

Multiple repeat regions within mouse DUX recruit chromatin regulators to facilitate an embryonic gene expression program

Reviewed Preprint

Revised by authors after peer review.

About eLife's process

Reviewed preprint version 2

July 21, 2023 (this version)

Reviewed preprint version 1

May 19, 2023

Posted to bioRxiv

April 1, 2023

Sent for peer review

March 29, 2023

Christina M. Smith, Edward J. Grow, Sean C. Shadle, Bradley R. Cairns 

Howard Hughes Medical Institute, Department of Oncological Sciences and Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA • Green Center for Reproductive Biological Sciences, Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, TX, USA

 https://en.wikipedia.org/wiki/Open_access

 <https://creativecommons.org/licenses/by/4.0/>

Abstract

The embryonic transcription factor DUX regulates chromatin opening and gene expression in totipotent cleavage-stage mouse embryos, and its expression in embryonic stem cells promotes their conversion to 2-cell embryo-like cells (2CLCs) with extraembryonic potential. However, little is known regarding which domains within mouse DUX interact with particular chromatin and transcription regulators. Here, we reveal that the C-terminus of mouse DUX contains five uncharacterized ~100 amino acid (aa) repeats followed by an acidic 14 amino acid tail. Unexpectedly, structure-function approaches classify two repeats as ‘active’ and three as ‘inactive’ in cleavage/2CLC transcription program enhancement, with differences narrowed to a key 6 amino acid section. Our proximity dependent biotin ligation (BioID) approach identified factors selectively associated with active DUX repeat derivatives (including the 14aa ‘tail’), including transcription and chromatin factors such as SWI/SNF (BAF) complex, as well as nucleolar factors that have been previously implicated in regulating the Dux locus. Finally, our mechanistic studies reveal cooperativity between DUX active repeats and the acidic tail in cofactor recruitment, DUX target opening, and transcription. Taken together, we provide several new insights into DUX structure-function, and mechanisms of chromatin and gene regulation.

eLife assessment

This **useful** study provides a systematic, **solid** mutational analysis to elucidate mechanisms involved in transcriptional activation by the murine DUX protein, DUX is a master transcription factor regulating mammalian early embryonic gene activation and its human homolog DUX4 is also involved in a muscular disease, fascioscapulohumeral dystrophy (FSHD). The data are **solid** and the interpretations of the findings are reasonable. The work will be of interest to colleagues studying early embryonic development or FSHD.

Introduction

Mouse pre-implantation development is a dynamic and complex process, which begins with a single-cell zygote and results in an implanted embryo bearing cells with different lineage specifications and fates. Within this period, numerous molecular processes must be properly completed, which are driven (in part) by gene expression changes and chromatin modifications (1,2). A hallmark of pre-implantation development is major embryonic genome activation (EGA), which occurs at the 2-cell stage in the mouse embryo (3). This critical process includes decreases in maternally-deposited mRNAs and proteins, combined with the onset of both embryo-derived transcription and translation – effectively turning developmental control over to the embryo from maternal resources in the egg (4). During EGA, chromatin undergoes considerable remodeling to enable enhancer and promoter accessibility, allowing for the onset of major embryonic transcription, and the marking of promoters and enhancers for future regulation (5–7).

Mouse embryonic stem cells (mESCs) are derived from the inner cell mass (ICM) of the blastocyst, and have enabled studies focused on the characterization of the blastocyst and downstream embryonic differentiation (8). In addition, the mESC system has been extended to examine properties of earlier cleavage-stage totipotent-related cells. Specifically, a small (<1%) population of spontaneously fluctuating mESCs has been identified, which closely resemble cells within the 2-cell embryo, termed 2-cell-like cells (2CLC) (9–12). Our lab and others showed that the expression of *Dux*, which encodes the double homeodomain transcription factor DUX, is sufficient to confer this 2C-like state to mESCs (12–14). However, we currently have an incomplete understanding regarding how DUX establishes the 2C transcriptome. DUX4, the human *Dux* ortholog is similarly expressed during EGA, which occurs at the 4-cell stage in humans. Likewise, DUX4 expression in human ESCs converts these cells to transcription programs resembling the 4–8 cell human cleavage-stage embryo (13,15–17). DUX4 has additionally been identified as the causal gene in facioscapulohumeral dystrophy (FSHD), leading to the protein being intensely studied to identify its pathological traits (18–20).

Mouse DUX is known to be targeted to specific sites in the genome (including genes and MERVL LTRs) by its two N-terminal homeodomains, whereas its C-terminal region drives transcriptional activation (12,21,22). However, little is known about the nature of this C-terminal region, which is much more divergent among DUX orthologs than are the homeodomains (23,24). One possible contributor to transcriptional activation involves the interaction of the C-terminus of DUX4 with the related histone acetyltransferases p300 and CBP (22,25). However, these or other interactions have not been identified with mouse DUX. Regarding chromatin remodeling during EGA, the abundant ATP-dependent chromatin remodeler SWI/SNF (BAF) complex plays a central role (26,27). Specifically, SMARCA4, the catalytic ATPase within SWI/SNF complex, is required for important for EGA in the 2-cell mouse embryo (28). SMARCC1 is a core structural scaffold for SWI/SNF complexes, and its absence results in loss of complex activity *in vivo* (29). However, possible physical links between mouse DUX and the chromatin remodeler SWI/SNF complex during EGA or during the mESC to 2C-like transition is currently unknown.

DUX activates about 25% of EGA genes and DUX expression is sufficient to revert mESCs to an earlier totipotent stage (12). Consistent with DUX impacting early development, *Dux*^{-/-} crosses yield smaller litter sizes (21). However, as viable *Dux*^{-/-} pups are generated, factor(s) redundant with DUX appear to contribute to DUX target activation and chromatin reprogramming in 2-cell embryos (21,30). Recently, oocyte-specific homeobox 4 (OBOX4), a factor highly expressed during EGA and in 2C-like mESCs, has been identified as partly redundant with DUX in 2-cell embryos (31). Taken together, DUX and OBOX4 appear to cooperate to ensure proper EGA in the early mouse embryo.

In this study, we sought to better characterize DUX domains, identify proteins interacting with DUX domains, and understand how DUX domains contribute to DUX target gene activation, DUX promoter recruitment, and chromatin accessibility at binding sites. Specifically, we reveal a unique C-terminal repeat structure in mouse DUX, and characterize two functionally distinct repeat types (termed ‘active’ and ‘inactive’) and cooperativity of the active repeats with the C-terminal acidic tail. Additionally, we identify DUX domains that interact with the chromatin remodeler subunit SMARCC1, which alongside the ability of DUX to confer histone acetylation, reveals new mechanistic insight into how the C-terminus of DUX confers transcriptional activation.

Results

The mouse DUX C-terminus has five conserved repeats followed by a 14 amino acid ‘tail’

Curiously, mouse DUX protein (673 amino acids (aa)) is considerably longer than DUX orthologs such as human DUX4 (424 aa). Through manual inspection and alignments of mouse DUX with other orthologs, we identified five protein repeats within DUX of ~100 aa each, followed by a single 14aa highly-acidic C-terminal ‘tail’ (**Figure 1a** [↗](#)). However, human DUX4 contains only a single copy of this ‘repeat’ domain, which is likewise followed by a 14aa tail that closely resembles the acidic tail in mouse DUX (**Figure 1b** [↗](#)). Prior work has revealed that the acidic DUX4 C-terminal tail helps activate transcription, in part by binding to CBP and p300 ([22](#) [↗](#)). To determine whether other placental mammals (e.g. rat, bovine, and human) contain a C-terminal repeat structure similar to mouse DUX, we conducted pairwise alignments using the algorithm BLAST, which revealed repeats only in the mouse lineage (**Figure 1c** [↗](#), **Supplementary Figure 1a,b,c** [↗](#)). Interestingly, although all five repeats generate highly significant pairwise similarity scores, two pairs with the highest similarity scores involve the C3-C5 pair and the C2-C4 pair – results supported by our earlier phylogenetic analyses (**Figure 1c** [↗](#)). Furthermore, the human DUX4 C-terminal region displays significant similarity to both the 5th and 2nd repeat of mouse DUX (E-Values 0.022 and 0.03, respectively). The bovine DUXC C-terminal region was most similar to the 5th repeat whereas the rat C-terminus was most similar to the 1st repeat of mouse DUX (**Supplementary Figure 1d** [↗](#)). However, as the comparisons DUX C-terminal ‘single repeat’ orthologs to the mouse repeats yield different mouse repeats with highest similarity, these results do not point clearly to a particular repeat representing the parental/original repeat in mouse DUX.

Next, we assessed the contribution of mouse DUX repeats to transcriptional activity by examining *Dux*^{-/-} mESCs containing a MERV1::GFP reporter. The long-terminal repeat (LTR) of MERV1 is bound by DUX, and MERV1::GFP is commonly used to assess DUX activation and conversion of mESCs to a 2CLC state ([9](#) [↗](#), [12](#) [↗](#)). First, *Dux*^{-/-} mESCs were transiently transfected with mCherry-tagged DUX repeat domain constructs encoding zero to five C-terminal repeats, without the 14aa tail (**Figure 1d** [↗](#)). Cells were induced with doxycycline for 18 hours, gated on mCherry positivity, and MERV1::GFP expression was quantified by flow cytometry (**Supplementary Figure 1e** [↗](#)). Expression of each derivative was confirmed by Western analysis (**Supplementary Figure 1f** [↗](#)). The flow cytometry results reveal that the first two repeats have minimal activity, but with the addition of the third repeat, low MERV1::GFP expression is observed. Activity is further improved by the addition of the fourth and fifth repeat. The results also support a potential role for the 14aa ‘tail’ in activation, as its omission modestly lowered MERV1::GFP expression (**Figure 1d** [↗](#)). Thus, our initial deletion series revealed activity with repeats 3-5, but not with repeats 1-2 alone. However, this experiment did not determine whether the repeats act in a simple additive manner, or if individual repeats confer distinct transcriptional activity in a singular manner, or how they cooperate with the acidic tail.

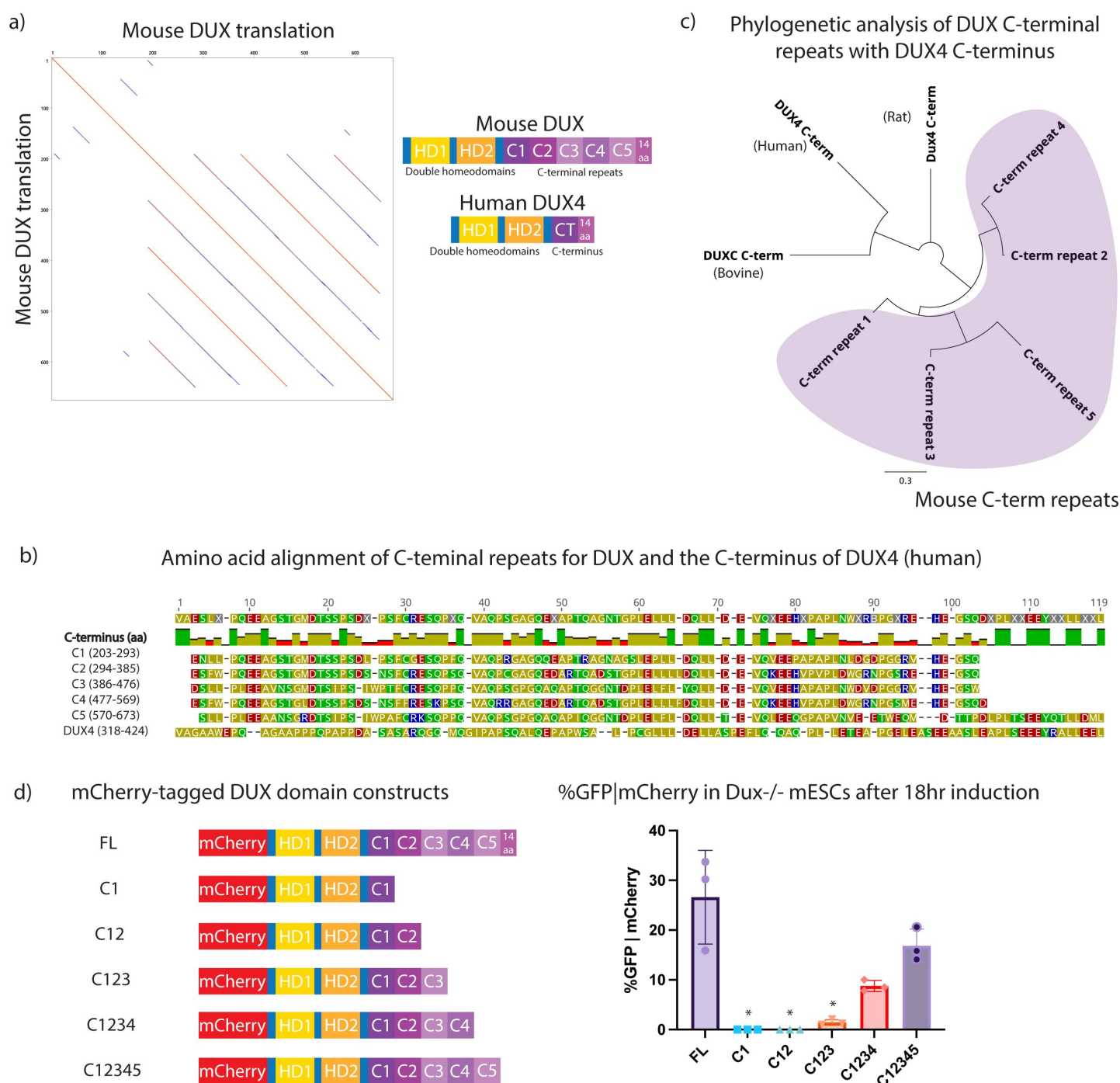


Figure 1

Identification of five repeats within the mouse DUX C-terminal region, and truncation analysis of transcriptional activity.

- a. (Left) Mouse DUX aligned to itself, revealing five repeats. (Right) Diagram of mouse DUX and human DUX4.
- b. Amino acid alignment of mouse DUX C-terminal repeats and comparison to human DUX4. A 14 amino acid tail follows both the terminal mouse C5 repeat and the sole DUX4 repeat.
- c. Phylogenetic analysis showing amino acid alignments of mouse DUX C-terminal repeats vs. the C-terminus of bovine DUXC, the C-terminus of rat DUX4, and the C-terminus of human DUX4. Line length unit is substitutions per base. Purple shading denotes mouse DUX C-terminal repeats.
- d. (Left) Schematic of constructs used for flow cytometry: mCherry-tagged full-length DUX (FL), homeodomains with C1, C12, C123, C1234, C12345. (Right) Flow cytometry data for MERVL::GFP reporter given mCherry expression in *Dux*^{-/-} mESCs following 18hr expression of indicated constructs. *p-value < 0.05, student's t-test. (n=3 biological replicates)

DUX C-terminal repeats C3 and C5 contribute to transcriptional activation when combined with the 14aa tail

We next explored the contributions of individual repeats, along with their cooperation with the C-terminal acidic ‘tail’. Here, we generated N-terminal mCherry-tagged DUX derivatives with varying numbers of repeats either with or without the 14aa tail, as well as with or without single or both homeodomains (**Supplementary Figure 2a** [↗](#)). All constructs contain a GGGGS₂ linker present in constructs with a single internal deletion and GGGGS₂ and GAGAS₂ with constructs with two internal deletions. First, as expected, derivatives lacking either homeodomain failed to activate the MERVL::GFP reporter (**Supplementary Figure 2a** [↗](#)). Second, combinations involving C1, C2 or C4 with the acidic ‘tail’ lack activation of the MERL::GFP reporter, whereas combinations of C3 or C5 with the ‘tail’ display potent reporter activation (**Figure 2a** [↗](#)). The latter combinations require the 14aa tail as constructs lacking the tail, but containing the C3 or C5 repeat alone, did not activate the reporter. Notably, a construct containing C1 and C2 with the 14aa tail was inert, whereas C1, C2 and C3 with the 14aa tail was highly active (**Supplementary Figure 2a,b** [↗](#)). Together, these results suggest that strong DUX target gene activation requires a combination of two C-terminal domains: a transcriptionally ‘active’ repeat (C3 or C5) and the 14aa tail (**Figure 2b** [↗](#)).

