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 (15–18). DUX4 has additionally
53 been identified as the causal gene in facioscapulohumeral dystrophy (FSHD), leading to the
54 protein being intensely studied to identify its pathological traits (19–21).
55 DUX is known to be targeted to specific sites in the genome (including genes and MERVL
56 LTRs) by its two N-terminal homeodomains, whereas its C-terminal region drives transcriptional
57 activation(12,22,23). However, little is known about the nature of this C-terminal region, which
58 is much more divergent among *DUX* orthologs than are the homeodomains (24,25). One
59 possible contributor to transcriptional activation involves the interaction of the C-terminus of
60 DUX4 with the related histone acetyltransferases p300 and CBP (23,26). However, these or
61 other interactions have not been identified with mouse DUX. Regarding chromatin remodeling
62 during EGA, the abundant ATP-dependent chromatin remodeler SWI/SNF (BAF) complex plays a
63 central role(27,28). Specifically, SMARCA4, the catalytic ATPase within SWI/SNF complex, is
64 required for important for EGA in the 2-cell mouse embryo (29). SMARCC1 is a core structural
65 scaffold for SWI/SNF complexes, and its absence results in loss of complex activity *in vivo(30)*.

66 However, possible physical links between mouse DUX and the chromatin remodeler SWI/SNF
67 complex during EGA or during the mESC to 2C-like transition is currently unknown.
68 DUX activates about 25% of EGA genes and DUX expression is sufficient to revert mESCs
69 to an earlier totipotent stage(12). Consistent with DUX impacting early development, *Dux-/-*
70 crosses yield smaller litter sizes (22). However, as viable *Dux -/-* pups are generated, factor(s)
71 redundant with DUX appear to contribute to DUX target activation and chromatin
72 reprogramming in 2-cell embryos (22,31). Recently, oocyte-specific homeobox 4 (OBOX4), a
73 factor highly expressed during EGA and in 2C-like mESCs, has been identified as partly
74 redundant with DUX in 2-cell embryos(32). Taken together, DUX and OBOX4 appear to
75 cooperate to ensure proper EGA in the early mouse embryo.
76 In this study, we sought to better characterize DUX domains, identify proteins
77 interacting with DUX domains, and understand how DUX domains contribute to DUX target
78 gene activation, DUX promoter recruitment, and chromatin accessibility at binding sites.
79 Specifically, we reveal a unique C-terminal repeat structure in mouse DUX, and characterize two
80 functionally distinct repeat types (termed 'active' and 'inactive') and cooperativity of the active
81 repeats with the C-terminal acidic tail. Additionally, we identify DUX domains that interact with
82 the chromatin remodeler subunit SMARCC1, which alongside the ability of DUX to confer
83 histone acetylation, reveals new mechanistic insight into how the C-terminus of DUX confers
84 transcriptional activation.
85
86 **Results**
87

88 **The mouse DUX C-terminus has five conserved repeats followed by a 14 amino acid 'tail'**

89 Curiously, mouse DUX protein (673 amino acids (aa)) is considerably longer than DUX
90 orthologs such as human DUX4 (424 aa). Through manual inspection and alignments of mouse
91 DUX with other orthologs, we identified five protein repeats within DUX of ~100 aa each,
92 followed by a single 14aa highly-acidic C-terminal 'tail' (Figure 1a). However, human DUX4
93 contains only a single copy of this 'repeat' domain, which is likewise followed by a 14aa tail that
94 closely resembles the acidic tail in mouse DUX (Figure 1b). Prior work has revealed that the
95 acidic DUX4 C-terminal tail helps activate transcription, in part by binding to CBP and p300 (23).
96 To determine whether other placental mammals (e.g. rat, bovine, and human) contain a C-
97 terminal repeat structure similar to mouse DUX, we conducted pairwise alignments, which
98 revealed repeats only in the mouse lineage (Figure 1c, Supplementary Figure 1a,b,c). The
99 human DUX4 C-terminal region is most similar to the 3rd repeat of mouse DUX whereas the
100 bovine DUXC C-terminal region was most similar to the 5th repeat (Supplementary Figure 1d).
101 These findings suggest either repeat C3 or repeat C5 as the parental repeat for the other four.

102 Next, we assessed the contribution of mouse DUX repeats to transcriptional activity by
103 examining *Dux*-/- mESCs containing a MERVL::GFP reporter. The long-terminal repeat (LTR) of
104 MERVL is bound by DUX, and MERVL:GFP is commonly used to assess DUX activation and
105 conversion of mESCs to a 2CLC state (9,12). First, *Dux*-/- mESCs were transiently transfected
106 with mCherry-tagged DUX repeat domain constructs encoding zero to five C-terminal repeats,
107 without the 14aa tail (Figure 1d). Cells were induced with doxycycline for 18 hours, gated on
108 mCherry positivity, and MERVL::GFP expression was quantified by flow cytometry
109 (Supplementary Figure 1e). Expression of each derivative was confirmed by Western analysis

110 (Supplementary Figure 1f). The flow cytometry results reveal that the first two repeats have
111 minimal activity, but with the addition of the third repeat, low MERVL::GFP expression is
112 observed. The activity is further improved by the addition of the fourth and fifth repeat. The
113 results also support a potential role for the 14aa 'tail' in activation, as its omission appeared to
114 modestly lower MERVL::GFP expression (Figure 1d). Thus, our initial deletion series revealed
115 activity with repeats 3-5, but not with repeats 1-2 alone. However, this experiment did not
116 determine whether the repeats act in a simple additive manner, or if individual repeats confer
117 distinct transcriptional activity in a singular manner, or how they cooperate with the acidic tail.
118

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

121 We next explored the contributions of individual repeats, along with their cooperation
122 with the C-terminal acidic 'tail'. Here, we generated N-terminal mCherry-tagged DUX
123 derivatives with varying numbers of repeats either with or without the 14aa tail, as well as with
124 or without single or both homeodomains (Supplementary Figure 2a). All constructs contain a
125 $GGGGS_2$ linker present in constructs with a single internal deletion and $GGGGS_2$ and $GAGAS_2$
126 with constructs with two internal deletions. First, as expected, derivatives lacking either
127 homeodomain failed to activate the MERVL::GFP reporter (Supplementary Figure 2a). Second,
128 combinations involving C1, C2 or C4 with the acidic 'tail' lack activation of the MERL::GFP
129 reporter, whereas combinations of C3 or C5 with the 'tail' display potent reporter activation
130 (Figure 2a). The latter combinations require the 14aa tail as constructs lacking the 14aa tail, but
131 containing the C3 or C5 repeat alone did not activate the reporter. Notably, a construct

132 containing C1 and C2 with the 14aa tail was inert, whereas C1, C2 and C3 with the 14aa tail was
133 highly active (Supplementary Figure 2a). Together, these results suggest that strong DUX target
134 gene activation requires a combination of two C-terminal domains: a transcriptionally 'active'
135 repeat (C3 or C5) and the 14aa tail (Figure 2b).

