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

Christina M. Smith[1], Edward J. Grow[1,2], Sean C. Shadle[1], and Bradley R. Cairns[1,*]

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

[2]Green Center for Reproductive Biological Sciences, Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, TX, USA

[*]Corresponding Author and Lead Contact: brad.cairns@hci.utah.edu

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

23 mechanistic studies reveal cooperativity between DUX active repeats and the acidic tail in
24 cofactor recruitment, DUX target opening, and transcription. Taken together, we provide
25 several new insights into DUX structure-function, and mechanisms of chromatin and gene
26 regulation.
27
28 **Introduction**
29 Mouse pre-implantation development is a dynamic and complex process, which begins
30 with a single-cell zygote and results in an implanted embryo bearing cells with different lineage
31 specifications and fates. Within this period, numerous molecular processes must be properly
32 completed, which are driven (in part) by gene expression changes and chromatin modifications
33 (1,2). A hallmark of pre-implantation development is major embryonic genome activation
34 (EGA), which occurs at the 2-cell stage in the mouse embryo(3). This critical process includes
35 decreases in maternally-deposited mRNAs and proteins, combined with the onset of both
36 embryo-derived transcription and translation – effectively turning developmental control over
37 to the embryo from maternal resources in the egg (4). During EGA, chromatin undergoes
38 considerable remodeling to enable enhancer and promoter accessibility, allowing for the onset
39 of major embryonic transcription, and the marking of promoters and enhancers for future
40 regulation (5–7).
41 Mouse embryonic stem cells (mESCs) are derived from the inner cell mass (ICM) of the
42 blastocyst, and have enabled studies focused on the characterization of the blastocyst and
43 downstream embryonic differentiation(8). In addition, the mESC system has been extended to
44 examine properties of earlier cleavage-stage totipotent-related cells. Specifically, a small (<1%)

45 population of spontaneously fluctuating mESCs has been identified, which closely resemble
46 cells within the 2-cell embryo, termed 2-cell-like cells (2CLC) (9–12). Our lab and others showed
47 that the expression of *Dux*, which encodes the double homeodomain transcription factor DUX,
48 is sufficient to confer this 2C-like state to mESCs (12–14). However, we currently have an
49 incomplete understanding regarding how DUX establishes the 2C transcriptome. *DUX4*, the
50 human *Dux* ortholog is similarly expressed during EGA, which occurs at the 4-cell stage in
51 humans. Likewise, *DUX4* expression in human ESCs converts these cells to transcription
52 programs resembling the 4-8 cell human cleavage-stage embryo (13,15–17). *DUX4* has
53 additionally been identified as the causal gene in facioscapulohumeral dystrophy (FSHD),
54 leading to the protein being intensely studied to identify its pathological traits (18–20).
55 Mouse DUX is known to be targeted to specific sites in the genome (including genes and
56 MERVL LTRs) by its two N-terminal homeodomains, whereas its C-terminal region drives
57 transcriptional activation(12,21,22). However, little is known about the nature of this C-
58 terminal region, which is much more divergent among *DUX* orthologs than are the
59 homeodomains (23,24). One possible contributor to transcriptional activation involves the
60 interaction of the C-terminus of DUX4 with the related histone acetyltransferases p300 and CBP
61 (22,25). However, these or other interactions have not been identified with mouse DUX.
62 Regarding chromatin remodeling during EGA, the abundant ATP-dependent chromatin
63 remodeler SWI/SNF (BAF) complex plays a central role(26,27). Specifically, SMARCA4, the
64 catalytic ATPase within SWI/SNF complex, is required for important for EGA in the 2-cell mouse
65 embryo (28). SMARCC1 is a core structural scaffold for SWI/SNF complexes, and its absence
66 results in loss of complex activity *in vivo* (29). However, possible physical links between mouse

67 DUX and the chromatin remodeler SWI/SNF complex during EGA or during the mESC to 2C-like
68 transition is currently unknown.
69 DUX activates about 25% of EGA genes and DUX expression is sufficient to revert mESCs
70 to an earlier totipotent stage(12). Consistent with DUX impacting early development, *Dux-/-*
71 crosses yield smaller litter sizes (21). However, as viable *Dux -/-* pups are generated, factor(s)
72 redundant with DUX appear to contribute to DUX target activation and chromatin
73 reprogramming in 2-cell embryos (21,30). Recently, oocyte-specific homeobox 4 (OBOX4), a
74 factor highly expressed during EGA and in 2C-like mESCs, has been identified as partly
75 redundant with DUX in 2-cell embryos(31). Taken together, DUX and OBOX4 appear to
76 cooperate to ensure proper EGA in the early mouse embryo.
77 In this study, we sought to better characterize DUX domains, identify proteins
78 interacting with DUX domains, and understand how DUX domains contribute to DUX target
79 gene activation, DUX promoter recruitment, and chromatin accessibility at binding sites.
80 Specifically, we reveal a unique C-terminal repeat structure in mouse DUX, and characterize two
81 functionally distinct repeat types (termed 'active' and 'inactive') and cooperativity of the active
82 repeats with the C-terminal acidic tail. Additionally, we identify DUX domains that interact with
83 the chromatin remodeler subunit SMARCC1, which alongside the ability of DUX to confer
84 histone acetylation, reveals new mechanistic insight into how the C-terminus of DUX confers
85 transcriptional activation.
86
87 **Results**
88

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.

110 Next, we assessed the contribution of mouse DUX repeats to transcriptional activity by
111 examining *Dux*-/- mESCs containing a MERVL::GFP reporter. The long-terminal repeat (LTR) of
112 MERVL is bound by DUX, and MERVL:GFP is commonly used to assess DUX activation and
113 conversion of mESCs to a 2CLC state (9,12). First, *Dux*-/- mESCs were transiently transfected
114 with mCherry-tagged DUX repeat domain constructs encoding zero to five C-terminal repeats,
115 without the 14aa tail (Figure 1d). Cells were induced with doxycycline for 18 hours, gated on
116 mCherry positivity, and MERVL::GFP expression was quantified by flow cytometry
117 (Supplementary Figure 1e). Expression of each derivative was confirmed by Western analysis
118 (Supplementary Figure 1f). The flow cytometry results reveal that the first two repeats have
119 minimal activity, but with the addition of the third repeat, low MERVL::GFP expression is
120 observed. Activity is further improved by the addition of the fourth and fifth repeat. The results
121 also support a potential role for the 14aa 'tail' in activation, as its omission modestly lowered
122 MERVL::GFP expression (Figure 1d). Thus, our initial deletion series revealed activity with
123 repeats 3-5, but not with repeats 1-2 alone. However, this experiment did not determine
124 whether the repeats act in a simple additive manner, or if individual repeats confer distinct
125 transcriptional activity in a singular manner, or how they cooperate with the acidic tail.
126

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

129 We next explored the contributions of individual repeats, along with their cooperation
130 with the C-terminal acidic 'tail'. Here, we generated N-terminal mCherry-tagged DUX
131 derivatives with varying numbers of repeats either with or without the 14aa tail, as well as with

132 or without single or both homeodomains (Supplementary Figure 2a). All constructs contain a
133 $GGGGS_2$ linker present in constructs with a single internal deletion and $GGGGS_2$ and $GAGAS_2$
134 with constructs with two internal deletions. First, as expected, derivatives lacking either
135 homeodomain failed to activate the MERVL::GFP reporter (Supplementary Figure 2a). Second,
136 combinations involving C1, C2 or C4 with the acidic 'tail' lack activation of the MERL::GFP
137 reporter, whereas combinations of C3 or C5 with the 'tail' display potent reporter activation
138 (Figure 2a). The latter combinations require the 14aa tail as constructs lacking the tail, but
139 containing the C3 or C5 repeat alone, did not activate the reporter. Notably, a construct
140 containing C1 and C2 with the 14aa tail was inert, whereas C1, C2 and C3 with the 14aa tail was
141 highly active (Supplementary Figure 2a,b). Together, these results suggest that strong DUX
142 target gene activation requires a combination of two C-terminal domains: a transcriptionally
143 'active' repeat (C3 or C5) and the 14aa tail (Figure 2b).
144 To determine the amino acids selectively involved in DUX target activation within the
145 active C3 and C5 repeats, chimeric constructs between C1 (a transcriptionally 'inactive' repeat)
146 and C3 (a transcriptionally 'active' repeat) were made, all containing the 14aa tail. The
147 cutoff/transition point between chimera regions occurred in 3 aa increments along the 2
148 repeats (Figure 2c, Supplementary Figure 2c). This narrowed down the functional difference to
149 a 6 aa region which is conserved between repeats C3 and C5, but not conserved in C1, C2, and
150 C4. Importantly, replacement solely of this 6 aa sequence from the active C3 repeat into
151 inactive C1 repeat (termed C1DPLELF+14aa) conferred robust activation of the MERVL::GFP
152 reporter (Figure 2d).

153 In order to expand these analyses to global transcription and all DUX targets (beyond
154 the MERVL::GFP reporter), we performed RNA-seq on full-length (FL) DUX and selected chimera
155 constructs. This approach was designed to test whether the 'active' constructs might differ in
156 their gene targets (Figure 2e and Supplementary Figure 2d). First, the PCA plot for these
157 samples separated samples largely into two groups – active versus inactive – with one sample
158 in an intermediate position (Figure 2e). Interestingly, the DUX constructs which activate the
159 MERVL::GFP reporter (FL, C1C3a, and C3+14a) all showed strong activation of DUX target genes,
160 whereas the 'inactive' constructs that do not activate the MERVL::GFP reporter (C1+14aa,
161 C1C3c, and mCherry alone) did not activate DUX target genes (Figure 2f). Notably, the
162 intermediate position was occupied by the C1C3b derivative, reinforcing the notion that the 6aa
163 region is important for regulating target activation. K-means clustering of the RNA-seq for the
164 chimera constructs at DUX target generated three clusters of genes: the first two of which
165 largely contained genes activated by FL DUX, whereas the cluster #3 contains genes that have
166 increased expression in DUX domain derivatives compared to FL DUX. The 'intermediate'
167 derivative C1C3b, which is a hybrid of an 'active' and an 'inactive' repeat, was capable of
168 activating the majority of DUX targets in cluster #2, but failed to activate those in cluster #1,
169 whereas it activated (or failed to attenuate) genes in cluster #3. Genes activated by C1C3b in
170 cluster #3 generate GO terms such as 'upregulated include synapse assembly', 'sex
171 differentiation', 'reproductive structure development', and 'neuron differentiation'. Here,
172 cluster 3 shows examples of known DUX target genes that have a higher relative expression in
173 C1C3b compared to full-length DUX include *Obox1*, *Olfr1277*, *Obox2*, and *Zfp296*. These
174 combined observations, especially those involving derivatives that cause the precocious

175 activation of genes in cluster #3, raise the interesting possibility that the 'inactive' repeats may
176 serve to attenuate the active repeats at particular targets, such as *Obox* genes. Finally, when
177 considering amino acid similarity between the five repeats, we note that C3 and C5 are most
178 similar to one another, compared to C1, C2, and C4 (Supplementary Figure 2e).
179
180 **Chromatin alterations at DUX binding sites are associated with a transcriptionally active DUX**
181 **repeat**
182 To investigate the chromatin landscape effects conferred by *Dux* derivative expression,
183 we performed CUT&Tag in *Dux*-/-mESCs on selected *Dux* domain constructs: FL, C12345△14aa,
184 C3+14aa, C3△14aa, △C12345+14aa, and mCherry alone. Because all of the constructs had the
185 same N-terminal mCherry tag, an mCherry antibody was used throughout. We isolated and
186 analyzed clonal stable cell lines which expressed species of the expected protein size and
187 demonstrated comparable expression to one another (Figure 3a). We note that mCherry-DUX
188 protein is expected to run at roughly 100kDa, but runs anomalously on polyacrylamide gels – at
189 150kDa. This is likely due to a property of the C-terminal repeats as a DUX derivative consisting
190 of the homeodomains alone runs at the expected molecular weight. To benchmark DUX
191 CUT&Tag against previous chromatin occupancy data, we first generated a CUT&Tag dataset for
192 mCherry-tagged full-length DUX protein, and found that it overlapped highly with previously
193 published HA-tagged DUX ChIP-seq results (Supplementary Figure 3a) (12). Extension of
194 CUT&Tag to the DUX derivatives revealed their moderate-to-high occupancy at DUX sites,
195 except for C3△14aa, which displayed reduced occupancy at DUX binding sites (Figure 3b, 3c).
196 Next, we determined the ability of our DUX domain constructs to alter H3K9ac, as H3K9ac

197 globally increases in 2C-like cells (32). Notably, H3K9ac occurs at DUX binding sites correlating
198 with the transcriptional activity of each *Dux* construct (Figure 3d,e, Supplementary Figure 3b).
199 Taken together, our data demonstrate that the 14aa acidic tail is not strictly required for the
200 occupancy of DUX targets. Furthermore, our data suggest that the 14aa tail is important for
201 conferring H3K9ac at bound sites and might also be involved in retention of DUX at target sites
202 since removal of the 14aa tail (in the context of a single active repeat) diminished target
203 occupancy.
204
205 **Proximity-labeling with full-length DUX reveals interactions with proteins involved in**
206 **chromatin de-repression.**
207 Next, we identified proteins that interact with DUX derivatives using the BioID approach,
208 which identifies proteins that interact directly, indirectly, or transiently with a protein of
209 interest (33). Here, we utilized BioID to define protein interactions with DUX in the 2C-like
210 cellular environment. To accomplish this, we fused BirA*, a promiscuous biotinylating enzyme,
211 to the N-terminus of full-length DUX. When biotin is added to the growth media, BirA* will
212 biotinylate lysine residues of proteins within a 10nm radius. Our approach involved the
213 isolation of biotinylated proteins and the use of liquid chromatography mass spectrometry (LC
214 MS/MS) to identify associated proteins in a quantitative manner (Figure 4a). Our experimental
215 design compared expression of BirA*-DUX (full-length) in mESCs (which creates a 2C-like state)
216 to expression of unfused BirA* alone (control) expressed in a 2C-like environment, in order to
217 mimic a DUX-expressing 2C-like proteome in both conditions (Figure 4b). Here, we created
218 clonal stable cell lines harboring our constructs, and expressed the proteins for 12hr with 0.25

219 µg/mL doxycycline and 5µM biotin. DESeq2 analysis of the sum intensity values from LC MS/MS
220 revealed clustering between replicates (Supplementary Figure 4a) and a volcano plot revealed
221 an enrichment in protein interactions for BirA*-DUX compared to BirA* alone, including those
222 involved in chromatin de-repression such as the SWI/SNF (BAF) complex members SMARCC1
223 and SMARCA5 (Figure 4c, 4d). In addition to DESeq2 analysis, which is normally used for RNA-
224 seq data, we applied the SAINTexpress package, which is designed for protein interactome
225 analysis, to additionally analyze this dataset. DESeq2 appeared to be slightly less stringent than
226 SAINTexpress analysis, with a larger number passing statistical threshold (p_{adj}<0.05)
227 (Supplementary Figure 4b). SAINTexpress uses a statistical score called SAINTscore ranging from
228 0-1, with a significance threshold established as ≥ 0.73, which parallels an estimated protein-
229 level Bayesian FDR of <0.05(34).
230 SWI/SNF subunit SMARCC1 and the transcription factor ZSCAN4D, two statistically
231 significant hits from both analyses, were then selected for validation by testing for interaction
232 *in vivo* via immunoprecipitation. We expressed N-terminally FLAG-tagged versions of SMARCC1
233 and ZSCAN4 in HEK293 cells co-expressing the BirA*-DUX fusion in the presence of 50 µM biotin
234 (50µM biotin for BioID was empirically determined as the optimal dose in HEK293Ts). SMARCC1
235 and ZSCAN4 were immunoprecipitated with anti-FLAG conjugated beads, and analysis with an
236 avidin-HRP blot showed they are both biotinylated and co-IP with DUX protein, validating BioID
237 protein interactions (Figure 4e). Additionally, CHAF1A, an additional BioID hit, was also
238 validated through FLAG-tag co-IP and DUX protein interaction, however KDM4C did not
239 validate, suggesting that this hit was a false positive (Supplementary Figure 4c), and highlighting
240 the importance of validating the BioID results.