To determine the amino acids selectively involved in DUX target activation within the active C3 and C5 repeats, chimeric constructs between C1 (a transcriptionally ‘inactive’ repeat) and C3 (a transcriptionally ‘active’ repeat) were made, all containing the 14aa tail. The cutoff/transition point between chimera regions occurred in 3 aa increments along the 2 repeats (**Figure 2c** [↗](#), **Supplementary Figure 2c** [↗](#)). This narrowed down the functional difference to a 6 aa region which is conserved between repeats C3 and C5, but not conserved in C1, C2, and C4. Importantly, replacement solely of this 6 aa sequence from the active C3 repeat into inactive C1 repeat (termed C1DPLELF+14aa) conferred robust activation of the MERVL::GFP reporter (**Figure 2d** [↗](#)).

In order to expand these analyses to global transcription and all DUX targets (beyond the MERVL::GFP reporter), we performed RNA-seq on full-length (FL) DUX and selected chimera constructs. This approach was designed to test whether the ‘active’ constructs might differ in their gene targets (**Figure 2e** [↗](#) and **Supplementary Figure 2d** [↗](#)). First, the PCA plot for these samples separated samples largely into two groups – active versus inactive – with one sample in an intermediate position (**Figure 2e** [↗](#)). Interestingly, the DUX constructs which activate the MERVL::GFP reporter (FL, C1C3a, and C3+14a) all showed strong activation of DUX target genes, whereas the ‘inactive’ constructs that do not activate the MERVL::GFP reporter (C1+14aa, C1C3c, and mCherry alone) did not activate DUX target genes (**Figure 2f** [↗](#)). Notably, the intermediate position was occupied by the C1C3b derivative, reinforcing the notion that the 6aa region is important for regulating target activation. K-means clustering of the RNA-seq for the chimera constructs at DUX target generated three clusters of genes: the first two of which largely contained genes activated by FL DUX, whereas the cluster #3 contains genes that have increased expression in DUX domain derivatives compared to FL DUX. The ‘intermediate’ derivative C1C3b, which is a hybrid of an ‘active’ and an ‘inactive’ repeat, was capable of activating the majority of DUX targets in cluster #2, but failed to activate those in cluster #1, whereas it activated (or failed to attenuate) genes in cluster #3. Genes activated by C1C3b in cluster #3 generate GO terms such as ‘upregulated include synapse assembly’, ‘sex differentiation’, ‘reproductive structure development’, and ‘neuron differentiation’. Here, cluster 3 shows examples of known DUX target genes that have a higher relative expression in C1C3b compared to full-length DUX include *Obox1*, *Olfr1277*, *Obox2*, and *Zfp296*. These combined observations, especially those involving derivatives that cause the precocious activation of genes in cluster #3, raise the interesting possibility that the ‘inactive’ repeats may serve to attenuate the active repeats at particular targets, such as *Obox* genes. Finally, when considering amino acid similarity between the five repeats, we note that C3 and C5 are most similar to one another, compared to C1, C2, and C4 (**Supplementary Figure 2e** [↗](#)).

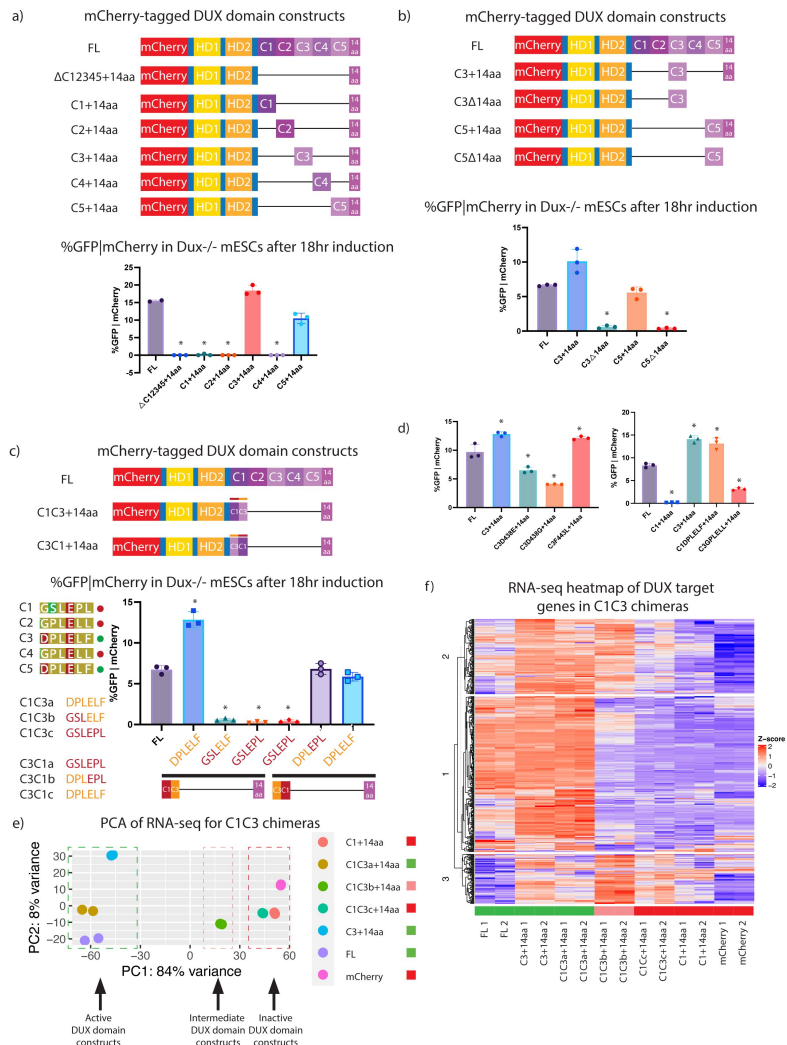


Figure 2

Structure function-analysis of DUX by 2C-like cell conversion.

- a. (Top) Schematic of constructs used for flow cytometry: mCherry-tagged FL, ΔC12345+14aa, C1+14aa, C2+14aa, C3+14aa, C4+14aa, and C5+14aa. (Bottom) Flow cytometry for MERVL::GFP reporter given mCherry expression in *Dux*^{-/-} mESCs with 18hr overexpression of indicated constructs. *p-value < 0.05, student's t-test. (n=3 biological replicates). Constructs with a single black line have the linker GGGS₂ and constructs with two black lines have linkers with GGGS₂ and GAGAS₂, respectively.
- b. (Top) Schematic of constructs used for flow cytometry: mCherry-tagged FL, C3+14aa, C3Δ14aa, C5+14aa, C5Δ14aa. (Bottom) Flow cytometry for MERVL::GFP reporter given mCherry expression in *Dux*^{-/-} mESCs following 18hr expression of indicated constructs. *p-value < 0.05, student's t-test. (n=3 biological replicates)
- c. (Top) DUX domain chimera construct design for C1-C3 fusions with (top) schematic of constructs and (bottom) specific chimera cut-offs for C1-C3 a,b,c (red to orange and C3-C1 a,b,c (orange to red) constructs. (Bottom) Schematic of DUX domain chimera constructs used for flow cytometry: mCherry-tagged full-length DUX (FL), C1C3a, C1C3b, C1C3c, C3C1a, C3C1b, C3C1c. Symbols adjacent to alignments: red circle denotes transcriptionally inactive repeats and the green circle denotes transcriptionally active repeats. (Right) Flow cytometry for MERVL::GFP reporter given mCherry expression in *Dux*^{-/-} mESCs following 18hr expression of indicated constructs. *p-value < 0.05, student's t-test. (n=3 biological replicates)
- d. Flow cytometry for MERVL::GFP reporter given mCherry expression in *Dux*^{-/-} mESCs following 18hr expression of point mutation constructs: (Left) FL-DUX, C3+14aa, C3 D438E+14aa, C3 D438G+14aa, and C3F443L+14aa. (Right) FL-DUX, C1+14aa, C3+14aa, C1DPLELF+14aa (substitution of C3 amino acids into C1+14aa), and C3GPLELL+14aa (substitution of C2 and C4 6aa sequence into C3+14aa). *p-value < 0.05, student's t-test. (n=3 biological replicates)
- e. Principal component analysis (PCA) of RNA-seq analysis for 18hr overexpression of DUX domain chimera constructs, full-length DUX, and mCherry alone (n=2). Colored squares beside sample legends denote the inability to activate MERVL::GFP reporter (red) or ability to activate MERVL::GFP reporter (green).
- f. Heat map of RNA-seq at DUX target genes (n=456) from 18hr expression of DUX domain chimera constructs, full-length DUX, and mCherry alone (n=2). Colored rectangles above samples denote inability to activate MERVL::GFP reporter (red) or ability to activate MERVL::GFP reporter (green).

Chromatin alterations at DUX binding sites are associated with a transcriptionally active DUX repeat

To investigate the chromatin landscape effects conferred by Dux derivative expression, we performed CUT&Tag in *Dux*^{-/-} mESCs on selected Dux domain constructs: FL, C12345 Δ 14aa, C3+14aa, C3 Δ 14aa, Δ C12345+14aa, and mCherry alone. Because all of the constructs had the same N-terminal mCherry tag, an mCherry antibody was used throughout. We isolated and analyzed clonal stable cell lines which expressed species of the expected protein size and demonstrated comparable expression to one another (**Figure 3a**). We note that mCherry-DUX protein is expected to run at roughly 100kDa, but runs anomalously on polyacrylamide gels – at 150kDa. This is likely due to a property of the C-terminal repeats as a DUX derivative consisting of the homeodomains alone runs at the expected molecular weight. To benchmark DUX CUT&Tag against previous chromatin occupancy data, we first generated a CUT&Tag dataset for mCherry-tagged full-length DUX protein, and found that it overlapped highly with previously published HA-tagged DUX ChIP-seq results (**Supplementary Figure 3a**). Extension of CUT&Tag to the DUX derivatives revealed their moderate-to-high occupancy at DUX sites, except for C3 Δ 14aa, which displayed reduced occupancy at DUX binding sites (**Figure 3b, 3c**). Next, we determined the ability of our DUX domain constructs to alter H3K9ac, as H3K9ac globally increases in 2C-like cells (32). Notably, H3K9ac occurs at DUX binding sites correlating with the transcriptional activity of each Dux construct (**Figure 3d,e**, **Supplementary Figure 3b**). Taken together, our data demonstrate that the 14aa acidic tail is not strictly required for the occupancy of DUX targets. Furthermore, our data suggest that the 14aa tail is important for conferring H3K9ac at bound sites and might also be involved in retention of DUX at target sites since removal of the 14aa tail (in the context of a single active repeat) diminished target occupancy.

Proximity-labeling with full-length DUX reveals interactions with proteins involved in chromatin de-repression

Next, we identified proteins that interact with DUX derivatives using the BioID approach, which identifies proteins that interact directly, indirectly, or transiently with a protein of interest (33). Here, we utilized BioID to define protein interactions with DUX in the 2C-like cellular environment. To accomplish this, we fused BirA*, a promiscuous biotinylating enzyme, to the N-terminus of full-length DUX. When biotin is added to the growth media, BirA* will biotinylate lysine residues of proteins within a 10nm radius. Our approach involved the isolation of biotinylated proteins and the use of liquid chromatography mass spectrometry (LC MS/MS) to identify associated proteins in a quantitative manner (**Figure 4a**). Our experimental design compared expression of BirA*-DUX (full-length) in mESCs (which creates a 2C-like state) to expression of unfused BirA* alone (control) expressed in a 2C-like environment, in order to mimic a DUX-expressing 2C-like proteome in both conditions (**Figure 4b**). Here, we created clonal stable cell lines harboring our constructs, and expressed the proteins for 12hr with 0.25 μ g/mL doxycycline and 5 μ M biotin. DESeq2 analysis of the sum intensity values from LC MS/MS revealed clustering between replicates (**Supplementary Figure 4a**) and a volcano plot revealed an enrichment in protein interactions for BirA*-DUX compared to BirA* alone, including those involved in chromatin de-repression such as the SWI/SNF (BAF) complex members SMARCC1 and SMARCA5 (**Figure 4c, 4d**). In addition to DESeq2 analysis, which is normally used for RNA-seq data, we applied the SAINTexpress package, which is designed for protein interactome analysis, to additionally analyze this dataset. DESeq2 appeared to be slightly less stringent than SAINTexpress analysis, with a larger number passing statistical threshold ($p_{adj} < 0.05$) (**Supplementary Figure 4b**). SAINTexpress uses a statistical score called SAINTscore ranging from 0-1, with a significance threshold established as ≥ 0.73 , which parallels an estimated protein-level Bayesian FDR of < 0.05 (34).

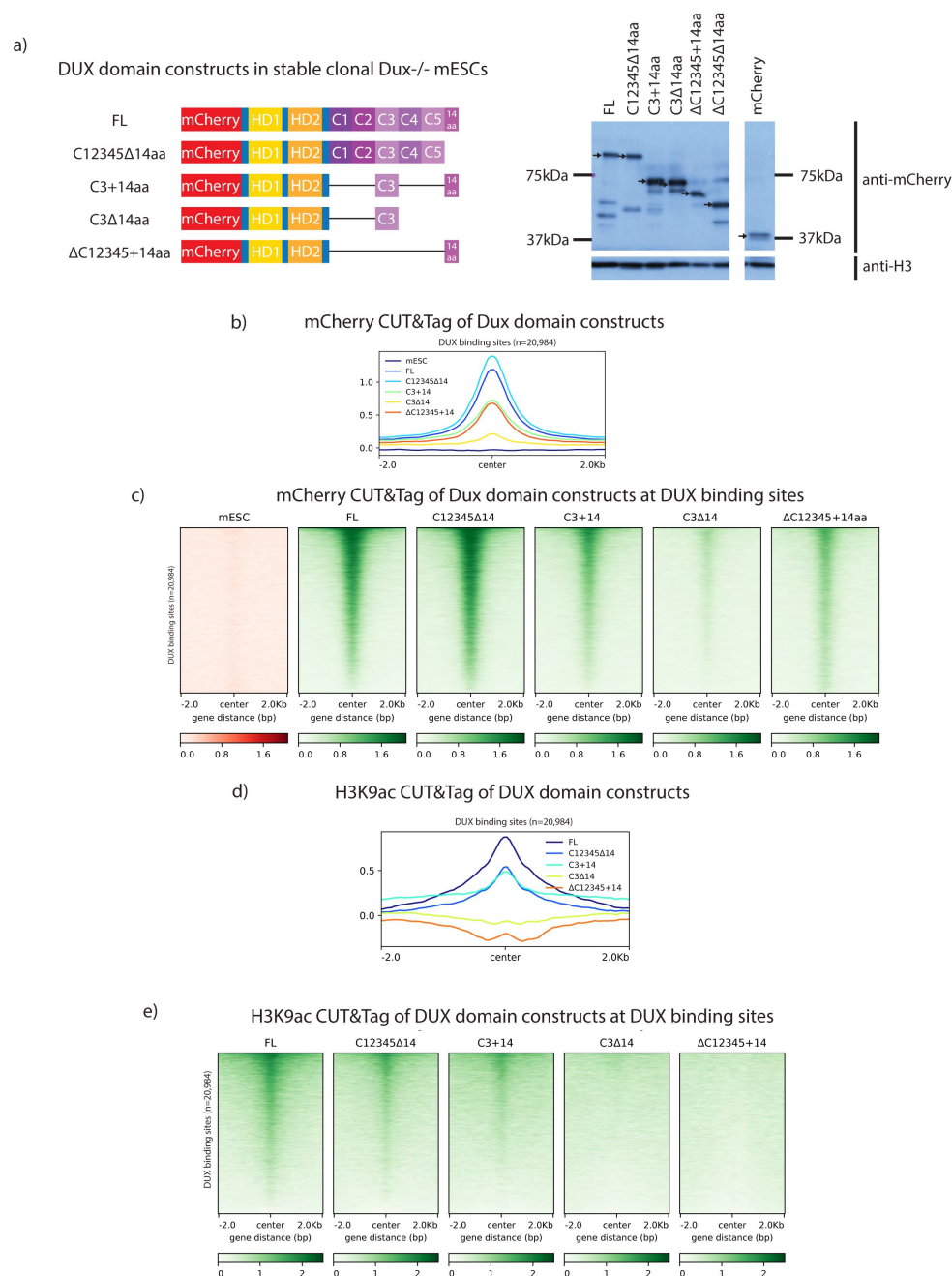


Figure 3

Chromatin modifications at DUX binding sites requires an active DUX repeat.

a. (Left) Schematic of mCherry-tagged DUX domain constructs in clonal *Dux*^{-/-} mESCs used for CUT&TAG experiments. (Right) anti-mCherry Western blot for indicated constructs after 18hr expression. Black arrows indicate construct of interest. Predicted protein sizes are 100 kDa (FL), 95 kDa (C12345Δ14aa), 65 kDa (C3+14aa), 59 kDa (C3Δ14aa), 48 kDa (ΔC12345+14aa), and 25 kDa (mCherry).

b. mCherry CUT&TAG class average map centered at DUX binding sites after 12-hour overexpression of Dux constructs illustrated in [Figure 2a](#) (n=2 biological replicates).

c. mCherry CUT&TAG heatmap centered at DUX binding sites after 12-hour expression of Dux constructs illustrated in [Figure 2a](#) (n=2 biological replicates).

d. H3K9ac CUT&TAG class average map centered at DUX binding sites after 12-hour expression of Dux constructs illustrated in [Figure 2a](#) (n=2 biological replicates).

e. H3K9ac CUT&TAG heatmap centered at DUX binding sites after 12-hour expression of Dux constructs illustrated in [Figure 2a](#) (n=2 biological replicates).