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

145 In order to expand these analyses to global transcription and all DUX targets (beyond
146 the MERVL::GFP reporter), we performed RNA-seq on full-length (FL) DUX and selected chimera
147 constructs. This approach was designed to test whether the 'active' constructs might differ in
148 their gene targets (Figure 2e and Supplementary Figure 2c). First, the PCA plot for these
149 samples separated samples largely into two groups – active versus inactive – with one sample
150 in an intermediate position (Figure 2e). Interestingly, the DUX constructs which activate the
151 MERVL::GFP reporter (FL, C1C3a, and C3+14a) all showed strong activation of DUX target genes,
152 whereas the 'inactive' constructs that do not activate the MERVL::GFP reporter (C1+14aa,
153 C1C3c, and mCherry alone) did not activate DUX target genes (Figure 2f). Notably, the

154 intermediate position was occupied by the C1C3b derivative, reinforcing the notion that the 6aa
155 region is important for regulating target activation. K-means clustering of the RNA-seq for the
156 chimera constructs at DUX target generated three clusters of genes: the first two of which
157 largely contained genes activated by FL DUX, whereas the cluster #3 contains genes that have
158 increased expression in DUX domain derivatives compared to FL DUX. The 'intermediate'
159 derivative C1C3b, which is a hybrid of an 'active' and an 'inactive' repeat, was capable of
160 activating the majority of DUX targets in cluster #1, but failed to activate those in cluster #2,
161 whereas it activated (or failed to attenuate) genes in cluster #3. Genes activated by C1C3b in
162 cluster #3 generate GO terms such as 'upregulated include synapse assembly', 'sex
163 differentiation', 'reproductive structure development', and 'neuron differentiation'. Here,
164 cluster 3 shows examples of known DUX target genes that have a higher relative expression in
165 C1C3b compared to full-length DUX include *Obox1*, *Olfr1277*, *Obox2*, and *Zfp296*. These
166 combined observations, especially those involving derivatives that cause the precocious
167 activation of genes in cluster #3, raise the interesting possibility that the 'inactive' repeats may
168 serve to attenuate the active repeats at particular targets, such as *Obox* genes. Finally, when
169 considering amino acid similarity between the five repeats, we note that C3 and C5 are most
170 similar to one another, compared to C1, C2, and C4 (Supplementary Figure 2d).
171

172 **Chromatin alterations at DUX binding sites require a transcriptionally active DUX repeat**

173 To investigate the chromatin landscape effects conferred by *Dux* derivative expression,
174 we performed CUT&Tag in *Dux*-/-mESCs on selected *Dux* domain constructs: FL, C12345△14aa,
175 C3+14aa, C3△14aa, △C12345+14aa, and mCherry alone. Because all of the constructs had the

176 same N-terminal mCherry tag, an mCherry antibody was used throughout. We isolated and
177 analyzed clonal stable cell lines which expressed species of the expected protein size and
178 demonstrated comparable expression to one another (Figure 3a). It is of note that mCherry-
179 DUX protein is expected to run at roughly 100kDa, but runs anomalously on polyacrylamide gels
180 – at 150kDa. This is likely due to a property of the C-terminal repeats as the homeodomains
181 alone run at the expected molecular weight. To benchmark DUX CUT&Tag against previous
182 chromatin occupancy data, we first generated a CUT&Tag dataset for mCherry-tagged full-
183 length DUX protein, and found that it overlapped highly with previously published HA-tagged
184 DUX ChIP-seq results (Supplementary Figure 3a) (12). Extension of CUT&Tag to the DUX
185 derivatives revealed their moderate-to-high occupancy at DUX sites, except for C3△14aa, which
186 displayed reduced occupancy at DUX binding sites (Figure 3b, 3c). Next, we determined the
187 ability of our DUX domain constructs to alter H3K9ac, as H3K9ac globally increases in 2C-like
188 cells (33). Notably, H3K9ac occurs at DUX binding sites correlating with the of transcriptional
189 activity of each *Dux* construct (Figure 3d,e, Supplementary Figure 3b). Taken together, our data
190 demonstrate that the 14aa acidic tail is not strictly required for the occupancy of DUX targets.
191 Further, our data suggest that the 14aa tail is important for conferring H3K9ac aids at bound
192 sites and might also be involved in retention of DUX at target sites since removal of the 14aa
193 tail in the context of a single active repeat appeared to diminish target occupancy.
194
195 **Proximity-labeling with full-length DUX reveals interactions with proteins involved in**
196 **chromatin de-repression.**

197 Next, we identified proteins that interact with DUX derivatives using the BioID approach,
198 which identifies proteins that interact directly, indirectly, or transiently with a protein of
199 interest (34). Here, we utilized BioID to define protein interactions with DUX in the 2C-like
200 cellular environment. To accomplish this, we fused BirA*, a promiscuous biotinylating enzyme,
201 to the N-terminus of full length DUX. When biotin is added to the growth media, BirA* will
202 biotinylate lysine residues of proteins within a 10nm radius. Our approach involved the
203 isolation of biotinylated proteins and the use of liquid chromatography mass spectrometry (LC
204 MS/MS) to identify associated proteins in a quantitative manner (Figure 4a). Our experimental
205 design compared expression of BirA*-DUX (full-length) in mESCs (which creates a 2C-like state)
206 to expression of unfused BirA* alone (control) expressed in a 2C-like environment, in order to
207 mimic a DUX-expressing 2C-like proteome in both conditions (Figure 4b). Here, we created
208 clonal stable cell lines harboring our constructs, and expressed the proteins for 12hr with 0.25
209 µg/mL doxycycline and 5µM biotin. DESeq2 analysis of the sum intensity values from LC MS/MS
210 revealed clustering between replicates (Supplementary Figure 4a) and a volcano plot revealed
211 an enrichment in protein interactions for BirA*-DUX compared to BirA* alone, including those
212 involved in chromatin de-repression such as the SWI/SNF (BAF) complex members SMARCC1
213 and SMARCA4 (Figure 4c, 4d). In addition to DESeq2 analysis, which is normally used for RNA-
214 seq data, we applied the SAINTexpress package, which is designed for protein interactome
215 analysis, to additionally analyze this dataset. DESeq2 appeared to be slightly less stringent than
216 SAINTexpress analysis, with a larger number passing statistical threshold (p_{adj} <0.05)
217 (Supplementary Figure 4b). SAINTexpress uses a statistical score called SAINTscore ranging from

218 0-1, with a significance threshold established as ≥ 0.73, which parallels an estimated protein-
219 level Bayesian FDR of < 0.05(35).
220 SWI/SNF subunit SMARCC1 and transcription factor ZSCAN4D, two statistically
221 significant hits from both analyses, were then selected for validation by testing for interaction
222 *in vivo* via immunoprecipitation. We expressed N-terminally FLAG-tagged versions of SMARCC1
223 and ZSCAN4 in HEK293 cells co-expressing the BirA*-DUX fusion in the presence of 50 µM biotin
224 (50µM biotin for BioID was empirically determined as the optimal dose in HEK293Ts). SMARCC1
225 and ZSCAN4 were immunoprecipitated with anti-FLAG conjugated beads, and analysis with an
226 avidin-HRP blot showed they are both biotinylated and co-IP with DUX protein, validating BioID
227 protein interactions (Figure 4e). Additionally, CHAF1A, an additional BioID hit, was also
228 validated through FLAG-tag co-IP and DUX protein interaction, however KDM4C did not
229 validate, suggesting that this hit was a false positive (Supplementary Figure 4c), and highlighting
230 the importance of validating the BioID results.
231 Lastly, it has been previously reported that human DUX4 interacts with p300 (23,26),
232 however p300 was not identified as a significant interactor in our dataset. This could be due to
233 differences in the interactomes of human DUX4 and mouse DUX, or instead be technical, due
234 perhaps to a lack of sensitivity to detect P300 peptides. To determine if mouse DUX similarly
235 interacts with P300, we performed a BioID-like experiment in mESCs, which revealed p300 as
236 modestly biotinylated, suggesting that a portion of p300 comes in close proximity with DUX, but
237 was not annotated in the dataset, representing a possible false negative (Supplementary Figure
238 4d). In summary, our BioID experiment included chromatin- and transcription-related factors, a
239 portion of which were validated using co-immunoprecipitation, which support prior