241 Lastly, it has been previously reported that human DUX4 interacts with p300 (22,25),
242 however p300 was not identified as a significant interactor in our dataset. This could be due to
243 differences in the interactomes of human DUX4 and mouse DUX, or instead be technical, due
244 perhaps to a lack of sensitivity to detect P300 peptides. To determine if mouse DUX similarly
245 interacts with P300, we performed a BioID-like experiment in mESCs, which revealed p300 as
246 modestly biotinylated, suggesting that a portion of p300 comes in close proximity with DUX, but
247 was not annotated in the dataset, representing a possible false negative (Supplementary Figure
248 4d). In summary, our BioID experiment included chromatin- and transcription-related factors, a
249 portion of which were validated using co-immunoprecipitation, which support prior
250 observations of a strong impact of DUX occupancy on chromatin opening and transcriptional
251 activity.
252
253 **Transcriptionally active DUX C-terminal repeat C3 binds to proteins involved in chromatin**
254 **remodeling**
255 Our work above presents the opportunity to identify proteins that interact differentially with
256 a transcriptionally 'active' DUX derivative compared to a highly similar 'inactive' DUX derivative.
257 Here, we separately N-terminally tagged our C1+14aa and C3+14aa derivatives with BirA*, to
258 represent transcriptionally 'inactive' and 'active' DUX constructs, respectively (Figure 5a). The
259 experimental design and setup was identical to that described in Figure 4a, but the mESC line
260 utilized bore a *Dux*-/- genotype. Furthermore, in place of the BirA* alone control, the inactive
261 BirA*-C1 (BirA* fused to C1+14aa) was used as the control construct for comparison to the
262 active BirA*-C3 (BirA* fused to C3+14aa).

263 PCA and clustering analysis reveals that the three replicates of the 'active' C3 derivative
264 group together and are more similar to each other (via clustering) than are the C1 replicates,
265 and vice versa (Supplementary Figure 4e,f). Using SAINTexpress, we observed many more
266 statistically-enriched (SAINTscore ≥ 0.73) protein interactions occurring with the DUX C3
267 'active' repeat compared to the 'inactive' C1 repeat. Specifically, we observed 12 proteins with
268 the C1 derivative, and 154 proteins with C3 (Figure 5b). For C3 derivative interactors, GO-term
269 analysis identified two overarching functional categories: chromatin and the nucleolus. Within
270 the chromatin category, proteins within the NuRD, SWI/SNF and ISWI chromatin remodeling
271 complexes were prominent (Figure 5c). Notably, proper formation of the nucleolus has been
272 found to be required for *Dux* repeat locus silencing (35,36), and our data strongly suggests a
273 direct physical interaction with nucleolar components, as DUX can bind to its own genetic locus
274 (37).
275 To further investigate chromatin-related interactors of interest, we performed co-IPs for
276 FLAG-tagged SMARCC1 (BAF155) (described above) using selected DUX domain derivatives.
277 First, SMARCC1 binding to DUX appears only slightly diminished with the loss of the 14aa tail
278 (Figure 5d). SMARCC1 also binds well to a DUX derivative containing both an active C3 repeat
279 and the acidic tail (C3+14aa), and binding is largely maintained when the tail is removed
280 (C3△14aa). Finally, a major reduction in SMARCC1 binding is observed when all repeats are
281 removed and only the acidic tail remains. This suggests a hierarchy of interaction for DUX and
282 SMARCC1: an 'active' repeat alone binds SMARCC1 whereas the acidic 'tail' alone does not bind
283 SMARCC1, however the acidic tail slightly improves SMARCC1 interaction when active and
284 inactive repeats are both present (full-length DUX).

285 Lastly, we wanted to determine whether the SWI/SNF (BAF) component SMARCC1 is
286 recruited to DUX binding sites by DUX, and if so, which DUX domains confer this function. To
287 address this, we performed CUT&Tag for endogenous SMARCC1 in the DUX derivatives and
288 found, similar to our co-IP results, that SMARCC1 is recruited to DUX binding sites by DUX
289 derivates that contain active C-terminal repeats but not by the 14aa tail (Figure 5e). Thus,
290 binding and recruitment of SWI/SNF to DUX targets *in vivo* appears to rely on interaction with
291 active DUX C-terminal repeats.
292 To continue investigating DUX interaction with SWI/SNF, we performed ATAC-seq in the DUX
293 derivatives to determine whether SMARCC1 recruitment to DUX binding sites correlates with
294 chromatin accessibility changes (Figure 5f, Supplementary Figure 4g). The results suggest a
295 hierarchy of interaction similar to that observed in our SMARCC1 CUT&Tag experiments. At
296 DUX binding sites, full-length DUX leads to increased opening, as do C12345△14aa and
297 C3+14aa, the two derivatives which confer transcriptional activity. However, DUX derivatives
298 which lack transcriptional activity or have decreased binding to SMARCC1 display a
299 corresponding decrease in chromatin opening, though some opening does occur. This evidence
300 further supports a combinatorial domain model for proper DUX protein interaction, occupancy,
301 and target activation – revealing a major/required contribution from the C-terminal repeats and
302 augmentation by the 14 aa C-terminal tail.
303
304 **Discussion**
305 Embryonic genome activation (EGA) creates a transcriptome and proteome that replaces
306 maternally-inherited components, reprograms chromatin and drives cleavage-stage processes

307 (12). The transcription factor DUX is highly but transiently expressed during EGA (specifically
308 the early 2-cell embryo stage) from a large locus containing ~30 *Dux* repeats. DUX plays a key
309 role in opening many promoters and enhancers, and in activating endogenous genes and repeat
310 elements in the 2-cell embryo (37). DUX targets account for about 25% of the cleavage-stage
311 transcriptome, and partly overlap with targets of OBOX4, underlying the observation that
312 combined DUX and OBOX4 omission/depletion strongly affects embryo development (31).
313 Although DUX targeting by its N-terminal homeodomains is well understood, how domains in
314 the C-terminus of DUX contribute to chromatin reprogramming and transcription has remained
315 largely unknown (23). Here, we conducted an analysis combined with proximity ligation to
316 reveal the protein partners of DUX domains and the roles of DUX domains in chromatin
317 remodeling and transcriptional activation.
318 First, we found that the mouse DUX C-terminal region contains five ~100aa repeats (termed
319 C1-5) followed by a 14aa acidic C-terminal tail, whereas DUX orthologs contain only one copy of
320 this repeat, followed by a similar acidic tail. Our phylogenetic and DUX domain analyses
321 suggest that two of the repeats (C3 and C5) perform a more active role than the 'inactive'
322 C1/2/4 repeats. Interestingly, single 'active' mouse repeats show no transcriptional stimulation
323 on their own, but show potent activity in combination with the acidic tail, essentially mimicking
324 the domain-activity relationships of their human, rat or bovine DUX4 orthologs. Together,
325 these observations suggest that additional DUX repeats in the mouse lineage may have derived
326 from duplication of the original 'active' repeat, with subsequent sub-functionalization.
327 These observations raise two interesting questions – why does DUX have multiple repeats
328 with different activities, and what distinguishes active from inactive repeats? Our work has

329 localized the primary difference between active and inactive repeats to a small 6 aa region.
330 Furthermore, active versus inactive repeats (in combination with the acidic tail) show very
331 different profiles of protein interaction – with active repeats displaying more robust interaction
332 with chromatin and transcription partners. Interestingly, we observe a cohort of genes (Figure
333 2f, cluster #3) that are silent with full-length DUX, silent with an 'inactive' repeat (C1), but
334 active with a single 'active' repeat (C3). This raises the possibility that the 'inactive' repeats
335 may serve to attenuate the activity of the active repeats under certain conditions or at certain
336 locations. Of note, DUX binds to its own promoter to help activate a strong positive feedback
337 loop; however, *Dux* activation is brief (half a cell cycle), with silencing accompanying the
338 association of the large *Dux* locus with the nucleolar envelope(37,38). Interestingly, we
339 observe high enrichment of nucleolar components with the 'active' C3 repeat, raising the
340 possibility that the 'active' repeat may initially recruit powerful chromatin and transcription
341 factors, and then subsequently contribute to localizing the *Dux* locus to a compartment that
342 overcomes and/or reverses those associations to enable *Dux* locus silencing(35). Furthermore,
343 interactions of DUX with histone deacetylase components were obtained, which remain to be
344 fully explored in this silencing process. These observations provide new mechanistic
345 information as well as opening new avenues for future exploration regarding the ability of DUX
346 to function as a strong activator, while also helping to subsequently silence the repetitive *Dux*
347 locus. Here, we envision that temporal and/or locus-specific post-translational modifications
348 on DUX may mediate the temporal associations of DUX with target factors. Indeed, the
349 presence of both 'active' and 'inactive' DUX repeats may help in coordination of these
350 associations.

351 Regarding the functions of the DUX C-terminus, prior work of others has shown that the
352 acidic C-terminal tail of DUX4 promotes transcriptional activation and interaction with the
353 histone acetyltransferase p300(22). Our validation experiments confirmed those results with
354 mouse DUX, and we extended our studies to explore the roles and relationships of the repeat
355 regions, both alone and in combination with the acidic tail. Our examination of *Dux* domain
356 constructs was extensive, and included RNA-seq, CUT&Tag, BioID, ATAC-seq and the ability to
357 convert mouse ES cells to 2C-like cells. As DUX opens chromatin, and as three SWI/SNF
358 complex components were present in our BioID results (SMARCC1, SMARCA4, ACL6A), these
359 results prompted our direct assessment and validation of the recruitment of SWI/SNF complex
360 (monitoring the core SMARCC1 subunit) to DUX binding sites (Figure 5c, Figure 6a).
361 An additional goal was to explore whether the functions and associations of the active
362 repeats versus the acidic tail are largely separable/independent, or instead primarily additive or
363 cooperative. First, regarding the activation of cleavage-specific genes (including MERVL), the
364 target repertoire observe with full-length DUX is largely recapitulated with DUX derivatives
365 containing the homeodomains and both active C-terminal repeats – even in the absence of the
366 acidic tail. Thus, the tail accentuates activity, but is not strictly required for activity at the vast
367 majority of targets. Regarding domain contributions to occupying DUX targets, all derivatives
368 were able to bind DUX targets, though the absence of the acidic tail lowers occupancy if only
369 one active repeat is present. This contrasts with the recruitment of H3K9ac activity; here,
370 neither the acidic tail nor an active repeat is sufficient for recruitment; instead, either multiple
371 repeats or repeat-tail combinations appear needed. For SMARCC1 (SWI/SNF), an active repeat
372 is sufficient for recruitment, but recruitment is enhanced by the addition of the tail, whereas

373 the tail alone appears ineffective. Finally, and of particular interest, a DUX derivative lacking
374 repeat regions but containing the acidic tail (△C12345+14) occupies DUX targets, but does not
375 activate targets – and does not recruit either H3K9ac or SWI/SNF complex. Therefore, the
376 acidic tail appears to be effective only when combined with an active C-terminal repeat (Figure
377 6a). Overall, our data is consistent with largely cooperative functions for active repeat(s) and
378 the acidic tail – as the absence of either compromises most functions and interactions – but
379 places the C-terminal domains as more important domain in the hierarchy, as key factors like
380 SWI/SNF complex are fully reliant on an active repeat for interaction. In keeping, our work
381 predicts interactions with chromatin and transcription factors involving combined surfaces on
382 repeats and the acidic tail (a two-domain model).
383 As previewed above, our proximity-labeling (BioID) revealed a variety of protein complex
384 interactions, and we validated a subset of interactors with co-IP experiments, including
385 particular chromatin and transcription-related factors. We note that certain interactions may
386 be indirect and involve DUX association with component(s) involved in a larger multi-protein
387 structures, such as the highly repetitive *Dux* locus (to which DUX binds(12,37)) in either its
388 active or inactive form, and the association of the *Dux* locus with the nucleolus or other
389 compartments. Thus, a complete understanding of the full repertoire of interactions will
390 require future studies. However, beyond the factors noted above, we will emphasize two
391 additional factors prominent in the BioID data that might be prioritized. First, the p53
392 interaction verifies prior work supporting p53 association with and activation of the *Dux*
393 locus(37), and prompts examination of a direct interaction with DUX. Second, the presence in
394 our dataset of SMCHD1, a cohesin-like protein known to help confer repression of the mouse

395 *Dux* and human *DUX4* repetitive loci, and to be mutated in a subset of FSHD patients, likewise
396 prompts considerable future attention (39,40).
397 Taken together our work provides considerable new mechanistic information regarding the
398 structure-function relationships and factor associations conducted by particular domains within
399 the mouse DUX C-terminal region. Key findings include the characterization of 'active' repeat
400 domains, their interaction with transcription and chromatin remodeling factors, and the
401 cooperativity between DUX active repeats and the acidic tail in cofactor recruitment and
402 transcriptional activation at DUX targets. In addition to factors for activation, factors related to
403 repression and nucleolar association were also isolated, and future studies should help
404 determine whether their temporal associations with DUX may help early embryos execute a
405 quick 'burst' of *Dux* activation followed by repression at the *Dux* locus in 2-cell embryos.
406
407

408 **References**

409 1. Jukam D, Shariati SAM, Skotheim JM. Zygotic Genome Activation in Vertebrates. Dev Cell.
410 2017 Aug;42(4):316–32.

411 2. Gao Y, Liu X, Tang B, Li C, Kou Z, Li L, et al. Protein Expression Landscape of Mouse Embryos
412 during Pre-implantation Development. Cell Rep. 2017 Dec 26;21(13):3957–69.

413 3. Pálfy M, Joseph SR, Vastenhouw NL. The timing of zygotic genome activation. Curr Opin
414 Genet Dev. 2017 Apr 1;43:53–60.

415 4. Lee MT, Bonneau AR, Takacs CM, Bazzini AA, DiVito KR, Fleming ES, et al. Nanog, Pou5f1 and
416 SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature.
417 2013 Nov;503(7476):360–4.

418 5. Vallot A, Tachibana K. The emergence of genome architecture and zygotic genome
419 activation. Curr Opin Cell Biol. 2020 Jun 1;64:50–7.

420 6. Saitou M, Kagiwada S, Kurimoto K. Epigenetic reprogramming in mouse pre-implantation
421 development and primordial germ cells. Development. 2012 Jan 1;139(1):15–31.

422 7. Xu R, Li S, Wu Q, Li C, Jiang M, Guo L, et al. Stage-specific H3K9me3 occupancy ensures
423 retrotransposon silencing in human pre-implantation embryos. Cell Stem Cell. 2022 Jul
424 7;29(7):1051-1066.e8.

425 8. Bryja V, Bonilla S, Arenas E. Derivation of mouse embryonic stem cells. Nat Protoc. 2006
426 Nov;1(4):2082–7.

427 9. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, et al. Embryonic stem
428 cell potency fluctuates with endogenous retrovirus activity. Nature. 2012 Jul;487(7405):57–
429 63.

430 10. Rodriguez-Terrones D, Hartleben G, Gaume X, Eid A, Guthmann M, Iturbide A, et al. A
431 distinct metabolic state arises during the emergence of 2-cell-like cells. EMBO Rep. 2020 Jan
432 7;21(1):e48354.

433 11. Genet M, Torres-Padilla ME. The molecular and cellular features of 2-cell-like cells: a
434 reference guide. Development. 2020 Aug 26;147(16):dev189688.

435 12. Hendrickson PG, Doráis JA, Grow EJ, Whiddon JL, Lim JW, Wike CL, et al. Conserved roles
436 of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL
437 retrotransposons. Nat Genet. 2017 Jun;49(6):925–34.

438 13. De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D. DUX-family transcription factors
439 regulate zygotic genome activation in placental mammals. Nat Genet. 2017 Jun;49(6):941–5.

440 14. Whiddon JL, Langford AT, Wong CJ, Zhong JW, Tapscott SJ. Conservation and innovation
441 in the DUX4-family gene network. Nat Genet. 2017 Jun;49(6):935–40.

442 15. Campbell AE, Belleville AE, Resnick R, Shadle SC, Tapscott SJ. Facioscapulohumeral
443 dystrophy: activating an early embryonic transcriptional program in human skeletal muscle.
444 Hum Mol Genet. 2018 Aug 1;27(R2):R153–62.

445 16. Yoshihara M, Kirjanov I, Nykänen S, Sokka J, Weltner J, Lundin K, et al. Transient DUX4
446 expression in human embryonic stem cells induces blastomere-like expression program that
447 is marked by SLC34A2. Stem Cell Rep. 2022 Jun 30;17(7):1743–56.

448 17. Taubenschmid-Stowers J, Rostovskaya M, Santos F, Ljung S, Argelaguet R, Krueger F, et
449 al. 8C-like cells capture the human zygotic genome activation program in vitro. Cell Stem Cell.
450 2022 Mar 3;29(3):449-459.e6.

