**LETTERS**

**DUX-family transcription factors regulate zygotic genome activation in placental mammals**

Alberto De Iaco, Evarist Planet, Andrea Coluccio, Sonia Verp, Julien Duc & Didier Trono

In animal embryos, transcription is mostly silent for several cell divisions, until the release of the first major wave of embryonic transcripts through so-called zygotic genome activation (ZGA)\(^1\). Maternally provided ZGA-triggering factors have been identified in *Drosophila melanogaster* and *Danio rerio*\(^2,3\), but their mammalian homologs are still undefined. Here, we provide evidence that the DUX family of transcription factors\(^4,5\) is essential to this process in mice and potentially in humans. First, human DUX4 and mouse Dux are both expressed before ZGA in their respective species. Second, both orthologous proteins bind the promoters of ZGA-associated genes and activate their transcription. Third, Dux knockout in mouse embryonic stem cells (mESCs) prevents the cells from cycling through a 2-cell-like state. Finally, zygotic depletion of Dux leads to impaired early embryonic development and defective ZGA. We conclude that DUX-family proteins are key inducers of zygotic genome activation in placental mammals.

Dux genes encode double-homeodomain proteins and are conserved throughout placental mammals\(^4,5\). Human DUX4, the intronless product of an ancestral DUXC, is nested within the D4Z4 macrosatellite repeat of chromosome 4 as an array of 10 to 100 units (ref. 6). DUX4, DUXC, and Dux genes from other placental mammals display the same repetitive structure: DUX4 from primates and Afrotheria and DUXC from cows and other Laurasiatheria localize at telomeric or pericentromeric regions, and mouse Dux tandem repeats lie adjacent to a mouse-specific chromosomal fusion point that resembles a subtelomeric structure\(^4,5\).

Overexpression-inducing mutations in DUX4 are associated with facioscapulohumeral muscular dystrophy (FSHD), the third most common muscular dystrophy\(^7,8\), and forced DUX4 production in human primary myoblasts leads to upregulation of genes active during early embryonic development\(^9\). On the basis of this premise, we analyzed publicly available RNA-seq data sets corresponding to this period, focusing on DUX4 and the 100 genes most upregulated in DUX4-overexpressing muscle cells\(^10,11\) (Fig. 1a and Supplementary Table 1). DUX4 RNA was detected from the oocyte to 4-cell (4C) stages, whereas transcripts from putative DUX4 targets emerged on average at the 2-cell (2C) stage and peaked at the 8-cell (8C) stage, as previously defined for human ZGA\(^12\). Transcripts upregulated in DUX4-overexpressing muscle cells\(^11\) were also enriched at the 8C stage (Supplementary Fig. 1a, b), and after clustering of genes according to their patterns of early embryonic expression (Fig. 1b), we delineated

---

**Figure 1** DUX4 promotes transcription of genes expressed during early ZGA. (a) Comparative expression during early human embryonic development of DUX4 (red) and the top 100 genes upregulated after DUX4 overexpression in human primary myoblasts (blue; solid line, average; dashed lines, 95% confidence interval around the mean). Oo, oocyte; Zy, zygote; Mo, morula; Bl, blastocyst. (b) Cluster of genes differentially expressed during early embryonic development were selected from the previously identified subsets of genes (Supplementary Fig. 1a) on the basis of high expression at 4C (top) and 8C (bottom). Blue solid and dotted lines delineate the mean and 95% confidence interval, respectively. (c) Expression of genes from each cluster illustrated in b, after ectopic expression of DUX4 in human primary myoblasts. Lower parts of the panels depict the fold change in expression of genes within these clusters, all randomly distributed along the y-axes, with kernel density plotted in the upper part.

---

School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. Correspondence should be addressed to D.T. (didier.trono@epfl.ch).

