puromycin (Thermo Fisher Scientific, A11138-03) at 0.5–1 $\mu g\,ml^{-1}$, geneticin (Life Technologies, 10131-035) at 100–200 $\mu g\,ml^{-1}$ or blasticidin (Fisher Scientific, B12150-0.1) at 1.5–3 $\mu g\,ml^{-1}$.

Transfection of mESCs. mESCs were transfected in Opti-MEM medium (Thermo Fisher Scientific, 31985070) using RNAiMAX (Thermo Fisher Scientific, 13778150) for siRNA or Lipofectamine 3000 (Thermo Fisher Scientific, L3000-015) for DNA transfections.

siRNA transfections for mESCs. siRNA pools were generated with *Giardia* Dicer. Briefly, primers were designed to amplify two ~100–300-bp fragments of the indicated ORF from genomic mouse DNA and to add T7 handles (Supplementary Table 1). Purified PCR products were then used as a template for in vitro transcription with a MEGAscript T7 Transcription kit (Thermo Fisher, AM1334). Template DNA was then degraded, and the dsRNA was allowed to anneal before dicing. Diced siRNA was purified with a PureLink Micro-to-Midi Total RNA purification kit (Invitrogen, 12183-018) with modifications. siRNA concentration was measured with a Qubit RNA HS Assay kit (Thermo Fisher, Q32852). mESCs containing the MERVL–GFP reporter were transfected with 20 pmol (10 pmol of each) total siRNA with RNAiMAX (Life Technologies). After siRNA transfection (24 h), mESCs were treated with doxorubicin or vehicle control for 6 h and then washed, and the medium was replaced with 2iLIF for 18 h. MERVL–GFP$^+$ cells were measured as described below.

Human pluripotent stem cell culture. Human iPSCs or ESCs were maintained on Matrigel (BD Corning Matrigel Matrix, VWR, 356231)-coated plates using StemMACS iPS-Brew (Miltenyi Biotec, 130-104-368) and passaged using ReLeSR (Stemcell Technologies, 05872).

Transfection of human pluripotent stem cells. Cells were transfected in Opti-MEM supplemented with ROCK-i (Selleck Chemicals, S1049) at 5 µM with FuGENE 6 (Promega, E2691) for 1–2 h, and then hPSC medium was added.

Human fibroblast collection. Dermal fibroblasts were obtained from skin biopsies taken from patients with FSHD under the University of Utah IRB-approved protocol 40092. Fibroblasts were expanded in DMEM, high glucose (Thermo Fisher Scientific, 11965118), supplemented with nonessential amino acids, 2-mercaptoethanol and dipeptide glutamine and were supplemented with 20% ESC-grade FBS. Skin punch biopsies from patients diagnosed with FSHD were dissociated with sterile forceps and scalpels and placed on gelatin-coated six-well plates in high-glucose-containing DMEM with 20% FBS. Primary outgrowth was monitored for 4 d, and, subsequently, medium was changed every 2–3 d until fibroblasts cultures were visible and ~90% confluent. Fibroblasts were then passaged at a ratio of 1:6 onto gelatin-coated six-well plates. Primary fibroblast cultures were passaged at 70% confluence four to five times before iPSC reprogramming.

Human induced pluripotent stem cell reprogramming. Patient fibroblasts were grown on gelatin-coated six-well plates to roughly 50% confluence before reprogramming. Sendai-virus-mediated iPSC reprogramming was performed using the CytoTune-iPS 2.0 Sendai Reprogramming kit from Thermo Fisher (A16518) according to the manufacturer's instructions for reprogramming fibroblasts. Independent colonies were picked to establish clonal cell lines from each patient, which were assessed for pluripotency by immunofluorescence, as well as EB formation. iPSC lines were karyotyped using GeneChip and KaryoStat assays from Thermo Fisher. All lines tested were karyotypically euploid, other than a 105.274-kb duplication of chr1:q21.1 in one patient's iPSCs (Extended Data Fig. 6b). Multiple independently reprogrammed iPSC clones from this patient's fibroblast cultures exhibited this abnormality, strongly suggesting that this alteration was not a reprogramming artifact and was instead a pre-existing chromosomal abnormality in the skin sample from this patient.

Reprogrammed iPSC lines were subsequently grown on vitronectin (Gibco, A14700)-coated six-well plates in Essential 8 medium (Thermo Fisher, A1517001), with medium changes every day and passaging every 3–4 d, on average, and then adapted to iPS-Brew medium and Matrigel coating. Pluripotency marker staining for OCT4 and SOX2 was performed (Extended Data Fig. 5a).

Human myoblast culture. Cell lines and their *D4Z4* repeat-contraction status are described elsewhere[19]. Myoblasts were maintained in Ham's F-10 Nutrient Mix (Gibco) supplemented with 20% HyClone FBS (GE Healthcare Life Sciences), 100 U per 100 µg penicillin–streptomycin (Gibco), 10 $ng\,ml^{-1}$ recombinant human basic FGF (Promega) and 1 µM dexamethasone (Sigma-Aldrich, D4902-25MG).

Drug treatment of cells. Small molecules were dissolved in DMSO or water. Zeocin (InvivoGen, ant-zn-05) and doxorubicin (Selleck, S1208) were used at 1 µM; hydroxyurea (VWR, IC10202325) was used at 5 mM; aphidicolin (Abcam, ab142400) was used at 6 µM; hydrogen peroxide (Sigma, 216763) was used at 10 µM; ATM-I (KU-55933, Selleck, S1092) was used at 10 µM; ATR-I (Selleck, VE-821, S8007) was used at 10 µM; CHK1-I (LY2603618, Selleck, S2626) and CHK2-I (Selleck, MK-8776 SCH 900776) were used at 10 µM; doxycycline (Clontech, 631311) was used at 0.25–2 $\mu g\,ml^{-1}$. For inhibition of ATM, ATR, CHK1 and CKH2 in mESCs, cells were pretreated with inhibitors or vehicle for 1 h and then treated with doxorubicin at 1 µM in combination with ATR, ATM, CHK1 and CHK2 inhibitors for 6 h and then washed, and these drugs were replaced either with kinase inhibitors or vehicle control for 18 h.

Anti-DUX antibody production. We used the KLH-conjugated peptide PQEEAGSTGMDTSSPSD for rabbit injections using Covance. Serum was purified using SulfoLink Coupling Resin (Thermo Fisher) following the manufacturer's instructions. The antigen-purified fraction was evaluated on an SDS–PAGE gel for purity and dialyzed against PBS with 0.1% Tween-20.

Flow cytometry. A BD LSRFortessa instrument with lasers for 351 nm, 488 nm, 561 nm and 640 nm was used to quantify MERVL–GFP$^+$ mESCs. Live-versus-dead discrimination was routinely performed using propidium iodide (Invitrogen, P1304MP), and samples were gated using forward FSA and side-scatter SSA to isolate cells from debris, and then double discrimination was performed using FSH × FSW and SSH × SSW. Flow cytometry sorting was performed on a BD Aria instrument with comparable gating strategies to those used for flow cytometry. Purity checks were performed after sorting, and samples were routinely >97% pure.

DNA damage using PpoI endonuclease. Stable cell lines using the PpoI coding sequence from Addgene plasmid 46963 or mCherry control Piggyback vectors were created by transfecting E14 mESCs and selecting with 3 $\mu g\,ml^{-1}$ blasticidin for 7 d. Resultant polyclonal cell cultures were induced using doxycycline at 2 $\mu g\,ml^{-1}$ for 24 h before quantification using flow cytometry.

dCas9–KRAB-mediated repression of the LTR10C in human iPSCs. Transgenic cells from patient 1 (FSHD1) were first created using integration of the sequence for a constitutively expressed GFP and the dCas9-KRAB Piggyback plasmid. After antibiotic selection of these cells with 0.5 $\mu g\,ml^{-1}$ puromycin and expansion for 2 weeks, we transfected cells with Piggyback plasmids containing an mCherry cassette and encoding gRNA species targeting the LTR10C locus or control gRNA. After selection with blasticidin at 1.2 $\mu g\,ml^{-1}$ and 1 week of expansion in culture, we obtained a mixed population of cells. This mixed population of cells was used for doxorubicin treatment, and, after 6 h of treatment with 1 µM doxorubicin, we sorted double-positive GFP$^+$mCherry$^+$ cells by flow cytometry from either LTR10C gRNA$^+$dCas9–KRAB$^+$ cells or gRNA control$^+$dCas–KRAB$^+$ cells. RNA was extracted using TRIzol (Thermo Fisher Life Sciences), and RT–qPCR was performed as described for other human cells in the paper measuring *DUX4* transcript levels and normalizing to values from 18S rRNA.