451 18. Ferreboeuf M, Mariot V, Bessières B, Vasiljevic A, Attié-Bitach T, Collardeau S, et al.
452 DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles. Hum Mol
453 Genet. 2014 Jan 1;23(1):171–81.

454 19. Vanderplanck C, Ansseau E, Charron S, Stricwant N, Tassin A, Laoudj-Chenivesse D, et al.
455 The FSHD Atrophic Myotube Phenotype Is Caused by DUX4 Expression. PLOS ONE. 2011 Oct
456 28;6(10):e26820.

457 20. Le Gall L, Sidlauskaite E, Mariot V, Dumonceaux J. Therapeutic Strategies Targeting DUX4
458 in FSHD. J Clin Med. 2020 Sep;9(9):2886.

459 21. Chen Z, Zhang Y. Loss of DUX causes minor defects in zygotic genome activation and is
460 compatible with mouse development. Nat Genet. 2019 Jun;51(6):947–51.

461 22. Choi SH, Gearhart MD, Cui Z, Bosnakovski D, Kim M, Schennum N, et al. DUX4 recruits
462 p300/CBP through its C-terminus and induces global H3K27 acetylation changes. Nucleic
463 Acids Res. 2016 Jun 20;44(11):5161–73.

464 23. Leidenroth A, Hewitt JE. A family history of DUX4: phylogenetic analysis of DUXA, B, C
465 and Duxbl reveals the ancestral DUX gene. BMC Evol Biol. 2010;10(1):364.

466 24. Leidenroth A, Clapp J, Mitchell LM, Coneyworth D, Dearden FL, Iannuzzi L, et al.
467 Evolution of DUX gene macrosatellites in placental mammals. Chromosoma. 2012 Oct
468 1;121(5):489–97.

469 25. Vuoristo S, Bhagat S, Hydén-Granskog C, Yoshihara M, Gawriyski L, Jouhilahti EM, et al.
470 DUX4 is a multifunctional factor priming human embryonic genome activation. iScience.
471 2022 Apr 15;25(4):104137.

472 26. Clapier CR, Cairns BR. The biology of chromatin remodeling complexes. Annu Rev
473 Biochem. 2009;78:273–304.

474 27. Romero OA, Sanchez-Cespedes M. The SWI/SNF genetic blockade: effects in cell
475 differentiation, cancer and developmental diseases. Oncogene. 2014 May;33(21):2681–9.

476 28. Bultman SJ, Gebuhr TC, Pan H, Svoboda P, Schultz RM, Magnuson T. Maternal BRG1
477 regulates zygotic genome activation in the mouse. Genes Dev. 2006 Jul 1;20(13):1744–54.

478 29. Schaniel C, Ang YS, Ratnakumar K, Cormier C, James T, Bernstein E, et al.
479 Smarcc1/Baf155 Couples Self-Renewal Gene Repression with Changes in Chromatin
480 Structure in Mouse Embryonic Stem Cells. Stem Cells. 2009 Dec 1;27(12):2979–91.

481 30. De Iaco A, Verp S, Offner S, Grun D, Trono D. DUX is a non-essential synchronizer of
482 zygotic genome activation. Development. 2019 Jan 1;dev.177725.

483 31. Guo Y, Kitano T, Inoue K, Murano K, Hirose M, Li TD, et al. Obox4 secures zygotic
484 genome activation upon loss of Dux [Internet]. bioRxiv; 2022 [cited 2023 Jan 5]. p.
485 2022.07.04.498763. Available from:
486 https://www.biorxiv.org/content/10.1101/2022.07.04.498763v2

487 32. Akiyama T, Xin L, Oda M, Sharov AA, Amano M, Piao Y, et al. Transient bursts of Zscan4
488 expression are accompanied by the rapid derepression of heterochromatin in mouse
489 embryonic stem cells. DNA Res Int J Rapid Publ Rep Genes Genomes. 2015 Oct;22(5):307–18.

490 33. Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies
491 proximal and interacting proteins in mammalian cells. J Cell Biol. 2012 Mar 19;196(6):801–
492 10.

493 34. Liu X, Salokas K, Tamene F, Jiu Y, Weldatsadik RG, Öhman T, et al. An AP-MS- and BioID-
494 compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular
495 localizations. Nat Commun. 2018 Mar 22;9(1):1188.

496 35. Xie SQ, Leeke BJ, Whilding C, Wagner RT, Garcia-Llagostera F, Low Y, et al. Nucleolar-
497 based Dux repression is essential for embryonic two-cell stage exit. Genes Dev. 2022 Mar
498 1;36(5–6):331–47.

499 36. Yu H, Sun Z, Tan T, Pan H, Zhao J, Zhang L, et al. rRNA biogenesis regulates mouse 2C-like
500 state by 3D structure reorganization of peri-nucleolar heterochromatin. Nat Commun. 2021
501 Dec;12(1):6365.

502 37. Grow EJ, Weaver BD, Smith CM, Guo J, Stein P, Shadle SC, et al. p53 convergently
503 activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy
504 cell models. Nat Genet. 2021 Aug;53(8):1207–20.

505 38. Vega-Sendino M, Olbrich T, Stein P, Tillo D, Carey GI, Savy V, et al. The homeobox
506 transcription factor DUXBL controls exit from totipotency [Internet]. bioRxiv; 2022 [cited
507 2023 Feb 8]. p. 2022.09.19.508541. Available from:
508 https://www.biorxiv.org/content/10.1101/2022.09.19.508541v1

509 39. Ruebel ML, Vincent KA, Schall PZ, Wang K, Latham KE. SMCHD1 terminates the first
510 embryonic genome activation event in mouse two-cell embryos and contributes to a
511 transcriptionally repressive state. Am J Physiol-Cell Physiol. 2019 Oct;317(4):C655–64.

512 40. Lemmers RJLF, Tawil R, Petek LM, Balog J, Block GJ, Santen GWE, et al. Digenic
513 inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes
514 facioscapulohumeral muscular dystrophy type 2. Nat Genet. 2012 Dec;44(12):1370–4.

515 41. Kaya-Okur HS, Janssens DH, Henikoff JG, Ahmad K, Henikoff S. Efficient low-cost
516 chromatin profiling with CUT&Tag. Nat Protoc. 2020 Oct;15(10):3264–83.

517

518 **Conflict of interest**

519 The authors declare no competing interests.

520 **Specific author contribution**

521 Experiments and analyses were conducted by C.M.S. with assistance from E.J.G. and S.C.S.

522 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.

523 **Funding**

524 This work was supported by the Howard Hughes Medical Institute, NIH/USDA R01HD095883

525 Dual Purpose Grant, the NICHD (F32HD104442) to S.C.S., and the NICHD (F32HD094500) and

526 Labor Foundation Fellowship 10041116 to E.J.G.

527 **Acknowledgements**

528 We thank members of the Cairns laboratory and M.B. Chandrasekhran for inspiring discussions.

529 We thank Brian Dalley for assistance with preparing RNA-seq libraries and for sequencing

530 services, and T. Parnell for bioinformatic assistance. We give special thanks to Ross Tomaino at

531 Taplin Mass Spectrometry Core (Harvard University) for mass spectrometry assistance for BioID

532 experiments. Lastly, we thank Brian Patten at Novogene for Illumina sequencing services for

533 CUT&Tag and ATAC-seq libraries.

534 **Materials Availability**

535 Plasmids used in this study will be available by request.

536 **Data accessibility**

537 RNA-sequencing, CUT&Tag, and ATAC-seq data are available through GEO GSE224300. Standard

538 packages were used for RNA-seq, CUT&Tag, and ATAC-seq analyses (see "Materials and

539 Methods").

540

541 **Materials and Methods**

542 **Amino acid alignment**

543 DUX ortholog protein translations were aligned using Geneious Local Alignment (Smith-

544 Waterman), Blosum62 cost matrix, gap open penalty of 12, and gap extension penalty of 4.

545 **Mouse embryonic stem cell culture**

546 Mycoplasma-free E14 mESCs were cultured on gelatin in 2i+LIF medium containing Gibco KO-

547 DMEM with nonessential amino acids, 2-mercaptoethanol and dipeptide glutamine and were

548 supplemented with 15% ESC-grade FBS, leukemia inhibitory factor (Thermo Fisher), 1 µM

549 PD0325901 (Sigma-Aldrich) and 3 µM CHIR99021 (Sigma-Aldrich). Stable cell lines were

550 blasticidin (Fisher Scientific, B12150-0.1) at 3 µg/mL.

551 **HEK293T cell culture**

552 HEK293Ts were cultured in high-glucose growth media DMEM with nonessential amino acids, 2-

553 mercaptoethanol and dipeptide glutamine and were supplemented with 10% FBS.

554 **Transfection of mESCs and HEK293Ts**

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

559 **Doxycycline treatment of cells**

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

563 **Western blot**

564 RIPA buffer (50 mM TRIS-HCl [pH 8.0], 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate,
565 0.1% SDS) with 1X protease inhibitors was used to lyse mESCs. A 100X protease inhibitor
566 cocktail consists of 0.03 mg/mL leupeptin, 0.14 mg/mL pepstatin A, 0.02 mg/mL chymostatin,
567 8.5 mg/mL phenylmethanesulfonylfluoride fluoride, 33mg/mL benzamidine and solubilized in
568 ethanol. Lysed cell suspensions were then rotated for 30 min at 4°C and were further sonicated
569 for 10 min with a 30-s on-off cycle in a Diagenode Bioruptor Pico device to solubilize chromatin-
570 bound proteins. Protein lysates were quantified with the Bio-Rad (500-0001EDU) Bradford
571 reagent and loaded on SDS-PAGE gel. After transfer to nitrocellulose membranes (VWR, 95040-
572 108), membranes were blocked with TSBT (TBS with 0.1% Tween-20) and 5% milk and then
573 incubated with antibodies overnight (mCherry ab167453, DUX (used in (37))and signals were
574 detected using HRP-conjugated secondary antibodies and Western Lightning Plus-ECL,
575 Enhanced Chemiluminescence (Perkin Elmer Health Sciences, NEL 105001EA).

576 **Flow cytometry**

577 A Cytoflex LX instrument with lasers for 405 nm, 488 nm, 561 nm and 640 nm was used to
578 quantify mCherry+ and MERVL–GFP+ mESCs. Samples were gated using forward FSA and side-
579 scatter SSA to isolate cells from debris, and then double discrimination was performed using
580 FSH Å~ FSW and SSH Å~ SSW.

581 **RNA-seq**

582 1 million mESCs were grown for 24 hours on a 100mm dish and then induced with doxycycline
583 (0.25 ug/mL) for 12 hours. Cells were washed two times with PBS, TRIzol (Thermo Fisher) was
584 added to tissue culture wells and incubated for 1 min at room temperature. TRIzol/cells were
585 added to Eppendorf tubes and RNA was extracted using the TRIzol extraction method. Libraries
586 were generated according to the manufacturer's instructions: polyA-selected RNA was isolated
587 and libraries were prepared using the NEBNext kit (New England Biolabs, e7500s). Purified
588 libraries were quantified on an Agilent Technologies 2200 TapeStation with a D1000
589 ScreenTape assay. The molarity of adaptor-modified molecules was defined by a qPCR with a
590 KAPA Library Quantification kit (Kapa Biosystems). Individual libraries were normalized to
591 10nM, and equal volumes were pooled in preparation for Ilumina sequence analysis.
592 Sequencing libraries (2pM) were chemically denatured and applied to an Illumina HiSeq paired-
593 end flow cells with an Illumina cBot. Flow cells were then transferred to an Ilumina HiSeq 2000
594 instruct and sequenced in the 125-bp paired-end mode. RNA-seq reads were trimmed and
595 filtered for quality using FastQC (version 0.5.0). Processed reads were aligned using STAR
596 (version 2.7.2c). Differential expression analysis was performed using DESeq2 (version 3.11).
597 Hierarchical clustering was performed using the R package 'Complex Heatmap,' with inputs
598 being differentially expressed transcripts (DESeq FDR < 0.05) z-score normalized.

599 **GO term analysis**

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

601 **CUT&Tag**

602 1 million mESCs were grown for 24 hours and then induced with doxycycline (0.25 ug/mL) for
603 12 hours. Cell were washed 1X with PBS and nuclei were isolated from mESCs using nuclear
604 extraction buffer (20 mM HEPES-KOH [7.9], 10 mM KCl, 0.1% Triton X-100, 20% glycerol, 0.5
605 mM spermidine, 1X protease inhibitors (see above). ConA-streptavidin beads were paired with
606 antibodies of interest 1:100 (mCherry ab167453, H3K9ac ab 32129, SMARCC1 ab172638) and
607 50,000 cells were added and incubated overnight in binding buffer (20 mM, HEPES [pH 7.9], 10
608 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$). Secondary guinea pig anti-rabbit 1:100 (ABIN101961) was
609 incubated for 1 hr at room temperature. pA-Tn5 (pre-loaded with Mosaic End A/B adaptors)
610 was incubated for 1 hour at room temperature and used at 1:100. After washing steps and
611 release from the beads, PCR was performed for 12 cycles using NEBnext High-Fidelity 2X PCR
612 Master Mix (New England Biolabs, M0541L) (41). Samples were 1:1.3 AMPure XP (Beckman
613 Coulter) purified and sent for sequencing. All libraries were sequenced on the Illumina HiSeq
614 6000 platform in 150-bp, paired-end format. Paired-end, raw read files were first processed by
615 Trim Galore (Babraham Institute) to trim low-quality reads and remove adapters. Processed
616 reads were then aligned to mm10 with the following parameters: (--end-to-end –very sensitive
617 –no-mixed). CUT&Tag libraries were trimmed and filtered for quality using FastQC (version
618 0.5.0) and Picard (version 2.26.3). Processed reads were aligned using Bowtie2 (version 2.4.2)
619 with the parameters: (--end-to-end –very sensitive –no-mixed) and indexed using Samtools
620 (version 1.15). Duplicate samples were merged and normalized to control samples (mESCs for

621 H3K9ac dataset and mCherry alone for mCherry and SMARCC1 datasets). DeepTools was used
622 to generate heatmaps and profile plots using DUX binding sites as previously published (21).
623 **BioID**
624 15 million BirA* mESC derivatives were plated on gelatin-coated 15-cm plates and recovered
625 for 24-hours. 0.25 µg/mL doxycycline and 0.5 µM biotin (Sigma Aldrich, B4639) was added to
626 cells and incubated for 18 hours. Cells were washed with ice cold PBS 2 times and 1 mL lysis
627 buffer (2 mM Tris-HCl [pH 8.0], 137 mM NaCl, 1% NP-40, 2 mM EDTA, 1x protease inhibitors)
628 was added to each plate and incubated for 1 min (33). Cells were scraped of off the plate and
629 transferred to a 2-mL tube. Cells then rotated for 30 min at 4°C and were further sonicated for
630 10 min with a 30-s on-off cycle in a Diagenode Bioruptor Pico device to solubilize chromatin-
631 bound proteins. Next, samples were centrifuged at 4°C at 12,000 rpm for 20 minutes and the
632 supernatant was transferred to new tubes. Protein lysates were quantified with the Bio-Rad
633 (500-0001EDU) Bradford reagent. For streptavidin pulldown, 2 µg of protein were added to pre-
634 washed 50 µL streptavidin Sepharose bead conjugates (Cell Signaling Technology, 3419S) and
635 rocked overnight at 4°C. Cells were pelleted and then washed 5X with wash buffer (10 mM TRIS
636 [pH 7.4], 1 mM EDTA, 1 mM EGTA, 150 mM NaCl, 1% Triton X-100, 0.2 mM sodium
637 orthovanadate, 1x protease inhibitors), boiled at 95°C in 30 µL 4X Laemmli buffer, and ran on a
638 SDS-page gel until the samples were 1-inch into the gel. Protein bands/smears were cut out of
639 the gel with razor blades and transferred to individual 1.7-mL Eppendorf tubes filled with
640 ddH_2O. Samples were overnight shipped on dry for liquid chromatography-tandem mass
641 spectrometry sequencing (LC-MS/MS) to Taplin Mass Spectrometry Facility, Cell Biology
642 Department, Harvard Medical School. Gel smears were reduced using 1 mM DTT in 50 mM

643 ammonium bicarbonate for 30 min at 60°C and cooled, and 5 mM iodoacetamide in 50 mM
644 ammonium bicarbonate was added for 15 min and incubated at room temperature. 5 mM DTT
645 was added to quench the reaction, and then 5 ng/µL of trypsin (Promega) was added and
646 samples incubated over night at 37 °C. Samples were then desalinated and resuspended in 10
647 µL of high-pressure liquids chromatography (HPLC) solvent A (2.5% acetonitrile, 0.1% formic
648 acid). A reverse-phase HPLC capillary column was made by placing 2.6 µL M C_{18} spherical silica
649 beads into a fused silica capillary (100-µL M inner diameter, ~30-cm length). Next, each sample
650 was loaded using a Famos autosampler (LC Packings) onto the HPLC capillary column. A linear
651 gradient was created to elute peptides with increasing amounts of solvent B (97.5%
652 acetonitrile, 0.1% formic acid). Eluted peptides were subjected to electrospray ionization
653 couple to a mass spectrometer (Thermo Fisher Scientific) to produce tandem mass spectra
654 whose *b* and *y* ion series patterns were compared to Sequest (Thermo Fisher Scientific) to
655 establish protein identity. The sequencing data were filtered to 1 to 2% peptide false-discovery
656 rate. To analyze BioID datasets, two approaches were taken. First, we utilized DESeq2 (version
657 3.11) to provide candidate genes for BirA*-DUX vs. BirA* and BirA*-C3 vs. BirA*-C1. Then, we
658 used SAINTexpress (version 3.5.3) with parameter: (-L6). Formal candidates exhibited > 2
659 Log_2FC and FDR < 0.05 and SAINT score > 0.74, for DESeq2 and SAINTexpress analyses
660 respectively.