Received 2 February; accepted 7 April; published online 1 May 2017; doi:10.1038/ng.3858
Figure 2 DUX4 binds TSSs of genes expressed during early ZGA and activates their expression in hESCs. (a) Average coverage normalized for sequencing depth of the ChIP–seq signal of DUX4 (blue) overexpressed in hESCs in a window of 5 kb from the annotated TSSs of genes belonging to the 2C–4C and 2C–8C clusters from Figure 1b. Total input is represented in gray (line, average; shading, s.e.m.). (b) Fraction of genes belonging to each cluster from Figure 1b with a DUX4 peak within 5 kb of their annotated TSSs. Two-sided Fisher’s exact test was performed to compare maternal versus 2C–4C and 2C–8C (***P = 3.54 × 10−13 and ***P = 2.23 × 10−13, respectively). (c) Average coverage of ChIP–seq signal of DUX4 (blue) overexpressed in hESCs within 5 kb of TFEs of transcripts specifically upregulated at oocyte–4C and 4C–8C transitions. Total input is represented in gray (line, average; shading, s.e.m.). (d) Fraction of TFEs from oocyte–4C (n = 32) and 4C–8C (n = 128) transitions with a DUX4 peak overlapping with the 5’ end. Two-sided Fisher’s exact test was performed to compare 4C–8C versus oocyte–4C TFEs (***P = 4.48 × 10−17). (e) Comparative expression in hESCs of three genes activated at ZGA (ZSCAN4, MBD3L2, and DUXA), and two control housekeeping genes (ACTB and TBP) 24 h after transfection with plasmids expressing LacZ (green squares) or DUX4 (blue circles). Expression was normalized to that of ACTB. Horizontal lines represent the mean. ***P ≤ 0.001, two-sided unpaired t-test.

(i) 1,517 genes whose transcript expression was already detected in oocytes, plateaud up to 4C and abruptly decreased afterward (maternal gene cluster); (ii) 94 genes and 124 genes whose expression started at 2C and peaked at 4C and 8C, respectively, before rapidly decreasing, a pattern consistent with that of early-ZGA genes (2C–4C and 2C–8C gene clusters); and (iii) 1,352 genes expressed only from 4C, peaking at 8C, and then progressively decreasing, as expected for late-ZGA genes (4C–8C gene cluster). Only the two early-ZGA clusters (2C–4C and 2C–8C) were highly enriched in genes upregulated in DUX4-overexpressing myoblasts (Fig. 1c and Supplementary Fig. 1c).

It is technically challenging to reliably analyze chromatin from the very low number of cells that make up an early embryo, but chromatin immunoprecipitation (ChIP)–seq data obtained in DUX4-overexpressing human embryonic stem cells (hESCs) (Fig. 2a,b and Supplementary Fig. 2) and myoblasts9 (Supplementary Fig. 3) showed a marked enrichment of the transcription factor around the annotated transcription start site (TSS) regions of early-ZGA genes (2C–4C and 2C–8C clusters) but not zygotic (maternal) and late-ZGA (4C–8C) genes. Interestingly, several genes were bound not on their annotated TSSs, but on neighboring sequences, and their transcription was found to start near this DUX4-binding site (Supplementary Fig. 4). DUX4 has been demonstrated to drive expression of many of its target genes from alternative promoters11. After examining publicly available single-cell RNA-seq data quantifying the far 5′ ends of transcripts (TFEs) in early human development13, we found that the TFEs of 24 out of 31 early-ZGA genes overlapped with DUX4-binding sites (Fig. 2c,d and Supplementary Fig. 3c,d). DUX4 was also recruited to several groups of transposable elements (TEs), notably endogenous retroviruses such as HERVL, MER11B, and MER11C, whose expression increased at ZGA (Supplementary Fig. 2c). Furthermore, DUX4 overexpression in hESCs led to early-ZGA-gene induction, as previously observed in myoblasts11 (Fig. 2e).