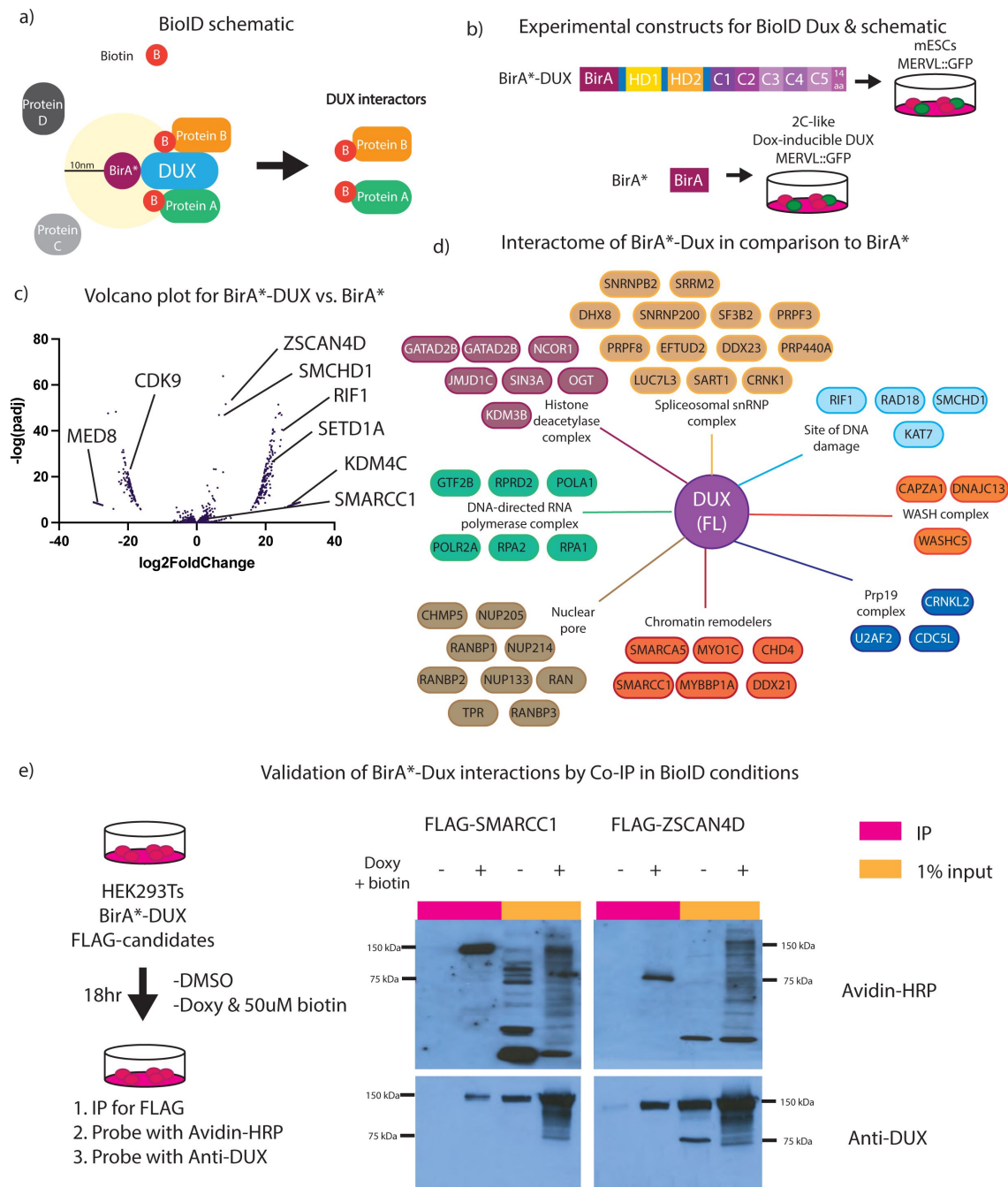


Figure 4

BioID for full-length DUX reveals interaction with proteins involved in chromatin de-repression.

- Experimental concept for BioID (proximity labeling assay) for BirA*-DUX to identify candidate DUX interacting proteins.
- Experimental setup for full-length DUX BioID. The control construct, BirA*-only, transfected into 2C-like mESCs, and BirA*-DUX, transfected into mESCs (n=2 biological replicates).
- Volcano plot using DeSeq2 analysis for BirA*-only vs. BirA*-DUX with log2FoldChange on the x-axis and -log(padj) on the y-axis (n=2 biological replicates).
- Interactome map of proteins enriched to interact with full-length DUX, vs. BirA*-only, split into specific cellular components and protein complexes (n=2 biological replicates). Full dataset available in Supplementary Table 1.
- (Left) Experimental setup for BioID-like experiment in HEK283T cells with BirA*-DUX and flag-tagged candidates SMARCC1 and ZSCAN4D. (Right) Co-immunoprecipitation of FLAG and Western blot after 18-hour expression of transiently transfected FLAG-SMARCC1 or FLAG-ZSCAN4D and BirA*-DUX in HEK293Ts (n=3 biological replicates). Predicted protein size for BirA*-DUX is 125 kDa.

SWI/SNF subunit SMARCC1 and the transcription factor ZSCAN4D, two statistically significant hits from both analyses, were then selected for validation by testing for interaction *in vivo* via immunoprecipitation. We expressed N-terminally FLAG-tagged versions of SMARCC1 and ZSCAN4 in HEK293 cells co-expressing the BirA*-DUX fusion in the presence of 50 μ M biotin (50 μ M biotin for BioID was empirically determined as the optimal dose in HEK293Ts). SMARCC1 and ZSCAN4 were immunoprecipitated with anti-FLAG conjugated beads, and analysis with an avidin-HRP blot showed they are both biotinylated and co-IP with DUX protein, validating BioID protein interactions (**Figure 4e**). Additionally, CHAF1A, an additional BioID hit, was also validated through FLAG-tag co-IP and DUX protein interaction, however KDM4C did not validate, suggesting that this hit was a false positive (**Supplementary Figure 4c**), and highlighting the importance of validating the BioID results.

Lastly, it has been previously reported that human DUX4 interacts with p300 ([22](#), [25](#)), however p300 was not identified as a significant interactor in our dataset. This could be due to differences in the interactomes of human DUX4 and mouse DUX, or instead be technical, due perhaps to a lack of sensitivity to detect P300 peptides. To determine if mouse DUX similarly interacts with P300, we performed a BioID-like experiment in mESCs, which revealed p300 as modestly biotinylated, suggesting that a portion of p300 comes in close proximity with DUX, but was not annotated in the dataset, representing a possible false negative (**Supplementary Figure 4d**). In summary, our BioID experiment included chromatin- and transcription-related factors, a portion of which were validated using co-immunoprecipitation, which support prior observations of a strong impact of DUX occupancy on chromatin opening and transcriptional activity.

Transcriptionally active DUX C-terminal repeat C3 binds to proteins involved in chromatin remodeling

Our work above presents the opportunity to identify proteins that interact differentially with a transcriptionally ‘active’ DUX derivative compared to a highly similar ‘inactive’ DUX derivative. Here, we separately N-terminally tagged our C1+14aa and C3+14aa derivatives with BirA*, to represent transcriptionally ‘inactive’ and ‘active’ DUX constructs, respectively (**Figure 5a**). The experimental design and setup was identical to that described in **Figure 4a**, but the mESC line utilized bore a *Dux*^{-/-} genotype. Furthermore, in place of the BirA* alone control, the inactive BirA*-C1 (BirA* fused to C1+14aa) was used as the control construct for comparison to the active BirA*-C3 (BirA* fused to C3+14aa).

PCA and clustering analysis reveals that the three replicates of the ‘active’ C3 derivative group together and are more similar to each other (via clustering) than are the C1 replicates, and vice versa (**Supplementary Figure 4e,f**). Using SAINTexpress, we observed many more statistically-enriched (SAINTscore ≥ 0.73) protein interactions occurring with the DUX C3 ‘active’ repeat compared to the ‘inactive’ C1 repeat. Specifically, we observed 12 proteins with the C1 derivative, and 154 proteins with C3 (**Figure 5b**). For C3 derivative interactors, GO-term analysis identified two overarching functional categories: chromatin and the nucleolus. Within the chromatin category, proteins within the NuRD, SWI/SNF and ISWI chromatin remodeling complexes were prominent (**Figure 5c**). Notably, proper formation of the nucleolus has been found to be required for Dux repeat locus silencing ([35](#), [36](#)), and our data strongly suggests a direct physical interaction with nucleolar components, as DUX can bind to its own genetic locus ([37](#)).

To further investigate chromatin-related interactors of interest, we performed co-IPs for FLAG-tagged SMARCC1 (BAF155) (described above) using selected DUX domain derivatives. First, SMARCC1 binding to DUX appears only slightly diminished with the loss of the 14aa tail (**Figure 5d**). SMARCC1 also binds well to a DUX derivative containing both an active C3 repeat and the acidic tail (C3+14aa), and binding is largely maintained when the tail is removed (C3 Δ 14aa). Finally, a major reduction in SMARCC1 binding is observed when all repeats are removed and only the acidic tail remains. This suggests a hierarchy of interaction for DUX and SMARCC1: an ‘active’

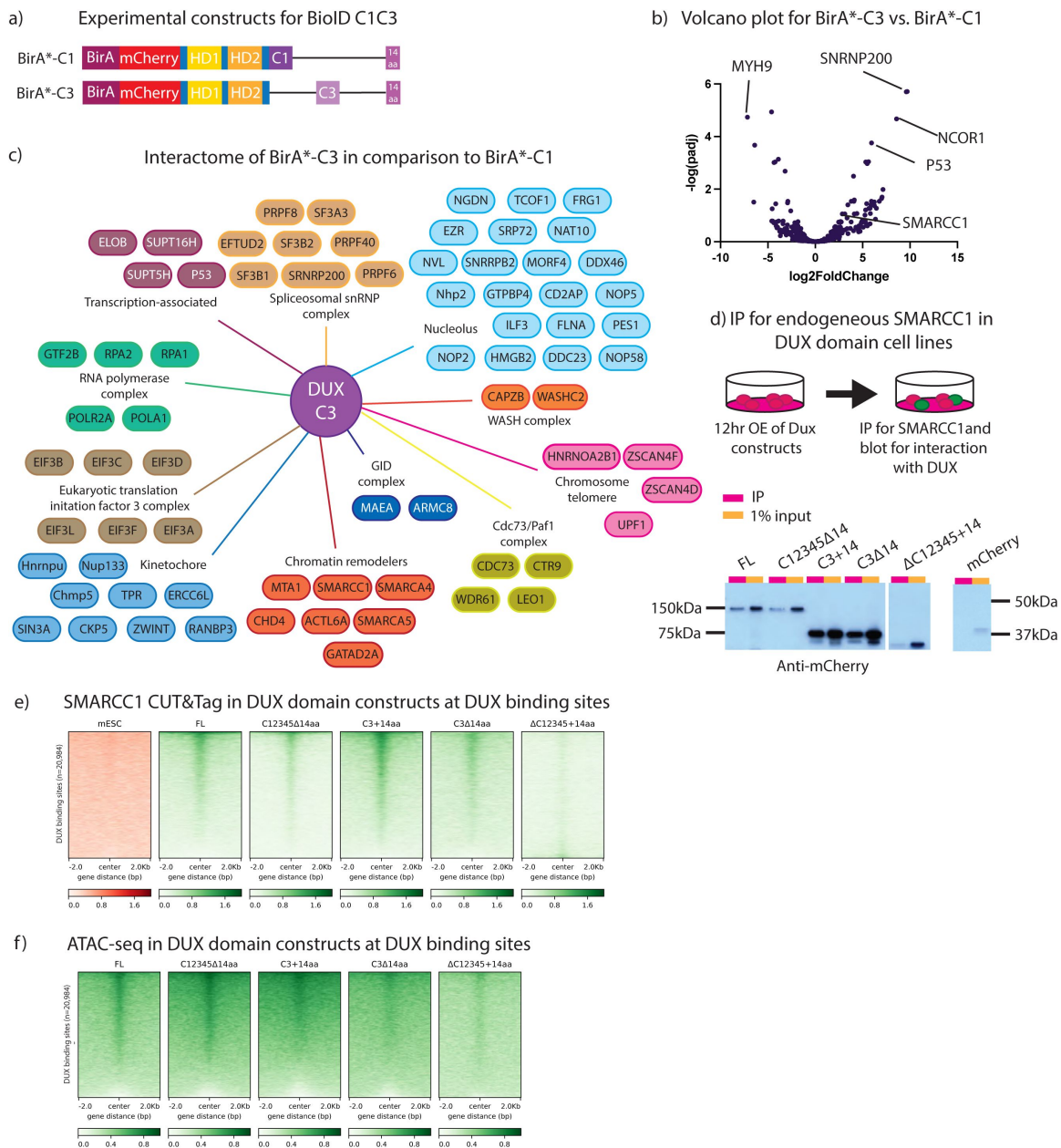


Figure 5

BioID for active C-terminal DUX repeat reveals interactors, including Smarcc1.

- Experimental constructs BioID of individual DUX repeats. Experimental constructs are N-terminally labelled BirA* linked to either mCherry-HDs-C1+14aa or mCherry-HDs-C3+14aa.
- Volcano plot using DeSeq2 analysis for BirA*mC-C1+14aa and BirA*mC-C3+14aa with log2FoldChange on the x-axis and -log(padj) on the y-axis (n=3 biological replicates).
- Interactome map of proteins enriched with BirA*mC-C3+14aa vs. BirA*mC-C1+14aa, split into specific cellular components and protein complexes (n=2 biological replicates). Full dataset available in Table 2.
- Co-immunoprecipitation of endogenous SMARCC1 after 12hr overexpression of DUX domain constructs in stable clonal cell lines illustrated in **Figure 2a** (n=3 biological replicates). Predicted protein sizes are 100 kDa (FL), 95 kDa (C12345Δ14aa), 65 kDa (C3+14aa), 59 kDa (C3Δ14aa), 48 kDa (ΔC12345+14aa), and 25 kDa (mCherry).
- SMARCC1 CUT&TAG heatmap at DUX binding sites after 12-hour overexpression of Dux constructs illustrated in **Figure 2a** (n=2 biological replicates).
- ATAC-seq heatmap at DUX binding sites after 12-hour overexpression of Dux constructs illustrated in **Figure 2a** (n=2 biological replicates, except full-length DUX only has 1 biological replicate).

repeat alone binds SMARCC1 whereas the acidic ‘tail’ alone does not bind SMARCC1, however the acidic tail slightly improves SMARCC1 interaction when active and inactive repeats are both present (full-length DUX).

Lastly, we wanted to determine whether the SWI/SNF (BAF) component SMARCC1 is recruited to DUX binding sites by DUX, and if so, which DUX domains confer this function. To address this, we performed CUT&Tag for endogenous SMARCC1 in the DUX derivatives and found, similar to our co-IP results, that SMARCC1 is recruited to DUX binding sites by DUX derivatives that contain active C-terminal repeats but not by the 14aa tail (**Figure 5e** [↗](#)). Thus, binding and recruitment of SWI/SNF to DUX targets *in vivo* appears to rely on interaction with active DUX C-terminal repeats.

To continue investigating DUX interaction with SWI/SNF, we performed ATAC-seq in the DUX derivatives to determine whether SMARCC1 recruitment to DUX binding sites correlates with chromatin accessibility changes (**Figure 5f** [↗](#), Supplementary Figure 4g). The results suggest a hierarchy of interaction similar to that observed in our SMARCC1 CUT&Tag experiments. At DUX binding sites, full-length DUX leads to increased opening, as do C12345 Δ 14aa and C3+14aa, the two derivatives which confer transcriptional activity. However, DUX derivatives which lack transcriptional activity or have decreased binding to SMARCC1 display a corresponding decrease in chromatin opening, though some opening does occur. This evidence further supports a combinatorial domain model for proper DUX protein interaction, occupancy, and target activation – revealing a major/required contribution from the C-terminal repeats and augmentation by the 14 aa C-terminal tail.