661 **Co-immunoprecipitation**

662 1 million mESCs were grown for 24 hours on 100mm dishes and then induced with doxycycline
663 (0.25 ug/mL) for 12 hours. Cells were trypsinized, rinsed once with PBS, and lysis buffer (2 mM
664 Tris-HCl, pH=8.0, 137 mM NaCl, 1% NP-40, 2 mM EDTA, 1x protease inhibitors) was added to

665 lyse cells. Cells then rotated for 30 min at 4°C and were further sonicated for 10 min with a 30-s
666 on-off cycle in a Diagenode Bioruptor Pico device to solubilize chromatin-bound proteins.
667 Protein lysates were quantified with the Bio-Rad (500-0001EDU) Bradford reagent. 3X FLAG
668 Peptide-coupled beads (Sigma Aldrich, F4799-4MG) were pre-washed with lysis buffer, 1 mg of
669 protein was added to 40 µL of pre-washed beads, and samples were incubated at room
670 temperature for 4 hours. Beads were pelleted by low-speed centrifugation. Samples were
671 washed 5X with wash buffer (10 mM TRIS [pH 7.4], 1 mM EDTA, 1 mM EGTA; pH=8.0, 150 mM
672 NaCl, 1% Triton X-100, 0.2 mM sodium orthovanadate, 1x protease inhibitors [see above]),
673 boiled at 95°C in 30 µL 4X Laemelli buffer, and run on an SDS-page gel for western blotting
674 analysis.
675 **ATAC-seq**
676 mESCs were grown for 24 hours and then induced with doxycycline (0.25 ug/mL) for 12 hours.
677 Cells were washed 1X with PBS, pelleted, resuspended, and lysed in cold lysis buffer (10 mM
678 Tris-HCl [pH 7.4], 10 mM NaCl, 3 mM MgCl2, and 0.1% IGEPAL CA-630). Nuclei were pelleted
679 and resuspended in transposase buffer. The Tn5 (Illumina FC-121-1030), and the transposition
680 reaction was carried out for 30 min at 37°C. After purification, the Nextera libraries were
681 amplified for 15 cycles with NEBnext PCR master mix and purified with a Qiagen PCR cleanup
682 kit. All libraries were sequenced on the Illumina HiSeq 6000 platform in 150-bp, paired-end
683 format. Paired-end, raw read files were first processed by Trim Galore (Babraham Institute) to
684 trim low-quality reads and remove adapters. Processed reads were then aligned to mm10 with
685 the following parameters: (--end-to-end –very sensitive –no-mixed).
686

687 **Figure Legends**

688

689 Figure 1: Identification of five repeats within the mouse DUX C-terminal region, and truncation
690 analysis of transcriptional activity.

691 a. (Left) Mouse DUX aligned to itself, revealing five repeats. (Right) Diagram of mouse DUX
692 and human DUX4.

693 b. Amino acid alignment of mouse DUX C-terminal repeats and comparison to human
694 DUX4. A 14 amino acid tail follows both the terminal mouse C5 repeat and the sole
695 DUX4 repeat.

696 c. Phylogenetic analysis showing amino acid alignments of mouse DUX C-terminal repeats
697 vs. the C-terminus of bovine DUXC, the C-terminus of rat DUX4, and the C-terminus of
698 human DUX4. Line length unit is substitutions per base. Purple shading denotes mouse
699 DUX C-terminal repeats.

700 d. (Left) Schematic of constructs used for flow cytometry: mCherry-tagged full-length DUX
701 (FL), homeodomains with C1, C12, C123, C1234, C12345. (Right) Flow cytometry data for
702 MERVL::GFP reporter given mCherry expression in *Dux*-/- mESCs following 18hr
703 expression of indicated constructs. *p-value < 0.05, student's t-test. (n=3 biological
704 replicates)

705

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

707 a. (Top) Schematic of constructs used for flow cytometry: mCherry-tagged FL,
708 ΔC12345+14aa, C1+14aa, C2+14aa, C3+14aa, C4+14aa, and C5+14aa. (Bottom) Flow

709 cytometry for MERVL::GFP reporter given mCherry expression in *Dux*-/- mESCs with
710 18hr overexpression of indicated constructs. *p-value < 0.05, student's t-test. (n=3
711 biological replicates). Constructs with a single black line have the linker $GGGGS_2$ and
712 constructs with two black lines have linkers with $GGGGS_2$ and $GAGAS_2$, respectively.
713 b. (Top) Schematic of constructs used for flow cytometry: mCherry-tagged FL, C3+14aa,
714 C3Δ14aa, C5+14aa, C5Δ14aa. (Bottom) Flow cytometry for MERVL::GFP reporter given
715 mCherry expression in *Dux*-/- mESCs following 18hr expression of indicated constructs.
716 *p-value < 0.05, student's t-test. (n=3 biological replicates)
717 c. (Top) DUX domain chimera construct design for C1-C3 fusions with (top) schematic of
718 constructs and (bottom) specific chimera cut-offs for C1-C3 a,b,c (red to orange and C3-
719 C1 a,b,c (orange to red) constructs. (Bottom) Schematic of DUX domain chimera
720 constructs used for flow cytometry: mCherry-tagged full-length DUX (FL), C1C3a, C1C3b,
721 C1C3c, C3C1a, C3C1b, C3C1c. Symbols adjacent to alignments: red circle denotes
722 transcriptionally inactive repeats and the green circle denotes transcriptionally active
723 repeats. (Right) Flow cytometry for MERVL::GFP reporter given mCherry expression in
724 *Dux*-/- mESCs following 18hr expression of indicated constructs. *p-value < 0.05,
725 student's t-test. (n=3 biological replicates)
726 d. Flow cytometry for MERVL::GFP reporter given mCherry expression in *Dux*-/- mESCs
727 following 18hr expression of point mutation constructs: (Left) FL-DUX, C3+14aa, C3
728 D438E+14aa, C3 D438G+14aa, and C3F443L+14aa. (Right) FL-DUX, C1+14aa, C3+14aa,
729 C1DPLELF+14aa (substation of C3 amino acids into C1+14aa), and C3GPLELL+14aa

730 (substitution of C2 and C4 6aa sequence into C3+14aa). *p-value < 0.05, student's t-test.
731 (n=3 biological replicates)

732 e. Principal component analysis (PCA) of RNA-seq analysis for 18hr overexpression of DUX
733 domain chimera constructs, full-length DUX, and mCherry alone (n=2). Colored squares
734 beside sample legends denote the inability to activate MERVL::GFP reporter (red) or
735 ability to activate MERVL::GFP reporter (green).

736 f. Heat map of RNA-seq at DUX target genes (n=456) from 18hr expression of DUX domain
737 chimera constructs, full-length DUX, and mCherry alone (n=2). Colored rectangles above
738 samples denote inability to activate MERVL::GFP reporter (red) or ability to activate
739 MERVL::GFP reporter (green).

740

741 Figure 3: Chromatin modifications at DUX binding sites requires an active DUX repeat.

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

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

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

751 d. H3K9ac CUT&TAG class average map centered at DUX binding sites after 12-hour
752 expression of *Dux* constructs illustrated in Figure 2a (n=2 biological replicates).
753 e. H3K9ac CUT&TAG heatmap centered at DUX binding sites after 12-hour expression of
754 *Dux* constructs illustrated in Figure 2a (n=2 biological replicates).
755
756 Figure 4: BioID for full-length DUX reveals interaction with proteins involved in chromatin de-
757 repression.
758 a. Experimental concept for BioID (proximity labeling assay) for BirA*-DUX to identify
759 candidate DUX interacting proteins.
760 b. Experimental setup for full-length DUX BioID. The control construct, BirA*-only,
761 transfected into 2C-like mESCs, and BirA*-DUX, transfected into mESCs (n=2 biological
762 replicates).
763 c. Volcano plot using DeSeq2 analysis for BirA*-only vs. BirA*-DUX with log2FoldChange on
764 the x-axis and -log(padj) on the y-axis (n=2 biological replicates).
765 d. Interactome map of proteins enriched to interact with full-length DUX, vs. BirA*-only,
766 split into specific cellular components and protein complexes (n=2 biological replicates).
767 Full dataset available in Supplementary Table 1.
768 e. (Left) Experimental setup for BioID-like experiment in HEK283T cells with BirA*DUX and
769 flag-tagged candidates SMARCC1 and ZSCAN4D. (Right) Co-immunoprecipitation of FLAG
770 and Western blot after 18-hour expression of transiently transfected FLAG-SMARCC1 or
771 FLAG-ZSCAN4D and BirA*-DUX in HEK293Ts (n=3 biological replicates). Predicted protein
772 size for BirA*-DUX is 125 kDa.

773

774 Figure 5: BioID for active C-terminal DUX repeat reveals interactors, including Smarcc1.

775 a. Experimental constructs BioID of individual DUX repeats. Experimental constructs are N-
776 terminally labelled BirA* linked to either mCherry-HDs-C1+14aa or mCherry-HDs-
777 C3+14aa.

778 b. Volcano plot using DeSeq2 analysis for BirA*mC-C1+14aa and BirA*mC-C3+14aa with
779 log2FoldChange on the x-axis and -log(padj) on the y-axis (n=3 biological replicates).

780 c. Interactome map of proteins enriched with BirA*mC-C3+14aa vs. BirA*mC-C1+14aa,
781 split into specific cellular components and protein complexes (n=2 biological replicates).
782 Full dataset available in Table 2.

783 d. Co-immunoprecipitation of endogenous SMARCC1 after 12hr overexpression of DUX
784 domain constructs in stable clonal cell lines illustration in Figure 2a (n=3 biological
785 replicates). Predicted protein sizes are 100 kDa (FL), 95 kDa (C12345Δ14aa), 65 kDa
786 (C3+14aa), 59 kDa (C3Δ14aa), 48 kDa (ΔC12345+14aa), and 25 kDa (mCherry).

787 e. SMARCC1 CUT&TAG heatmap at DUX binding sites after 12-hour overexpression of *Dux*
788 constructs illustrated in Figure 2a (n=2 biological replicates).

789 f. ATAC-seq heatmap at DUX binding sites after 12-hour overexpression of *Dux* constructs
790 illustrated in Figure 2a (n=2 biological replicates, except full-length DUX only has 1
791 biological replicate).

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

794 a. Summary table of DUX domain constructs (illustrated in Figure 2a) and their
795 transcription, recruitment, chromatin modification, and protein interaction activities.
796 b. Model for the contributions of DUX domains on DUX interactions and activity
797
798 **Supplementary Figure Legends**
799
800 Supplementary Figure 1**:** DUX orthologs do not contain a repeat structure similar to mouse
801 DUX.
802 a) Human DUX4 aligned to itself.
803 b) Rat DUX4 aligned to itself.
804 c) Bovine DUXC aligned to itself.
805 d) Table with a pairwise comparison of e-values from a BLAST-based amino acid similarity
806 comparing mouse DUX C-terminal repeats, the human DUX4 C-terminus, the rat DUX4
807 C-terminus, and the bovine DUXC C-terminus. Parameters are PAM250 matrix and gap
808 costs of existence: 15 and extension: 3. *No significant similarities found (>0.05).
809 e) Experimental design for expression of mCherry-tagged DUX domain constructs and flow
810 cytometry of GFP% | mCherry tag expression.
811 f) Western blot for mCherry tag after 18hr expression of C1, C12, C123, C1234, C12345.
812 Predicted protein sizes are 59 kDa (C1), 69 kDa (C12), 79 kDa (C123), 89 kDa (C1234), 99
813 kDa (C12345), 48 kDa (ΔC12345+14aa), and 100 kDa (FL).
814
815 Supplementary Figure 2: DUX homeodomains are required for transcriptional activity.

816 a) (Left) Schematic of constructs used for flow cytometry: mCherry-tagged full-length DUX
817 (FL), ΔC12345+14aa, C12345Δ14aa, C1+14aa, C12+14aa, C123+14aa, C1234+14aa,
818 C5+14aa, C45+14aa, C345+14aa, C2345+14aa, ΔHDs, ΔHD1, ΔHD2, and *Dux*-/- alone.
819 (Right) Flow cytometry for MERVL::GFP reporter given mCherry expression in *Dux*-/-
820 mESCs following 18hr expression of indicated constructs. *p-value < 0.05, student's t-
821 test. (n=3 biological replicates).
822 b) (Left) Schematic of additional constructs used for flow cytometry: mCherry-tagged full-
823 length DUX (FL), C3+C5+14aa, C3+C5Δ14aa, C124+14aa. (Right) Flow cytometry for
824 MERVL::GFP reporter given mCherry expression in *Dux*-/- mESCs following 18hr
825 expression of indicated constructs. *p-value < 0.05, student's t-test. (n=3 biological
826 replicates)
827 c) Schematic of C1C3 chimera DUX constructs for amino acid cut offs for C1C3a, C3C3b,
828 C1C3c, C3C1a, C3C1b, and C3C1c aligned to the 5 C-terminal repeats
829 d) Heat map of RNA-seq at DUX target genes (n=456) (21) from 18hr expression of DUX
830 domain constructs of FL, C12345Δ14aa, C12345+14aa, C12345Δ14aa(n=2). Colored
831 rectangles above samples denote inability to activate MERVL::GFP reporter (red) and
832 ability to activate MERVL::GFP reporter (green).
833 e) Phylogenetic analysis of the five mouse DUX C-terminal repeats. Line length unit is
834 substitutions per base.
835
836 Supplementary Figure 3: DUX CUT&Tag data strongly resembles DUX ChIP-seq, and DUX binding
837 sites gain H3K9ac upon DUX expression.