Dux and DUX4 have largely conserved amino acid sequences, particularly within the two DNA-binding homeodomains and the...
Dux is necessary for formation of 2C-like mESCs. (a) Comparative expression of the two alternative transcripts of Dux, Dux (pink) and Gm4981 (orange), with genes (blue) and TEs (MERVL; green) specifically expressed during mouse ZGA. Solid lines, average; dashed lines, 95% confidence interval around the mean. Zygote; e, early; m, middle; l, late; Bl, blastocyst. (b) Single-cell RNA-seq comparison between mESCs sorted for expression of both Tomato and GFP reporters driven by MERVL and Zscan4 promoters, respectively (revealers of 2C-like cells), and the double-negative population. Average gene expression was quantified, and the fold change between positive and negative cells is plotted. Dots are randomly distributed along the y axis. The top plot represents the kernel-density estimate of the middle-2C stage (blue line) and the rest of the genes (gray line). The Dux macrosatellite repeat was deleted in mESCs carrying a MERVL-GFP reporter by CRISPR-Cas9-mediated excision. (c) Fraction of GFP+ cells in wild type (WT) or Dux-deleted cells. (d) RNA-seq analysis of WT and Dux-KO mESC clones. The dot plot displays the average gene expression of three independent clones from each cell type. (e) GFP expression in Dux-KO (blue circles) and WT (green squares) mESC clones carrying an integrated MERVL-GFP reporter and transiently expressing LacZ, DUX4, Dux, or Gm4981 transgenes. Data are shown in log10 scale. (f) RNA-seq analysis of Dux-KO mESC clones transiently expressing Dux or control. The dot plot displays the average gene expression of two independent clones from each cell type. (g) Dux-KO mESCs carrying an integrated MERVL-GFP reporter and transiently expressing a hemagglutinin (HA)-tagged form of Dux were stained for HA, and immunofluorescence was detected by confocal microscopy. DAPI nuclear stain, blue; GFP, green; HA, red. Horizontal bars in c and e represent the mean. ***P < 0.001, two-sided unpaired t-test.
that Dux-depleted embryos presented a major differentiation defect, in which most failed to reach the morula/blastocyst stage and did not exhibit transcriptional changes typical of ZGA, such as the induction of MERVL, Zscan4, and several other tested early-ZGA genes, and the decrease in the Mpo maternal transcript (Fig. 5b,c and Supplementary Fig. 12).
In sum, our data highlight DUX genes as key regulators of early embryonic development. The demonstrated ability of DUX4 to recruit the p300–CBP complex and to induce local chromatin relaxation\(^2\), as well as the mechanism of action of Zelda, a master inducer of ZGA in Drosophila\(^3\),\(^4\), suggests that DUX proteins may act as pioneer factors in transcriptional activation, by opening chromatin around the TSSs of early-ZGA genes, thereby facilitating access by other transcription factors. Nonetheless, the genomic recruitment of pioneer factors such as OCT4, NANOG, and KLF4 may be hindered if heterochromatin marks are overly abundant at their target loci\(^5\). Many mouse ZGA genes are expressed from the long terminal repeats of endogenous retroviruses, which in mESC cells, are typically enriched in repressive marks\(^6\). At any given time, these marks may be relieved in only a small percentage of mESCs in culture. What drives this fluctuation remains to be determined. Furthermore, what controls expression of DUX genes themselves remains to be defined, although the conserved genomic localization of all placental-mammal DUX orthologs close to telomeric and subcentromeric regions suggests that this genomic context, characterized by high levels of repression, might be of primary relevance\(^6,\(^7,\(^22\). The 2C state after maternal-to-zygotic transition is characterized by high levels of repression, might be of primary relevance\(^4,\(^5,\(^22.\)

The genomic recruitment of pioneer factors such as OCT4, NANOG, and KLF4 may be hindered if heterochromatin marks are overly abundant at their target loci\(^5\). Many mouse ZGA genes are expressed from the long terminal repeats of endogenous retroviruses, which in mESC cells, are typically enriched in repressive marks\(^6\). At any given time, these marks may be relieved in only a small percentage of mESCs in culture. What drives this fluctuation remains to be determined. Furthermore, what controls expression of DUX genes themselves remains to be defined, although the conserved genomic localization of all placental-mammal DUX orthologs close to telomeric and subcentromeric regions suggests that this genomic context, characterized by high levels of repression, might be of primary relevance\(^6,\(^5,\(^22\). The 2C state after maternal-to-zygotic transition is characterized by high levels of repression, might be of primary relevance\(^4,\(^5,\(^22.\)

**METHODS**

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

**ACKNOWLEDGMENTS**

We thank T. Macfarlan (NIH) and M.E. Torres-Padilla (Helmholtz Zentrum Muenchen) for sharing reagents and for helpful discussions, and we thank the Transgenic and Gene Expression Core Facility (EPFL) and S. Oppen for technical assistance. This work was financed through grants from the Swiss National Science Foundation, the Gebert-Rüf Foundation, an INGENIUM grant (FP7 MC-ITN INGENIUM 290123), and the European Research Council (ERC 268721 and ERC 306390) to D.T.