240 observations of a strong impact of DUX occupancy on chromatin opening and transcriptional
241 activity.
242
243 **Transcriptionally active DUX C-terminal repeat C3 binds to proteins involved in chromatin**
244 **remodeling**
245 Our work above presents the opportunity to identify proteins that interact differentially with
246 a transcriptionally 'active' DUX derivative compared to a highly similar 'inactive' DUX derivative.
247 Here, we separately N-terminally tagged our C1+14aa and C3+14aa derivatives with BirA*, to
248 represent transcriptionally 'inactive' and 'active' DUX constructs, respectively (Figure 5a). The
249 experimental design and setup was identical to that described in Figure 4a, but the mESC line
250 utilized bore a *Dux*-/- genotype. Furthermore, in place of the BirA* alone control, the inactive
251 BirA*-C1 (BirA* fused to C1+14aa) was used as the control construct for comparison to the
252 active BirA*-C3 (BirA* fused to C3+14aa).
253 PCA and clustering analysis reveals that the three replicates of the 'active' C3 derivative
254 group together and are more similar to each other (via clustering) than are the C1 replicates,
255 and vice versa (Supplementary Figure 4e,f). Using SAINTexpress, we observed many more
256 statistically-enriched (SAINTscore ≥ 0.73) protein interactions occurring with the DUX C3
257 'active' repeat compared to the 'inactive' C1 repeat. Specifically, we observed 12 proteins with
258 the C1 derivative, and 154 proteins with C3 (Figure 5b). For C3 derivative interactors, GO-term
259 analysis identified two overarching functional categories: chromatin and the nucleolus. Within
260 the chromatin category, proteins within the NuRD, SWI/SNF and ISWI chromatin remodeling
261 complexes were prominent (Figure 5c). Notably, proper formation of the nucleolus has been

262 found to be required for *Dux* repeat locus silencing (36,37), and our data strongly suggests a
263 direct physical interaction with nucleolar components, as DUX can bind to its own genetic locus
264 (38).

265 To further investigate chromatin-related interactors of interest, we performed co-IPs for
266 FLAG-tagged SMARCC1 (BAF155) (described above) using selected DUX domain derivatives.
267 First, SMARCC1 binding to DUX appears only slightly diminished with the loss of the 14aa tail
268 (Figure 5d). SMARCC1 also binds well to a DUX derivative containing both an active C3 repeat
269 and the acidic tail (C3+14aa), and binding is largely maintained when the tail is removed
270 (C3△14aa). Finally, a major reduction in SMARCC1 binding is observed when all repeats are
271 removed and only the acidic tail remains. This suggests a hierarchy of interaction for DUX and
272 SMARCC1: an 'active' repeat alone binds SMARCC1 whereas the acidic 'tail' alone does not bind
273 SMARCC1, however the acidic tail slightly improves SMARCC1 interaction when active and
274 inactive repeats are both present (full-length DUX).

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

282 To continue investigating DUX interaction with SWI/SNF, we performed ATAC-seq in the DUX
283 derivatives to determine whether SMARCC1 recruitment to DUX binding sites correlates with

284 chromatin accessibility changes (Figure 5f). The results suggest a similar hierarchy of interaction
285 observed in the SMARCC1 CUT&Tag experiment for DUX domain constructs. At DUX binding
286 sites, full-length DUX leads to increased opening, as do C12345△14aa and C3+14aa, the two
287 derivatives which confer transcriptional activity. However, DUX derivatives which lack
288 transcriptional activity or have decreased binding to SMARCC1 display a corresponding
289 decrease in chromatin opening, though some opening does occur. This evidence further
290 supports a combinatorial domain model for proper DUX protein interaction, occupancy, and
291 target activation – revealing a major/required contribution from the C-terminal repeats and
292 augmentation by the C-terminal tail.
293
294 **Discussion**
295 Embryonic genome activation (EGA) creates a transcriptome and proteome that replaces
296 maternally-inherited components, reprograms chromatin and drives cleavage-stage processes
297 (12). The transcription factor DUX is highly but transiently expressed during EGA (specifically
298 the early 2-cell embryo stage) from a large locus containing ~30 *Dux* repeats. DUX plays a key
299 role in opening many promoters and enhancers, and in activating endogenous genes and repeat
300 elements in the 2-cell embryo (38). DUX targets account for about 25% of the cleavage-stage
301 transcriptome, and partly overlap with targets of OBOX4, underlying the observation that
302 combined DUX and OBOX4 omission/depletion strongly affects embryo development (32).
303 Although DUX targeting by its N-terminal homeodomains is well understood, how domains in
304 the C-terminus of DUX contribute to chromatin reprogramming and transcription has remained
305 largely unknown (24). Here, we conducted an analysis combined with proximity ligation to

306 reveal the protein partners of DUX domains and the roles of DUX domains in chromatin
307 remodeling and transcriptional activation.
308 First, we found that the mouse DUX C-terminal region contains five ~100aa repeats (termed
309 C1-5) followed by a 14aa acidic C-terminal tail, whereas DUX orthologs contain only one copy of
310 this repeat, followed by a similar acidic tail. Our phylogenetic and DUX domain analyses
311 suggest that two of the repeats (C3 and C5) perform a more active role – and we note that the
312 single repeat present in DUX orthologs more highly resembles the 'active' repeats C3 and C5,
313 than the 'inactive' C1/2/4 repeats. Interestingly, single 'active' repeats show no transcriptional
314 stimulation on their own, but show potent activity in combination with the acidic tail,
315 essentially mimicking the domain-activity relationships of their human, rat or bovine DUX4
316 orthologs. Together, these observations suggest that additional DUX repeats may have derived
317 from duplication of either C3 or C5 active repeat, with subsequent sub-functionalization.
318 These observations raise two interesting questions – why does DUX have multiple repeats
319 with different activities, and what distinguishes active from inactive repeats? Our work has
320 localized the differences between active and inactive repeats to a small 6 aa region, and active
321 versus inactive repeats (in combination with the acidic tail) show very different profiles of
322 protein interaction – with active repeats displaying more robust interaction with chromatin and
323 transcription partners. Interestingly, we observe a cohort of genes (Figure 2f, cluster #3) that
324 are silent with full-length DUX, silent with an 'inactive' repeat (C1), but active with a single
325 'active' repeat (C3). This raises the possibility that the 'inactive' repeats may serve to attenuate
326 the activity of the active repeats under certain conditions or at certain locations. Of note, DUX
327 binds to its own promoter to help activate a strong positive feedback loop; however, *Dux*

328 activation is brief (half a cell cycle), with silencing accompanying the association of the large
329 *Dux* locus with the nucleolar envelope(38,39). Interestingly, we observe high enrichment of
330 nucleolar components with the 'active' C3 repeat, raising the possibility that the 'active' repeat
331 may initially recruit powerful chromatin and transcription factors, and then subsequently
332 contribute to localizing the *Dux* locus to a compartment that overcomes and/or reverses those
333 associations to enable *Dux* locus silencing(36). Furthermore, interactions of DUX with histone
334 deacetylase components were obtained, which remain to be fully explored in this silencing
335 process. These observations provide new mechanistic information as well as opening new
336 avenues for future exploration regarding the ability of DUX to function as a strong activator,
337 while also helping to subsequently silence the repetitive *Dux* locus. Here, we envision that
338 temporal and/or locus-specific post-translational modifications on DUX may mediate the
339 temporal associations of DUX with target factors. Indeed, the presence of both 'active' and
340 'inactive' DUX repeats may help in coordination of these associations.
341 Regarding the functions of the DUX C-terminus, prior work of others has shown that the
342 acidic C-terminal tail of DUX4 promotes transcriptional activation and interaction with the
343 histone acetyltransferase p300(23). Our validation experiments confirmed those results with
344 mouse DUX, and we extended our studies to explore the roles and relationships of the repeat
345 regions, both alone and in combination with the acidic tail. Our examination of *Dux* domain
346 constructs was extensive, and included RNA-seq, CUT&Tag, BioID, ATAC-seq and the ability to
347 convert mouse ES cells to 2C-like cells. As DUX opens chromatin, and as three SWI/SNF
348 complex components were present in our BioID results (SMARCC1, SMARCA4, ACL6A), these