838 a. Profile plot of mCherry-FL DUX CUT&TAG centered at DUX binding sites after 12 hour
839 expression vs. HA-DUX ChIP-seq (n=2 biological replicates) (12)
840 b. IGV screenshots at TCSTV3, ZSCAN4F, USP17LB, and ZSCAN4C (DUX targets genes) for
841 H3K9ac CUT&Tag in mESCs and mCherry-DUX 12hr expression and mCherry CUT&TAG
842 for mCherry tag alone and mCherry-DUX (n=2 biological replicates)
843
844 Supplementary Figure 4: The proximity-labeling assay BioID identifies DUX protein interactors.
845 a) BioID DeSeq2 analysis replicate structure of BirA*-only and BirA*-DUX duplicates for
846 mass spectrometry datasets.
847 b) Venn diagram of statistically significant enriched proteins for BirA*-DUX interaction for
848 DeSeq2 and SAINT*express*
849 c) Streptavidin pulldown and Western blot after 18-hour BioID induction in clonal stable
850 BirA*-DUX mESC line expressing FLAG-tagged SMARCC1, CHAF1A, ZSCAN4D, or KDM4C
851 with blots for DUX and FLAG-tag (n=3 biological replicates). KDM4C was a BioID hit, but
852 does not interact with DUX, indicating a possible false positive. Predicted protein size is
853 125 kDa (BirA*-DUX).
854 d) Streptavidin pulldown and Western blot after 18-hour BioID induction in clonal stable
855 BirA*-DUX mESC line, with blot for N-terminal myc-BirA*-DUX, P300, and Avidin-HRP
856 (n=2 biological replicates). Predicted protein sizes are 125 kDa (myc-BirA*-DUX) and 263
857 kDa (p300).
858 e) BioID individual repeat DeSeq2 analysis replicate structure of BirA*-C1 and BirA*-C3
859 triplicates for mass spectrometry (n=3 biological replicates)

860 f) Principle component analysis of DeSeq2 analysis for BirA*mC-C1+14aa and BirA*mC-

861 C3+14aa with PC1: 54% variance and PC2: 17% variance

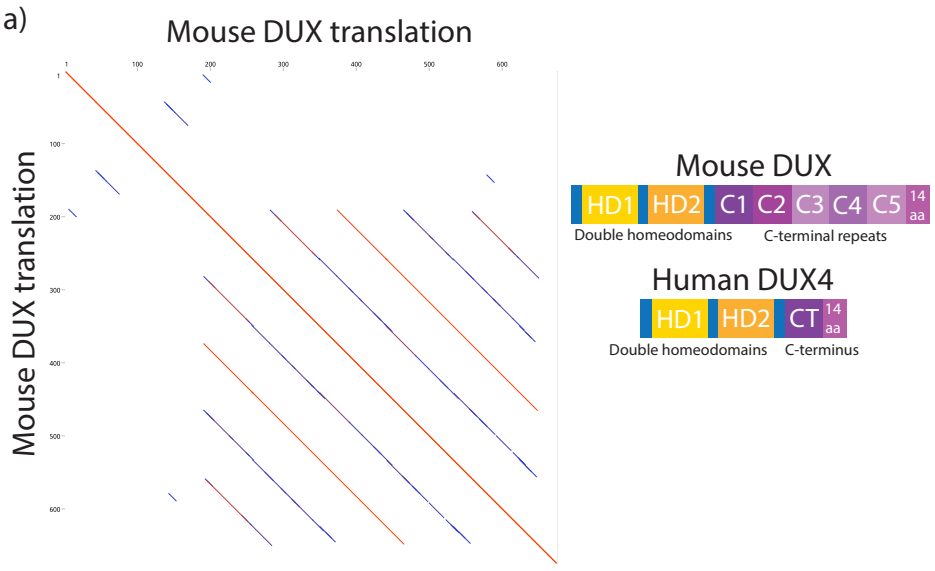
862 g) ATAC-seq heatmap at DUX binding sites after 12-hour overexpression of duplicate *Dux*

863 constructs illustrated in Figure 2a and C1+14aa

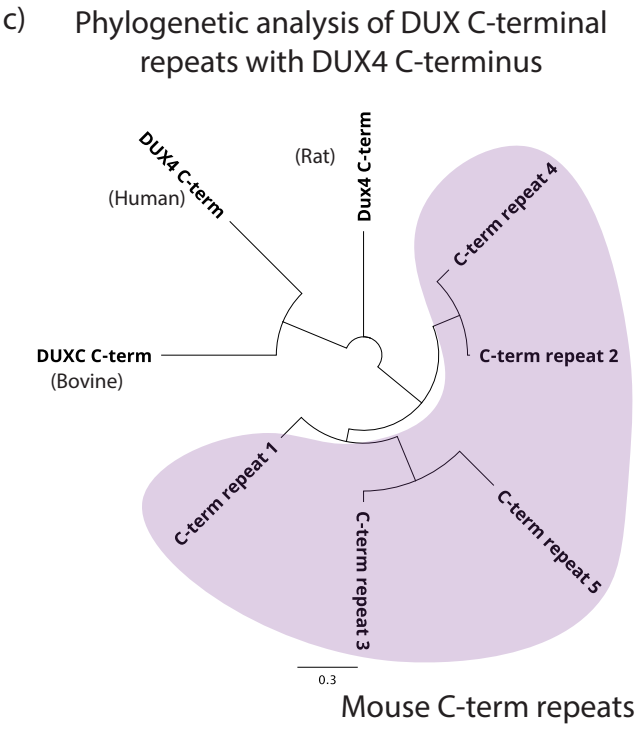
864

Figure 1

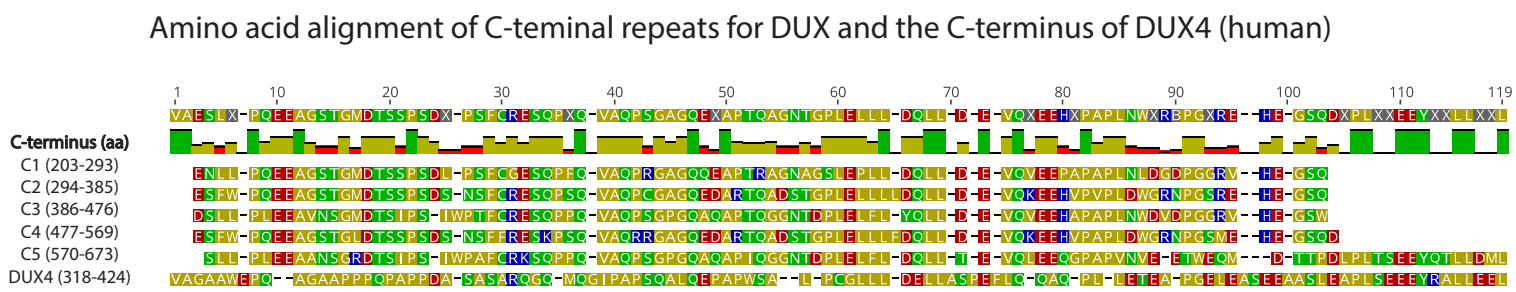
a)



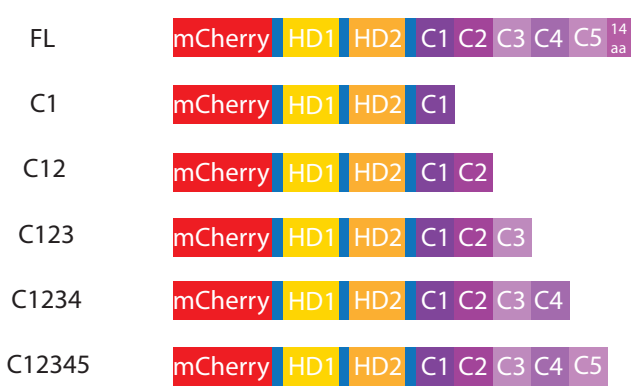
c)



b)



d)



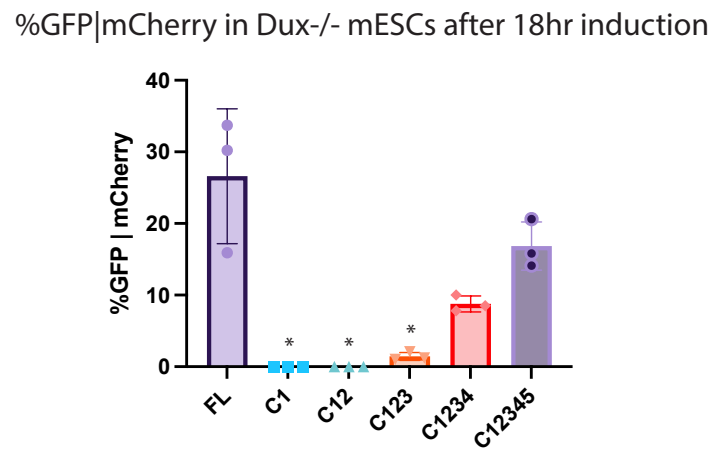


Figure 2

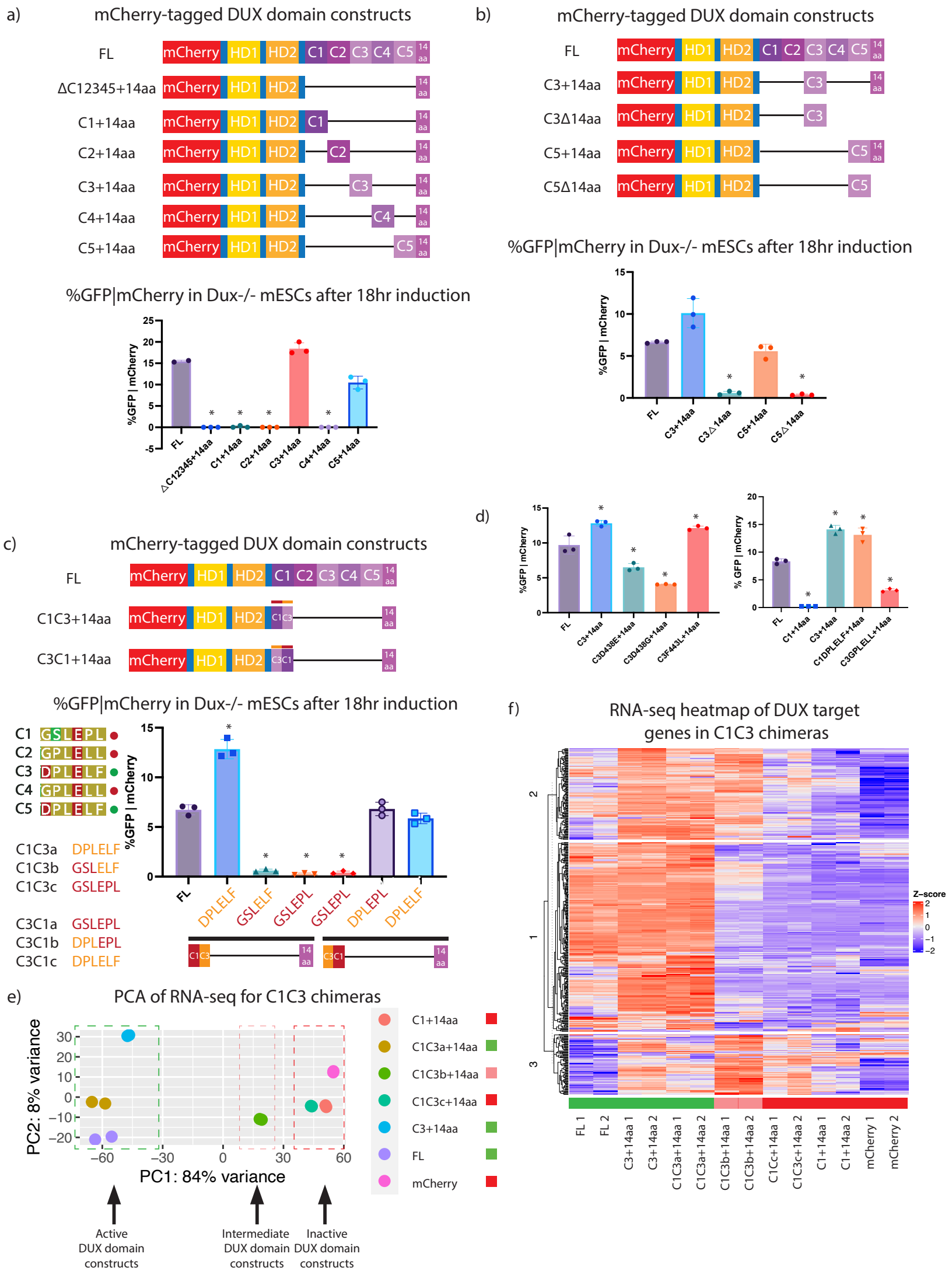
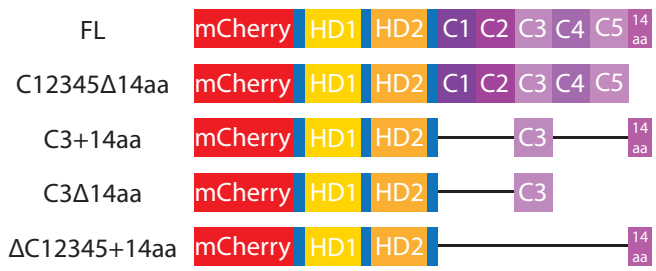
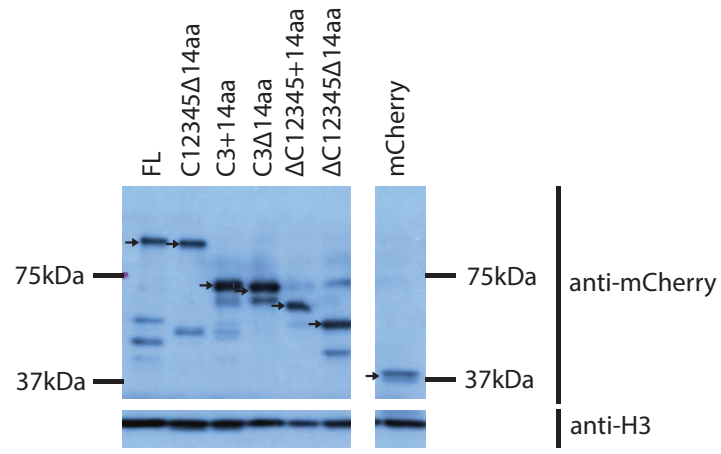

a)
mCherry-tagged DUX domain constructs
FL
ΔC12345+14aa
C1+14aa
C2+14aa
C3+14aa
C4+14aa
C5+14aa
mCherry
HD1
HD2
C1
C2
C3
C4
C5
14 aa
%GFP|mCherry in Dux-/- mESCs after 18hr induction
%GFP | mCherry
b)
mCherry-tagged DUX domain constructs
C3+14aa
C3Δ14aa
C5+14aa
C5Δ14aa
%GFP|mCherry in Dux-/- mESCs after 18hr induction
c)
mCherry-tagged DUX domain constructs
C1C3+14aa
C3C1+14aa
%GFP|mCherry in Dux-/- mESCs after 18hr induction
C1 GSLEPL
C2 GPLELL
C3 DPLELF
C4 GPLELL
C5 DPLELF
C1C3a DPLELF
C1C3b GSLELF
C1C3c GSLEPL
C3C1a GSLEPL
C3C1b DPLEPL
C3C1c DPLELF
d)
%GFP | mCherry
% GFP | mCherry
e)
PCA of RNA-seq for C1C3 chimeras
PC2: 8% variance
PC1: 84% variance
C1+14aa
C1C3a+14aa
C1C3b+14aa
C1C3c+14aa
C3+14aa
FL
mCherry
Active DUX domain constructs
Intermediate DUX domain constructs
Inactive DUX domain constructs
f)
RNA-seq heatmap of DUX target genes in C1C3 chimeras
Z-score
FL 1
FL 2
C3+14aa 1
C3+14aa 2
C1C3a+14aa 1
C1C3a+14aa 2
C1C3b+14aa 1
C1C3b+14aa 2
C1Cc+14aa 1
C1C3c+14aa 2
C1+14aa 1
C1+14aa 2
mCherry 1
mCherry 2

Figure 3

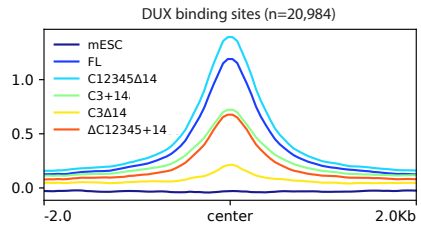
a)