CRISPR activation using dCas9–p300HAT or dCas9–VP160. For Extended Data Fig. 2b, E14 mESCs were transfected with Piggyback transgenes on plasmids encoding eIF1a–GFP and eIF1a–dCas9–p300 or eIF1a–GFP and eIF1a–dCas9–VP160. After selection with 3 $\mu g\,ml^{-1}$ blasticidin for 7 d, cells were expanded and transfected with a plasmid encoding either a control gRNA or gRNA species (two to three gRNA species per regions 1–4 in the *Dux* locus linker or three gRNA species targeting the *Dux* promoter). After transfection (36 h), cells were collected and treated with TRIzol (Thermo Fisher), and RNA was purified and subjected to reverse transcription as described above, and qPCR was performed (as described above) to quantify the levels of *Dux* RNA normalized to those of *Actin* (as described above).

CRISPR knockout and knock-in clone screening. Please see Supplementary Methods.

PCR. PCR primers for genotyping *Trp53*-KO mice or amplifying CRIPSR–Cas9-edited alleles are found in Supplementary Table 1. PCR was performed using Phusion High-Fidelity PCR Master Mix (New England Biolabs, M0531S).

Rescue of *Dppa2*- and *Dppa4*-dKO mESCs with p53 expression. *Dppa2*- and *Dppa4*-dKO mESCs were transfected with mCherry control or mouse *Trp53*-mCherry piggyback plasmids. Rescue clones were screened for *Trp53*-mCherry expression using fluorescence monitoring for high p53 expression levels. *Dppa2*- and *Dppa4*-dKO cells with *Trp53*-mCherry or mCherry control

rescue plasmids were induced with doxycycline for 24 h, and then their DNA was damaged with 1 µM doxorubicin for 6 h, followed by medium washout for 12 h. RNA was then extracted with TRIzol (Life Sciences).

DPPA2- and DPPA4-overexpression experiments in WT mESCs. Doxycycline-inducible piggyback vectors that encode DPPA2–IRES–GFP or DPPA4–IRES–GFP were delivered into E14 mESCs by cotransfection (as described above). After doxycycline induction (48 h), cells were fixed with PFA, and immunofluorescence was performed as described above using DPPA2-, DPPA4- and DUX-specific antibodies.

Western blotting. RIPA buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1% NP-40, 0.1% sodium deoxycholate) with protease inhibitors was used to lyse cells. Lysates were further sonicated for 15 min with a 30-s on–off cycle in a Diagenode Biorupter Pico device to solubilize chromatin-bound proteins. Protein lysates were quantified with the Bio-Rad (500-0001EDU) Bradford reagent and loaded on SDS–PAGE gels. After transfer to nitrocellulose membranes (VWR, 95040-108), membranes were blocked with TBST (TBS with 0.1% Tween-20) and 5% milk and then incubated with antibodies overnight (Supplementary Table 3), and signals were detected using HRP-conjugated secondary antibodies and Western Lightning Plus-ECL, Enhanced Chemiluminescence (PerkinElmer Health Sciences, NEL 105001EA).

Quantitative PCR with reverse transcription. RNA was isolated using TRIzol extraction. Unless indicated, all RNA samples were treated with DNase to eliminate genomic DNA contamination. Reverse transcription was performed with SuperScript IV (Invitrogen) with random hexamer primer (Invitrogen), and qPCR (see Supplementary Table 1 for qPCR primer pairs) was performed with iTaq Universal SYBR Green Supermix (Bio-Rad) using a Bio-Rad qPCR CFX Connect instrument. Experiments were performed in biological replicates as indicated in figure legends. *Actin* RNA was used to normalize data from RT–qPCR mouse samples, and 18S rRNA was used to normalize data from RT–qPCR human samples.

Immunofluorescence of cultured cells. Cells were fixed in 4% paraformaldehyde at room temperature for 15 min and extensively washed with PBS. Cells were permeabilized using PBS with 5% BSA and 0.3% Triton X-100 at room temperature for 1 h. Primary antibodies (Supplementary Table 3) were diluted in PBS with 1% BSA and 0.1% Triton X-100 and incubated overnight at 4 °C, washed three times in PBS with 0.1% Triton and incubated with secondary antibodies (see Supplementary Table 3 for details) diluted in PBS with 1% BSA and 0.1% Triton X-100 and incubated for 1 h at room temperature. Samples were washed in PBS with 0.1% Triton X-100 and mounted with ProLong Gold Anti-fade with DAPI mounting medium (Fisher Scientific, P-36931).

Immunofluorescence of embryos. Zonae pellucidae were removed immediately before fixation using acid Tyrode's solution (137 mM NaCl, 2.68 mM KCl, 1.63 mM $CaCl_2$, 0.49 mM $MgCl_2$, 5.55 mM glucose, 0.01 mM polyvinylpyrrolidone, pH 2.5). Embryos were fixed for 30 min in 2.5% paraformaldehyde in PBS and then permeabilized for 20 min in 0.1% Triton X-100 in PBS. Embryos were then incubated at 4 °C overnight with primary antibody, washed and then incubated with secondary antibody (Thermo Fisher, 1:500) for 1 h at room temperature. All embryos were mounted in VECTASHIELD -containing 1.5 µg ml^{-1} DAPI (Vector Laboratories), and slides were scanned using a Zeiss LSM 510 UV confocal microscope (p53 staining) or a Leica SP8 confocal microscope (embryos stained for DUX). All immunofluorescence experiments were repeated a minimum of three different times using at least ten embryos per group.

Microscopy of cultured cells and embryos. Positive cells were counted using the ImageJ plugin 'Cell Counter'. Fluorescence signal for mouse eggs or embryos was quantified using ImageJ.

RNA-seq. For RNA-seq of mESCs, RNA was extracted using TRIzol, and libraries were generated according to the manufacturer's instructions: polyA-selected RNA was isolated, and libraries were prepared using the NEBNext kit (New England Biolabs, e7500s). For RNA-seq of embryos, the zona pellucida was removed using acid Tyrode's solution, and cells were lysed directly in lysis buffer using the SMART-Seq version 4 Ultra Low Input RNA Kit for Sequencing (Takara Bio, 634890). Fifteen cycles of pre-amplification were performed before following the manufacturer's instructions for tagmentation into double-stranded cDNA using the Nextera XT DNA Library Preparation kit (Illumina, FC-131-1024). Purified libraries were quantified on an Agilent Technologies 2200 TapeStation with a D1000 ScreenTape assay. The molarity of adaptor-modified molecules was defined by qPCR with a KAPA Library Quantification kit (Kapa Biosystems). Individual libraries were normalized to 10 nM, and equal volumes were pooled in preparation for Illumina sequence analysis. Sequencing libraries (25 pM) were chemically denatured and applied to an Illumina HiSeq paired-end flow cell with an Illumina cBot. Flow cells were then transferred to an Illumina HiSeq 2000 instrument and sequenced in the 125-bp paired-end mode.

RNA-seq analysis. RNA-seq reads were trimmed and filtered for quality using BBDuk and FastQC (version 0.10.1). Processed reads were aligned using TopHat2 version 2.1.0 (--t --q --N1 --L 25 --X 2000 --no-mixed --no-discordant), and counts for each transcript were generated using 'Get datasets' (https://metacpan.org/pod/distribution/Bio-ToolBox/scripts/get_datasets.pl). Differential expression analysis was performed using DESeq2 (version 3.11). Hierarchical clustering was performed using the R package 'ComplexHeatmap', with inputs being differentially expressed transcripts (DESeq FDR < 0.05) *z*-score normalized from the MARSS R package (version 3.10.12).

Repetitive element analysis. TEtranscripts (version 2.1.4) was used to identify differentially expressed repetitive elements in mouse embryo SMART-seq libraries.

GO term analysis. GO terms were identified using the Gene Ontology Resource (http://geneontology.org/).

Two-cell embryo-like cell and DUX-direct target identification. First, 2CLC-enriched transcripts were defined by comparing MERVL–GFP$^+$-sorted mESCs and GFP$^-$ mESCs ($n = 2{,}385$ (ref. [8])). DUX-direct targets were defined as genes bound by DUX–3×HA ChIP–seq[2] and upregulated after *Dux* transgene expression[2].

Embryoid body differentiation. Please see Supplementary Methods.

Single-cell RNA-seq and analysis. Please see Supplementary Methods.

Extra-embryonic endoderm cell differentiation. Differentiation was performed as described in ref. [43]. Briefly, *Rhox5*-mCherry knock-in mESCs either pulsed with doxycycline or treated with vehicle in 2iLIF medium and then transferred to XEN medium (RPMI 1640 supplemented with 15% (vol/vol) FBS and 0.1 mM β-mercaptoethanol, with 1% (vol/vol) penicillin–streptomycin). Standard XEN medium was supplemented with 0.01 µM all-*trans* retinoic acid (Sigma-Aldrich, R2625) dissolved in DMSO with activin A (R&D Systems, 338-AC-010) at 10 ng ml^{-1}.

Active transposition into active chromatin sequencing. Please see Supplementary Methods.

Luciferase assay. Please see Supplementary Methods.

Myoblast DNA damage. Human myoblasts were treated with 0.25 µM doxorubicin or vehicle control for 24 h before collecting RNA in TRIzol and processing it for RT–qPCR.