**AUTHOR CONTRIBUTIONS**

A.D.I. and D.T. conceived the project, designed the experiments, analyzed the data, and wrote the manuscript; A.D.I., A.C., and S.V. carried out the experiments; E.P. and J.D. performed bioinformatic and statistical analyses.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

---

1. Lee, M.T., Bonneau, A.R. & Giraldes, A.J. Zygotic genome activation during the maternal-to-zygotic transition. *Annu. Rev. Cell Dev. Biol.* 30, 581–613 (2014).
2. Ishiuchi, T. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature* 456, 400–403 (2008).
3. Lee, M.T. et al. Nanog, Pou5f1 and Sox2 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* 503, 360–364 (2013).
4. Leidenroth, A. et al. Evolution of DUX gene macrosatellites in placental mammals. *Chromosoma* 121, 489–497 (2012).
5. 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).
6. Hewitt, J.E. et al. Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.* 3, 1287–1295 (1994).
7. Wijmenga, C. et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat. Genet.* 2, 26–30 (1992).
8. Gabriels, J. et al. Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene* 102, 29–32 (1992).
9. Geng, L.N. et al. DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell* 22, 38–51 (2012).
10. Yan, L. et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1131–1139 (2013).
11. Young, J.M. et al. DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. *PLoS Genet.* 9, e1003947 (2013).
12. Vassena, R. et al. Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development* 138, 3699–3709 (2011).
13. Töhönen, V. et al. Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nat. Commun.* 6, 8207 (2015).
14. Choi, S.H. et al. DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res.* 44, 5161–5173 (2016).
15. Deng, Q., Ramsköld, D., Reinuis, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* 343, 193–196 (2014).
16. Ishiiuchi, T. et al. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* 22, 662–671 (2015).
17. Macfarlan, T.S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57–63 (2012).
18. Eckerley-Maslin, M.A. et al. MERVL/2scan4 network activation results in transient genome-wide DNA demethylation of mESCs. *Cell Rep.* 17, 179–192 (2015).
19. Sun, Y. et al. ZELDA overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. *Genome Res.* 25, 1703–1714 (2015).
20. Schulz, K.N. et al. ZELDA is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Res.* 25, 1715–1726 (2015).
21. Soufi, A., Donahue, G. & Zaret, K.S. Facilitators and impediments of the pluripotency reprogramming factors’ initial engagement with the genome. *Cell* 151, 994–1004 (2012).
22. Perrero, S. & Gasser, S.M. Long-range silencing and position effects at telomeres and centromeres: parallels and differences. *Cell. Mol. Life Sci.* 60, 2303–2318 (2003).
23. van der Maarel, S.M., Tawil, R. & Tapscott, S.J. Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. *Trends Mol. Med.* 17, 252–258 (2011).
24. Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* 534, 652–657 (2016).
25. Maksakova, I.A. et al. Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics Chromatin* 6, 15 (2013).
26. Lu, F., Liu, J., Jiang, L., Yanaguchi, S. & Zhang, Y. Role of Tet proteins in enhancer activity and telomere elongation. *Gene* 560, 2103–2119 (2014).
27. Schoollemmer, J., Pérez-Palacios, R., Climent, M., Guallar, D. & Muniesa, P. Regulation of mouse retroelement MuERV-L/MERVL expression by REX1 and epigenetic control of stem cell potency. *Front. Oncol.* 4, 14 (2014).
28. Walter, M., Teissandier, A., Pérez-Palacios, R. & Bourchis, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife* 5, e11418 (2016).
29. Macfarlan, T.S. et al. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* 25, 594–607 (2011).
Plasmids and lentiviral vectors. The MT2/gag sequence was amplified from the pGL3 plasmid39, and the human PGK promoter was amplified from pRRSIN.c.PPT.R12.PKG-GFP.WPRE30 and was cloned upstream of luciferase in pGL4.20. Supplementary Table 2 shows the primers used to obtain truncations of the MT2/gag sequence. sgRNAs targeting sequences flanking the 5′ and 3′ ends of the Dux-containing macrosatellite repeat were cloned into px459 (version 2) through a standard protocol31. Supplementary Table 2 shows the primers used to clone the sgRNAs. The pLKO.1-puroycin shRNA vector was used for the Trim28 KD30. The pLKO.1 vector was further modified to express a blasticidin-S-deaminase drug-resistance cassette in place of the puromycin N-acetyltransferase. The resulting pLKO.1-blastidin backbone was used to clone shRNAs against the mouse Dux transcript. The sequence of the primers used to clone the Dux shRNA is shown in Supplementary Table 2. The Gm4981 1DNA was cloned from the genome of E14 mESCs, and codon-optimized DUX4 and Dux were synthesized (Invitrogen). Gm4981, DUX4, Dux, and LacZ cDNAs were cloned in the pAIB HIV-1-based transfer vector, which also confers blastidin resistance, by using an In-Fusion HD Cloning Kit (Clontech)31. pmD2-G encodes the vesicular stomatitis virus G protein (VSV-G). The minimal HIV-1 packaging plasmid 8.9NdSB, carrying the MERVL regulatory sequence driving expression of a 3×tagged DUX4 cDNA, was used to transduce Trim28 KO mESC lines.