349 results prompted our direct assessment and validation of the recruitment of SWI/SNF complex
350 (monitoring the core SMARCC1 subunit) to DUX binding sites (Figure 5c, Figure 6a).
351 An additional goal was to explore whether the functions and associations of the active
352 repeats versus the acidic tail are largely separable/independent, or instead primarily additive or
353 cooperative. First, regarding the activation of cleavage-specific genes (including MERVL), the
354 target repertoire observe with full-length DUX is largely recapitulated with DUX derivatives
355 containing the homeodomains and both active C-terminal repeats – even in the absence of the
356 acidic tail. Thus, the tail accentuates activity, but is not strictly required for activity at the vast
357 majority of targets. Regarding domain contributions to occupying DUX targets, all derivatives
358 were able to bind DUX targets, though the absence of the acidic tail lowers occupancy if only
359 one active repeat is present. This contrasts with the recruitment of H3K9ac activity; here,
360 neither the acidic tail nor an active repeat is sufficient for recruitment; instead, either multiple
361 repeats or repeat-tail combinations appear needed. For SMARCC1 (SWI/SNF), an active repeat
362 is sufficient for recruitment, but recruitment is enhanced by the addition of the tail, whereas
363 the tail alone appears ineffective. Finally, and of particular interest, a DUX derivative lacking
364 repeat regions but containing the acidic tail (△C12345+14) occupies DUX targets, but does not
365 activate targets – and does not recruit either H3K9ac or SWI/SNF complex. Therefore, the
366 acidic tail appears to be effective only when combined with an active C-terminal repeat.
367 Overall, our data is consistent with largely cooperative functions for active repeat(s) and the
368 acidic tail – as the absence of either compromises most functions and interactions – but places
369 the C-terminal domains as more important domain in the hierarchy, as key factors like SWI/SNF
370 complex are fully reliant on an active repeat for interaction. In keeping, our work predicts

371 interactions with chromatin and transcription factors involving combined surfaces on repeats
372 and the acidic tail (a two-domain model).
373 As previewed above, our proximity-labeling (BioID) revealed a variety of protein complex
374 interactions, and we validated a subset of interactors with co-IP experiments, including
375 particular chromatin and transcription-related factors. We note that certain interactions may
376 be indirect and involve DUX association with component(s) involved in a larger multi-protein
377 structures, such as the highly repetitive *Dux* locus (to which DUX binds(12,38)) in either its
378 active or inactive form, and the association of the *Dux* locus with the nucleolus or other
379 compartments. Thus, a complete understanding of the full repertoire of interactions will
380 require future studies. However, beyond the factors noted above, we will emphasize two
381 additional factors prominent in the BioID data that might be prioritized. First, the p53
382 interaction verifies prior work supporting p53 association with and activation of the *Dux* locus,
383 and prompts examination of a direct interaction with DUX(38). Second, the presence in our
384 dataset of SMCHD1, a cohesin-like protein known to help confer repression of the mouse *Dux*
385 and human *DUX4* repetitive loci, and to be mutated in a subset of FSHD patients, likewise
386 prompts considerable future attention (40,41).
387 Taken together our work provides considerable new mechanistic information regarding the
388 structure-function relationships and factor associations conducted by particular domains within
389 the mouse DUX C-terminal region. Key findings include the characterization of 'active' repeat
390 domains, their interaction with transcription and chromatin remodeling factors, and the
391 cooperativity between DUX active repeats and the acidic tail in cofactor recruitment and
392 transcriptional activation at DUX targets. In addition to factors for activation, factors related to

393 repression and nucleolar association were also isolated, and future studies should help

394 determine whether their temporal associations with DUX may help early embryos execute a

395 quick ‘burst’ of *Dux* activation followed by repression at the *Dux* locus in 2-cell embryos.

396

397

398 **References**

399 1. Jukam D, Shariati SAM, Skotheim JM. Zygotic Genome Activation in Vertebrates. Dev Cell.
400 2017 Aug;42(4):316–32.

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

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

405 4. Lee MT, Bonneau AR, Takacs CM, Bazzini AA, DiVito KR, Fleming ES, et al. Nanog, Pou5f1 and
406 SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature.
407 2013 Nov;503(7476):360–4.

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

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

412 7. Xu R, Li S, Wu Q, Li C, Jiang M, Guo L, et al. Stage-specific H3K9me3 occupancy ensures
413 retrotransposon silencing in human pre-implantation embryos. Cell Stem Cell. 2022 Jul
414 7;29(7):1051-1066.e8.

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

417 9. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, et al. Embryonic stem
418 cell potency fluctuates with endogenous retrovirus activity. Nature. 2012 Jul;487(7405):57–
419 63.

420 10. Rodriguez-Terrones D, Hartleben G, Gaume X, Eid A, Guthmann M, Iturbide A, et al. A
421 distinct metabolic state arises during the emergence of 2-cell-like cells. EMBO Rep. 2020 Jan
422 7;21(1):e48354.

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

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

428 13. Zuo F, Jiang J, Fu H, Yan K, Liefke R, Zhang J, et al. A TRIM66/DAX1/Dux axis suppresses
429 the totipotent 2-cell-like state in murine embryonic stem cells. Cell Stem Cell. 2022 Jun
430 2;29(6):948-961.e6.

431 14. Guo M, Zhang Y, Zhou J, Bi Y, Xu J, Xu C, et al. Precise temporal regulation of Dux is
432 important for embryo development. Cell Res. 2019 Nov;29(11):956–9.

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

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

439 17. Taubenschmid-Stowers J, Rostovskaya M, Santos F, Ljung S, Argelaguet R, Krueger F, et
440 al. 8C-like cells capture the human zygotic genome activation program in vitro. Cell Stem Cell.
441 2022 Mar 3;29(3):449-459.e6.

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

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

447 20. Vanderplanck C, Ansseau E, Charron S, Stricwant N, Tassin A, Laoudj-Chenivesse D, et al.
448 The FSHD Atrophic Myotube Phenotype Is Caused by DUX4 Expression. PLOS ONE. 2011 Oct
449 28;6(10):e26820.

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

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

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

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

459 25. Leidenroth A, Clapp J, Mitchell LM, Coneyworth D, Dearden FL, Iannuzzi L, et al.
460 Evolution of DUX gene macrosatellites in placental mammals. Chromosoma. 2012 Oct
461 1;121(5):489–97.

462 26. Vuoristo S, Bhagat S, Hydén-Granskog C, Yoshihara M, Gawriyski L, Jouhilahti EM, et al.
463 DUX4 is a multifunctional factor priming human embryonic genome activation. iScience.
464 2022 Apr 15;25(4):104137.

465 27. Clapier CR, Cairns BR. The biology of chromatin remodeling complexes. Annu Rev
466 Biochem. 2009;78:273–304.

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

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

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

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

476 32. Guo Y, Kitano T, Inoue K, Murano K, Hirose M, Li TD, et al. Obox4 secures zygotic
477 genome activation upon loss of Dux [Internet]. bioRxiv; 2022 [cited 2023 Jan 5]. p.
478 2022.07.04.498763. Available from:
479 https://www.biorxiv.org/content/10.1101/2022.07.04.498763v2

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

483 34. Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies
484 proximal and interacting proteins in mammalian cells. J Cell Biol. 2012 Mar 19;196(6):801–
485 10.