PacBio sequencing and *Dux* locus assembly. DNA was extracted from mixed-background BL6 mESCs using the Genomic-tip 100/G kit (Qiagen), and DNA integrity was assessed with the Femto Pulse system (Agilent). Shearing, SMRT library preparation and BluePippin size selection was performed following the PacBio Procedure & Checklist—Preparing HiFi SMRTbell Libraries using the SMRTbell Express Template Prep Kit 2.0 (PN 101-853-100 version 01, September 2019).

DNA was sheared with a g-TUBE (Covaris), treated with RNase and purified with AMPure XP beads. After assessment of the fragmented DNA on the Femto Pulse system, 10 µg sheared DNA was converted into a SMRT library. The library was size selected using the BluePippin system (Sage Science), collecting 9–13-kb and >15-kb fractions. The >15-kb fraction was loaded at a concentration of 50 pM on two Sequel II cells using the binding kit 2.0, sequencing reagent 2.0, sequencing primer V2 and a movie collection time of 30 h with a pre-extension of 2 h.

Consensus reads (CCS reads) were generated using CCS software version 4.0.0 in SMRT Link version 8 with '--minPasses 3 --minPredictedAccuracy 0.99 --maxLength 50000'. The total CCS read yield was 50 Gb.

The mouse genome was assembled using HiCanu version 2 with default parameters and a genome size of 2.7 Gb. Starting from a total of 2,997,947 HiFi reads, (~18.5-fold coverage of the genome), the final assembly had a total length of 2.79 Gb resolved in 1,551 contigs with an N_{50} of 5.7 Mb.

Contigs from the primary assembly were aligned to the mm10 genome using minimap2, and the *Dux* locus was visually inspected on the IGV browser to find the best matching contigs.

Chromatin immunoprecipitation followed by sequencing. Please see Supplementary Methods.

HTPS data availability. High-throughput sequencing (HTPS) data from this study were deposited at the Gene Expression Omnibus under accession number GSE149267 and summarized in Supplementary Table 4.

Public datasets reanalyzed or remapped to the new *Dux* locus contig. Datasets are available under the following accession numbers: H3K4me1, H3K4me2, H3K4me3, H3K27me3 and p300 ChIP–seq in mESCs, GSE98063; H4K20me3 ChIP–seq in mESCs, GSE130721; H3K122ac and H3K64ac ChIP–seq in mESCs,

GSE66023; H3.3S31ph ChIP–seq in mESCs, SRR8310851; H3.3 ChIP–seq in mESCs, GSE42152; DPPA2 and DPPA4 ChIP–seq in mESCs, GSE117173; ZSCAN4 ChIP–seq in mESCs, GSE140621; DUX ChIP–seq in mESCs, GSE95519; NELFA ChIP–seq in mESCs, GSE113671; H3K4me3 ChIP–seq in *Zscan4*$^{+}$ mESCs, GSE164486; p53 ChIP–seq in human fibroblasts, GSE115940; p53 ChIP–seq in human fibroblasts, GSE111009; p53 ChIP–seq in hESCs, GSE39912.

The *Dux* locus contig sequence and assembly were deposited in GenBank as submission BankIt2443327 tig00005009 MW810794.

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

Adjustment for multiple comparisons. When appropriate, multiple comparisons were reported as statistically different if the Bonferroni-corrected *P* value (stated as FDR in figure legends) met the indicated significance threshold. DESeq2 analysis was reported using adjusted *P* values (FDR) that were corrected using the Benjamini–Hochberg procedure.

Mycoplasma testing. Cells were routinely tested for mycoplasma using the Lonza MycoAlert test kit, and the results were negative.

Box plot conventions. The median is shown as a line in the box, and the outline of the box is depicted at the 25th and 75th percentiles. Extended whiskers depict $Q1 - 1.5 \times IQR$ and $Q3 + 1.5 \times IQR$. Outliers are depicted as dots.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data, cell lines, reagents and unique materials are available upon request. ChIP–seq, ATAC-seq, scRNA-seq and RNA-seq data were deposited under GSE149267 and are detailed in Supplementary Table 4. Source data are provided with this paper.

Acknowledgements

We thank members of the Cairns laboratory and M.B. Chandrasekharan for fruitful discussions. We are grateful to the patients who made this work possible. We thank T. Oliver for the *Trp53*$^{fl/fl}$ mouse, S.J. Tapscott for FSHD1 and FSHD2 myoblasts, F. Zhang for the px330-Cas9 plasmid, S. Jackson for the pICE-HA-NLS-I-PpoI plasmid, A. Chavez and G. Church for the dCas9-KRAB-MeCP2 plasmid and C. Gersbach for the pcDNA-dCas9-p300(HAT) plasmid. We also thank B. Dalley in the HCI High-Throughput Genomics and Bioinformatic Analysis Shared Resource (NCI grant P30CA042014), the CCTS Stem Cell Facility (National Institutes of Health (NIH), UL1TR002538), J. Marvin and the University of Utah Flow Cytometry Facility (NIH, 1S10RR026802-01; NCI, 5P30CA042014-24) and the University of Utah Cell Imaging Core. Funding for this work supported C.J.W. in part from the Intramural Research Program of the NIH (NIEHS, 1ZIAES102985), the NCI (P30 CA015704-45S6) and W.OP.14-01 from the Prinses Beatrix Spierfonds to S.M.v.d.M., and additional funding was from Wellstone Center from UMass (NICHD, P50HD060848) to R.J.B., the NIH (F30HD098000) to B.D.W., the NICHD (F32HD104442) to S.C.S., the NICHD (F32HD094500) and Lalor Foundation Fellowship 10041116 to E.J.G. and the Howard Hughes Medical Institute and the NICHD (1R01HD095833) to B.R.C. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author contributions

IRB processing, patient consent, patient management and sample selection and processing were overseen by N.E.J. and R.J.B. Experiments and analyses were conducted by E.J.G., B.D.W., C.M.S., J.G., S.C.S., P.G.H., P.S., R.M. and S.L.K. with contributions by C.J.W. and S.M.v.d.M. E.J.G. and B.R.C. designed the study and wrote the manuscript with input from co-authors.

Competing interests

The authors declare no competing interests.

Additional information

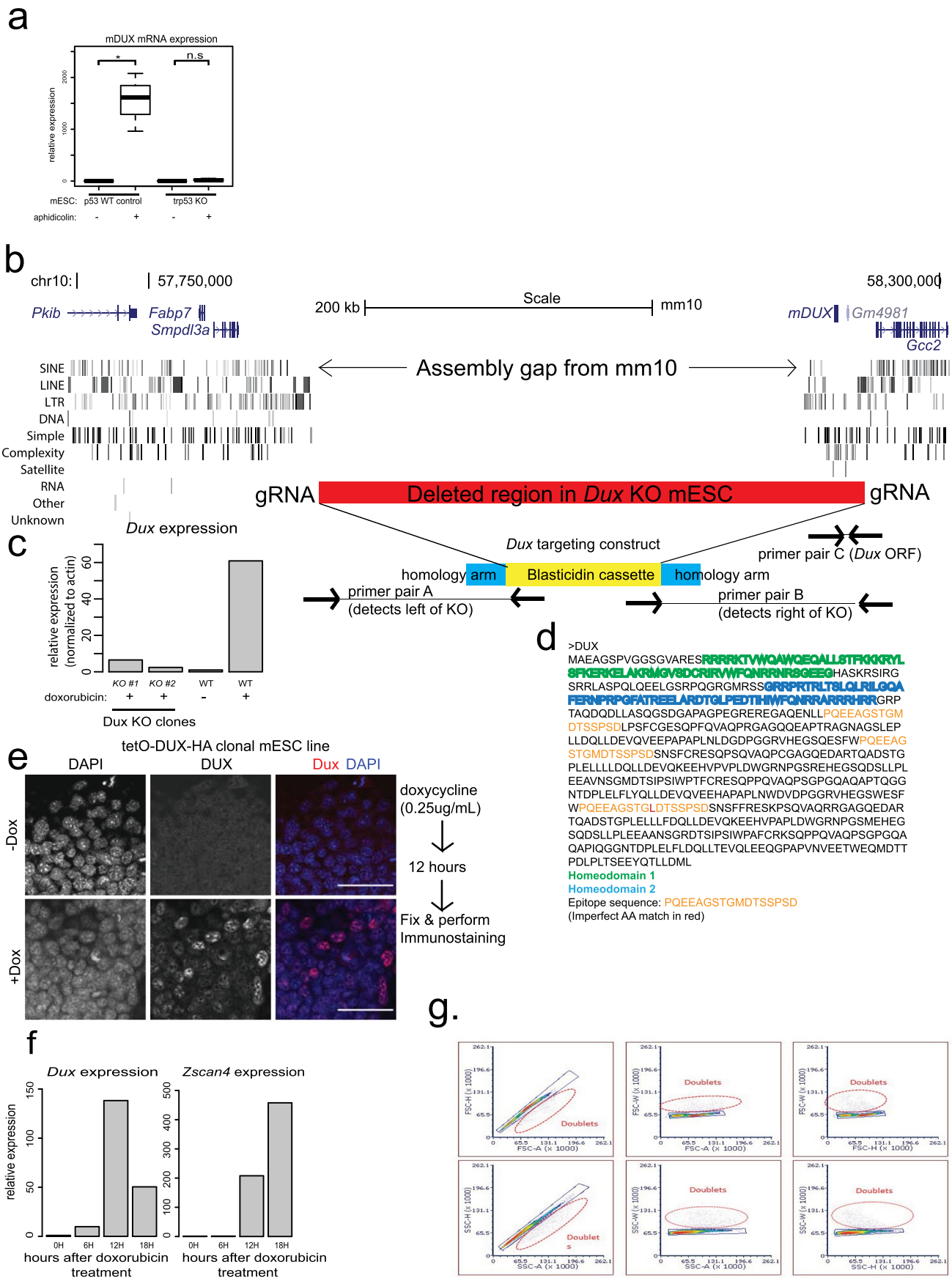
Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00893-0.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00893-0.