Production of lentiviral vectors, and transduction and transfection of mammalian cells. Lentiviral vectors were produced by transfection of 293T cells with polyethylenimine (PEI) (Sigma)32. To generate stable KD, mESCs were transduced with empty pLKO.1 vector or vectors for expression of shRNA targeting Trim28 or Dux transcripts30. Cells were selected with 1 µg/ml puroycin or 3 µg/ml blastidin starting 1 d after transduction. hESCs expressing the 5′′-KO mESC lines.

Creation of Dux-KO mESC lines. E14 mESCs containing the MERVL regulatory sequence driving expression of a 3×tagged GFP-PEST were cotransfected with px459 plasmids encoding Cas9, the appropriate sgRNAs and a puromycin-resistance cassette33, through nucleofection (Amaxa P3 Primary Cell 4D-Nucleofector X Kit) was used to engineer mESCs expressing LacZ, DUX4, Dux, and Gm4981.

Luciferase assay. 293T cells or E14 mESCs were cotransfected with the various pGL4.20 derivatives, the Renilla plasmid and the pAIB transfer vector encoding LacZ, Dux, Gm4981, or DUX4, by using Lipofectamine 3000 (Invitrogen). Luciferase activity was quantified 24 h after transfection. Firefly luciferase activity was normalized to the activity of Renilla luciferase. Light emission was measured on a luminescence plate reader.

Immunofluorescence assay. mESC clones expressing an HA-tagged DUX protein were fixed for 20 min with 4% paraformaldehyde, permeabilized for 5 min with 0.1% Triton X-100, and blocked for 30 min with 1% BSA in PBS. Cells were then incubated for 1 h with anti-HA.11 (Covance), anti-NANOG (61628, Active Motif), or anti-SOX2 (39824, Active Motif) antibodies diluted in PBS with 1% BSA. Validation information is available on the manufacturers’ websites. After being washed three times, the cells were incubated with anti-mouse (for HA) or anti-rabbit (for NANOG and SOX2) Alexa Fluor 647-conjugated secondary antibodies for 1 h and washed again three times. Every step until this point was carried out with cells in suspension. Pelleted cells were then resuspended in VECTASHIELD Mounting Medium with DAPI (Vector Laboratories) and mounted on the coverslip. The slides were viewed with a Zeiss LSM700 confocal microscope.

Fluorescence-activated cell sorting (FACS). FACS analysis was performed with a BD FACSscan system. Trim28-KD mESCs containing the MT2/gag-GFP reporter were subjected to FACS sorting with an AriAII instrument (BD Biosciences).

Standard PCR, RT–PCR and RNA sequencing. For the genotyping of Dux WT and KO alleles, genomic DNA was extracted with a DNeasy Blood & Tissue Kit (Qiagen), and the specific PCR products were amplified with PCR Master Mix 2× (Thermo Scientific) combined with the appropriate primers (design in Supplementary Fig. 6a; primer sequences in Supplementary Table 2).

Total RNA from cell lines was isolated with a High Pure RNA Isolation Kit (Roche). cdNA was prepared with SuperScript II reverse transcriptase (Invitrogen). An Ambion Single Cell-to-CT Kit (Thermo Fisher) was used for RNA extraction, cdNA conversion, and cdNA preamplification of 2C-stage embryos. Primers listed in Supplementary Table 2 were used for SYBR Green qPCR (Applied Biosystems). Library preparation and 150-bp paired-end RNA-seq were performed with standard Illumina procedures for the NextSeq 500 platform (GSE94325).