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

489 36. Xie SQ, Leeke BJ, Whilding C, Wagner RT, Garcia-Llagostera F, Low Y, et al. Nucleolar-
490 based Dux repression is essential for embryonic two-cell stage exit. Genes Dev. 2022 Mar
491 1;36(5–6):331–47.

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

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

498 39. Vega-Sendino M, Olbrich T, Stein P, Tillo D, Carey GI, Savy V, et al. The homeobox
499 transcription factor DUXBL controls exit from totipotency [Internet]. bioRxiv; 2022 [cited
500 2023 Feb 8]. p. 2022.09.19.508541. Available from:
501 https://www.biorxiv.org/content/10.1101/2022.09.19.508541v1

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

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

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

510

511 **Conflict of interest**

512 The authors declare no competing interests.

513 **Specific author contribution**

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

515 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.

516 **Funding**

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

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

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

520 **Acknowledgements**

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

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

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

524 Taplin Mass Spectrometry Core (Harvard University) for mass spectrometry assistance for BioID
525 experiments. Lastly, we thank Brian Patten at Novogene for Illumina sequencing services for
526 CUT&Tag and ATAC-seq libraries.

527 **Materials Availability**

528 Plasmids used in this study will be deposited to Addgene or available by request.

529 **Data accessibility**

530 RNA-sequencing, CUT&Tag, and ATAC-seq data are available through GEO GSEXXXXXX.
531 Standard packages were used for RNA-seq, CUT&Tag, and ATAC-seq analyses (see "Materials
532 and Methods").

533

534 **Materials and Methods**

535 **Amino acid alignment**

536 DUX ortholog protein translations were aligned using Geneious Local Alignment (Smith-
537 Waterman), Blosum62 cost matrix, gap open penalty of 12, and gap extension penalty of 4.

538 **Mouse embryonic stem cell culture**

539 Mycoplasma-free E14 mESCs were cultured on gelatin in 2i+LIF medium containing Gibco KO-
540 DMEM with nonessential amino acids, 2-mercaptoethanol and dipeptide glutamine and were
541 supplemented with 15% ESC-grade FBS, leukemia inhibitory factor (Thermo Fisher), 1 mM
542 PD0325901 (Sigma-Aldrich) and 3 mM CHIR99021 (Sigma-Aldrich). Stable cell lines were
543 blasticidin (Fisher Scientific, B12150-0.1) at 3 µg/mL.

544 **HEK293T cell culture**

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

547 **Transfection of mESCs and HEK293Ts**

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

552 **Doxycycline treatment of cells**

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

556 **Western blot**

557 RIPA buffer (50 mM TRIS-HCl [pH 8.0], 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate,
558 0.1% SDS) with 1X protease inhibitors was used to lyse mESCs. A 100X protease inhibitor
559 cocktail consists of 0.03 mg/mL leupeptin, 0.14 mg/mL pepstatin A, 0.02 mg/mL chymostatin,
560 8.5 mg/mL phenylmethanesulfonylfluoride fluoride, 33mg/mL benzamidine and solubilized in
561 ethanol. Lysed cell suspensions were then rotated for 30 min at 4°C and were further sonicated
562 for 10 min with a 30-s on-off cycle in a Diagenode Bioruptor Pico device to solubilize chromatin-
563 bound proteins. Protein lysates were quantified with the Bio-Rad (500-0001EDU) Bradford
564 reagent and loaded on SDS-PAGE gel. After transfer to nitrocellulose membranes (VWR, 95040-
565 108), membranes were blocked with TSBT (TBS with 0.1% Tween-20) and 5% milk and then
566 incubated with antibodies overnight (mCherry ab167453, DUX (used in (38))and signals were

567 detected using HRP-conjugated secondary antibodies and Western Lightning Plus-ECL,
568 Enhanced Chemiluminescence (Perkin Elmer Health Sciences, NEL 105001EA).

569 **Flow cytometry**

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

574 **RNA-seq**

575 1 million mESCs were grown for 24 hours on a 100mm dish and then induced with doxycycline
576 (0.25 ug/mL) for 12 hours. Cells were washed two times with PBS, TRIzol (Thermo Fisher) was
577 added to tissue culture wells and incubated for 1 min at room temperature. TRIzol/cells were
578 added to Eppendorf tubes and RNA was extracted using the TRIzol extraction method. Libraries
579 were generated according to the manufacturer's instructions: polyA-selected RNA was isolated
580 and libraries were prepared using the NEBNext kit (New England Biolabs, e7500s). Purified
581 libraries were quantified on an Agilent Technologies 2200 TapeStation with a D1000
582 ScreenTape assay. The molarity of adaptor-modified molecules was defined by a qPCR with a
583 KAPA Library Quantification kit (Kapa Biosystems). Individual libraries were normalized to
584 10nM, and equal volumes were pooled in preparation for Ilumina sequence analysis.
585 Sequencing libraries (2pM) were chemically denatured and applied to an Illumina HiSeq paired-
586 end flow cells with an Illumina cBot. Flow cells were then transferred to an Ilumina HiSeq 2000
587 instruct and sequenced in the 125-bp paired-end mode. RNA-seq reads were trimmed and
588 filtered for quality using FastQC (version 0.5.0). Processed reads were aligned using STAR

589 (version 2.7.2c). Differential expression analysis was performed using DESeq2 (version 3.11).
590 Hierarchical clustering was performed using the R package 'Complex Heatmap,' with inputs
591 being differentially expressed transcripts (DESeq FDR < 0.05) z-score normalized.

592 **GO term analysis**

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

594 **CUT&Tag**

595 1 million mESCs were grown for 24 hours and then induced with doxycycline (0.25 ug/mL) for
596 12 hours. Cell were washed 1X with PBS and nuclei were isolated from mESCs using nuclear
597 extraction buffer (20 mM HEPES-KOH [7.9], 10 mM KCl, 0.1% Triton X-100, 20% glycerol, 0.5
598 mM spermidine, 1X protease inhibitors (see above). ConA-streptavidin beads were paired with
599 antibodies of interest 1:100 (mCherry ab167453, H3K9ac ab 32129, SMARCC1 ab172638) and
600 50,000 cells were added and incubated overnight in binding buffer (20 mM, HEPES [pH 7.9], 10
601 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$). Secondary guinea pig anti-rabbit 1:100 (ABIN101961) was
602 incubated for 1 hr at room tempoerature. pA-Tn5 (pre-loaded with Mosaic End A/B adaptors)
603 was incubated for 1 hour at room temperature and used at 1:100. After washing steps and
604 release from the beads, PCR was performed for 12 cycles using NEBnext High-Fidelity 2X PCR
605 Master Mix (New England Biolabs, M0541L) (42). Samples were 1:1.3 AMPure XP (Beckman
606 Coulter) purified and sent for sequencing. All libraries were sequenced on the Illumina HiSeq
607 6000 platform in 150-bp, paired-end format. Paired-end, raw read files were first processed by
608 Trim Galore (Babraham Institute) to trim low-quality reads and remove adapters. Processed
609 reads were then aligned to mm10 with the following parameters: (--end-to-end –very sensitive
610 –no-mixed). CUT&Tag libraries were trimmed and filtered for quality using FastQC (version

611 0.5.0) and Picard (version 2.26.3). Processed reads were aligned using Bowtie2 (version 2.4.2)
612 with the parameters: (--end-to-end –very sensitive –no-mixed) and indexed using Samtools
613 (version 1.15). Duplicate samples were merged and normalized to control samples (mESCs for
614 H3K9ac dataset and mCherry alone for mCherry and SMARCC1 datasets). DeepTools was used
615 to generate heatmaps and profile plots using DUX binding sites as previously published (22).