Correspondence and requests for materials should be addressed to B.R.C.

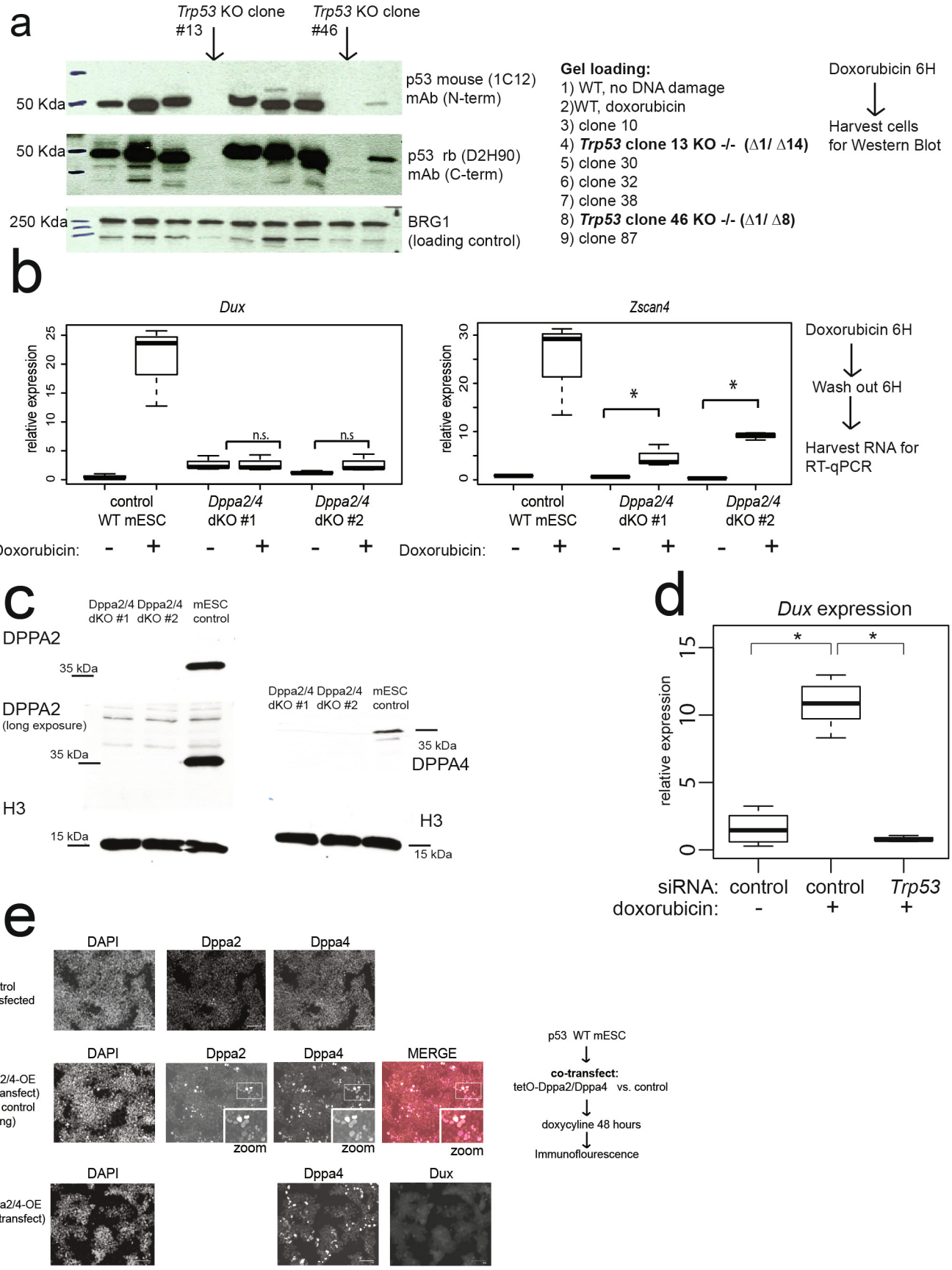
Peer review information *Nature Genetics* thanks Guillermina Lozano and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



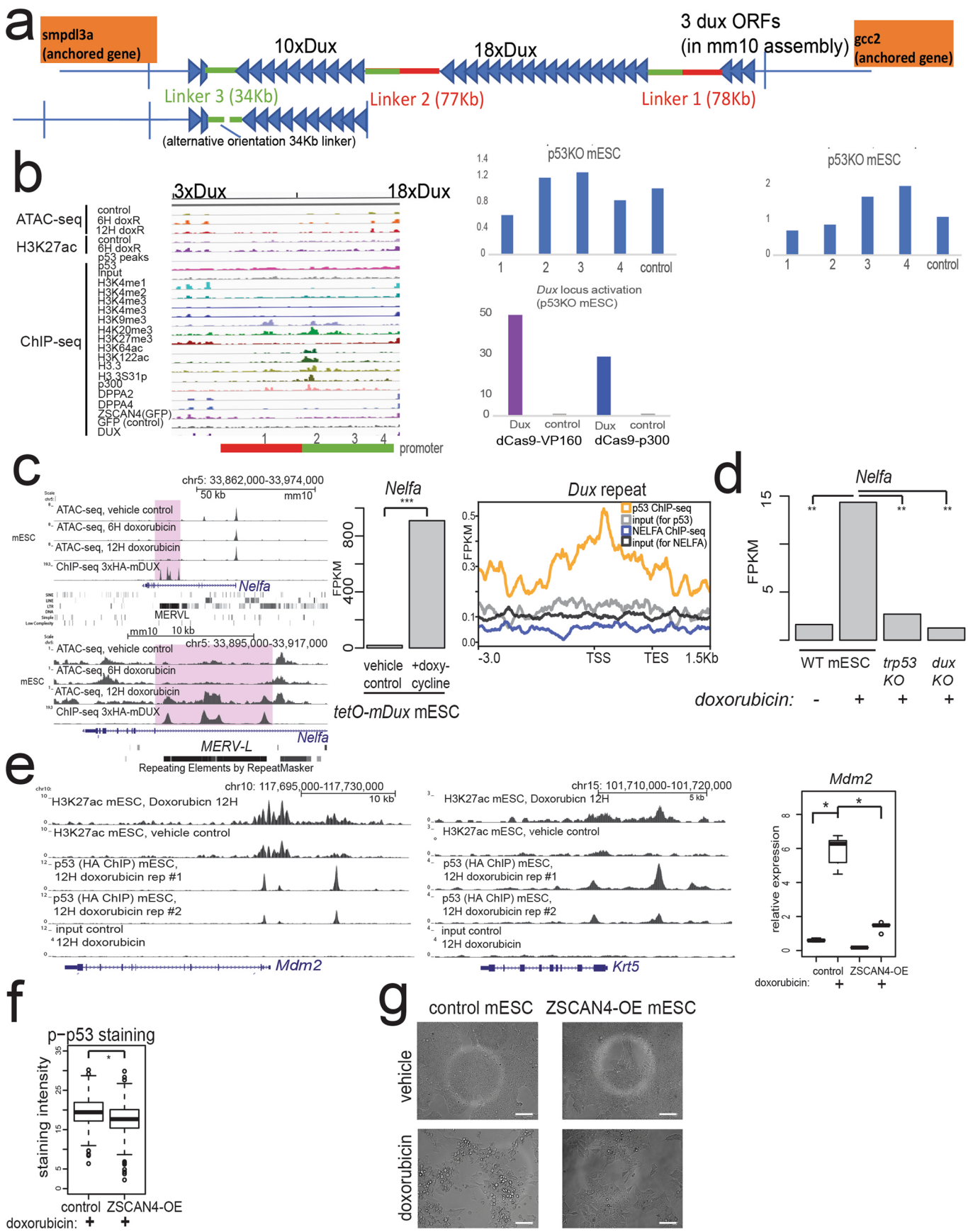
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Related to Fig. 1. a, *Dux* expression in aphidicolin or vehicle control treatment of *Trp53* WT mESC or *Trp53* KO mESC. RT-qPCR, n = 3 biological replicates, * pvalue<0.05, for p53WT p-value = 0.02041, for p53KO p-value= p-value = 0.05449, t-test, one sided. **b**, Schematic of *Dux* locus in mm10 genome assembly, the design of the targeting construct, location of gRNAs for *Dux* KO mESC line generation. Shown below are locations of genotyping primers. **c**, RT-qPCR analysis *Dux* KO mESC clones #1 and #2 treated with doxorubicin to induce endogenous *Dux* expression, which were used for experiments in Fig. 1. Shown is a representative analysis of three independent experiments. **d**, Design of DUX peptide antigen for antibody creation. **e**, immunofluorescence with the rabbit polyclonal anti-DUX antibody using mESC with a tetracycline inducible DUX-3xHA transgene. Representative image from 3 independent experiments. Merge: DAPI = blue, DUX = red. Scale bar = 125 micrometers. **f**, Kinetic analysis of *Dux* and *Zscan4* transcript induction in WT mESC treated with 1uM of doxorubicin for indicated times. Note earlier induction of *Dux* compared to *Zscan4*, the RNAseq from Fig1 and Fig2 is using the later time point 18H. FACS/flow cytometry gating scheme to exclude doublets. For Extended Data Fig 1a, the median is shown as a line in the box, and the outline of the box is depicted at the 25th and 75th percentile. The extended whiskers depict Q1-1.5*IQR and Q3 + 1.5*IQR. Outliers points are depicted as dots.