DUX domain constructs in stable clonal Dux-/- mESCs

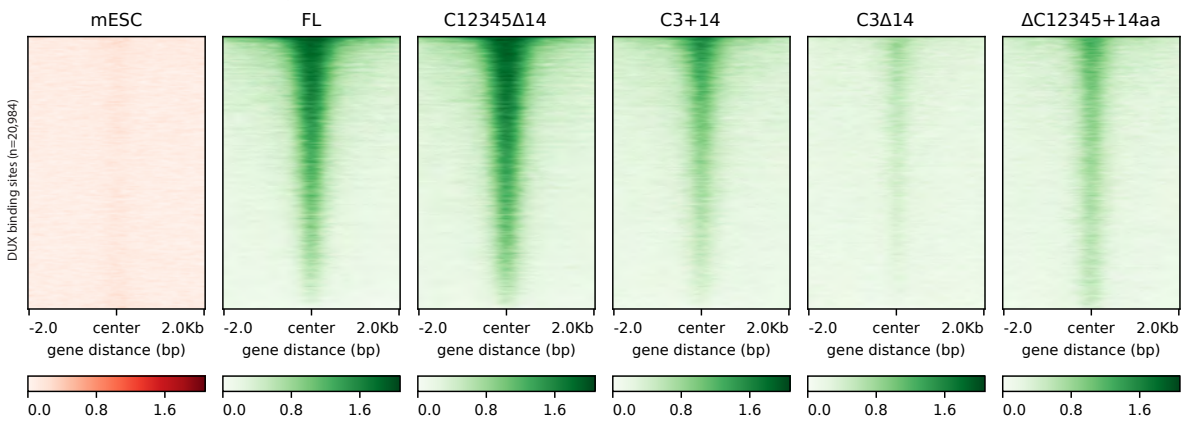




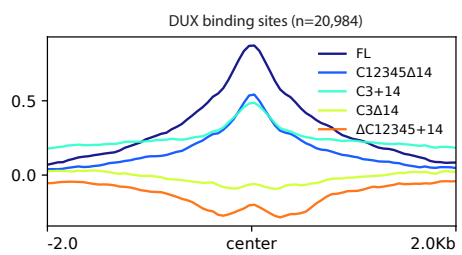
b) mCherry CUT&Tag of Dux domain constructs



c) mCherry CUT&Tag of Dux domain constructs at DUX binding sites



d) H3K9ac CUT&Tag of DUX domain constructs



e) H3K9ac CUT&Tag of DUX domain constructs at DUX binding sites

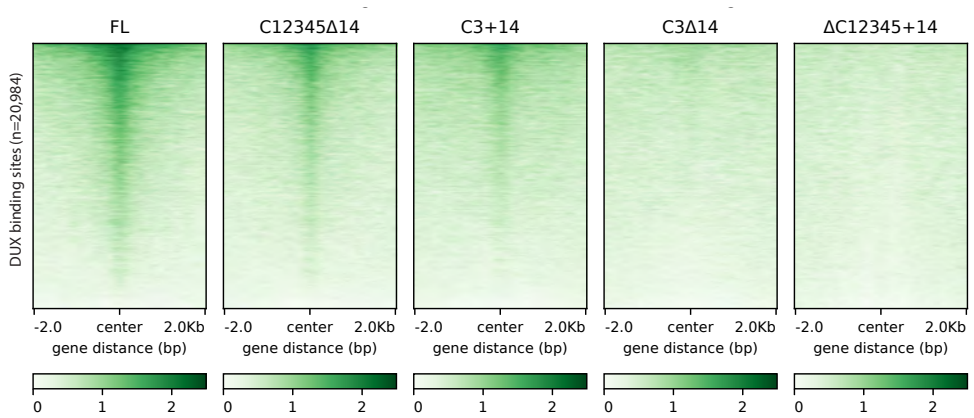
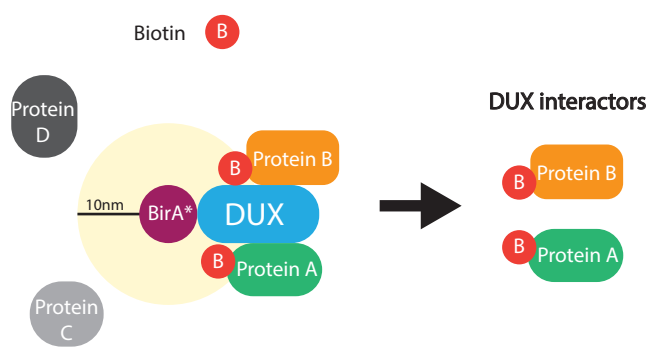
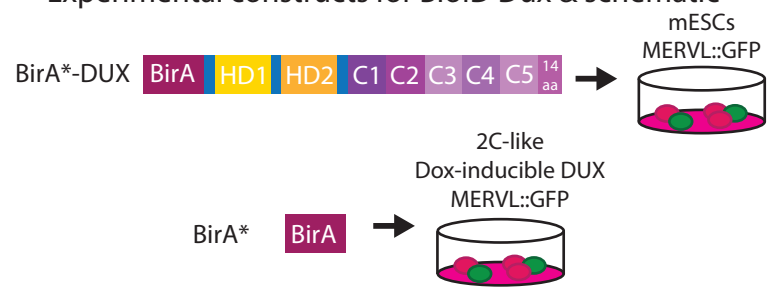


Figure 4

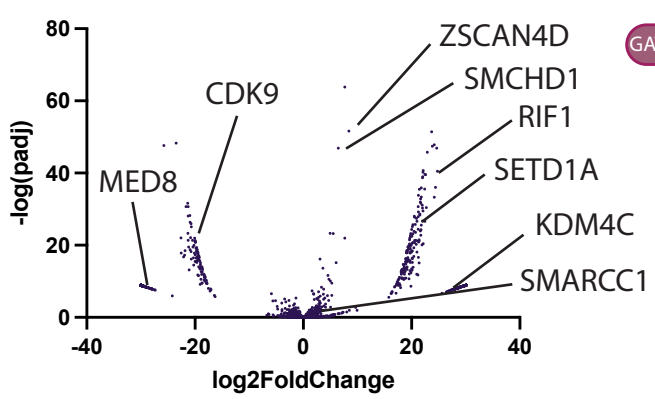
a) BioID schematic



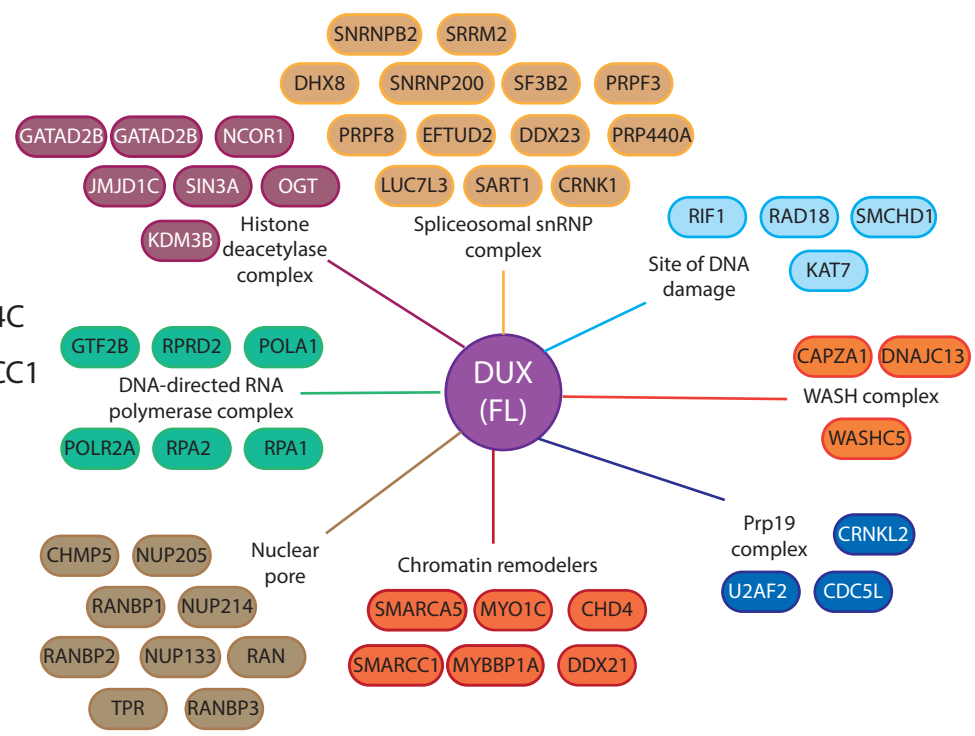
b) Experimental constructs for BioID Dux & schematic



c) Volcano plot for BirA*-DUX vs. BirA*



d) Interactome of BirA*-Dux in comparison to BirA*



e) Validation of BirA*-Dux interactions by Co-IP in BioID conditions

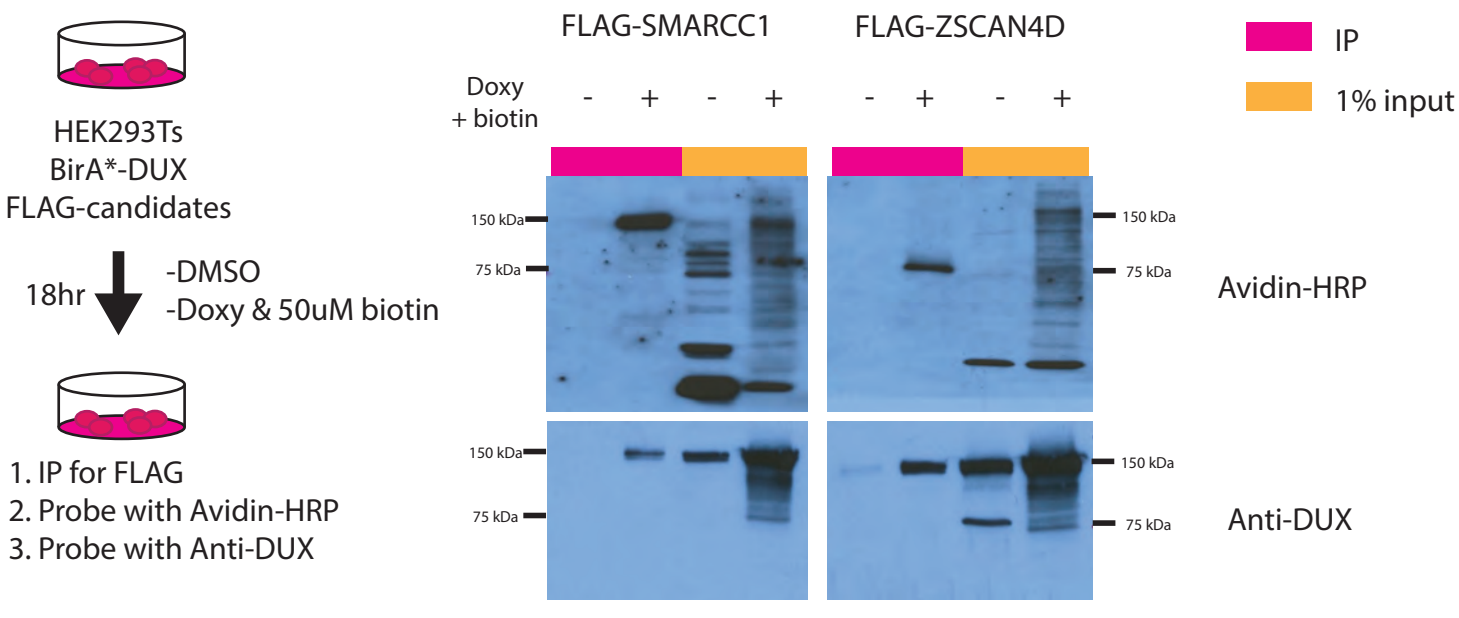
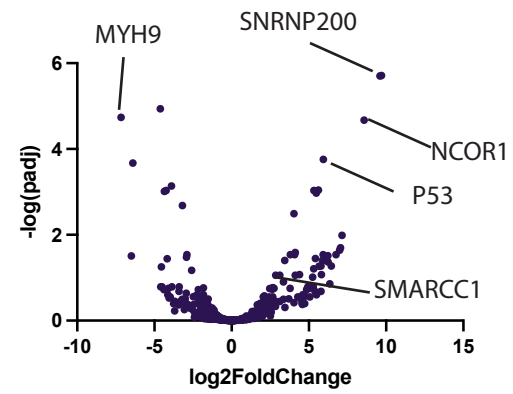


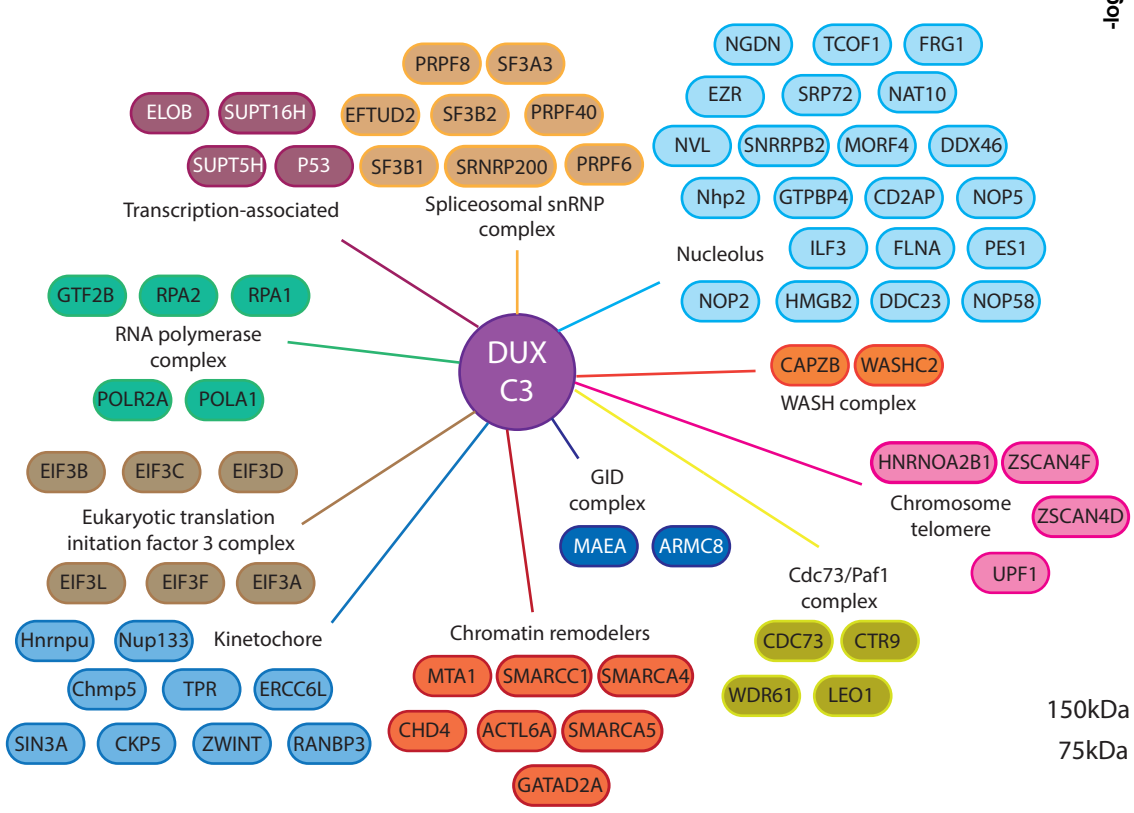
Figure 5

a) Experimental constructs for BioID C1C3

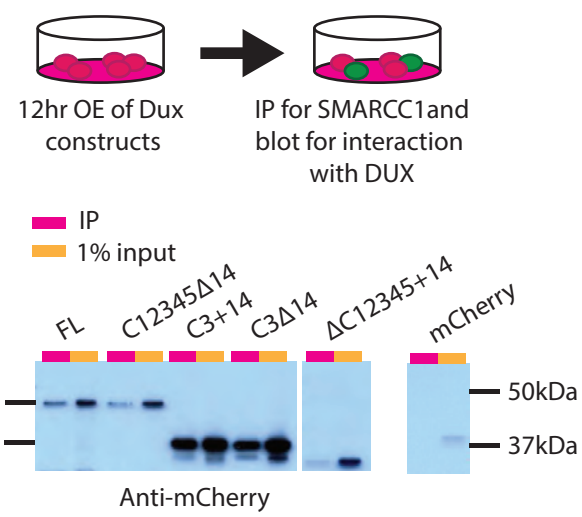
b) Volcano plot for BirA*-C3 vs. BirA*-C1