Discussion

Embryonic genome activation (EGA) creates a transcriptome and proteome that replaces maternally-inherited components, reprograms chromatin and drives cleavage-stage processes ([12](#) [↗](#)). The transcription factor DUX is highly but transiently expressed during EGA (specifically the early 2-cell embryo stage) from a large locus containing ~30 Dux repeats. DUX plays a key role in opening many promoters and enhancers, and in activating endogenous genes and repeat elements in the 2-cell embryo ([37](#) [↗](#)). DUX targets account for about 25% of the cleavage-stage transcriptome, and partly overlap with targets of OBOX4, underlying the observation that combined DUX and OBOX4 omission/depletion strongly affects embryo development ([31](#) [↗](#)). Although DUX targeting by its N-terminal homeodomains is well understood, how domains in the C-terminus of DUX contribute to chromatin reprogramming and transcription has remained largely unknown ([23](#) [↗](#)). Here, we conducted an analysis combined with proximity ligation to reveal the protein partners of DUX domains and the roles of DUX domains in chromatin remodeling and transcriptional activation.

First, we found that the mouse DUX C-terminal region contains five ~100aa repeats (termed C1-5) followed by a 14aa acidic C-terminal tail, whereas DUX orthologs contain only one copy of this repeat, followed by a similar acidic tail. Our phylogenetic and DUX domain analyses suggest that two of the repeats (C3 and C5) perform a more active role than the ‘inactive’ C1/2/4 repeats. Interestingly, single ‘active’ mouse repeats show no transcriptional stimulation on their own, but show potent activity in combination with the acidic tail, essentially mimicking the domain-activity relationships of their human, rat or bovine DUX4 orthologs. Together, these observations suggest that additional DUX repeats in the mouse lineage may have derived from duplication of the original ‘active’ repeat, with subsequent sub-functionalization.

These observations raise two interesting questions – why does DUX have multiple repeats with different activities, and what distinguishes active from inactive repeats? Our work has localized the primary difference between active and inactive repeats to a small 6 aa region. Furthermore, active versus inactive repeats (in combination with the acidic tail) show very different profiles of

protein interaction – with active repeats displaying more robust interaction with chromatin and transcription partners. Interestingly, we observe a cohort of genes (**Figure 2f** [↗](#), cluster #3) that are silent with full-length DUX, silent with an ‘inactive’ repeat (C1), but active with a single ‘active’ repeat (C3). This raises the possibility that the ‘inactive’ repeats may serve to attenuate the activity of the active repeats under certain conditions or at certain locations. Of note, DUX binds to its own promoter to help activate a strong positive feedback loop; however, Dux activation is brief (half a cell cycle), with silencing accompanying the association of the large Dux locus with the nucleolar envelope ([37](#) [↗](#), [38](#) [↗](#)). Interestingly, we observe high enrichment of nucleolar components with the ‘active’ C3 repeat, raising the possibility that the ‘active’ repeat may initially recruit powerful chromatin and transcription factors, and then subsequently contribute to localizing the Dux locus to a compartment that overcomes and/or reverses those associations to enable Dux locus silencing ([35](#) [↗](#)). Furthermore, interactions of DUX with histone deacetylase components were obtained, which remain to be fully explored in this silencing process. These observations provide new mechanistic information as well as opening new avenues for future exploration regarding the ability of DUX to function as a strong activator, while also helping to subsequently silence the repetitive Dux locus. Here, we envision that temporal and/or locus-specific post-translational modifications on DUX may mediate the temporal associations of DUX with target factors. Indeed, the presence of both ‘active’ and ‘inactive’ DUX repeats may help in coordination of these associations.

Regarding the functions of the DUX C-terminus, prior work of others has shown that the acidic C-terminal tail of DUX4 promotes transcriptional activation and interaction with the histone acetyltransferase p300 ([22](#) [↗](#)). Our validation experiments confirmed those results with mouse DUX, and we extended our studies to explore the roles and relationships of the repeat regions, both alone and in combination with the acidic tail. Our examination of Dux domain constructs was extensive, and included RNA-seq, CUT&Tag, BioID, ATAC-seq and the ability to convert mouse ES cells to 2C-like cells. As DUX opens chromatin, and as three SWI/SNF complex components were present in our BioID results (SMARCC1, SMARCA4, ACL6A), these results prompted our direct assessment and validation of the recruitment of SWI/SNF complex (monitoring the core SMARCC1 subunit) to DUX binding sites (**Figure 5c** [↗](#), **Figure 6a** [↗](#)).

An additional goal was to explore whether the functions and associations of the active repeats versus the acidic tail are largely separable/independent, or instead primarily additive or cooperative. First, regarding the activation of cleavage-specific genes (including MERV1), the target repertoire observed with full-length DUX is largely recapitulated with DUX derivatives containing the homeodomains and both active C-terminal repeats – even in the absence of the acidic tail. Thus, the tail accentuates activity, but is not strictly required for activity at the vast majority of targets. Regarding domain contributions to occupying DUX targets, all derivatives were able to bind DUX targets, though the absence of the acidic tail lowers occupancy if only one active repeat is present. This contrasts with the recruitment of H3K9ac activity; here, neither the acidic tail nor an active repeat is sufficient for recruitment; instead, either multiple repeats or repeat-tail combinations appear needed. For SMARCC1 (SWI/SNF), an active repeat is sufficient for recruitment, but recruitment is enhanced by the addition of the tail, whereas the tail alone appears ineffective. Finally, and of particular interest, a DUX derivative lacking repeat regions but containing the acidic tail (Δ C12345+14) occupies DUX targets, but does not activate targets – and does not recruit either H3K9ac or SWI/SNF complex. Therefore, the acidic tail appears to be effective only when combined with an active C-terminal repeat (**Figure 6a** [↗](#)). Overall, our data is consistent with largely cooperative functions for active repeat(s) and the acidic tail – as the absence of either compromises most functions and interactions – but places the C-terminal domains as more important domain in the hierarchy, as key factors like SWI/SNF complex are fully reliant on an active repeat for interaction. In keeping, our work predicts interactions with chromatin and transcription factors involving combined surfaces on repeats and the acidic tail (a two-domain model).

a) Summary table for DUX domain construct characteristics

	FL	C12345Δ14	C3+14	C3Δ14	ΔC12345+14
MERVL Activation	++	+	+++	-	-
DUX Target Activation	+++	+	++	-	-
DUX Target Occupancy	+++	+++	++	+	++
H3K9ac Recruitment	+++	++	++	-	-
SMARCC1/BAF Recruitment	++	+	+++	+	-
Chromatin Accessibility	+++	+++	++	+	+

b) Model for DUX mechanism of target gene activation

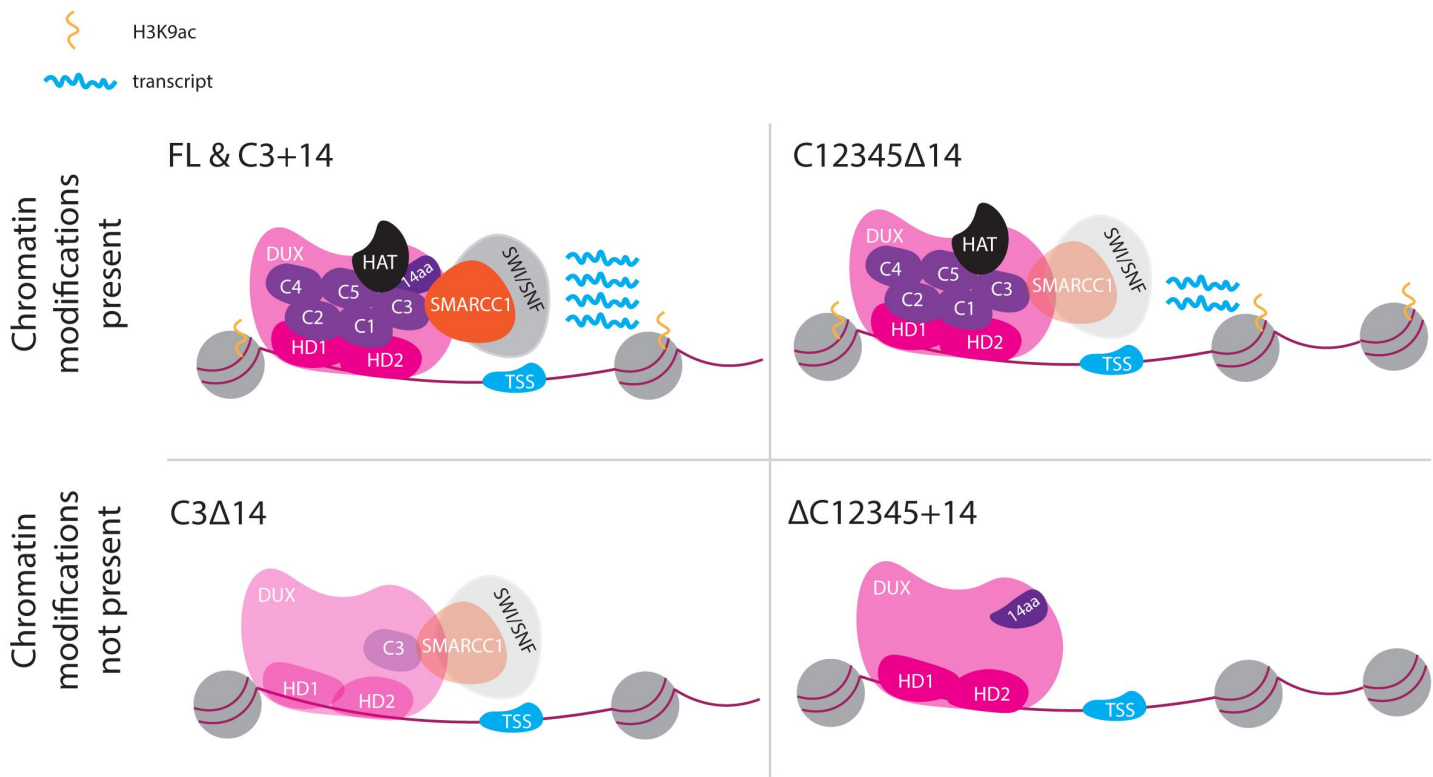


Figure 6 Model of DUX protein mechanism of activation based on C-terminal domain combinations.

a. Summary table of DUX domain constructs (illustrated in Figure 2a) and their transcription, recruitment, chromatin modification, and protein interaction activities.

b. Model for the contributions of DUX domains on DUX interactions and activity

As previewed above, our proximity-labeling (BioID) revealed a variety of protein complex interactions, and we validated a subset of interactors with co-IP experiments, including particular chromatin and transcription-related factors. We note that certain interactions may be indirect and involve DUX association with component(s) involved in a larger multi-protein structures, such as the highly repetitive Dux locus (to which DUX binds(12 [↗](#), 37 [↗](#))) in either its active or inactive form, and the association of the Dux locus with the nucleolus or other compartments. Thus, a complete understanding of the full repertoire of interactions will require future studies. However, beyond the factors noted above, we will emphasize two additional factors prominent in the BioID data that might be prioritized. First, the p53 interaction verifies prior work supporting p53 association with and activation of the Dux locus(37 [↗](#)), and prompts examination of a direct interaction with DUX. Second, the presence in our dataset of SMCHD1, a cohesin-like protein known to help confer repression of the mouse Dux and human DUX4 repetitive loci, and to be mutated in a subset of FSHD patients, likewise prompts considerable future attention (39 [↗](#), 40 [↗](#)).

Taken together our work provides considerable new mechanistic information regarding the structure-function relationships and factor associations conducted by particular domains within the mouse DUX C-terminal region. Key findings include the characterization of ‘active’ repeat domains, their interaction with transcription and chromatin remodeling factors, and the cooperativity between DUX active repeats and the acidic tail in cofactor recruitment and transcriptional activation at DUX targets. In addition to factors for activation, factors related to repression and nucleolar association were also isolated, and future studies should help determine whether their temporal associations with *DUX* may help early embryos execute a quick ‘burst’ of Dux activation followed by repression at the Dux locus in 2-cell embryos.

Conflict of interest

The authors declare no competing interests.

Specific author contribution

Experiments and analyses were conducted by C.M.S. with assistance from E.J.G. and S.C.S. C.M.S., E.J.G., S.C.S., and B.R.C. designed the study and C.M.S. and B.R.C. wrote the manuscript.

Funding

This work was supported by the Howard Hughes Medical Institute, NIH/USDA R01HD095883 Dual Purpose Grant, the NICHD (F32HD104442) to S.C.S., and the NICHD (F32HD094500) and Labor Foundation Fellowship 10041116 to E.J.G.

Acknowledgements

We thank members of the Cairns laboratory and M.B. Chandrasekhran for inspiring discussions. We thank Brian Dalley for assistance with preparing RNA-seq libraries and for sequencing services, and T. Parnell for bioinformatic assistance. We give special thanks to Ross Tomaino at Taplin Mass Spectrometry Core (Harvard University) for mass spectrometry assistance for BioID experiments. Lastly, we thank Brian Patten at Novogene for Illumina sequencing services for CUT&Tag and ATAC-seq libraries.

Materials Availability

Plasmids used in this study will be available by request.

Data accessibility

RNA-sequencing, CUT&Tag, and ATAC-seq data are available through GEO GSE224300. Standard packages were used for RNA-seq, CUT&Tag, and ATAC-seq analyses (see “Materials and Methods”).

Materials and Methods

Amino acid alignment

DUX ortholog protein translations were aligned using Geneious Local Alignment (Smith-Waterman), Blosum62 cost matrix, gap open penalty of 12, and gap extension penalty of 4. 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 μ M PD0325901 (Sigma-Aldrich) and 3 μ M CHIR99021 (Sigma-Aldrich). Stable cell lines were blasticidin (Fisher Scientific, B12150-0.1) at 3 μ g/mL.

HEK293T cell culture

HEK293Ts were cultured in high-glucose growth media DMEM with nonessential amino acids, 2-mercaptoethanol and dipeptide glutamine and were supplemented with 10% FBS.

Transfection of mESCs and HEK293Ts

Cells were transfected in Opti-MEM medium (Thermo Fisher Scientific, 31985070) using Lipofectamine 3000 (Thermo Fisher Scientific, L3000-015) for DNA transfections and the manufacturer's recommendations. If transfection were to create stable clonal cell lines, a PiggyBac transposase was included in the transfection.

Doxycycline treatment of cells

After cells have been plated for 24 hours (mESCs, or HEK293Ts), they were treated with doxycycline (0.25 μ g/mL; Clonetechn, 631311) for 12 h or 18 h to induce mCherry-DUX or DUX domain deletion construct expression, as indicated.

Western blot

RIPA buffer (50 mM TRIS-HCl [pH 8.0], 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) with 1X protease inhibitors was used to lyse mESCs. A 100X protease inhibitor cocktail consists of 0.03 mg/mL leupeptin, 0.14 mg/mL pepstatin A, 0.02 mg/mL chymostatin, 8.5 mg/mL phenylmethanesulfonylfluoride fluoride, 33mg/mL benzamidin and solubilized in ethanol. Lysed cell suspensions were then rotated for 30 min at 4°C and were further sonicated for 10 min with a 30-s on-off cycle in a Diagenode Bioruptor Pico device to solubilize chromatin-bound proteins. Protein lysates were quantified with the Bio-Rad (500-0001EDU) Bradford reagent and loaded on SDS-PAGE gel. After transfer to nitrocellulose membranes (VWR, 95040-108), membranes were blocked with TBS-T (TBS with 0.1% Tween-20) and 5% milk and then incubated with antibodies

overnight (mCherry ab167453, DUX (used in [\(37\)](#))) and signals were detected using HRP-conjugated secondary antibodies and Western Lightning Plus-ECL, Enhanced Chemiluminescence (Perkin Elmer Health Sciences, NEL 105001EA).

Flow cytometry

A Cytoflex LX instrument with lasers for 405 nm, 488 nm, 561 nm and 640 nm was used to quantify mCherry+ and MERVL-GFP+ mESCs. Samples were gated using forward FSA and side-scatter SSA to isolate cells from debris, and then double discrimination was performed using FSH $\dot{\text{A}}\sim$ FSW and SSH $\dot{\text{A}}\sim$ SSW.