616 **BioID**

617 15 million BirA* mESC derivatives were plated on gelatin-coated 15-cm plates and recovered
618 for 24-hours. 0.25 μg/mL doxycycline and 0.5 μM biotin (Sigma Aldrich, B4639) was added to
619 cells and incubated for 18 hours. Cells were washed with ice cold PBS 2 times and 1 mL lysis
620 buffer (2 mM Tris-HCl [pH 8.0], 137 mM NaCl, 1% NP-40, 2 mM EDTA, 1x protease inhibitors)
621 was added to each plate and incubated for 1 min (34). Cells were scraped of off the plate and
622 transferred to a 2-mL tube. Cells then rotated for 30 min at 4°C and were further sonicated for
623 10 min with a 30-s on-off cycle in a Diagenode Bioruptor Pico device to solubilize chromatin-
624 bound proteins. Next, samples were centrifuged at 4°C at 12,000 rpm for 20 minutes and the
625 supernatant was transferred to new tubes. Protein lysates were quantified with the Bio-Rad
626 (500-0001EDU) Bradford reagent. For streptavidin pulldown, 2 μg of protein were added to pre-
627 washed 50 μL streptavidin Sepharose bead conjugates (Cell Signaling Technology, 3419S) and
628 rocked overnight at 4°C. Cells were pelleted and then washed 5X with wash buffer (10 mM TRIS
629 [pH 7.4], 1 mM EDTA, 1 mM EGTA, 150 mM NaCl, 1% Triton X-100, 0.2 mM sodium
630 orthovanadate, 1x protease inhibitors), boiled at 95°C in 30 μL 4X Laemelli buffer, and ran on a
631 SDS-page gel until the samples were 1-inch into the gel. Protein bands/smears were cut out of
632 the gel with razor blades and transferred to individual 1.7-mL Eppendorf tubes filled with

633 ddH_20. Samples were overnight shipped on dry for liquid chromatography-tandem mass
634 spectrometry sequencing (LC-MS/MS) to Taplin Mass Spectrometry Facility, Cell Biology
635 Department, Harvard Medical School. Gel smears were reduced using 1 mM DTT in 50 mM
636 ammonium bicarbonate for 30 min at 60°C and cooled, and 5 mM iodoacetamide in 50 mM
637 ammonium bicarbonate was added for 15 min and incubated at room temperature. 5 mM DTT
638 was added to quench the reaction, and then 5 ng/μL of trypsin (Promega) was added and
639 samples incubated over night at 37 °C. Samples were then desalinated and resuspended in 10
640 μL of high-pressure liquids chromatography (HPLC) solvent A (2.5% acetonitrile, 0.1% formic
641 acid). A reverse-phase HPLC capillary column was made by placing 2.6 μL M C_{18} spherical silica
642 beads into a fused silica capillary (100-μL M inner diameter, ~30-cm length). Next, each sample
643 was loaded using a Famos autosampler (LC Packings) onto the HPLC capillary column. A linear
644 gradient was created to elute peptides with increasing amounts of solvent B (97.5%
645 acetonitrile, 0.1% formic acid). Eluted peptides were subjected to electrospray ionization
646 couple to a mass spectrometer (Thermo Fisher Scientific) to produce tandem mass spectra
647 whose *b* and *y* ion series patterns were compared to Sequest (Thermo Fisher Scientific) to
648 establish protein identity. The sequencing data were filtered to 1 to 2% peptide false-discovery
649 rate. To analyze BioID datasets, two approaches were taken. First, we utilized DESeq2 (version
650 3.11) to provide candidate genes for BirA*-DUX vs. BirA* and BirA*-C3 vs. BirA*-C1. Then, we
651 used SAINTexpress (version 3.5.3) with parameter: (-L6). Formal candidates exhibited > 2
652 Log_2FC and FDR < 0.05 and SAINT score > 0.74, for DESeq2 and SAINTexpress analyses
653 respectively.

654 **Co-immunoprecipitation**

655 1 million mESCs were grown for 24 hours on 100mm dishes and then induced with doxycycline
656 (0.25 ug/mL) for 12 hours. Cells were trypsinized, rinsed once with PBS, and lysis buffer (2 mM
657 Tris-HCl, pH=8.0, 137 mM NaCl, 1% NP-40, 2 mM EDTA, 1x protease inhibitors) was added to
658 lyse cells. Cells then rotated for 30 min at 4°C and were further sonicated for 10 min with a 30-s
659 on-off cycle in a Diagenode Bioruptor Pico device to solubilize chromatin-bound proteins.
660 Protein lysates were quantified with the Bio-Rad (500-0001EDU) Bradford reagent. 3X FLAG
661 Peptide-coupled beads (Sigma Aldrich, F4799-4MG) were pre-washed with lysis buffer, 1 mg of
662 protein was added to 40 μL of pre-washed beads, and samples were incubated at room
663 temperature for 4 hours. Beads were pelleted by low-speed centrifugation. Samples were
664 washed 5X with wash buffer (10 mM TRIS [pH 7.4], 1 mM EDTA, 1 mM EGTA; pH=8.0, 150 mM
665 NaCl, 1% Triton X-100, 0.2 mM sodium orthovanadate, 1x protease inhibitors [see above]),
666 boiled at 95°C in 30 μL 4X Laemelli buffer, and run on an SDS-page gel for western blotting
667 analysis.

668 **ATAC-seq**

669 mESCs were grown for 24 hours and then induced with doxycycline (0.25 ug/mL) for 12 hours.
670 Cells were washed 1X with PBS, pelleted, resuspended, and lysed in cold lysis buffer (10 mM
671 Tris-HCl [pH 7.4], 10 mM NaCl, 3 mM MgCl2, and 0.1% IGEPAL CA-630). Nuclei were pelleted
672 and resuspended in transposase buffer. The Tn5 (Illumina FC-121-1030), and the transposition
673 reaction was carried out for 30 min at 37°C. After purification, the Nextera libraries were
674 amplified for 15 cycles with NEBnext PCR master mix and purified with a Qiagen PCR cleanup
675 kit. All libraries were sequenced on the Illumina HiSeq 6000 platform in 150-bp, paired-end
676 format. Paired-end, raw read files were first processed by Trim Galore (Babraham Institute) to

677 trim low-quality reads and remove adapters. Processed reads were then aligned to mm10 with
678 the following parameters: (--end-to-end –very sensitive –no-mixed).
679
680 **Figure Legends**
681
682 Figure 1: Identification of five repeats within the mouse DUX C-terminal region, and truncation
683 analysis of transcriptional activity.
684 a. (Left) Mouse DUX aligned to itself, revealing five repeats. (Right) Diagram of mouse DUX
685 and human DUX4.
686 b. Amino acid alignment of mouse DUX C-terminal repeats and comparison to human
687 DUX4. A 14 amino acid tail follows both the terminal mouse C5 repeat and the sole
688 DUX4 repeat.
689 c. Phylogenetic analysis showing amino acid alignments of mouse DUX C-terminal repeats
690 vs. the C-terminus of bovine DUXC, the C-terminus of rat DUX4, and the C-terminus of
691 human DUX4. Line length unit is substitutions per base. Purple shading denotes mouse
692 DUX C-terminal repeats.
693 d. (Left) Schematic of constructs used for flow cytometry: mCherry-tagged full-length DUX
694 (FL), homeodomains with C1, C12, C123, C1234, C12345. (Right) Flow cytometry data for
695 MERVL::GFP reporter given mCherry expression in *Dux*-/- mESCs following 18hr
696 expression of indicated constructs. *p-value < 0.05, student's t-test. (n=3 biological
697 replicates)
698

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

700 a. (Top) Schematic of constructs used for flow cytometry: mCherry-tagged FL,
701 ΔC12345+14aa, C1+14aa, C2+14aa, C3+14aa, C4+14aa, and C5+14aa. (Bottom) Flow
702 cytometry for MERVL::GFP reporter given mCherry expression in *Dux*-/- mESCs with
703 18hr overexpression of indicated constructs. *p-value < 0.05, student's t-test. (n=3
704 biological replicates). Constructs with a single black line have the linker $GGGGS_2$ and
705 constructs with two black lines have linkers with $GGGGS_2$ and $GAGAS_2$, respectively.