ChIP and ChIP–seq. ChIP and library preparation were performed as previously described30. DUX4-HA ChIP was done with anti-HA.11 (Covance) antibody. Sequencing of Trim28 and H3K9me3 ChIP was performed with the Illumina HiSeq 2500 platform in 100-bp-read runs. Sequencing of DUX4 ChIP was performed with an Illumina NextSeq 500 in 75-bp-paired-end-read runs.

RNA-seq data-set preprocessing. Single-cell RNA-seq of human and mouse early embryonic development (GSE36552 and GSE45719 respectively), single-cell RNA-seq of 2C-like cells (E-MTAB-3058), DUX4 overexpression in mouse myoblasts (GSE45883), and TRIM28-KO (GSE74278) data sets were downloaded from different repositories (NCBI GEO and EBI ArrayExpress)34,35. Reads were mapped to the human genome (hg19) or mouse genome (mm9) with TopHat (v2.0.11)36 in sensitive mode (the exact parameters were: tophat -g 1 –no-novel-juncs –no-novel-indels -G G.sf –transcriptome-index 3 transcriptome –b2-sensitive -o Slocaldir Sindex Streads1 Streads2). Gene counts were generated with HTSeq-count. Normalization for sequencing depth and differential gene expression analysis was performed with Voom37, as implemented in the limma package of Bioconductor38. TEs overlapping exons were removed from the analysis. Counts per TE integrant (genomic loci) were generated with the multiBamCov tool from bedtools software39. Normalization for sequencing depth was performed with Voom, with the total number of reads on genes as the size factor. To compute the total number of reads per TE family, counts on all integrants of each family were summed.

Analysis of single-cell expression data from human and mouse embryonic stages. For every embryonic stage, we identified the genes that had different expression levels compared with those at other stages10, by using a moderated t test (comparing the interest group against every other), as implemented in the limma package of Bioconductor. Genes were selected as being expressed in a specific stage if they had a significant P value (<0.05 after adjustment for multiple testing with the Benjamini and Hochberg method) and an average fold change greater than ten, with respect to the other embryonic stages. We additionally removed all genes exhibiting a 1.1-fold higher expression in any of the embryonic stages compared with the stage analyzed (
Pvalue (<0.05 after adjustment for multiple testing with the Benjamini and Hochberg method) and an average fold change greater than ten, with respect to the other embryonic stages. We additionally removed all genes exhibiting a 1.1-fold higher expression in any of the embryonic stages compared with the stage analyzed (Table 2).

For the genotyping of Dux WT and KO alleles, genomic DNA was extracted with a DNeasy Blood & Tissue Kit (Qiagen), and the specific PCR products were amplified with PCR Master Mix 2× (Thermo Scientific) combined with the appropriate primers (design in Supplementary Fig. 6a; primer sequences in Supplementary Table 2).

Total RNA from cell lines was isolated with a High Pure RNA Isolation Kit (Roche). cdNA was prepared with SuperScript II reverse transcriptase (Invitrogen). An Ambion Single Cell-to-CT Kit (Thermo Fisher) was used for RNA extraction, cdNA conversion, and cdNA preamplification of 2C-stage embryos. Primers listed in Supplementary Table 2 were used for SYBR Green qPCR (Applied Biosystems). Library preparation and 150-bp paired-end RNA-seq were performed with standard Illumina procedures for the NextSeq 500 platform (GSE94325).

ChIP and ChIP–seq. ChIP and library preparation were performed as previously described30. DUX4-HA ChIP was done with anti-HA.11 (Covance) antibody. Sequencing of Trim28 and H3K9me3 ChIP was performed with the Illumina HiSeq 2500 platform in 100-bp-read runs. Sequencing of DUX4 ChIP was performed with an Illumina NextSeq 500 in 75-bp-paired-end-read runs.