RNA-seq

1 million mESCs were grown for 24 hours on a 100mm dish and then induced with doxycycline (0.25 ug/mL) for 12 hours. Cells were washed two times with PBS, TRIzol (Thermo Fisher) was added to tissue culture wells and incubated for 1 min at room temperature. TRIzol/cells were added to Eppendorf tubes and RNA was extracted using the TRIzol extraction method. 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). Purified libraries were quantified on an Agilent Technologies 2200 TapeStation with a D1000 ScreenTape assay. The molarity of adaptor-modified molecules was defined by a qPCR with a KAPA Library Quantification kit (Kapa Biosystems). Individual libraries were normalized to 10nM, and equal volumes were pooled in preparation for Illumina sequence analysis. Sequencing libraries (2pM) were chemically denatured and applied to an Illumina HiSeq paired-end flow cells 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 reads were trimmed and filtered for quality using FastQC (version 0.5.0). Processed reads were aligned using STAR (version 2.7.2c). Differential expression analysis was performed using DESeq2 (version 3.11). Hierarchical clustering was performed using the R package 'Complex Heatmap,' with inputs being differentially expressed transcripts (DESeq FDR < 0.05) z-score normalized.

GO term analysis

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

CUT&Tag

1 million mESCs were grown for 24 hours and then induced with doxycycline (0.25 ug/mL) for 12 hours. Cells were washed 1X with PBS and nuclei were isolated from mESCs using nuclear extraction buffer (20 mM HEPES-KOH [7.9], 10 mM KCl, 0.1% Triton X-100, 20% glycerol, 0.5 mM spermidine, 1X protease inhibitors (see above). ConA-streptavidin beads were paired with antibodies of interest 1:100 (mCherry ab167453, H3K9ac ab 32129, SMARCC1 ab172638) and 50,000 cells were added and incubated overnight in binding buffer (20 mM, HEPES [pH 7.9], 10 mM KCl, 1 mM CaCl_2 , 1 mM MnCl_2). Secondary guinea pig anti-rabbit 1:100 (ABIN101961) was incubated for 1 hr at room temperature. pA-Tn5 (pre-loaded with Mosaic End A/B adaptors) was incubated for 1 hour at room temperature and used at 1:100. After washing steps and release from the beads, PCR was performed for 12 cycles using NEBnext High-Fidelity 2X PCR Master Mix (New England Biolabs, M0541L) ([\(41\)](#)). Samples were 1:1.3 AMPure XP (Beckman Coulter) purified and sent for sequencing. All libraries were sequenced on the Illumina HiSeq 6000 platform in 150-bp, paired-end format. Paired-end, raw read files were first processed by Trim Galore (Babraham Institute) to trim low-quality reads and remove adapters. Processed reads were then aligned to mm10 with the following parameters: (–end-to-end –very sensitive –no-mixed). CUT&Tag libraries were trimmed and filtered for quality using FastQC (version 0.5.0) and Picard (version 2.26.3). Processed reads were aligned using Bowtie2 (version 2.4.2) with the parameters: (–end-to-end –very sensitive –no-mixed) and indexed using Samtools (version 1.15). Duplicate samples were merged and

normalized to control samples (mESCs for H3K9ac dataset and mCherry alone for mCherry and SMARCC1 datasets). DeepTools was used to generate heatmaps and profile plots using DUX binding sites as previously published (21 [DOI](#)).

BioID

15 million BirA* mESC derivatives were plated on gelatin-coated 15-cm plates and recovered for 24-hours. 0.25 µg/mL doxycycline and 0.5 µM biotin (Sigma Aldrich, B4639) was added to cells and incubated for 18 hours. Cells were washed with ice cold PBS 2 times and 1 mL lysis buffer (2 mM Tris-HCl [pH 8.0], 137 mM NaCl, 1% NP-40, 2 mM EDTA, 1x protease inhibitors) was added to each plate and incubated for 1 min (33 [DOI](#)). Cells were scraped off the plate and transferred to a 2-mL tube. Cells then rotated for 30 min at 4°C and were further sonicated for 10 min with a 30-s on-off cycle in a Diagenode Bioruptor Pico device to solubilize chromatin-bound proteins. Next, samples were centrifuged at 4°C at 12,000 rpm for 20 minutes and the supernatant was transferred to new tubes. Protein lysates were quantified with the Bio-Rad (500-0001EDU) Bradford reagent. For streptavidin pulldown, 2 µg of protein were added to pre-washed 50 µL streptavidin Sepharose bead conjugates (Cell Signaling Technology, 3419S) and rocked overnight at 4°C. Cells were pelleted and then washed 5X with wash buffer (10 mM TRIS [pH 7.4], 1 mM EDTA, 1 mM EGTA, 150 mM NaCl, 1% Triton X-100, 0.2 mM sodium orthovanadate, 1x protease inhibitors), boiled at 95°C in 30 µL 4X Laemmli buffer, and ran on a SDS-page gel until the samples were 1-inch into the gel. Protein bands/smears were cut out of the gel with razor blades and transferred to individual 1.7-mL Eppendorf tubes filled with ddH₂O. Samples were overnight shipped on dry for liquid chromatography-tandem mass spectrometry sequencing (LC-MS/MS) to Taplin Mass Spectrometry Facility, Cell Biology Department, Harvard Medical School. Gel smears were reduced using 1 mM DTT in 50 mM ammonium bicarbonate for 30 min at 60°C and cooled, and 5 mM iodoacetamide in 50 mM ammonium bicarbonate was added for 15 min and incubated at room temperature. 5 mM DTT was added to quench the reaction, and then 5 ng/µL of trypsin (Promega) was added and samples incubated overnight at 37 °C. Samples were then desalinated and resuspended in 10 µL of high-pressure liquids chromatography (HPLC) solvent A (2.5% acetonitrile, 0.1% formic acid). A reverse-phase HPLC capillary column was made by placing 2.6 µL M C₁₈ spherical silica beads into a fused silica capillary (100-µL M inner diameter, ~30-cm length). Next, each sample was loaded using a Famos autosampler (LC Packings) onto the HPLC capillary column. A linear gradient was created to elute peptides with increasing amounts of solvent B (97.5% acetonitrile, 0.1% formic acid). Eluted peptides were subjected to electrospray ionization couple to a mass spectrometer (Thermo Fisher Scientific) to produce tandem mass spectra whose b and y ion series patterns were compared to Sequest (Thermo Fisher Scientific) to establish protein identity. The sequencing data were filtered to 1 to 2% peptide false-discovery rate. To analyze BioID datasets, two approaches were taken. First, we utilized DESeq2 (version 3.11) to provide candidate genes for BirA*-DUX vs. BirA* and BirA*-C3 vs. BirA*-C1. Then, we used SAINTexpress (version 3.5.3) with parameter: (-L6). Formal candidates exhibited > 2 Log₂FC and FDR < 0.05 and SAINT score > 0.74, for DESeq2 and SAINTexpress analyses respectively.

Co-immunoprecipitation

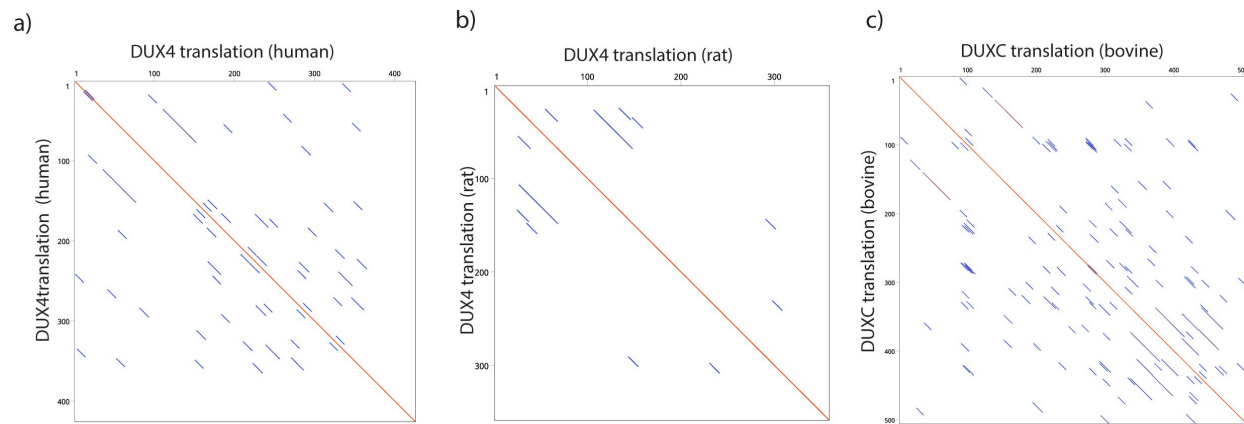
1 million mESCs were grown for 24 hours on 100mm dishes and then induced with doxycycline (0.25 µg/mL) for 12 hours. Cells were trypsinized, rinsed once with PBS, and lysis buffer (2 mM Tris-HCl, pH=8.0, 137 mM NaCl, 1% NP-40, 2 mM EDTA, 1x protease inhibitors) was added to lyse cells. Cells then rotated for 30 min at 4°C and were further sonicated for 10 min with a 30-s on-off cycle in a Diagenode Bioruptor Pico device to solubilize chromatin-bound proteins. Protein lysates were quantified with the Bio-Rad (500-0001EDU) Bradford reagent. 3X FLAG Peptide-coupled beads (Sigma Aldrich, F4799-4MG) were pre-washed with lysis buffer, 1 mg of protein was added to 40 µL of pre-washed beads, and samples were incubated at room temperature for 4 hours. Beads were pelleted by low-speed centrifugation. Samples were washed 5X with wash buffer (10

mM TRIS [pH 7.4], 1 mM EDTA, 1 mM EGTA; pH=8.0, 150 mM NaCl, 1% Triton X-100, 0.2 mM sodium orthovanadate, 1x protease inhibitors [see above]), boiled at 95°C in 30 µL 4X Laemmli buffer, and run on an SDS-page gel for western blotting analysis.

ATAC-seq

mESCs were grown for 24 hours and then induced with doxycycline (0.25 µg/mL) for 12 hours. Cells were washed 1X with PBS, pelleted, resuspended, and lysed in cold lysis buffer (10 mM Tris-HCl [pH 7.4], 10 mM NaCl, 3 mM MgCl₂, and 0.1% IGEPAL CA-630). Nuclei were pelleted and resuspended in transposase buffer. The Tn5 (Illumina FC-121-1030), and the transposition reaction was carried out for 30 min at 37°C. After purification, the Nextera libraries were amplified for 15 cycles with NEBnext PCR master mix and purified with a Qiagen PCR cleanup kit. All libraries were sequenced on the Illumina HiSeq 6000 platform in 150-bp, paired-end format. Paired-end, raw read files were first processed by Trim Galore (Babraham Institute) to trim low-quality reads and remove adapters. Processed reads were then aligned to mm10 with the following parameters: (--end-to-end --very sensitive --no-mixed).

Supplementary Figure Legends

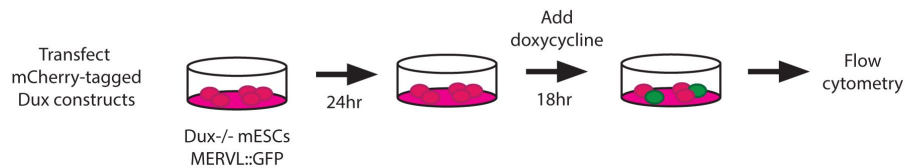


d)

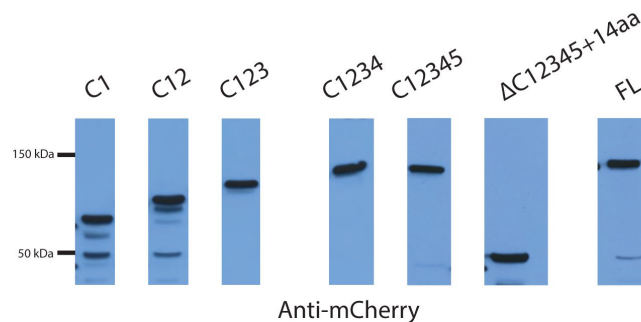
	C-term repeat 1	C-term repeat 2	C-term repeat 3	C-term repeat 4	C-term repeat 5	DUX4 C-term	DUXC C-term	Dux4 C-term
C-term repeat 1		3e-24	4e-36	7e-22	2e-25	*	0.024	0.0009
C-term repeat 2	3e-24		1e-34	1e-41	2e-23	0.030	*	*
C-term repeat 3	4e-36	1e-34		9e-31	7e-42	*	*	*
C-term repeat 4	7e-22	1e-41	9e-31		1e-09	*	*	0.002
C-term repeat 5	2e-25	2e-23	7e-42	1e-09		0.022	0.016	0.035
DUX4 C-term (H)	*	0.030	*	*	0.022		1e-06	*
DUXC C-term (B)	0.024	*	*	*	0.016	1e-06		*
DUX4 C-term (R)	0.0009	*	*	0.002	0.035	*	*	

* No significant similarities found (>0.05)

e) Schematic for experimental design for all flow cytometry in Dux domain constructs



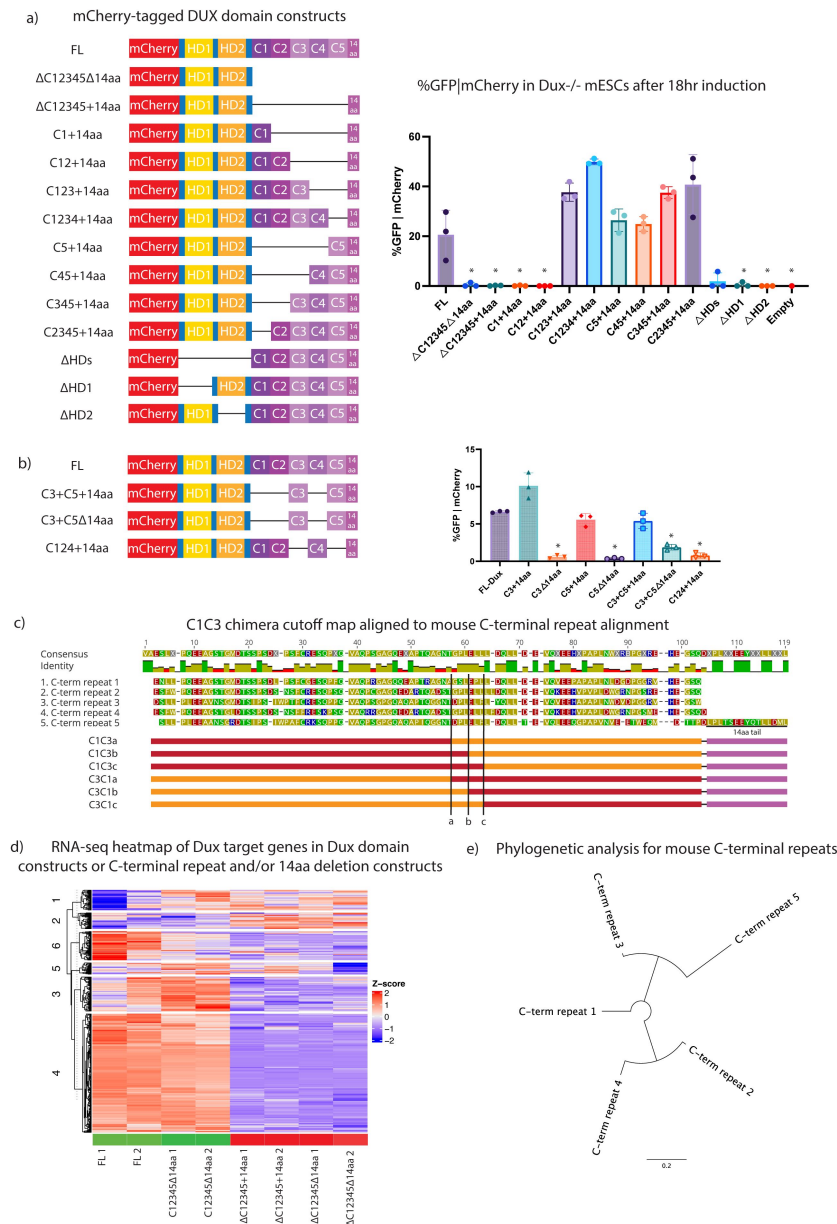
f) Western blot for mCherry-tagged Dux domain constructs



Supplementary Figure 1

DUX orthologs do not contain a repeat structure similar to mouse DUX.