706 b. (Top) Schematic of constructs used for flow cytometry: mCherry-tagged FL, C3+14aa,
707 C3Δ14aa, C5+14aa, C5Δ14aa. (Bottom) Flow cytometry for MERVL::GFP reporter given
708 mCherry expression in *Dux*-/- mESCs following 18hr expression of indicated constructs.
709 *p-value < 0.05, student's t-test. (n=3 biological replicates)

710 c. (Top) DUX domain chimera construct design for C1-C3 fusions with (top) schematic of
711 constructs and (bottom) specific chimera cut-offs for C1-C3 a,b,d (pink to blue) and C3-
712 C1 a,b,c (blue to pink) constructs. (Bottom) Schematic of DUX domain chimera
713 constructs used for flow cytometry: mCherry-tagged full-length DUX (FL), C1C3a, C1C3b,
714 C1C3c, C3C1a, C3C1b, C3C1c. Symbols adjacent to alignments: red circle denotes
715 transcriptionally inactive repeats and the green circle denotes transcriptionally active
716 repeats. (Right) Flow cytometry for MERVL::GFP reporter given mCherry expression in
717 *Dux*-/- mESCs folllowing 18hr expression of indicated constructs. *p-value < 0.05,
718 student's t-test. (n=3 biological replicates)

719 d. Flow cytometry for MERVL::GFP reporter given mCherry expression in *Dux*-/- mESCs
720 following 18hr expression of point mutation constructs: (Top) FL-DUX, C3+14aa, C3

721 D438E+14aa, C3 D438G+14aa, and C3F443L+14aa. (Bottom) FL-DUX, C1+14aa, C3+14aa,
722 C1DPLELF+14aa (substation of C3 amino acids into C1+14aa), and C3GPLELL+14aa
723 (substitution of C2 and C4 6aa sequence into C3+14aa). *p-value < 0.05, student's t-test.
724 (n=3 biological replicates)

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

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

733

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

735 a. (Left) Schematic of mCherry-tagged DUX domain constructs in clonal *Dux*-/- mESCs used
736 for CUT&TAG experiments. (Right) anti-mCherry Western blot for indicated constructs
737 after 18hr expression. Black arrows indicate construct of interest.

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

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

742 d. H3K9ac CUT&TAG class average map centered at DUX binding sites after 12-hour
743 expression of *Dux* constructs illustrated in Figure 2a (n=2 biological replicates).
744 e. H3K9ac CUT&TAG heatmap centered at DUX binding sites after 12-hour expression of
745 *Dux* constructs illustrated in Figure 2a (n=2 biological replicates).
746
747 Figure 4: BioID for full-length DUX reveals interaction with proteins involved in chromatin de-
748 repression.
749 a. Experimental concept for BioID (proximity labeling assay) for BirA*-DUX to identify
750 candidate DUX interacting proteins.
751 b. Experimental set-up for full-length DUX BioID. The control construct, BirA*-only,
752 transfected into 2C-like mESCs, and BirA*-DUX, transfected into mESCs (n=2 biological
753 replicates).
754 c. Volcano plot using DeSeq2 analysis for BirA*-only vs. BirA*-DUX with log2FoldChange on
755 the x-axis and -log(padj) on the y-axis (n=2 biological replicates).
756 d. Interactome map of proteins enriched to interact with full-length DUX, vs. BirA*-only,
757 split into specific cellular components and protein complexes (n=2 biological replicates).
758 Full dataset available in Supplementary Table 1.
759 e. (Left) Experimental set-up for BioID-like experiment in HEK283T cells with BirA*DUX and
760 flag-tagged candidates – SMARCC1 and ZSCAN4D. (Right) Co-immunoprecipitation of
761 FLAG and Western blot after 18-hour expression of transiently transfected FLAG-
762 SMARCC1 or FLAG-ZSCAN4D and BirA*-DUX in HEK293Ts (n=3 biological replicates).
763

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

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

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

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

773 d. Co-immunoprecipitation of endogenous SMARCC1 after 12hr overexpression of DUX
774 domain constructs in stable clonal cell lines illustration in Figure 2a (n=3 biological
775 replicates).

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

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

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

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

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

786

787 **Supplementary Figure Legends**

788

789 Supplementary Figure 1**:** DUX orthologs do not contain a repeat structure similar to mouse

790 DUX.

791 a) Human DUX4 aligned to itself.

792 b) Rat DUX4 aligned to itself.

793 c) Bovine DUXC aligned to itself.

794 d) Table with pairwise comparison of percent identity comparing mouse DUX C-terminal

795 repeats, the human DUX4 C-terminus, the rat DUX4 C-terminus, and the bovine DUXC C-

796 terminus.

797 e) Experimental design for expression of mCherry-tagged DUX domain constructs and flow

798 cytometry of GFP% | mCherry tag expression

799 f) Western blot for mCherry tag after 18hr expression of C1, C12, C123, C1234, C12345

800

801 Supplementary Figure 2: DUX homeodomains are required for transcriptional activity.

802 a) (Left) Schematic of constructs used for flow cytometry: mCherry-tagged full-length DUX

803 (FL), ΔC12345+14aa, C12345Δ14aa, C1+14aa, C12+14aa, C123+14aa, C1234+14aa,

804 C5+14aa, C45+14aa, C345+14aa, C2345+14aa, ΔHD1, ΔHD2, and *Dux*-/- alone. (Right)

805 Flow cytometry for MERVL::GFP reporter given mCherry expression in *Dux*-/- mESCs

806 following 18hr expression of indicated constructs. *p-value < 0.05, student's t-test. (n=3

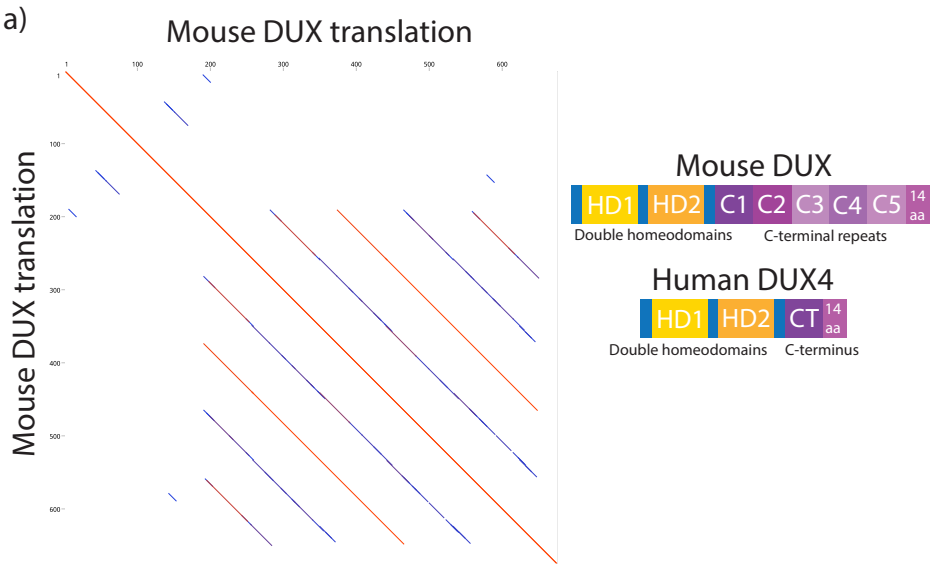
807 biological replicates)