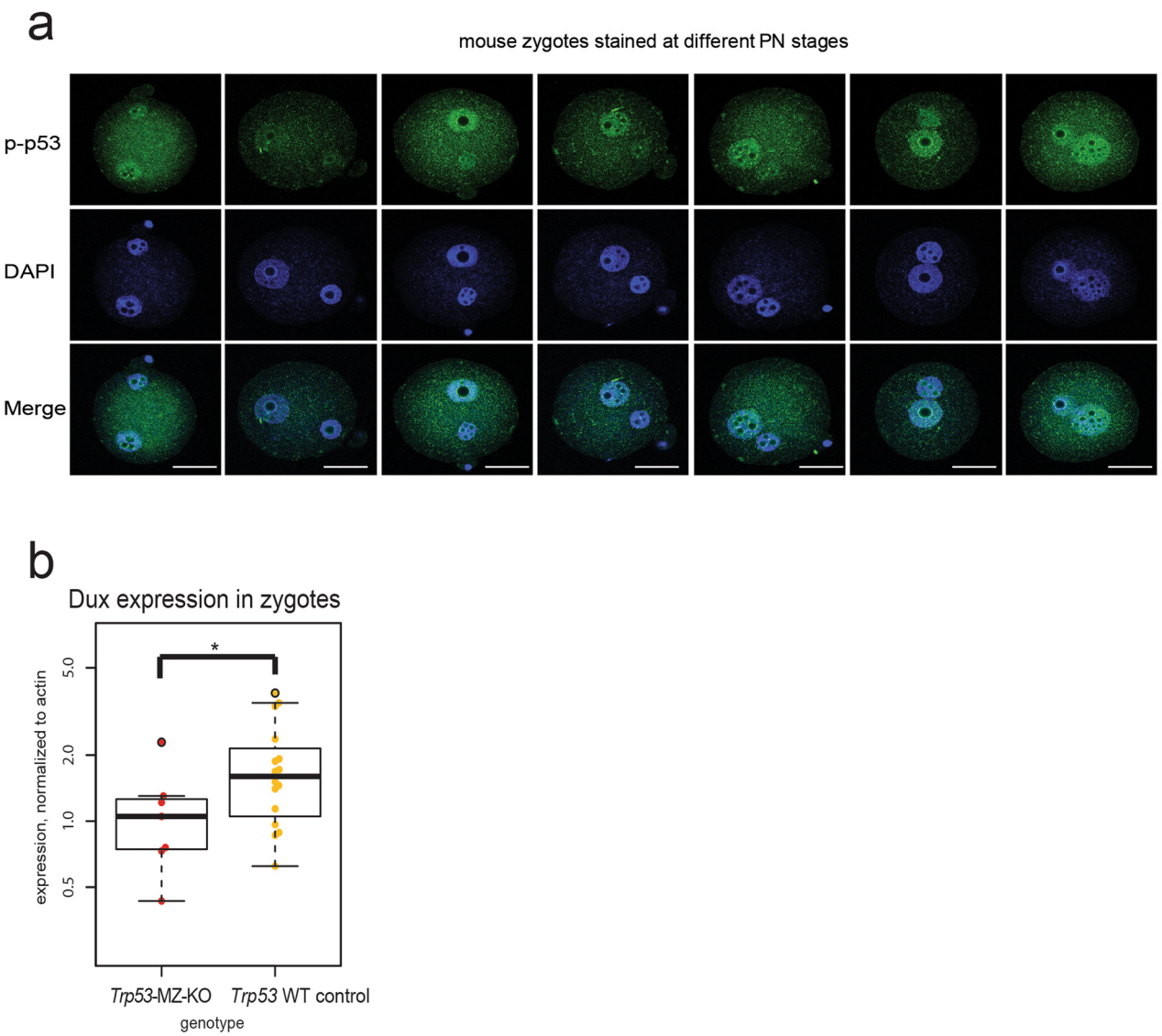
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Related to Fig. 2. a, Western blot analysis and KO deletion allele Sanger sequencing results for two independent *Trp53* KO mESC clones (used for Fig. 2). **b**, RT-qPCR measure of *Dux* or *Zscan4* expression levels in two independent *Dppa2/4* dKO mESC clones. *Dppa2/4* dKO clone #1 was used for Fig. 2e with p53 rescue experiments. *<0.05 FDR, One-way ANOVA. **c**, Western blot confirmation of dKO Dppa2/4 mESC clones #1, 2. **d**, *Dux* expression in *Trp53* WT mESC after control or *Trp53* siRNA knockdown, with vehicle or doxorubicin treatment. RT-qPCR, N = 6 biological replicates, *<0.05 FDR, One-way ANOVA. **e**, Co-overexpression of DPPA2/4 in *Trp53* WT mESC does not activate *Dux* expression. Merge: DAPI = cyan, DPPA2 = yellow, DPPA4 = magenta. Representative image from 3 independent experiments. Scale bar = 125 micrometers. For Extended Data Fig 2b, d, the median is shown as a line in the box, and the outline of the box is depicted at the 25th and 75th percentile. The extended whiskers depict Q1-1.5*IQR and Q3 + 1.5*IQR. Outliers points are depicted as dots.