- Human DUX4 aligned to itself.
- Rat DUX4 aligned to itself.
- Bovine DUXC aligned to itself.
- Table with a pairwise comparison of e-values from a BLAST-based amino acid similarity comparing mouse DUX C-terminal repeats, the human DUX4 C-terminus, the rat DUX4 C-terminus, and the bovine DUXC C-terminus. Parameters are PAM250 matrix and gap costs of existence: 15 and extension: 3. *No significant similarities found (>0.05).
- Experimental design for expression of mCherry-tagged DUX domain constructs and flow cytometry of GFP% | mCherry tag expression.
- Western blot for mCherry tag after 18hr expression of C1, C12, C123, C1234, C12345. Predicted protein sizes are 59 kDa (C1), 69 kDa (C12), 79 kDa (C123), 89 kDa (C1234), 99 kDa (C12345), 48 kDa (ΔC12345+14aa), and 100 kDa (FL).



Supplementary Figure 2

Supplementary Figure 2:

DUX homeodomains are required for transcriptional activity.

a) (Left) Schematic of constructs used for flow cytometry: mCherry-tagged full-length DUX (FL), ΔC12345+14aa, C12345Δ14aa, C1+14aa, C12+14aa, C123+14aa, C1234+14aa, C5+14aa, C45+14aa, C345+14aa, C2345+14aa, ΔHDs, ΔHD1, ΔHD2, and *Dux*^{-/-} alone. (Right) Flow cytometry for MERV::GFP reporter given mCherry expression in *Dux*^{-/-} mESCs following 18hr expression of indicated constructs. *p-value < 0.05, student's t-test. (n=3 biological replicates).

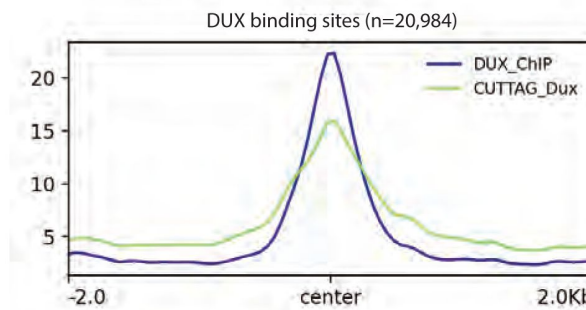
b) (Left) Schematic of additional constructs used for flow cytometry: mCherry-tagged full-length DUX (FL), C3+C5+14aa, C3+C5Δ14aa, C124+14aa. (Right) Flow cytometry for MERV::GFP reporter given mCherry expression in *Dux*^{-/-} mESCs following 18hr expression of indicated constructs. *p-value < 0.05, student's t-test. (n=3 biological replicates)

c) Schematic of C1C3 chimera DUX constructs for amino acid cut offs for C1C3a, C3C3b, C1C3c, C3C1a, C3C1b, and C3C1c aligned to the 5 C-terminal repeats

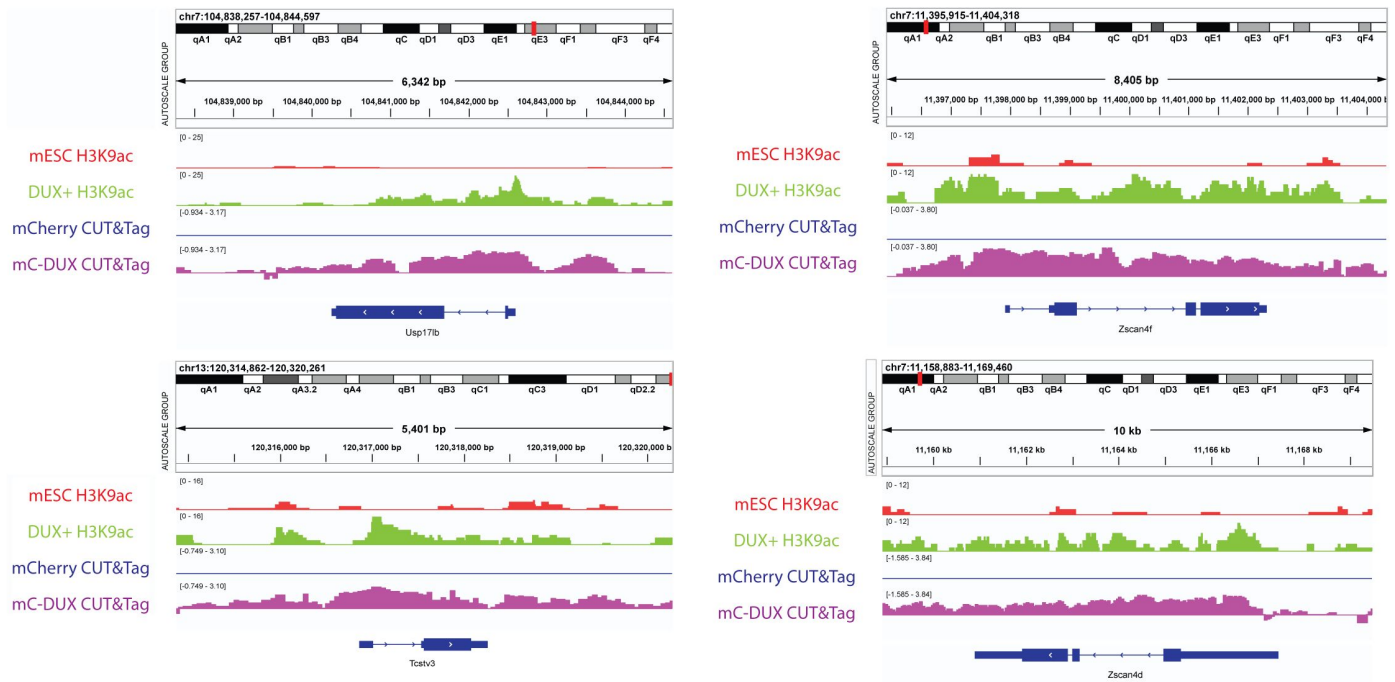
d) Heat map of RNA-seq at DUX target genes (n=456) (21) from 18hr expression of DUX domain constructs of FL, C12345Δ14aa, C12345+14aa, C12345Δ14aa(n=2). Colored rectangles above samples denote inability to activate MERV::GFP reporter (red) and ability to activate MERV::GFP reporter (green).

e) Phylogenetic analysis of the five mouse DUX C-terminal repeats. Line length unit is substitutions per base.

a) Profile plot of DUX ChIP-seq vs. mCherry-tagged DUX CUT&Tag at DUX binding sites



b) IGV screenshots of mESC H3K9ac and DUX-expressing cells H3K9ac at DUX target genes



Supplementary Figure 3

DUX CUT&Tag data strongly resembles DUX ChIP-seq, and DUX binding sites gain H3K9ac upon DUX expression.

- Profile plot of mCherry-FL DUX CUT&TAG centered at DUX binding sites after 12 hour expression vs. HA-DUX ChIP-seq (n=2 biological replicates) (12 [DOI](#))
- IGV screenshots at TCSTV3, ZSCAN4F, USP17LB, and ZSCAN4C (DUX targets genes) for H3K9ac CUT&Tag in mESCs and mCherry-DUX 12hr expression and mCherry CUT&TAG for mCherry tag alone and mCherry-DUX (n=2 biological replicates)

References

2. Jukam D, Shariati SAM, Skotheim JM (2017) **Zygotic Genome Activation in Vertebrates** *Dev Cell* **42**:316–32
3. Gao Y, Liu X, Tang B, Li C, Kou Z, Li L, et al. (2017) **Protein Expression Landscape of Mouse Embryos during Pre-implantation Development** *Cell Rep* **21**:3957–69
4. Pálffy M, Joseph SR, Vastenhouw NL (2017) **The timing of zygotic genome activation** *Curr Opin Genet Dev* **43**:53–60
5. Lee MT, Bonneau AR, Takacs CM, Bazzini AA, DiVito KR, Fleming ES (2013) **Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition** *Nature*
6. Vallot A, Tachibana K (2020) **The emergence of genome architecture and zygotic genome activation** *Curr Opin Cell Biol* **64**:50–7
7. Saitou M, Kagiwada S, Kurimoto K (2012) **Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells** *Development* **139**:15–31
8. Xu R, Li S, Wu Q, Li C, Jiang M, Guo L, et al. (2022) **Stage-specific H3K9me3 occupancy ensures retrotransposon silencing in human pre-implantation embryos** *Cell Stem Cell* **29**:1051–1066
9. Bryja V, Bonilla S, Arenas E (2006) **Derivation of mouse embryonic stem cells** *Nat Protoc* **1**:2082–7
10. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, et al. (2012) **Embryonic stem cell potency fluctuates with endogenous retrovirus activity** *Nature*
11. Rodriguez-Terrones D, Hartleben G, Gaume X, Eid A, Guthmann M, Iturbide A, et al. (2020) **A distinct metabolic state arises during the emergence of 2-cell-like cells** *EMBO Rep* **21**
12. Genet M, Torres-Padilla ME (2020) **The molecular and cellular features of 2-cell-like cells: a reference guide** *Development* **147**
13. Hendrickson PG, Doráis JA, Grow EJ, Whiddon JL, Lim JW, Wike CL, et al. (2017) **Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons** *Nat Genet* **49**:925–34
14. De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D (2017) **DUX-family transcription factors regulate zygotic genome activation in placental mammals** *Nat Genet* **49**:941–5
15. Whiddon JL, Langford AT, Wong CJ, Zhong JW, Tapscott SJ (2017) **Conservation and innovation in the DUX4-family gene network** *Nat Genet* **49**:935–40
16. Campbell AE, Belleville AE, Resnick R, Shadle SC, Tapscott SJ (2018) **Facioscapulohumeral dystrophy: activating an early embryonic transcriptional program in human skeletal muscle** *Hum Mol Genet* **27**:R153–62

17. Yoshihara M, Kirjanov I, Nykänen S, Sokka J, Weltner J, Lundin K, et al. (2022) **Transient DUX4 expression in human embryonic stem cells induces blastomere-like expression program that is marked by SLC34A2** *Stem Cell Rep* **17**:1743–56
18. Taubenschmid-Stowers J, Rostovskaya M, Santos F, Ljung S, Argelaguet R, Krueger F, et al. (2022) **8C-like cells capture the human zygotic genome activation program in vitro** *Cell Stem Cell* **29**:449–459
19. Ferreboeuf M, Mariot V, Bessières B, Vasiljevic A, Attié-Bitach T, Collardeau S, et al. (2014) **DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles** *Hum Mol Genet* **23**:171–81
20. Vanderplanck C, Ansseau E, Charron S, Stricwant N, Tassin A, Laoudj-Chenivresse D, et al. (2011) **The FSHD Atrophic Myotube Phenotype Is Caused by DUX4 Expression** *PLOS ONE* **6**
21. Le Gall L, Sidlauskaitė E, Mariot V, Dumonceaux J (2020) **Therapeutic Strategies Targeting DUX4 in FSHD** *J Clin Med* **9**
22. Chen Z, Zhang Y (2019) **Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development** *Nat Genet* **51**:947–51
23. Choi SH, Gearhart MD, Cui Z, Bosnakovski D, Kim M, Schennum N, et al. (2016) **DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes** *Nucleic Acids Res* **44**:5161–73
24. Leidenroth A, Hewitt JE (2010) **A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene** *BMC Evol Biol* **10**
25. Leidenroth A, Clapp J, Mitchell LM, Coneyworth D, Dearden FL, Iannuzzi L, et al. (2012) **Evolution of DUX gene macrosatellites in placental mammals** *Chromosoma* **121**:489–97
26. Vuoristo S, Bhagat S, Hydén-Granskog C, Yoshihara M, Gawrylski L, Jouhilahti EM, et al. (2022) **DUX4 is a multifunctional factor priming human embryonic genome activation** *iScience* **25**
27. Clapier CR, Cairns BR (2009) **The biology of chromatin remodeling complexes** *Annu Rev Biochem* **78**:273–304
28. Romero OA, Sanchez-Céspedes M (2014) **The SWI/SNF genetic blockade: effects in cell differentiation, cancer and developmental diseases** *Oncogene* **33**:2681–9
29. Bultman SJ, Gebuhr TC, Pan H, Svoboda P, Schultz RM, Magnuson T (2006) **Maternal BRG1 regulates zygotic genome activation in the mouse** *Genes Dev* **20**:1744–54
30. Schaniel C, Ang YS, Ratnakumar K, Cormier C, James T, Bernstein E, et al. (2009) **Smarcc1/Baf155 Couples Self-Renewal Gene Repression with Changes in Chromatin Structure in Mouse Embryonic Stem Cells** *Stem Cells* **27**:2979–91
31. De Iaco A, Verp S, Offner S, Grun D, Trono D (2019) **DUX is a non-essential synchronizer of zygotic genome activation** *Development*
31. Guo Y, Kitano T, Inoue K, Murano K, Hirose M, Li TD (2022) **32. Guo Y, Kitano T, Inoue K, Murano K, Hirose M, Li TD, et al. Obox4 secures zygotic genome activation upon loss of Dux** [Internet]. *bioRxiv*; 2022 [cited 2023 Jan 5]. p. 2022.07.04.498763. Available from: <https://www.biorxiv.org/content/10.1101/2022.07.04.498763v2>

33. Akiyama T, Xin L, Oda M, Sharov AA, Amano M, Piao Y, et al. (2015) **Transient bursts of Zscan4 expression are accompanied by the rapid derepression of heterochromatin in mouse embryonic stem cells** *DNA Res Int J Rapid Publ Rep Genes Genomes* **22**:307–18
34. Roux KJ, Kim DI, Raida M, Burke B (2012) **A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells** *J Cell Biol* **196**:801–10
35. Liu X, Salokas K, Tamene F, Jiu Y, Weldatsadik RG, Öhman T, et al. (2018) **An AP-MS-and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations** *Nat Commun* **9**
36. Xie SQ, Leeke BJ, Whilding C, Wagner RT, Garcia-Llagostera F, Low Y, et al. (2022) **Nucleolar-based Dux repression is essential for embryonic two-cell stage exit** *Genes Dev* **36**:331–47
37. Yu H, Sun Z, Tan T, Pan H, Zhao J, Zhang L, et al. (2021) **rRNA biogenesis regulates mouse 2C-like state by 3D structure reorganization of peri-nucleolar heterochromatin** *Nat Commun* **12**
38. Grow EJ, Weaver BD, Smith CM, Guo J, Stein P, Shadle SC, et al. (2021) **p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models** *Nat Genet* **53**:1207–20
38. Vega-Sendino M, Olbrich T, Stein P, Tillo D, Carey GI, Savy V (2022) **39. Vega-Sendino M, Olbrich T, Stein P, Tillo D, Carey GI, Savy V, et al. The homeobox transcription factor DUXBL controls exit from totipotency [Internet]. bioRxiv; 2022 [cited 2023 Feb 8]. p. 2022.09.19.508541. Available from: <https://www.biorxiv.org/content/10.1101/2022.09.19.508541v1>**
40. Ruebel ML, Vincent KA, Schall PZ, Wang K, Latham KE (2019) **SMCHD1 terminates the first embryonic genome activation event in mouse two-cell embryos and contributes to a transcriptionally repressive state** *Am J Physiol-Cell Physiol* **317**:C655–64
41. Lemmers RJLF, Tawil R, Petek LM, Balog J, Block GJ, Santen GWE, et al. (2012) **Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2** *Nat Genet* **44**:1370–4
42. Kaya-Okur HS, Janssens DH, Henikoff JG, Ahmad K, Henikoff S (2020) **Efficient low-cost chromatin profiling with CUT&Tag** *Nat Protoc* **15**:3264–83

Author information

Christina M. Smith

Howard Hughes Medical Institute, Department of Oncological Sciences and Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA
ORCID iD: [0000-0003-0403-0920](https://orcid.org/0000-0003-0403-0920)

Edward J. Grow

Howard Hughes Medical Institute, Department of Oncological Sciences and Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA, Green Center for Reproductive Biological Sciences, Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, TX, USA

Sean C. Shadle

Howard Hughes Medical Institute, Department of Oncological Sciences and Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA

Bradley R. Cairns

Howard Hughes Medical Institute, Department of Oncological Sciences and Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA

For correspondence: brad.cairns@hci.utah.edu

ORCID iD: [0000-0002-9864-8811](https://orcid.org/0000-0002-9864-8811)

Editors

Reviewing Editor

Jerry Workman

Stowers Institute for Medical Research, United States of America

Senior Editor

Kevin Struhl

Harvard Medical School, United States of America

Reviewer #1 (Public Review):

In this manuscript, the authors identified and characterized the five C-terminus repeats and a 14aa acidic tail of the mouse Dux protein. They found that repeat 3&5, but not other repeats, contribute to transcriptional activation when combined with the 14aa tail. Importantly, they were able to narrow down to a 6 aa region that can distinguish "active" repeats from "inactive" repeats. Using proximal labeling proteomics, the authors identified candidate proteins that are implicated in Dux-mediated gene activation. They were able to showcase that the C-terminal repeat 3 binds to some proteins, including Smarcc1, a component of SWI/SNF (BAF) complex. In addition, by overexpressing different Dux variants, the authors characterized how repeats in different combinations, with or without the 14aa tail, contribute to Dux binding, H3K9ac, chromatin accessibility, and transcription. In general, the data is of high quality and convincing. The identification of the functionally important two C-terminal repeats and the 6 aa tail is enlightening. The work shined light on the mechanism of Dux function.