RNA-seq data-set preprocessing. Single-cell RNA-seq of human and mouse early embryonic development (GSE36552 and GSE45719 respectively), single-cell RNA-seq of 2C-like cells (E-MTAB-3058), DUX4 overexpression in mouse myoblasts (GSE45883), and TRIM28-KO (GSE74278) data sets were downloaded from different repositories (NCBI GEO and EBI ArrayExpress)34,35. Reads were mapped to the human genome (hg19) or mouse genome (mm9) with TopHat (v2.0.11)36 in sensitive mode (the exact parameters were: tophat -g 1 –no-novel-juncs –no-novel-indels -G G.sf –transcriptome-index 3 transcriptome –b2-sensitive -o Slocaldir Sindex Streads1 Streads2). Gene counts were generated with HTSeq-count. Normalization for sequencing depth and differential gene expression analysis was performed with Voom37, as implemented in the limma package of Bioconductor38. TEs overlapping exons were removed from the analysis. Counts per TE integrant (genomic loci) were generated with the multiBamCov tool from bedtools software39. Normalization for sequencing depth was performed with Voom, with the total number of reads on genes as the size factor. To compute the total number of reads per TE family, counts on all integrants of each family were summed.

Analysis of single-cell expression data from human and mouse embryonic stages. For every embryonic stage, we identified the genes that had different expression levels compared with those at other stages10, by using a moderated F test (comparing the interest group against every other), as implemented in the limma package of Bioconductor. Genes were selected as being expressed in a specific stage if they had a significant P value (<0.05 after adjustment for multiple testing with the Benjamini and Hochberg method) and an average fold change greater than ten, with respect to the other embryonic stages. We additionally removed all genes exhibiting a 1.1-fold higher expression in any of the embryonic stages compared with the stage analyzed (Supplementary Fig. 1a). Notably, through this approach, a gene may be marked as being expressed in more than one stage. Codes are available upon request.

Correspondence between DUX4 overexpression and single-cell expression data from human embryonic stages. For every stage, we classified the genes in four patterns of expression by performing hierarchical clustering (with
Expression of the genes identified with this method was then compared between DUX4- and GFP-overexpressing human myoblast cells. For a gene to be considered differentially expressed, a P value (after multiple testing correction with the Benjamini and Hochberg method) < 0.05 and a fold change greater than two were imposed. A moderated t-test was used as the statistical test, as implemented in the limma package of Bioconductor.

ChIP–seq data processing. The ChIP–seq data set of DUX4 overexpressed in human myoblasts (GSE94325) was downloaded from GEO. Reads were mapped to the human genome assembly hg19 with Bowtie2 by using the sensitive-local mode. SICER was used to call histone-mark peaks. For the peaks that did not represent histone marks, we used MACS (with default parameters) for single-end data and used MACS2 (with the parameters macs2 callpeak -t Schipbam -c $ibam -f BAM -g Sorg -n $name -B -q 0.01 ---format BAMPE) for paired-end data. Both SICER peaks with an FDR > 0.05 and MACS peaks with a score < 50 were discarded. RSAT was used for motif discovery and to compute motif abundance. To compute the percentage of bound TE integrants in each family, we used the bedtools suite.

Coverage plots. ChIP–seq signals on features of interest were extracted from bigWig files beforehand and normalized for sequencing depth (in reads per hundred million). Each signal was then smoothed with a running window average of 75 bp for DUX4, 250 bp for Trim28, and 500 bp for H3K9me3. Finally, the mean and s.e.m. of the signals were computed and plotted for each set of features of interest. Scripts are available upon request.

Pronuclear injection of mouse embryos. Pronuclear injection was performed according to the standard protocol of the Transgenic Core Facility of EPFL. In brief, five-week-old B6D2F1 mice were used as egg donors. Mice were injected with PMSG (10 IU) and then with HCG (10 IU) 48 h later. After females were mated with B6D2F1 males, zygotes were collected and kept in KSOM pregassed under 5% CO2 at 37 °C. Embryos were then transferred to M2 medium and mated with B6D2F1 males, zygotes were collected and kept in KSOM pregassed with PMSG (10 IU) and then with HCG (10 IU) 48 h later. After females were injected, embryos were cultured in KSOM at 37 °C in 5% CO2 for 2–3 days. In each of three independent experiments, five embryos per condition were collected approximately 7 h after the first cell division (2C formation) for qPCR analysis, and the remaining embryos were differentiated.

Sample sizes and statistical tests. We used nonparametric statistical tests (two-sided Wilcoxon test), when we had sufficient sample sizes (low-cell-number qPCR). Otherwise, we used two-sided unpaired t-tests (standard qPCR and FACS). Fisher’s exact test was used to test for differences in proportions in contingency tables.

Data availability. RNA-seq and ChIP–seq data generated in this study have been deposited in the NCBI Gene Expression Omnibus (GEO) database under accession number GSE94325.