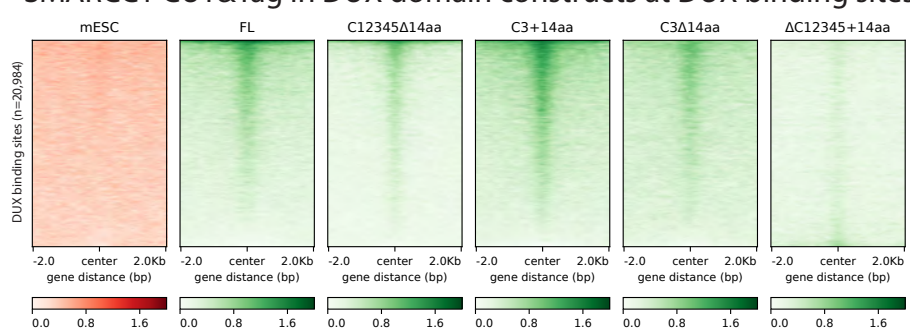
c) Interactome of BirA*-C3 in comparison to BirA*-C1



d) IP for endogeneous SMARCC1 in DUX domain cell lines



e) SMARCC1 CUT&Tag in DUX domain constructs at DUX binding sites



f) ATAC-seq in DUX domain constructs at DUX binding sites

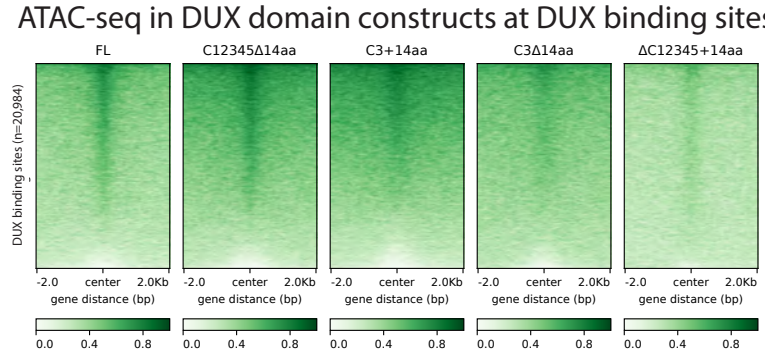


Figure 6

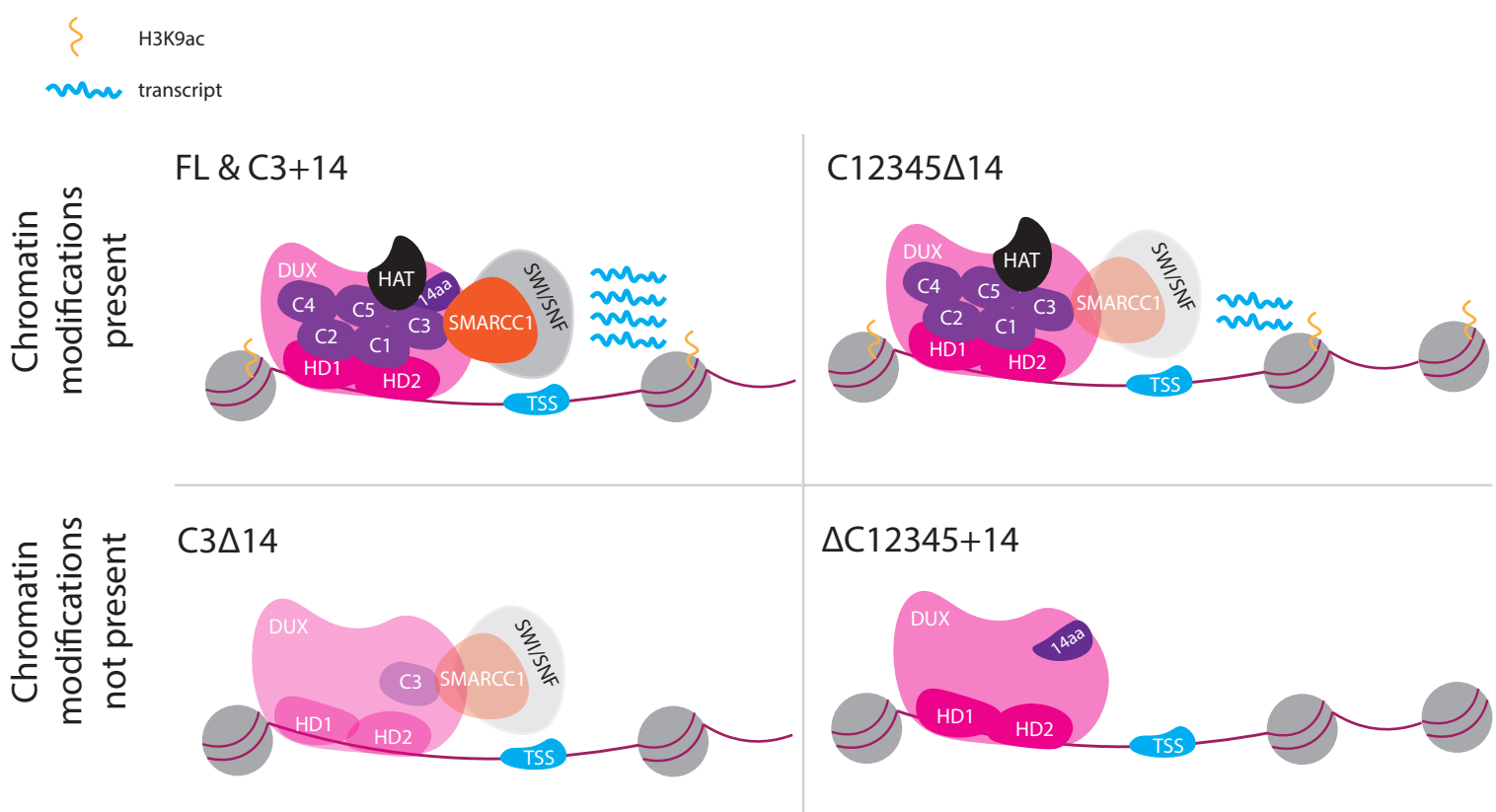
a)

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	+++	+++	++	+	+

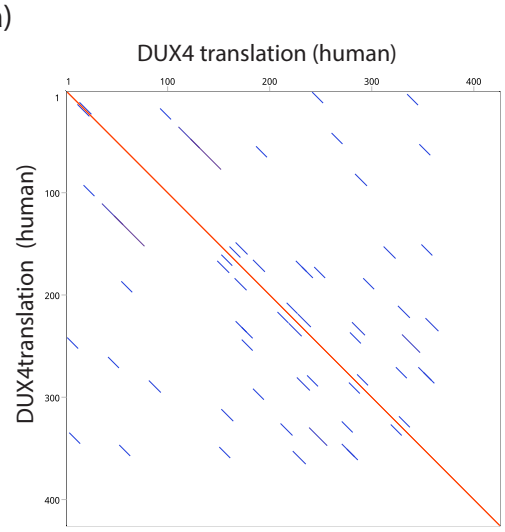
b)

Model for DUX mechanism of target gene activation

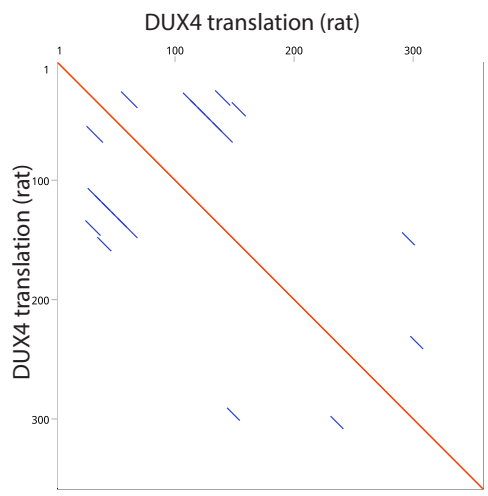


Supplementary Figure 1

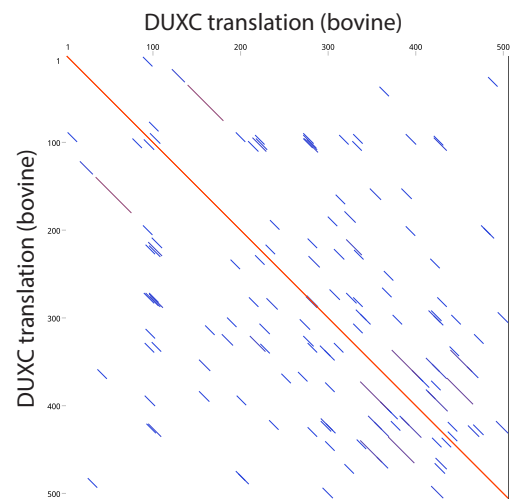
a)



b)



c)

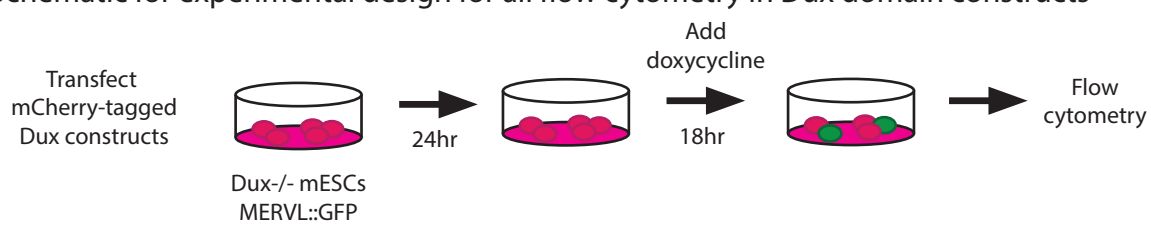


d)

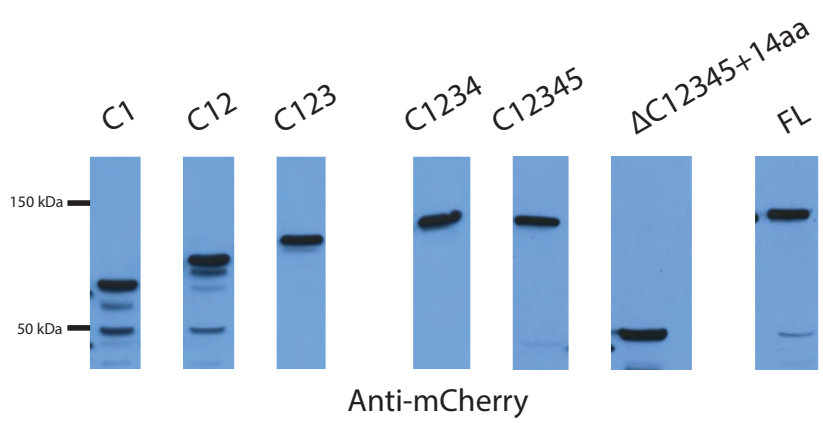
	C-term repeat 1	C-term repeat 2	C-term repeat 3	C-term repeat 4	C-term repeat 5	DUX4 C-term	DUXC C-term	Dux4 C-term
C-term repeat 1		3e-24	4e-36	7e-22	2e-25	*	0.024	0.0009
C-term repeat 2	3e-24		1e-34	1e-41	2e-23	0.030	*	*
C-term repeat 3	4e-36	1e-34		9e-31	7e-42	*	*	*
C-term repeat 4	7e-22	1e-41	9e-31		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

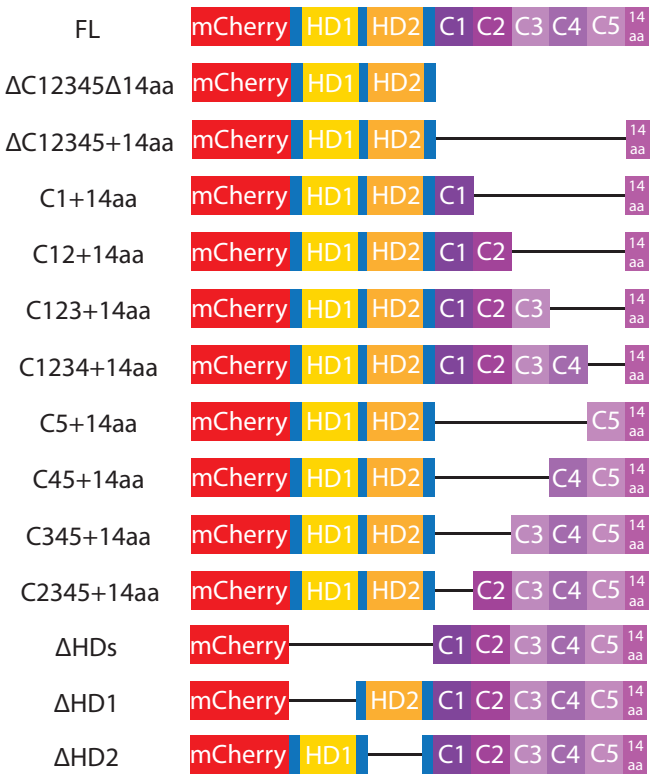


f) Western blot for mCherry-tagged Dux domain constructs

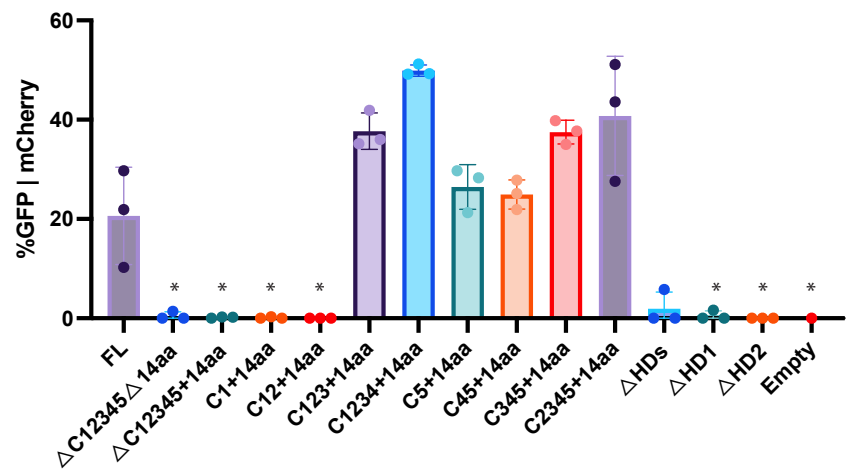


Supplementary Figure 2

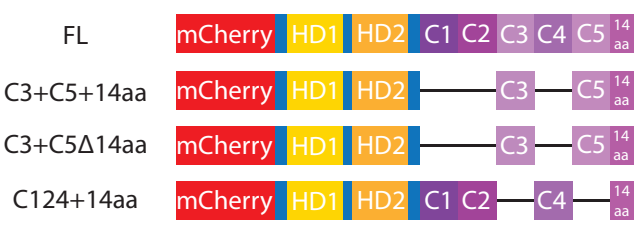
a) mCherry-tagged DUX domain constructs

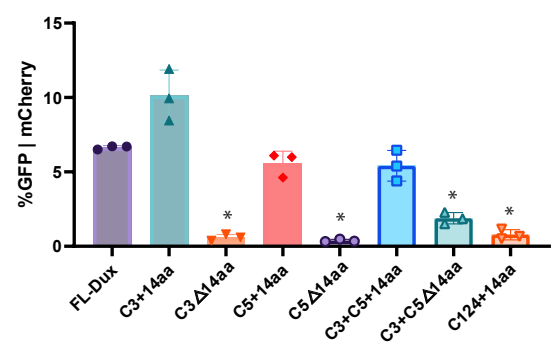


%GFP|mCherry in Dux-/- mESCs after 18hr induction

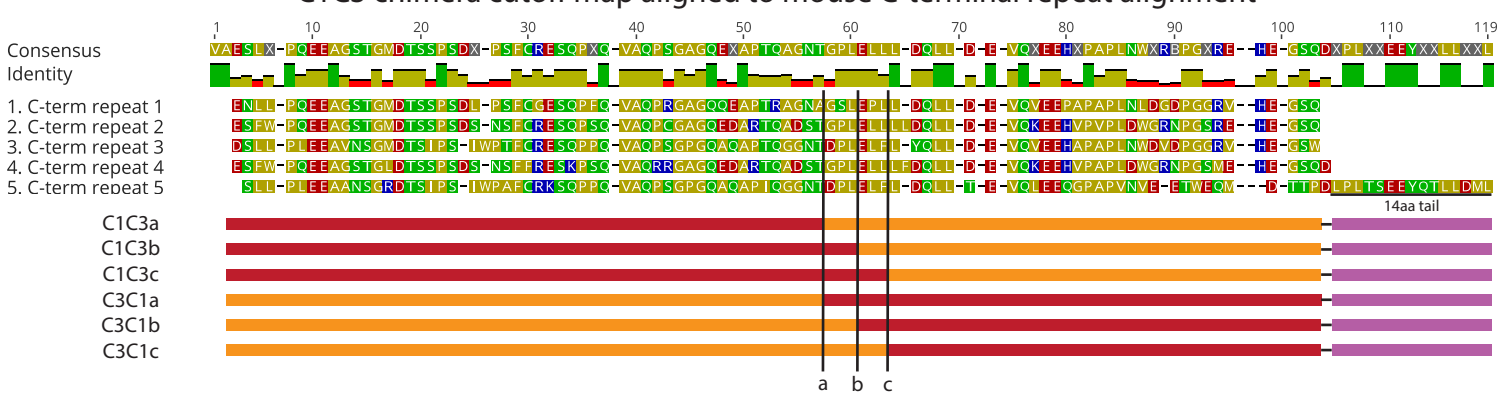


b)