Reviewer #2 (Public Review):

In this manuscript, Smith et al. delineated novel mechanistic insights into the structure-function relationships of the C-terminal repeat domains within the mouse DUX protein. Specifically, they identified and characterised the transcriptionally active repeat domains, and narrowed down to a critical 6aa region that is required for interacting with key transcription and chromatin regulators. The authors further showed how the DUX active repeats collaborate with the C-terminal acidic tail to facilitate chromatin opening and transcriptional activation at DUX genomic targets.

Reviewer #3 (Public Review):

Dux (or DUX4 in human) is a master transcription factor regulating early embryonic gene activation and has garnered much attention also for its involvement in reprogramming

pluripotent embryonic stem cells to totipotent "2C-like" cells. The presented work starts with the recognition that DUX contains five conserved c. 100-amino acid carboxy-terminal repeats (called C1-C5) in the murine protein but not in that of other mammals (e.g. human DUX4). Using state-of-the-art techniques and cell models (BioID, Cut&Tag; rescue experiments and functional reporter assays in ESCs), the authors dissect the activity of each repeat, concluding that repeats C3 and C5 possess the strongest transactivation potential in synergy with a short C-terminal 14 AA acidic motif. In agreement with these findings, the authors find that full-length and active (C3) repeat containing Dux leads to increased chromatin accessibility and active histone mark (H3K9Ac) signals at genomic Dux binding sites. A further significant conclusion of this mutational analysis is the proposal that the weakly activating repeats C2 and C4 may function as attenuators of C3+C5-driven activity.

By next pulling down and identifying proteins bound to Dux (or its repeat-deleted derivatives) using BioID-LC/MS/MS, the authors find a significant number of interactors, notably chromatin remodellers (SMARCC1), a histone chaperone (CHAF1A/p150) and transcription factors previously (ZSCAN4D) implicated in embryonic gene activation.

The experiments are of high quality, with appropriate controls, and thus provide a rich compendium of Dux interactors for future study. Indeed, a number of these (SMARCC1, SMCHD1, ZSCAN4) make biological sense, both for embryonic genome activation and for FSHD (SMCHD1).

The central question raised by this study, however, concerns the function of the Dux repeats, apparently unique to mice. While it is possible, as the authors propose, that the weak activating C1, C2 C4 repeats may exert an attenuating function ("sub-functionalization") on activation mediated by C3 and/or C5, it could similarly be argued that the different repeats are indeed expected to display different activation potentials, chromatin opening, cofactor recruitment, due to, simply, the differences in their sequences. The argument for an active attenuating function would have been strengthened, for example, by the finding of repressor recruitment by C1/C2/C4 (and not just less of everything). The possible biological relevance of these repeats thus remains to be established.

Author Response

The following is the authors' response to the original reviews.

First, the authors would like to thank the reviewers and editors for their thoughtful comments. The comments were used to guide our revision, which is substantially improved over our initial submission. We have addressed all comments in our responses below, through a combination of clarification, new analyses and new experimental data.

Reviewer #1 (Public Review):

In this manuscript, the authors identified and characterized the five C-terminus repeats and a 14aa acidic tail of the mouse Dux protein. They found that repeat 3&5, but not other repeats, contribute to transcriptional activation when combined with the 14aa tail. Importantly, they were able to narrow down to a 6 aa region that can distinguish "active" repeats from "inactive" repeats. Using proximal labeling proteomics, the authors identified candidate proteins that are implicated in Dux-mediated gene activation. They were able to showcase that the C-terminal repeat 3 binds to some proteins, including Smarcc1, a component of SWI/SNF (BAF) complex. In addition, by overexpressing different Dux variants, the authors characterized how repeats in different combinations, with or without the 14aa tail, contribute to Dux binding, H3K9ac, chromatin accessibility,

and transcription. In general, the data is of high quality and convincing. The identification of the functionally important two C-terminal repeats and the 6 aa tail is enlightening. The work shined light on the mechanism of DUX function.

A few major comments that the authors may want to address to further improve the work:

We thank the reviewer for their efforts and constructive comments, which have guided our revisions.

- 1. The summary table for the Dux domain construct characteristics in Fig. 6a could be more accurate. For example, C3+14 clearly showed moderate weaker Dux binding and H3K9ac enrichment in Fig 3c and 3e. However, this is not illustrated in Fig. 6a. The authors may consider applying statistical tests to more precisely determine how the different Dux constructs contribute to DNA binding (Fig. 3c), H3K9ac enrichment (Fig. 3e), Smarcc1 binding (Fig. 5e), and ATAC-seq signal (Fig. 5f).*

We thank the reviewer for this comment, and agree that there were some modest differences in construct characteristics that were not captured in the Summary Table (6a). To better reflect the differences between constructs, we added additional dynamic range to our depiction/scoring, and believe that the new scoring system provides sufficient qualitative range to capture the difference without imposing a statistical approach.

- 1. Another concern is that exogenous overexpressed Dux was used throughout the experiments. The authors may consider validating some of the protein-protein interactions using spontaneous or induced 2CLCs (where Dux is expressed).*

We agree that it would be helpful to determine endogenous DUX interaction with our BioID candidates. Here, we attempted co-IPs for endogenous DUX protein with the DUX antibody and were unsuccessful, which indicated that the DUX antibody is useful for detection but not efficient in the primary IP. This is why we utilized the mCherry tag for DUX IP experiments, which worked exceptionally well.

- 1. It could be technically challenging, but the authors may consider to validate Dux and Smarcc1 interaction in a biologically more relevant context such as mouse 2-cell embryos where both proteins are expressed. Whether Smarcc1 binding will be dramatically reduced at 4-cell embryos due to loss of Dux expression?*

While we agree that it would be interesting to validate the in vivo interaction of DUX and SMARCC1 in the early embryo, it is not technically feasible for us to conduct the experiment, as the IP would require thousands of two-cell embryos, and we have the issue of poor co-IP quality with the DUX antibody.

Reviewer #2 (Public Review):

In this manuscript, Smith et al. delineated novel mechanistic insights into the structure-function relationships of the C-terminal repeat domains within the mouse DUX protein. Specifically, they identified and characterised the transcriptionally active repeat domains, and narrowed down to a critical 6aa region that is required for interacting with key transcription and chromatin regulators. The authors further showed how the DUX active repeats collaborate with the C-terminal acidic tail to facilitate chromatin opening and transcriptional activation at DUX genomic targets.

Although this study attempts to provide mechanistic insights into how DUX4 works, the authors will need to perform a number of additional experiments and controls to bolster

We thank this reviewer for their constructive comments, and have conducted several new analyses, additional experiments and clarifications – which have strengthened the manuscript in several locations. Highlights include a statistical approach to the similarity of mouse repeats to themselves and to orthologs (Figure S1d) and clarified interpretations, a wider dynamic range to better reflect changes in DUX construct behaviors (Figure 6a), and additional data on construct behavior, including ‘inactive’ constructs (e.g C1+14aa in Figure 1a,d, new ATAC-seq in Figure S1g), and active constructs such as C3+C5+14aa and C3+C5Δ14aa (in Figure S1b).

Reviewer #3 (Public Review):

Dux (or DUX4 in human) is a master transcription factor regulating early embryonic gene activation and has garnered much attention also for its involvement in reprogramming pluripotent embryonic stem cells to totipotent "2C-like" cells. The presented work starts with the recognition that DUX contains five conserved c. 100-amino acid carboxy-terminal repeats (called C1-C5) in the murine protein but not in that of other mammals (e.g. human DUX4). Using state-of-the-art techniques and cell models (BioID, Cut&Tag; rescue experiments and functional reporter assays in ESCs), the authors dissect the activity of each repeat, concluding that repeats C3 and C5 possess the strongest transactivation potential in synergy with a short C-terminal 14 AA acidic motif. In agreement with these findings, the authors find that full-length and active (C3) repeat containing Dux leads to increased chromatin accessibility and active histone mark (H3K9Ac) signals at genomic Dux binding sites. A further significant conclusion of this mutational analysis is the proposal that the weakly activating repeats C2 and C4 may function as attenuators of C3+C5-driven activity.

By next pulling down and identifying proteins bound to Dux (or its repeat-deleted derivatives) using BioID-LC/MS/MS, the authors find a significant number of interactors, notably chromatin remodellers (SMARCC1), a histone chaperone (CHAF1A/p150) and transcription factors previously (ZSCAN4D) implicated in embryonic gene activation.

The experiments are of high quality, with appropriate controls, thus providing a rich compendium of Dux interactors for future study. Indeed, a number of these (SMARCC1, SMCHD1, ZSCAN4) make biological sense, both for embryonic genome activation and for FSHD (SMCHD1).

A critical question raised by this study, however, concerns the function of the Dux repeats, apparently unique to mice. While it is possible, as the authors propose, that the weak activating C1, C2 C4 repeats may exert an attenuating function on activation (and thus may have been selected for under an "adaptationist" paradigm), it is also possible that they are simply the result of Jacobian evolutionary bricolage (tinkering) that happens to work in mice. The finding that Dux itself is not essential, in fact appears to be redundant (or cooperates with) the OBOX4 factor, in addition to the absence of these repeats in the DUX protein of all other mammals (as pointed out by the authors), might indeed argue for the second, perhaps less attractive possibility.

In summary, while the present work provides a valuable resource for future study of Dux and its interactors, it fails, however, to tell a compelling story that could link the obtained data together.

We appreciated the reviewer's views regarding the high quality of the work and our generation of an important dataset of DUX interactors. We also appreciate the comments provided to improve the work, and have performed and included in the revised version a set

of clarifications, additional analyses and additional experiments that have served to reinforce our main points and provide additional mechanistic links. We also agree that more remains to be done to understand the function and evolution of repeats C1, C2 and C4.

Reviewer #1 (Recommendations For The Authors):

1. For immuno-blots, authors may indicate the expected bands to help readers better understand the results.

Agreed, and we have included the predicted molecular weight of proteins in the Figure Legends. We note that our work shows that the C-terminal domains confer anomalous migration in SDS-PAGE.

1. Fig. 5b, a blot missing for the mCherry group?

Figure 5b is a volcano blot, so we believe the reviewer is referring to Figure 5d, which is a coimmunoprecipitation experiment between SMARCC1 and mCherry-tagged DUX constructs. However, we are unsure of the comment as an anti mCherry sample is present in that panel.

1. Line 99-100, Fig. S1d, it seems that repeat2, but not repeat3, is more similar to human DUX4 C-terminal region.

This comment and one by another reviewer have prompted us to re-examine the similarities of the DUX repeats, and we have new analyses (Figure S1d) and an alternative framing in the manuscript as a result. We have expanded on this in our response to Reviewer #2, point #1 – and direct the reviewer there for our expanded treatment.

1. There are a few references are misplaced. For example, line 48, the studies that reported the role of Dux in inducing 2CLCs should be from Hendrickson et al., 2017, De Iaco et al., 2017, and Whiddon et al., 2017. The authors may want to double check all references.

Thanks for pointing these out. These issues have been corrected in the manuscript.

1. In the materials & methods section, a few potential errors are noticed. For example, concentrations of PD0325901 and CHIR99021 in mESC medium appear ~1000-fold higher than standards.

Thanks – corrected.

Reviewer #2 (Recommendations For The Authors):

Major Points

1. Line 99 - The authors claimed that the "human DUX4 C-terminal region is most similar to the 3rd repeat of mouse DUX", but based on Supp. Fig. 1d, the human DUX4 C-term should be most similar to the 2nd repeat of mouse DUX. If this is indeed the case, it will undermine the rest of this study, since the authors claim that the 3rd repeat is transcriptionally active, whereas the 2nd repeat is transcriptionally inactive, and the bulk of this study largely focused on how the active repeats, not the inactive repeats, are critical in recruiting key transcriptional and chromatin regulators to induce the embryonic gene expression program.

We thank the reviewer for their comments here. Since submission, and as mentioned above for reviewer #1 we have revisited the issue of similarity of the DUX4 C-terminal region to the

mouse C-terminal repeats, with a BLAST-based approach that is more rigorous and informed by statistics – which is in Author response table 1 and now in the manuscript as Figure S1d, and has affected our interpretation. Our prior work involved a simple % identity comparison table and we now appreciate that some of the similarity analyses did not meet statistical significance, and therefore we are unable to draw certain conclusions. We make the appropriate modifications in the text. For example, we no longer state that the DUX4 C-terminus appears to be most similar to mouse repeats 3 and 5. This does not affect the main conclusions of the paper regarding interactions of the C-terminus with chromatin-related proteins, only our speculation on which repeat might have represented the original single repeat in the mouse – an issue we think of some interest, but did not rise to the level of mentioning in the original or current abstract.

Author response table 1.

Parameters: PAM250 matrix. Gap costs of existence: 15 and extension: 3. Numbers represent e-value of each pairwise comparison

*No significant similarities found (>0.05).

	C1	C2	C3	C4	C5	DUX4 (human)	DUXC (bovine)	Dux4 (rat)
C1		3e-24	4e-36	7e-22	2e-25	*	0.024	0.0009
C2	3e-24		1e-34	1e-41	2e-23	0.030	*	*
C3	4e-36	1e-34		9e-31	7e-42	*	*	*
C4	7e-22	1e-41	9e-31		1e-09	*	*	0.002
C5	2e-25	2e-23	7e-42	1e-09		0.022	0.016	0.035
DUX4 (human)	*	0.030	*	*	0.022		1e-06	*
DUXC (bovine)	0.024	*	*	*	0.016	1e-06		*
Dux4 (rat)	0.0009	*	*	0.002	0.035	*	*	

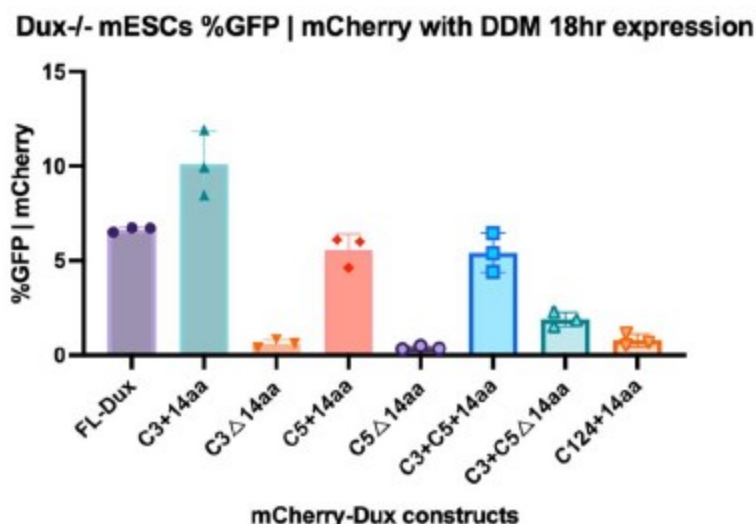
1. In Supp Fig 1d, it seems that the rat DUX4 C-terminal region is most similar to the 4th repeat of mouse DUX, which according to the author is supposedly transcriptionally inactive. This weakens the authors justification that the 3rd or 5th repeat is likely the "parental repeat for the other four", and further echoes my concern in point 1 where the human DUX4 C-term is most similar to the 2nd (inactive) repeat of mouse DUX.

The reviewer's point is well taken and is addressed in point #1 above.

1. In Fig. 1d, the authors showed that DUX4-containing C3 and C5, but lacking acidic tail, can promote MERVL::GFP expression, albeit to a slightly lower extent compared to FL. However, in Fig. 2b, C3 or C5 alone (lacking acidic tail) completely failed to promote MERVL::GFP expression. However, in the presence of the acidic tail, both versions were able to promote MERVL::GFP expression, similar to that of FL. The latter would suggest that it is the acidic tail that is crucial for MERVL::GFP expression, and this does not quite agree with Fig 1b, where C12345 (lacking acidic tail) was able to promote MERVL::GFP expression. Although C12345 did not activate MERVL to a similar level as FL, it is clearly proficient, compared to C3 or C5 alone (lacking acidic tail) where there is no increase in MERVL at all. Additional constructs will be helpful to clarify these points. For example, 'C3+C5 minus acidic tail' and 'HD1+HD2+acidic tail only' constructs.