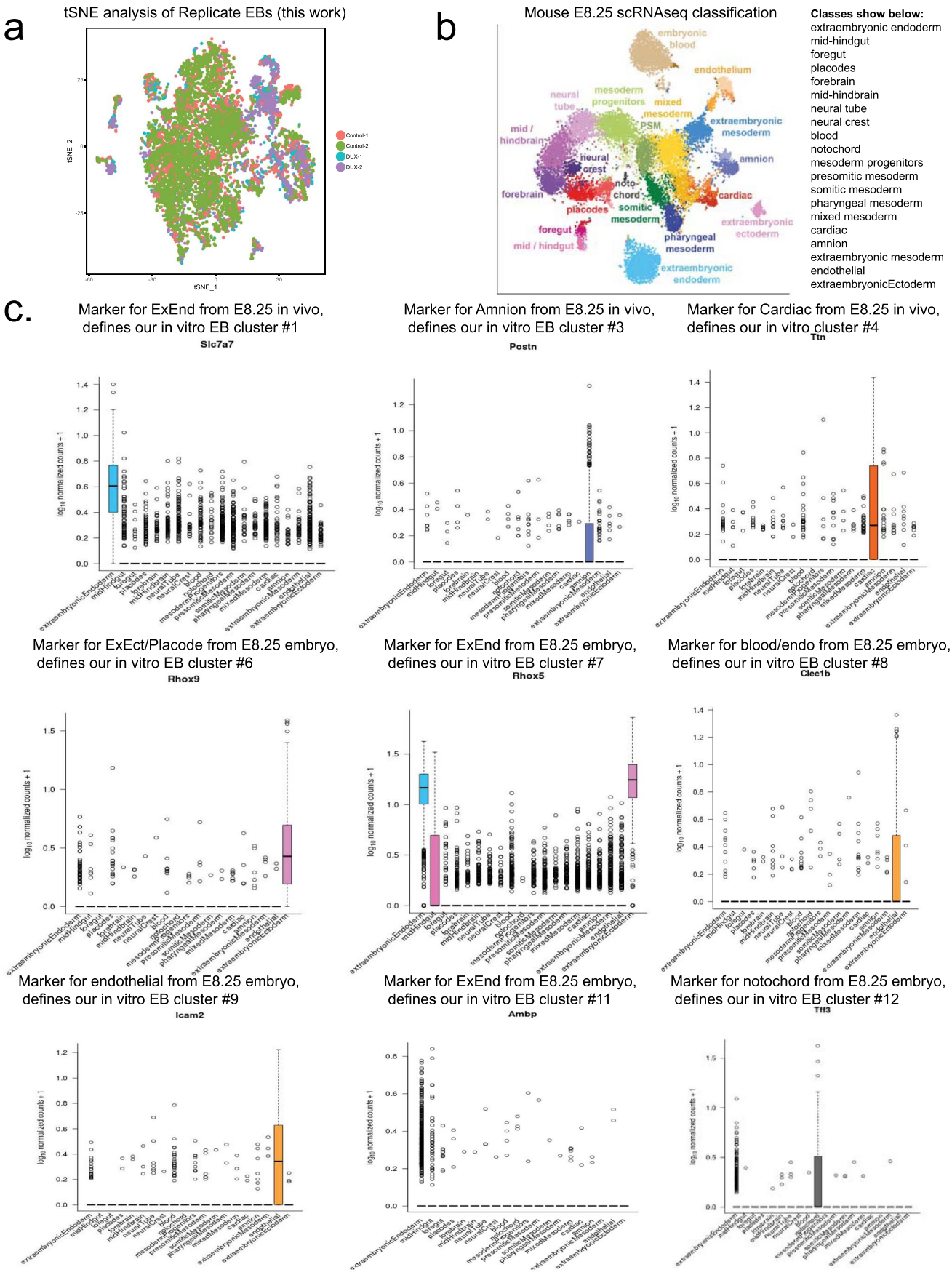


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Related to Fig. 3. a, Schematic of long-read (PacBio) sequenced and assembled mouse *Dux* locus. **b**, CRISPR-A experiment in *Trp53* KO mESC. CRIPSR-A with the *Dux* promoter-targeted gRNAs strongly activates *Dux* expression. RT-qPCR, n = 3 biological replicates. **c**, Mouse *Nelfa* locus showing enrichment of DUX binding at the 3' end of the gene at intronic MERVL element. Bottom panel is zoomed (n = 2 biological replicates for each condition). Barplot (middle panel) of doxycycline induced *Dux* transgenic mESC showing strong induction of *Nelfa* transcripts (RNA-seq, n = 2 biological replicates, * FDR < 0.05, DESeq2, data reprocessed from Hendrickson, et al. Nature Genetics 2017). Right-most panel: Metagene plot of the *28xDux* repeat units showing p53-ChIP-seq and input control and NELFA ChIP-seq and input control (blue and black lines respectively from Hu, et al. *Nature Cell Biology*, 2020—note lower NELFA ChIP-seq signal compared to the matched-control input). **d**, *Nelfa* is transcriptionally induced by doxorubicin treatment, and this requires both p53 and DUX. RNA-seq from this paper, n = 2 biological replicates, * FDR < 0.05 DESeq2. **e**, *Mdm2 and Krt5* are direct p53 targets. p53 ChIP-seq and H3K27ac ChIP-seq (n = 2 biological replicates). RT-qPCR measuring Mdm2 expression in control mESC or ZSCAN4-OE mESC, treated with vehicle control or doxorubicin; n = 5 biological replicates, *p-value <0.05, one-sides t-test. **f**, Immunofluorescence staining of control mESC (n = 135 cells) or clonal ZSCAN4-OE mESC (n = 487 cells) using phospho-p53 antibodies after doxorubicin treatment, n. *<0.001, one-sided Wilcox test. **g**, Brightfield image showing decreased cell death after doxorubicin treatment in ZSCAN4-OE mESC compared to control. Image is representative from 3 independent experiments. Scale bar = 125 micrometers. For Extended Data Fig 3e, f, the median is shown as a line in the box, and the outline of the box is depicted at the 25th and 75th percentile. The extended whiskers depict Q1-1.5*IQR and Q3 + 1.5*IQR. Outliers points are depicted as dots.