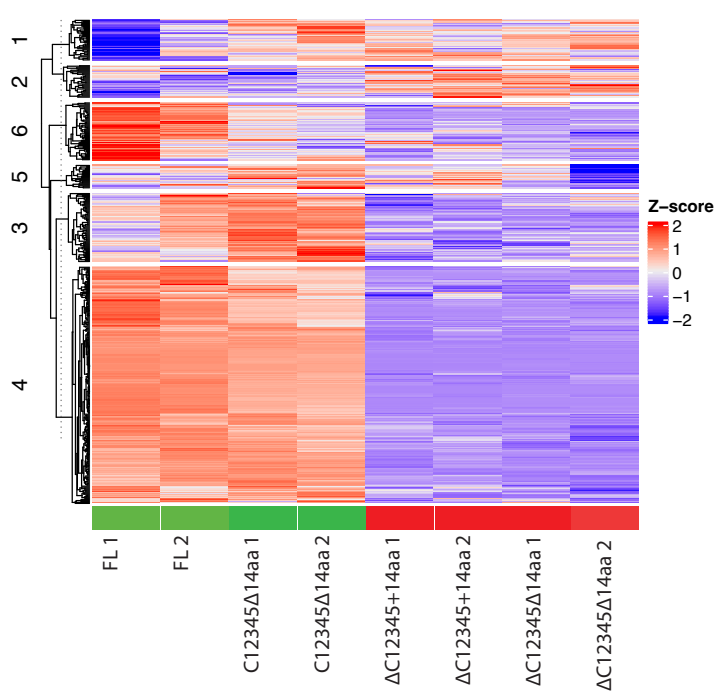




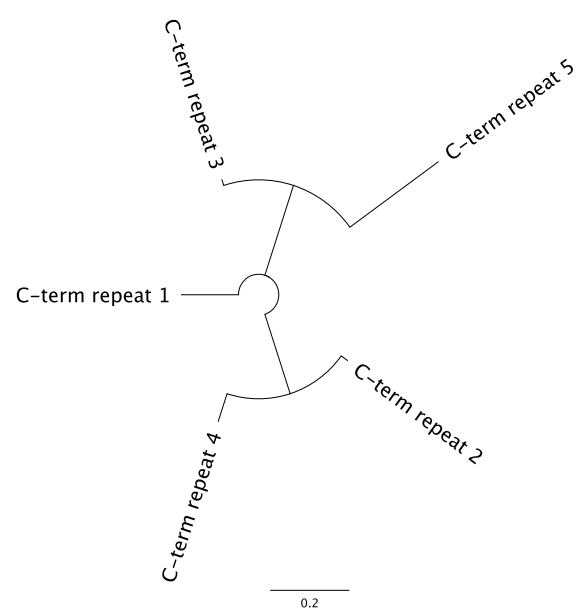
c) C1C3 chimera cutoff map aligned to mouse C-terminal repeat alignment



d) RNA-seq heatmap of Dux target genes in Dux domain constructs or C-terminal repeat and/or 14aa deletion constructs

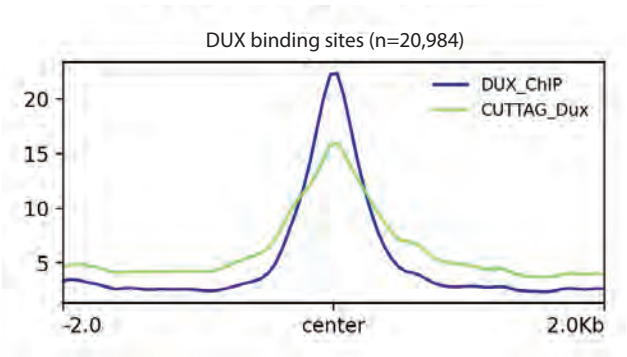


e) Phylogenetic analysis for mouse C-terminal repeats

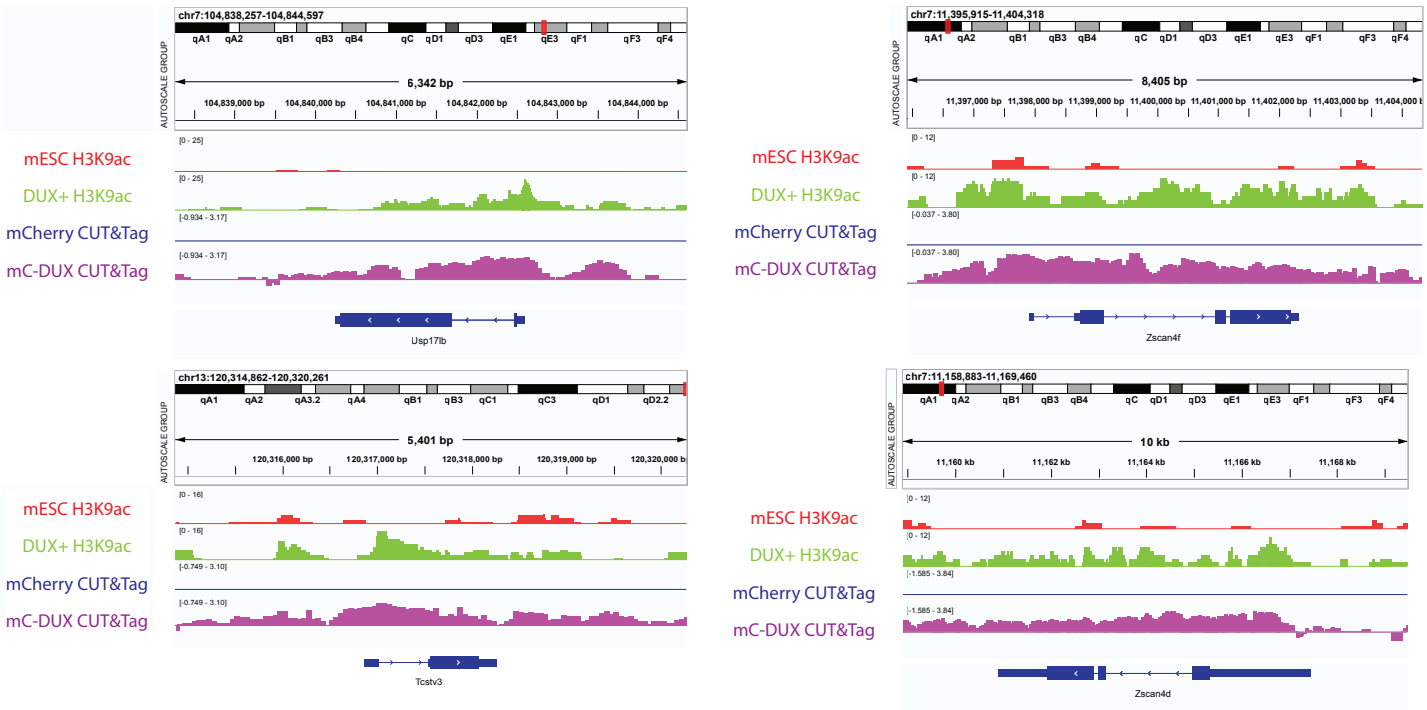


Supplementary Figure 3

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

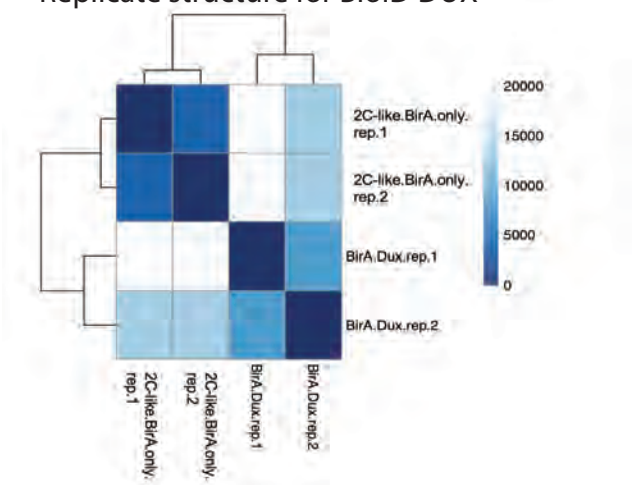


b) IGV screenshots of mESC H3K9ac and DUX-expressing cells H3K9ac at DUX target genes

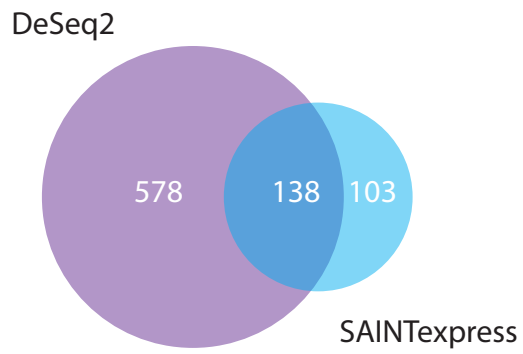


Supplementary Figure 4

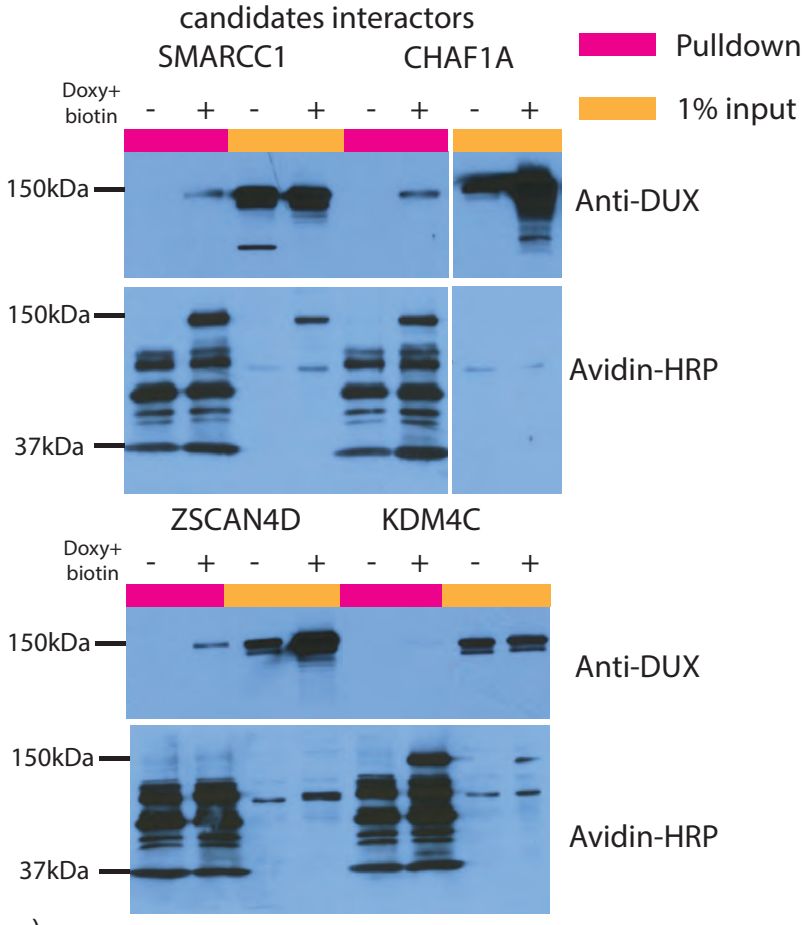
a) Replicate structure for BioID DUX



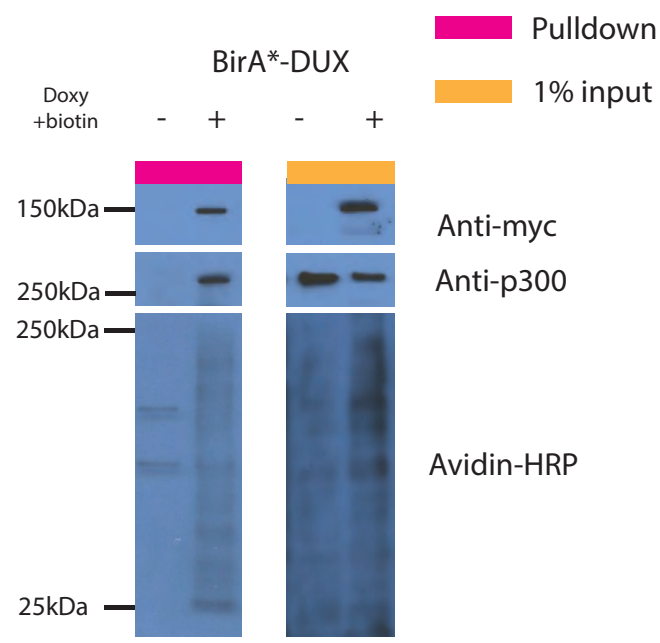
b) Number of statistically significant proteins in two different BioID DUX analysis packages



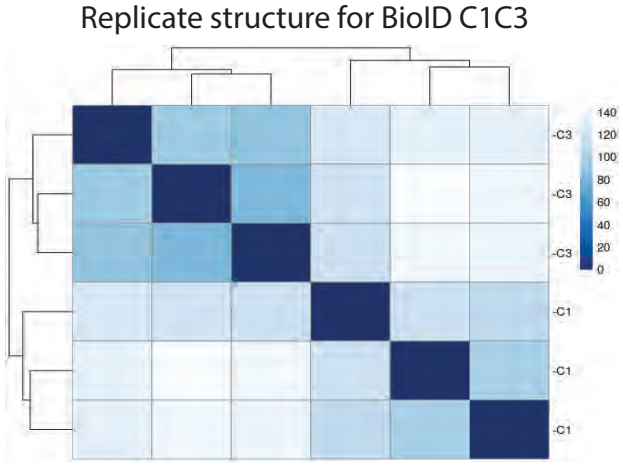
c) BioID experiment with FLAG-tagged candidates interactors



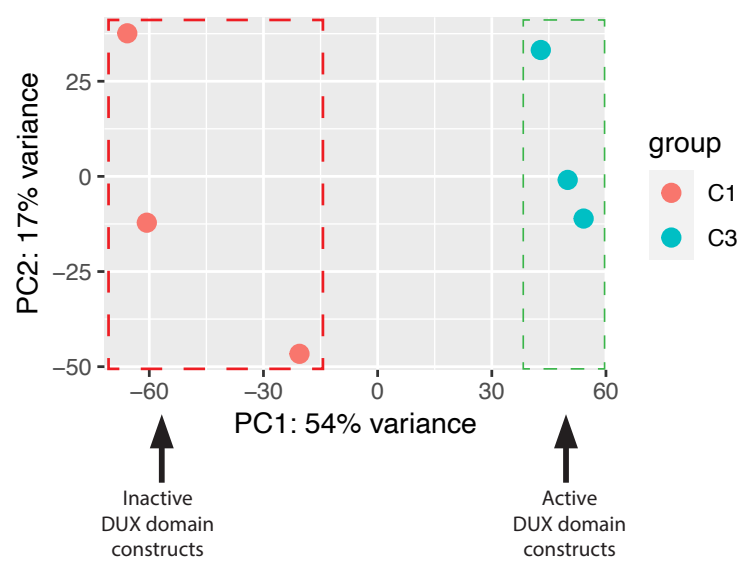
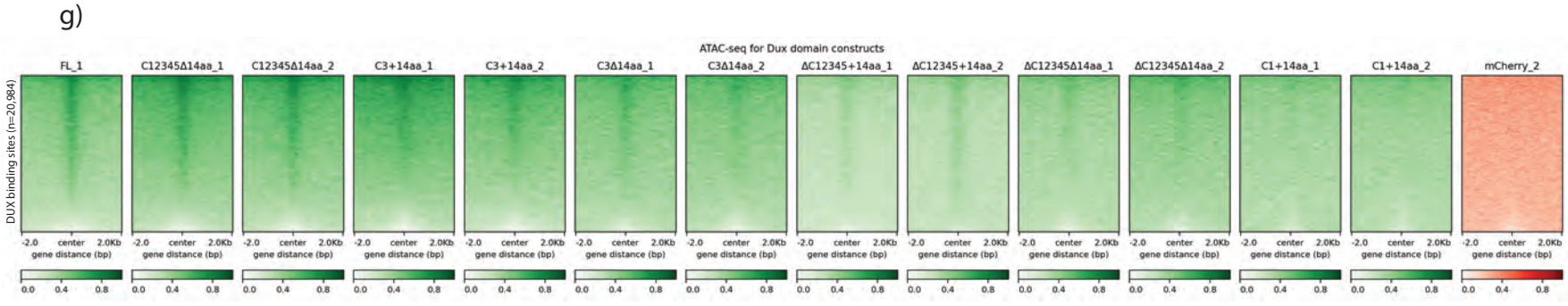
d) BioID DUX experiment with p300 Western blot



e) Replicate structure for BioID C1C3



f) Principle component analysis for BioID C1C3



g)