We agree that constructs such as those mentioned would add to the work. First, we have done the additional construct HD1+HD2+14aa tail, which is presented as $\Delta C12345+14aa$ in Figure 2a and in S2a. Additionally, we performed experiments on the requested C3+C5+14aa and C3+C5 Δ 14aa (see samples 6 and 7 in Author response image 1, which are now included in Supplemental Figure 2b). The results reinforce our hypothesis of an additive effect toward DUX target gene activation by increasing C-terminal repeats and including the 14aa tail.

Author response image1.



1. Related to the above, the flow cytometry data for the MERVL::GFP reporter as presented in Figures 1 and 2, as well as in Supp. Fig. 2, show a considerably large difference in the %GFP|mCherry for the FL construct, ranging from ~6-26%. This makes it difficult to convince the reader which of the different DUX domain constructs cannot or can partially induce GFP|mCherry signal when compared to FL, and hence it is tough to definitively ascertain the exact contribution of each of the 5 C-terminal repeats with high confidence, as it appears that there exists a significant amount of variability in this MERVL::GFP reporter system. The authors need to address this issue since this is their primary method to elucidate the transcriptional activity of each of the mouse DUX repeat domains.

We note that with the Dux^{-/-} cell lines we used throughout the timeline of the study, the percent of %GFP|mCherry expression progressively and slowly decreased – possibly due to slow/modest epigenetic silencing of the reporter. However, we always used the full-length DUX construct to establish the dynamic range. We emphasize that the relative differences between constructs over multiple cell line replicates remained relatively consistent. However, we elected to show absolute values in each experiment, rather than simply normalizing the full-length to 100% and showing relative.

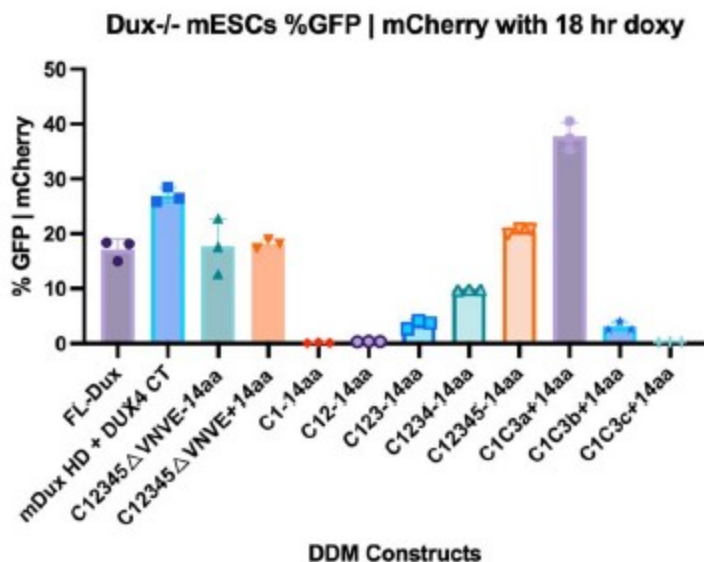
1. Lines 140-142 - The authors claimed that the functional difference between the transcriptionally active and inactive repeats could be narrowed down to a "6aa region which is conserved between repeats C3 and C5, but not conserved in C1, C2 and C4". Assuming the 6aa sequence is DPLELF, why does C1C3a elicit almost twice the intensity of GFP|mCherry signal compared to C3C1c, despite both constructs having the exact same 6aa sequence?

Indeed, C1C3a and C3C1c both containing the ‘active’ DPL sequence but having different relative levels of %GFP | mCherry. This is consistent with these sequences having a positive role in DUX target gene regulation – but likely in combination with other other regions which potentiate its affect, possibly through interacting proteins or post-translational modifications.

Why does DPLEPL (the intermediate C3C1b construct) induce a similar extent of GFP | mCherry signal as the FL construct, even though the former includes 3aa from a transcriptionally inactive repeat? In contrast, GSLELF (the other intermediate C1C3b construct) that also includes 3aa from a transcriptionally inactive repeat is almost completely deficient in inducing any GFP | mCherry signal. Why is that so? Is DPL the most crucial sequence? It will be important to mutate these 3 (or the above 6) residues on FL DUX4 to examine if its transcriptional activity is abolished.

These are interesting points. DPL does appear to be the most important region in the mouse DUX repeats. However, DPL is not shared in the C-terminus of human DUX4. Notably, the DUX4 C-terminus is sufficient to activate the mouse MERV1::GFP reporter when cloned to mouse homeodomains (see Author response image 2, second sample) and other DUX target genes (initially published in Whiddon et al. 2017). One clear possibility is that the DPL region is helping to coordinate the additive effects of multiple DUX repeats, which only exist in the mouse protein.

Author response image 2.



1. Line 154 - The intermediate DUX domain construct C1C3b occupied a different position on the PCA plot from the C1C3c construct that does not contain any of the critical 6aa sequence, as shown in Fig. 2e. However, both these constructs appear to be similarly deficient in inducing any GFP | mCherry signal, as seen in Fig. 2c. Why is that so?

The PCA plot assesses the impact on the whole transcriptome and not just the MERV1::GFP reporter, suggesting the 3aa region has transcriptional effects on the genome beyond what is detected in the MERV1::GFP reporter.

1. To strengthen the claim that "Chromatin alterations at DUX bindings sites require a transcriptionally active DUX repeat", the authors should also perform CUT&Tag for constructs containing transcriptionally inactive DUX repeats (e.g. C1+14aa), and show that such constructs fail to occupy DUX binding sites, as well as are deficient in H3K9ac accumulation.

This is a good comment. We elected to control this with constructs containing or lacking an active repeat. Although we have not pursued this by CUT&TAG, we have examined the impact of DUX constructs with inactive repeats (including the requested C1+14aa, new Figure S1g) by ATAC-seq (see #12, ATAC-seq section, below), and observe no chromatin opening, suggesting that the lack of transcriptional activity is rooted in the inability to open chromatin.

1. It would be good if the authors could also include CUT&Tag data for some of the C1C3 chimeric constructs that were used in Fig. 2, since the authors argued that the minimal 6aa region is sufficient to activate many of the DUX target genes. This would also strengthen the authors' case that the transcriptionally active, not inactive, repeats are critical for binding at DUX binding sites and ensuring H3K9ac occupancy.

We agree that these would be helpful, and have examined the inactive repeats in transcription and ATAC-seq formats during revision (new data in Figures 1d and S1g), but not yet the CUT&TAG format.

1. Line 213 - "SMARCA4" should have been "SMARCA5"? Based on Fig. 4d, SMARCA5 is picked up in the BirA*-DUX interactome, not SMARCA4.

Thanks – corrected.

1. Lines 250-252 - The authors compared the active BirA-C3 against the inactive BirA-C1 to elucidate the interactome of the transcriptionally active C3 repeat, as illustrated in Fig. 5c. They found 12 proteins more enriched in C1 and 154 proteins in C3. This information should be presented clearly as a separate tab in Supp Table 2. What are the proteins common to both constructs, i.e. enriched to a similar extent? Do they include chromatin remodellers too? Although the authors sought to identify differential interactors between the 2 constructs, it is also meaningful to perform 2 separate comparisons - active BirA-C3 against BirA alone control, and inactive BirA-C1 against BirA alone control - like in Fig. 4d, so as to more accurately define whether the active C3 repeat, and not the inactive C1 repeat, interacts with proteins involved in chromatin remodeling.

We thank the reviewer for this comment, and we have modified the manuscript by adding a second sheet in Supplementary Table 2 including the results for enriched proteins in BirA-C1 vs. C3. Additionally, due to limitations of annotation between BirA alone and BirA*-C3 being sequenced in different mass spectrometry experiments, it is difficult to quantitatively compare the two datasets with pairwise comparisons.

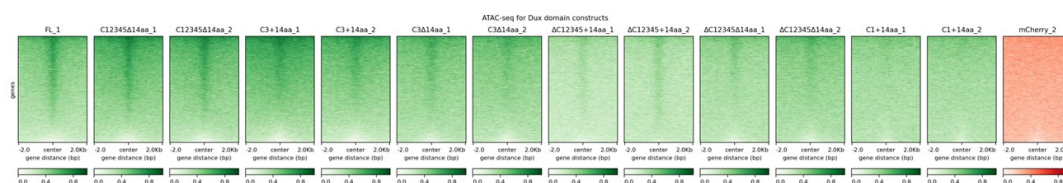
1. Fig 5d: The authors mentioned in the legend that endogenous IP was performed for SMARCC1. However, in line 266, they stated Flag-tagged SMARCC1. Is SMARCC1 overexpressed? The reciprocal IP should also be presented. More importantly, C1 constructs (e.g. C1+14aa and C1Δ14aa) should also be included.

To clarify, Figure 4e used exogenously overexpressed FLAG-SMARCC1 in HEK-293T cells to confirm the results of the full-length DUX BioID experiment. Figure 5d was performed with overexpressed DUX construct, but involved endogenous SMARCC1 in mESCs. This has now been made clearer in the revised manuscript.

1. For both the SMARCC1 CUT&Tag and ATAC-seq experiments shown in Figures 5e and 5f respectively, the authors need to include DUX derivatives that contain transcriptionally inactive repeats with and without the 14aa acidic tail, i.e. C1+14aa and C1Δ14aa, and show that these constructs prevent the binding/recruitment of SMARCC1 to DUX genomic targets, and correspondingly display a decrease in chromatin accessibility. Only then can they assert the requirement of the transcriptionally active repeat domains for proper DUX protein interaction, occupancy and target activation.

We agree that examination of an inactive repeat in certain approaches would improve the manuscript. Importantly, we have now included C1+14 in our ATAC-seq experiments, and in Author response image 3 two individual replicates, which constitute a new Figure S1g. Compared to the transcriptionally active DUX constructs, which see opening at DUX binding sites, we do not see chromatin opening at DUX binding sites with transcriptionally inactive C1+14.

Author response image 3.



1. To prove that DUX-interactors are important for embryonic gene expression, it will be important to perform loss of function studies. For instance, will the knockdown/knockout of SMARCC1 in cells expressing the active DUX repeat(s) lead to a loss of DUX target gene occupancy and activation?

We agree that it would be interesting to better understand SMARCC1 cooperation with DUX function in the embryo, but we believe this is beyond the scope of this paper.

Minor Points

1. Lines 124-126 - What is the reason/rationale for why the authors used one linker (GGGGS2) for constructs with a single internal deletion, but 2 different linkers (GGGGS2 and GAGAS2) for constructs with 2 internal deletions?

With Gibson cloning, there are homology overhang arms for each PCR amplicon that are required to be specific for each overlap. Additionally, each PCR amplicon needs to be specific enough from one another so that all inserts (up to 5 in this manuscript) are included and oriented in the right order. The linker sequences were included in the homology arm overlaps, so the nucleotide sequences for each linker needed to be specific enough to include all inserts. This is a general rule to Gibson cloning. Additionally, both GGGGS2 and GAGAS2 are common linker sequences used in molecular biology and the amino acids structures are similar to one another, suggesting there is no functional difference between linkers.

1. Line 704 - 705: In the figure legend, the authors stated that 'Constructs with a single black line have the linker GGGGS2 and constructs with two black lines have linkers with GGGGS2 and GAGAS2, respectively.' This was not obvious in the figures.

Constructs used for flow and genomics experiments that are depicted in Figure 2, Supplementary Figure 2, Figure 3, Figure 4, and Figure 5 have depicted black lines where

deletions are present. Where these deletions are present, there are linkers in order to preserve spacing and mobility for the protein.

1. Line 160 - Clusters #1 and #2 are likely written in the wrong order. It should have been "activating the majority of DUX targets in cluster #2, not cluster #1" and "failed to activate those in cluster #1, not cluster #2", based on the RNA-seq heatmap in Fig. 2f.

We thank the reviewer for this comment, and the error has been corrected in the manuscript.

1. Line 188 - Delete the word "of" in the following sentence fragment: "DUX binding sites correlating with the of transcriptional".

Thanks – corrected.

1. Line 191 - Delete the word "aids" in the following sentence fragment: "important for conferring H3K9ac aids at bound".

Thanks – corrected.

1. Line 711 - "C1-C3 a,b,d" should be "C1-C3 a,b,c".

Thanks – corrected.

1. Lines 711-712 - The colors "pink to blue" and "blue to pink" are likely written in the wrong order. Based on Fig. 2c, the blue to pink bar graphs should represent C1-C3 a,b,c in that order, and likewise the pink to blue bar graphs should represent C3-C1 a,b,c in that order.

Thanks – corrected.

1. There is an overload of data presented in Fig. 2c, such that it is difficult to follow which part of the figure represents each data segment as written in the figure legend. It is recommended that the data presented here is split into 2 sub-figures.

Figure 2c has a supporting figure in Supplementary Figure 2b. While there is both a graphical depiction of the constructions and the data both in the main panel of Figure 2C, we have depicted it as so to be as clear as possible for the reader to interpret the complexity and presence of amino acids in each of the constructs.

1. Line 717 - "following" is misspelt.

Thanks – corrected.

1. Lines 720-721 - "(Top)" and "(Bottom)" should be replaced with "(Left)" and "(Right)", as the 2 bar graphs presented in Fig. 2d are placed side by side to each other, not on the top and bottom.

Thanks – corrected.

1. Lines 725 and 839 - "Principle" is misspelt. It should be "Principal".

Thanks – corrected.

1. In Figures 3d and 3e, the sample labeled "C3+14_1" should be re-labeled to "C3+14", in accordance with the other sub-figures. Additionally, for the sake of consistency, "aa" should be appended to the relevant constructs, e.g. "C3+14aa" and "C3Δ14aa".

Thanks – corrected.

1. Line 773 - Were the DUX domain constructs over-expressed for 12hr (as written in the figure legend) or 18hr (as labeled in Fig. 5d)?

Thanks – corrected.

1. Related to minor point 19 above, is there a reason/rationale for why some of the experiments used 12hr over-expression of DUX domain constructs (e.g. for CUT&TAG in Fig. 3), whereas in other experiments 18hr over-expression was chosen instead (e.g. flow cytometry for MERVL::GFP reporter in Figures 1 and 2, and co-IP validations of BirA*-DUX interactions in Fig. 4)?

Thanks for the opportunity to explain. In this work, experiments that reported on proteins that are translated following DUX gene activation (e.g. MERVL:GFP via flow) were done at 18hr to allow for enough time for transcription and translation of GFP (or other DUX target genes). For experiments that report on the impact of DUX on chromatin and transcription, such as RNA-seq, CUT&Tag, and ATAC-seq, we induced DUX domain constructs for 12 hours.

1. Line 804 - "ΔHDs" is missing between "C2345+14aa" and "ΔHD1".

Thanks – corrected.

1. In Fig. 5c, "Chromatin remodelers" is misspelt.

Thanks – corrected.

1. There is no reference in the manuscript to the proposed model that is presented in Fig. 6b.

Thanks – corrected.

Reviewer #3 (Recommendations For The Authors):

Given the uncertainty of the function of the Dux peptide repeats in mice, could it not also be possible that the underlying repeated nature of the (coding) DNA? That is, could these DNA repeats exert a regulatory function on Dux transcription itself (also given the dire consequences of misregulated DUX4 expression as seen in FSHD, for example).

Yes, it remains possible that the internal coding repeats within Dux are playing a role in locus regulation, and might be interesting to examine. However, we consider this question as being outside the scope of the current paper.

*Finally, it would be interesting to know whether these repeats are, in fact, present in all mouse species. Already no longer present in rat, do they exist, or not, in more "distant" mice, e.g. *M. caroli*?*

Determining whether all mouse strains contain C-terminal repeats in DUX is a question we also considered. However, Dux and its orthologs are present in long and very complex repeat

arrays that are not present in the sequencing data or annotation in other mouse strains. Therefore, we are not unable to answer this question from existing sequencing data. Answering would require a considerable genome sequencing and bioinformatics effort, or alternatively a considerable effort aimed at cloning ortholog cDNAs from 2-cell embryos.

Minor points:

line 169: here it seems, in fact, that the 'inactive' C2, C4 repeats are more similar to each other (my calculation: 91 and 96% identity at the protein and DNA level, respectively) than the active C3 and C5 repeats (82 and 89% identity, resp.), the outlier being C1.

Thanks for this comment, which was mentioned by other reviewers as well and has been addressed through new statistical analyses and interpretation (see new Figure S1d).

line 191: I'm not sure this sentence parses correctly ("...14AA tail is important for conferring H3K9Ac aids at bound sites...")

We thank the reviewer for this comment, and we have corrected the sentence in the manuscript.