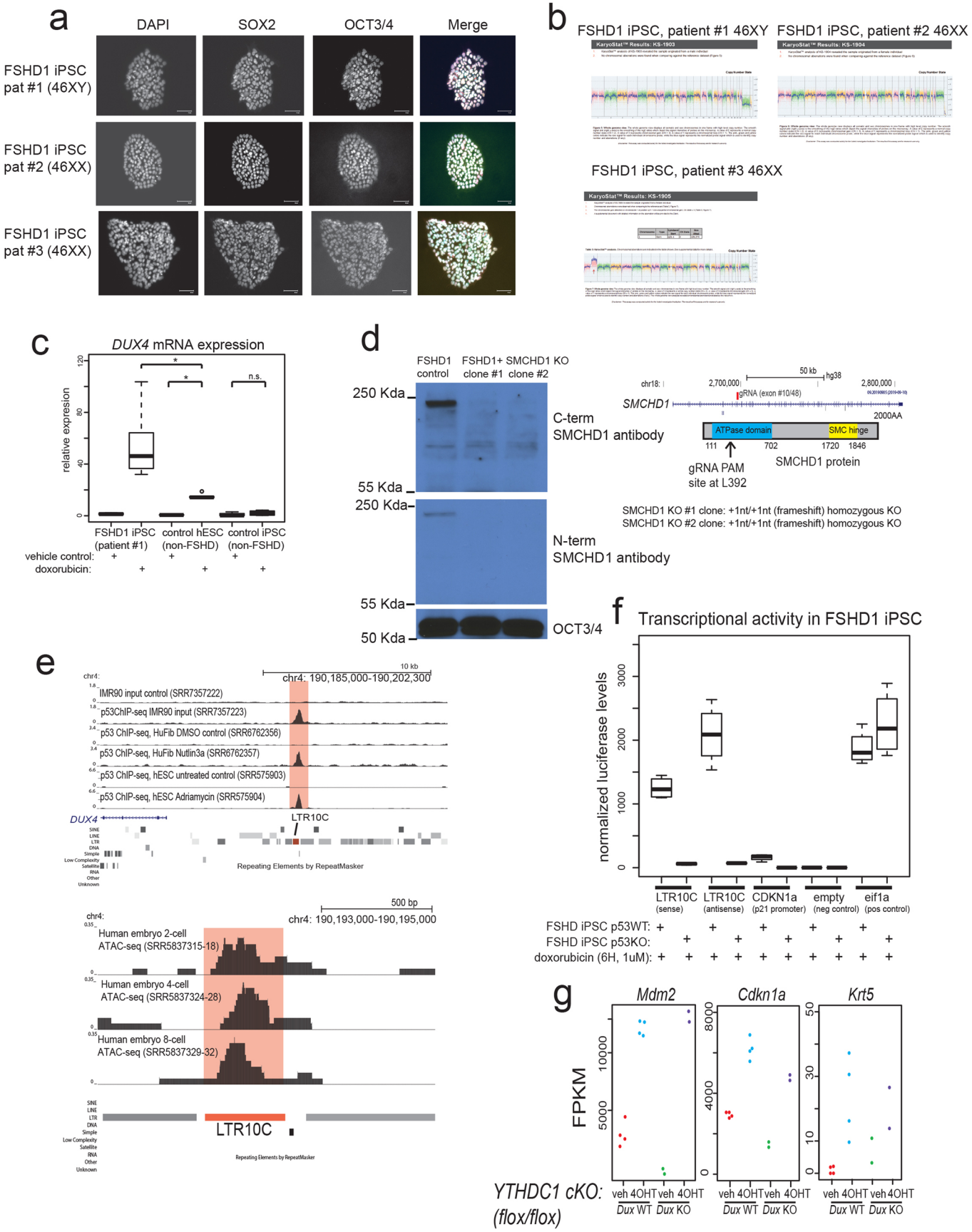


Extended Data Fig. 4 | Related to Fig. 4. a, Immunofluorescence of pronuclei (PN)-stage zygotes showing nuclear phospho-S15 p53 staining (quantified in Fig. 4a). Scale bar = 40 micrometers. **b**, Single mouse zygote RT-qPCR measuring *Dux* expression in PN5 stage zygotes (n = 7 p53 MZ-KO, n = 16 p53 WT), * p-value <0.05, p-value= 0.0409, one-sided t-test. For Extended Data Fig 4b, the median is shown as a line in the box, and the outline of the box is depicted at the 25th and 75th percentile. The extended whiskers depict Q1-1.5*IQR and Q3 + 1.5*IQR. Outliers points are depicted as dots.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Related to Fig. 5. a, Comparison of two biological replicates for EB scRNAseq (control vs Dux-pulsed) showing high concordance between samples. **b**, Ibarra-Soria, et al. Nature Cell Biology 'Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation' depicting different cell types defined in E8.25 mouse embryos, (data retrieved from https://marionilab.cruk.cam.ac.uk/organogenesis/ February 2020). **c**, Analysis of Ibarra-Soria, et al. data compared to our EB scRNAseq data with indicated markers identifying cell types (see Fig. 4c table). Each plot shows the marker identified in our Seurat analysis of EB scRNAseq as discriminating between other cell type clusters, and the data shows the distribution of that marker in E8.25 mouse *in vivo* cell types. For Extended Data Fig 5c, the median is shown as a line in the box, and the outline of the box is depicted at the 25th and 75th percentile. The extended whiskers depict Q1-1.5*IQR and Q3 + 1.5*IQR. Outliers points are depicted as dots.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Related to Fig. 6. a, FSHD1 iPSC (patient #1, 2, 3) immunostaining for pluripotency markers SOX2 and OCT3/4. Representative image from 3 independent experiments. Merge: DAPI = cyan, SOX2 = yellow, OCT4 = magenta. Scale bar = 125 micrometers. **b**, Thermo-Fisher Karyostat report for FSHD1 iPSC clones patients #1, 2,3. **c**, RT-qPCR after vehicle control or doxorubicin treatment measuring *DUX4* levels in FSHD1 iPSC patient #1, non-FSHD hESC female 'LSJ2', and non-FSHD iPSC 'WT33'. N = 6 biological replicates, *<0.05 FDR, One-way ANOVA. **d**, Western blot with N- and C-term SMCHD1 antibodies that the 2 independent clones show in Fig. 6c are KO, isogenically created in FSHD1 patient #1. CRISPR/Cas9 deletion strategy shown on right top, with the Sanger sequencing of KO clones shown on bottom right. **e**, Genome browser snap-shot of ATAC-seq performed in human embryos showing open chromatin signal at the 4qA LTR10C element. **f**, Luciferase assay testing directionality of the LTR10C element in FSHD1 patient #1 iPSC (p53 WT or isogenic p53 KO). N = 4 biological replicates. **g**, RNA-seq analysis from Liu, et al. *Nature* 2021 showing reactivation of direct p53 targets (Mdm2, Cdkn1a (p21), and Krt5) in *Ythdc1* conditional knockout (cKO) mESC ± *Dux* KO, treated with vehicle or 4-OHT (tamoxifen) to eliminate the YTHDC1 protein. For Extended Data Fig 6c, f, the median is shown as a line in the box, and the outline of the box is depicted at the 25th and 75th percentile. The extended whiskers depict Q1-1.5*IQR and Q3 + 1.5*IQR. Outliers points are depicted as dots.

nature research

Corresponding author(s): Cairns, Bradley

Last updated by author(s): Mar 5, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	BD FACSDiva v9.0 Software
Data analysis	Seurat program (http://satijalab.org/seurat/, R package, v.2.2.0) , Cell Ranger (v3.1), STAR (v2.5.4a), TEtranscripts (Version: 2.1.4), MARSS 3.10.12, Tophat2 v2.1.0, DESEq2 v3.11, Bowtie2 (v2.2.6) , MACS2 v 2.2.5, t-SNE (v0.1-3). R v3.6.2, HiCanu v2, Image J 1.52, DeepTools 2.0, ComplexHeatmap 1.10.2, IGV 2.9.4

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data is available upon request. ChIP-seq, ATAC-seq, scRNA-seq, RNA-seq are deposited under GSE149267 and are detailed in Supplementary Table 4. H3K4me1, H3K4me2, H3K4me3, H3K27me3, p300, in mESC: GSE98063; H4K20me3 mESC GSE130721; H3K122ac, H3K64ac in mESC GSE66023; H3.3S31phospho mESC SRR8310851; H3.3 in mESC GSE42152; DPPA2, DPPA4 ChIP-seq mESC GSE117173; ZSCAN4 ChIP-seq mESC GSE140621; DUX CHIP-seq mESC GSE95519; NELFA ChIP-seq mESC GSE113671; H3K4me3+ in Zscan4+ mESC GSE164486; p53 ChIP-seq in human fibroblasts GSE115940; p53 ChIP-seq human fibroblasts GSE111009; p53 ChIP-seq hESC GSE39912.

The Dux locus contig sequence/assembly was deposited in GENbank as Submission # BankIt2443327 tig00005009 MW810794.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For RNA-seq, ChIP-seq, and ATAC-seq datasets, most samples were measured in biological duplicate for analysis with DESeq2. No sample size calculation was performed. Based on the expected effect size and estimate of variance within groups, sample sizes were selected to detect modest effect sizes (2-3 fold).
Data exclusions	No data were excluded
Replication	All attempts at replication were successful. Experiments were repeated at least two independent times.
Randomization	N/A (allocation based on genotype)
Blinding	Where possible, samples were blinded to the the individual collecting raw data and performing analysis. Experiments involving RNA-seq, single-cell RNAseq, ATAC-seq, microscopy, luciferase assays, flow cytometry, RT-qPCR were blinded. Reanalysis of published ChIP-seq data was not blinded because MACS requires user identification of background (input ChIP, control ChIP, etc).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Supplemental table #3
p53 1C12, mouse mAb Cell Signaling 2524 Western blotting 1:1000
p53 D2H90, rb mAB Cell Signaling 32532 Western blotting 1:1000
phospho-S15-p53 rb pAb affinity purified, Cell Signaling 9284 Immunostaining: embryos 1:200
DPPA2 mAb clone 6C1.2, Millipore Sigma MAB4356 Western blotting 1:1000
DPPA4 goat pAb, affinity purified, R&D Systems AF3730 Western blotting 1:1000
RHOX5 rb pAb, Abcam ab31922 Immunostaining, cells: 1:1000
MCHERRY rb pAb, Abcam ab167453 Western blotting 1:1000, Immunostaining: 1:1000
SOX17 goat pAb, R&D Systems, AF1924-SP Immunostaining, cells: 1:250
GATA6 rb mAb, Cell Signaling Technology 5851T Immunostaining, cells 1:500
Histone H3 mAb, Active Motif 39763 Western blotting 1:3000
SMCHD1 abcam ab122555 (N-term) Western blotting 1:1000
SMCHD1 abcam ab31865 (C-term) Western blotting 1:1000
BRG1 mouse mAb santa Cruz (G-7): sc-17796 Western blotting 1:1000
SOX2 goat pAb R&D AF2018-SP Immunostaining, cells 1:500
anti-rabbit 594 Donkey anti rabbit Alexa 594, Thermo Scientific A-21207 Immunostaining, cells 1:500
anti-mouse 647 Donkey anti-Mouse Alexa 647, Thermo Fisher Scientific A-31571 Immunostaining, cells 1:500
anti-rabbit 488 Donkey-anti Rabbit alexa 488, Life Technologies A-21206 Immunostaining, cells 1:500
anti-goat 594 Donkey-anti Goat Alexa 594, Thermo Fisher Scientific A-11058 Immunostaining, cells 1:500

goat anti-rabbit HRP BioRad 1706515 Western blotting 1:3000
HA epitope tag (for p53-HA) abcam ab91110 ChIP, 5 ug/ChIP sample
H3K9me3 active motif 39161 ChIP, 3ug/sample
H3K27ac abcam 4729 ChIP, 3ug/sample

Validation

For commercially available primary antibodies: only previously cited antibodies with literature confirmed data were used. They were routinely validated with two methods: the ability to detect an over-expressed cDNA (Immuno-flourescence and western blotting: rhox5, gata6, sox2, sox17, H3, oct4, Zscan4, mcherry, GFP, anti-HA, BRG1) and the specificity was evaluated using knock-out cell lines (western-blotting/immunoflourescence: p53 (human and mouse), SMCHD1 (human), DPPA2 (mouse), DPPA4(mouse)) . For DUX antibodies developed from Covance, they were validated using immunoflourescence in mESC with a tet-inducible transgene and DUX KO mESC treated with doxorubicin (see Supplemental Figure 1 and Main Fig 1). The H3K9me3 and H3K27ac antibodies are validated in terms of their specificity by the company (active motif and abcam) using histone peptide arrays.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

E14 mESC were obtained from Maria Torres-Padilla, human myoblasts were obtained from Stephen Tapscott, control human hESC were obtained from Center for Human Embryonic Stem Cell Research and Education, control human iPSC were obtained from Rusty Gage, and human FSHD iPSC were derived in our lab at University of Utah.

Supplementary Table #2:
Cell line name species cell type used for Published
E14-MERVL-GFP Mus musculus mouse ESC DNA damage experiments, RNA-seq, EB formation, chimera generation Ishiuchi, et al. NSMB 2015
E14-MERVL-GFP, DUX KO clone #1 Mus musculus mouse ESC DNA damage experiments, RNA-seq, Immunoflourescence this study
E14-MERVL-GFP, DUX KO clone #2 Mus musculus mouse ESC DNA damage experiments, RNA-seq, Immunoflourescence this study
E14-MERVL-GFP, Trp53 KO clone #1 Mus musculus mouse ESC DNA damage experiments, RNA-seq, Immunoflourescence this study
E14-MERVL-GFP, Trp53 KO clone #2 Mus musculus mouse ESC DNA damage experiments, RNA-seq, Immunoflourescence this study
E14-MERVL-GFP, Dppa2/4 dKO clone #1 Mus musculus mouse ESC DNA damage experiments, RT-qPCR this study
E14-MERVL-GFP, Dppa2/4 dKO clone #2 Mus musculus mouse ESC DNA damage experiments, RT-qPCR this study
E14-MERVL-GFP, Rhox5-mCherry C-term knockin Mus musculus mouse ESC EB formation this study
E14-MERVL-GFP, tetO-3xHA-DUX Mus musculus mouse ESC QC of DUX antibody, EB formation, chimera generation Hendrickon, et al. Nature Genetics 2017.
FSHD iPSC, patient #1 Homo sapiens human iPSC DNA damage experiments this study
FSHD iPSC, patient #1 TP53 KO Homo sapiens human iPSC DNA damage experiments this study
FSHD iPSC, patient #1, SMCHD1 KO clone #1 Homo sapiens human iPSC DNA damage experiments this study
FSHD iPSC, patient #1, SMCHD1 KO clone #2 Homo sapiens human iPSC DNA damage experiments this study
FSHD iPSC, patient #1, SMCHD1 KO clone #3 Homo sapiens human iPSC DNA damage experiments this study
FSHD iPSC patient #2 Homo sapiens human iPSC DNA damage experiments this study
FSHD iPSC patient #3 Homo sapiens human iPSC DNA damage experiments this study
control iPSC (female WT33) Homo sapiens human iPSC DNA damage experiments Marchetto, et al. Cell 2010.
control hESC ("LSJ2" female 46XX) Homo sapiens human ESC DNA damage experiments https://med.stanford.edu/hesc/research/corefacility/distributions.html
MB135 Homo sapiens Control (non-FSHD) myoblast DNA damage experiments Campbell, et al. Elife 2018.
MB2401 Homo sapiens Control (non-FSHD) myoblast DNA damage experiments Campbell, et al. Elife 2018.
MB073 Homo sapiens FSHD1 myoblast DNA damage experiments Campbell, et al. Elife 2018.
54-2 Homo sapiens FSHD1 myoblast DNA damage experiments Campbell, et al. Elife 2018.
MB200 Homo sapiens FSHD2 myoblast DNA damage experiments Campbell, et al. Elife 2018.
2305 Homo sapiens FSHD2 myoblast DNA damage experiments Campbell, et al. Elife 2018.
2453 Homo sapiens FSHD2 myoblast DNA damage experiments Campbell, et al. Elife 2018.
54-1 Homo sapiens Control (non-FSHD) myoblast DNA damage experiments Campbell, et al. Elife 2018.

Authentication

For the human FSHD iPSC we derived in this study, we authenticated them by karyotyping using the GeneChip and KaryoStat assays by ThermoFisher (desribed in the methods and in Supplementary Figure 5). hIPSC derived in this study were also stained using pluripotency markers OCT4/SOX2. mESC and hPSC were routinely differentiated into EBs to evaluate their pluripotency. ESC cell lines were authenticated using RT-qPCR for species-specific cell type markers (such as pluripotency markers for NANOG, OCT4, SOX2). Myoblasts cell lines were authenticated using RT-qPCR for human myoblast specific markers.

Mycoplasma contamination

Cell lines were routinely tested for mycoplasma using the MycoAlert test kit and found to be mycoplasma negative.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified lines were used in this study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	For Mus musculus wildtype embryos used for p53 staining we used 6- to 10-week-old CF1 female mice (n=69 zygotes). To generate p53 KO females, we used a p53 flox/flox allele that flanks exons 2-10 maintained in a mixed Bl6/C57 background. We crossed these p53 flox/flox animals with a ddx4:Cre driver (maintained in a Bl6/C57 background) to obtain WT p53/floxed heterozygous knockout animals. These F1 p53 +/- animals were intercrossed to obtain F2 animals. To generate the p53 female KO animals, F2 p53 KO males were crossed to F2 p53+/- females, which were used between ages 6-10 week old for studies (n=4 females, n=20 embryos for Immunoflourescence, n=3 females for SMRT-seq (n=12 embryos, n=3 eggs)). p53 Wildtype controls were p53 flox/flox animals kept on a Bl6/C57 mixed background and used for experiments during 6-10 weeks of age (n=4 females, n=37 embryos for Immunoflourescence, n=3 females (n=12 embryos, n=3 eggs) for SMRT-seq). Mice were maintained on 12H light/dark cycles with food ad libitum, at 20–26 °C and 16-23% humidity.
Wild animals	No wild animals were used.
Field-collected samples	No field-collected samples were used.
Ethics oversight	Animal experiments performed at the University of Utah were first approved by IACUC and Institutional approval, the protocol # is 18-07014.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	One male and two female patients were selected. Average age of diagnosis is 17 years of age, all three patients were around those ages.
Recruitment	FSHD patients were selected for fibroblast biopsies and iPSC reprogramming from patients with symptomatic presentation (see Material and methods for clinical data): Subject Onset Genetics* Family History Age at last visit Severity at Last Visit 4034 18-19 yrs old, shoulder weakness 9 RU None 29 Upper extremity weakness with limited range of motion in shoulder, scapular winging Unable to raise arm. Ambulatory, no significant weakness in lower extremity 4050 17 yrs old, shoulder weakness, tripping 6 RU** Multiple affected family members including mother 41 Upper extremity weakness with limited range of motion and scapular winging, weakness elbow flexion. Ambulatory, mild hip flexion and dorsiflexion weakness 4059 Early childhood, unable to keep up with peers, formal dx at age 14, difficulty walking in later teens 6 RU** Multiple affected family members including mother 54 Weakness in shoulder girdle and upper arm. Unable to raise arm >180deg. Ambulatory but frequent falls with weakness in dorsiflexion and hip flexion, and knee extension. *number of D4Z4 repeat units (RU) **Part of original Utah FSHD Kindred (Flanigan et al. 2001) Potential self-selection biases: since all three of these patients were severely affected enough to be diagnosed around the average age (17), it is possible that they have more severe penetrance than other individuals with similar D4Z4 repeat unit copy number. Whether the selection of these individuals would affect the p53 responsiveness of their DUX4 allele compared to non-symptomatic, less-severely affected, or later-diagnosed FSHD patients is an open question for further study.
Ethics oversight	Dermal fibroblasts were obtained from skin biopsies taken from patients with FSHD under a University of Utah IRB approved protocol #40092.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	Previously published datasets for: H3K4me1, H3K4me2, H3K4me3, H3K27me3, p300, in mESC: GSE98063; H4K20me3 mESC GSE130721; H3K122ac, H3K64ac in mESC GSE66023; H3.3S31phospho mESC SRR8310851; H3.3 in mESC GSE42152; DPPA2, DPPA4 ChIP-seq mESC GSE117173; ZSCAN4 ChIP-seq mESC GSE140621; DUX CHIP-seq mESC GSE95519; NELFA ChIP-seq mESC GSE113671; H3K4me3+ in Zscan4+ mESC GSE164486; p53 ChIP-seq in human fibroblasts GSE115940; p53 ChIP-seq human fibroblasts GSE111009; p53 ChIP-seq hESC GSE39912. The Dux locus contig sequence/assembly was deposited in GENbank as Submission # BankIt2443327 tig00005009

	MW810794. Data generated in this paper: ChIP-seq, ATAC-seq, scRNA-seq, RNA-seq are deposited under GSE149267 and are detailed in Supplementary Table 4.
Files in database submission	GSE149267
Genome browser session (e.g. UCSC)	hg38/ncbi37/mm10+DUXlocus (custom contig)

Methodology

Replicates	public data (1-2 replicates), for data produced for this paper, we used 2 biological replicates/sample. The sequencing depth and alignment statistics are included in Supplementary Table #4
Sequencing depth	69,283,594 p53, input: 156,103,146
Antibodies	public: p53 (Abcam, Cambridge, MA, USA, ab80645, clone DO1), antibodies used in our paper: see supplementary table #3
Peak calling parameters	MACS2 (v2.2.5) -nomodel --keep-dup 1 --call-summits
Data quality	23,366 peaks at MACS2 FDR <0.05
Software	MACS2 (v2.2.5)

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	mESC were trypsinized, quenched with FBS, washed 2x with PBS, and resuspended in 2iLIF media
Instrument	A BD LSRFortessa instrument with lasers for 351nm, 488nm, 561nm, and 640nm
Software	BD FACSDiva 8.0.1
Cell population abundance	Flow cytometry: A minimum of 10,000 cells were quantified for the MERVL-GFP reporter or Rhox5-KI XEN experiment. For rhox5-KI EB experiment, all the cells in a single EB (typicall 5,000-10,000) were analyzed.
Gating strategy	FSC and SSC were used to gate cells and doublet discrimination was performed.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.