808 b) Schematic of C1C3 chimera DUX constructs for amino acid cut offs for C1C3a, C3C3b,
809 C1C3c, C3C1a, C3C1b, and C3C1c aligned to the 5 C-terminal repeats
810 c) Heat map of RNA-seq at DUX target genes (n=456) (22) from 18hr expression of DUX
811 domain constructs of FL, C12345Δ14aa, C12345+14aa, C12345Δ14aa(n=2). Colored
812 rectangles above samples denote inability to activate MERVL::GFP reporter (red) and
813 ability to activate MERVL::GFP reporter (green).
814 d) Phylogenetic analysis of the five mouse DUX C-terminal repeats. Line length unit is
815 substitutions per base.
816
817 Supplementary Figure 3: DUX CUT&Tag data strongly resembles DUX ChIP-seq, and DUX binding
818 sites gain H3K9ac upon DUX expression.
819 a. Profile plot of mCherry-FL DUX CUT&TAG centered at DUX binding sites after 12 hour
820 expression vs. HA-DUX ChIP-seq (n=2 biological replicates) (12)
821 b. IGV screenshots at TCSTV3, ZSCAN4F, USP17LB, and ZSCAN4C (DUX targets genes) for
822 H3K9ac CUT&Tag in mESCs and mCherry-DUX 12hr expression and mCherry CUT&TAG
823 for mCherry tag alone and mCherry-DUX (n=2 biological replicates)
824
825 Supplementary Figure 4: The proximity-labeling assay BioID identifies DUX protein interactors.
826 a) BioID DeSeq2 analysis replicate structure of BirA*-only and BirA*-DUX duplicates for
827 mass spectrometry datasets.
828 b) Venn diagram of statistically significant enriched proteins for BirA*-DUX interaction for
829 DeSeq2 and SAINT*express*

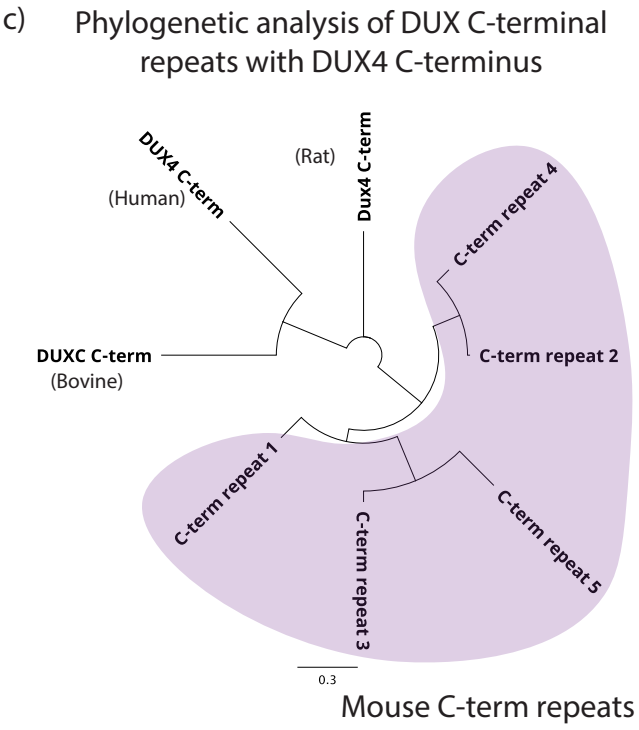
830 c) Streptavidin pulldown and Western blot after 18-hour BioID induction in clonal stable
831 BirA*-DUX mESC line expressing FLAG-tagged SMARCC1, CHAF1A, ZSCAN4D, or KDM4C
832 with blots for DUX and FLAG-tag (n=3 biological replicates). KDM4C was a BioID hit, but
833 does not interact with DUX, indicating a possible false positive.
834 d) Streptavidin pulldown and Western blot after 18-hour BioID induction in clonal stable
835 BirA*-DUX mESC line, with blot for N-terminal myc-BirA*-DUX, P300, and Avidin-HRP
836 (n=2 biological replicates)
837 e) BioID individual repeat DeSeq2 analysis replicate structure of BirA*-C1 and BirA*-C3
838 triplicates for mass spectrometry (n=3 biological replicates)
839 f) Principle component analysis of DeSeq2 analysis for BirA*mC-C1+14aa and BirA*mC-
840 C3+14aa with PC1: 54% variance and PC2: 17% variance
841

Figure 1

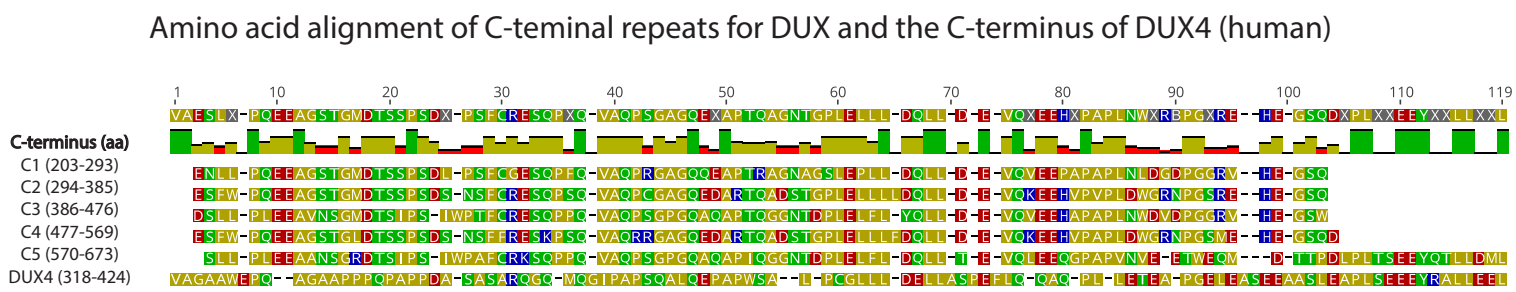
a)



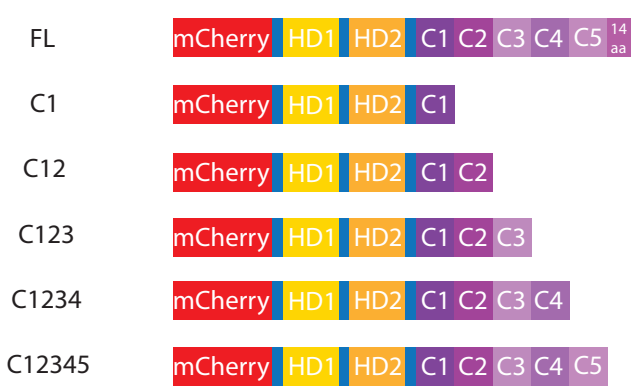
c)



b)



d)



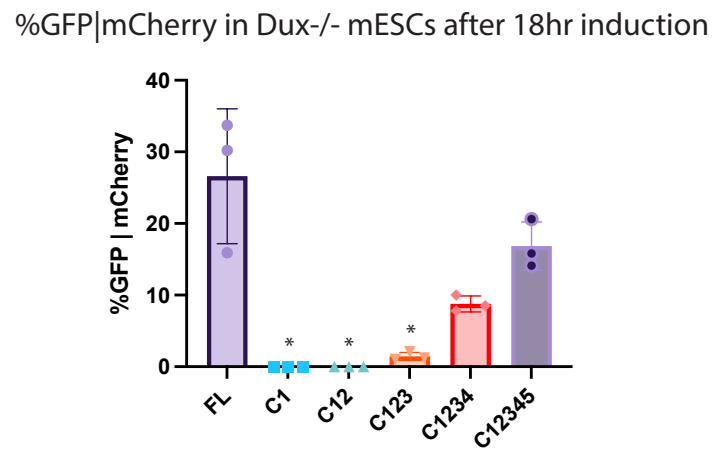


Figure 2

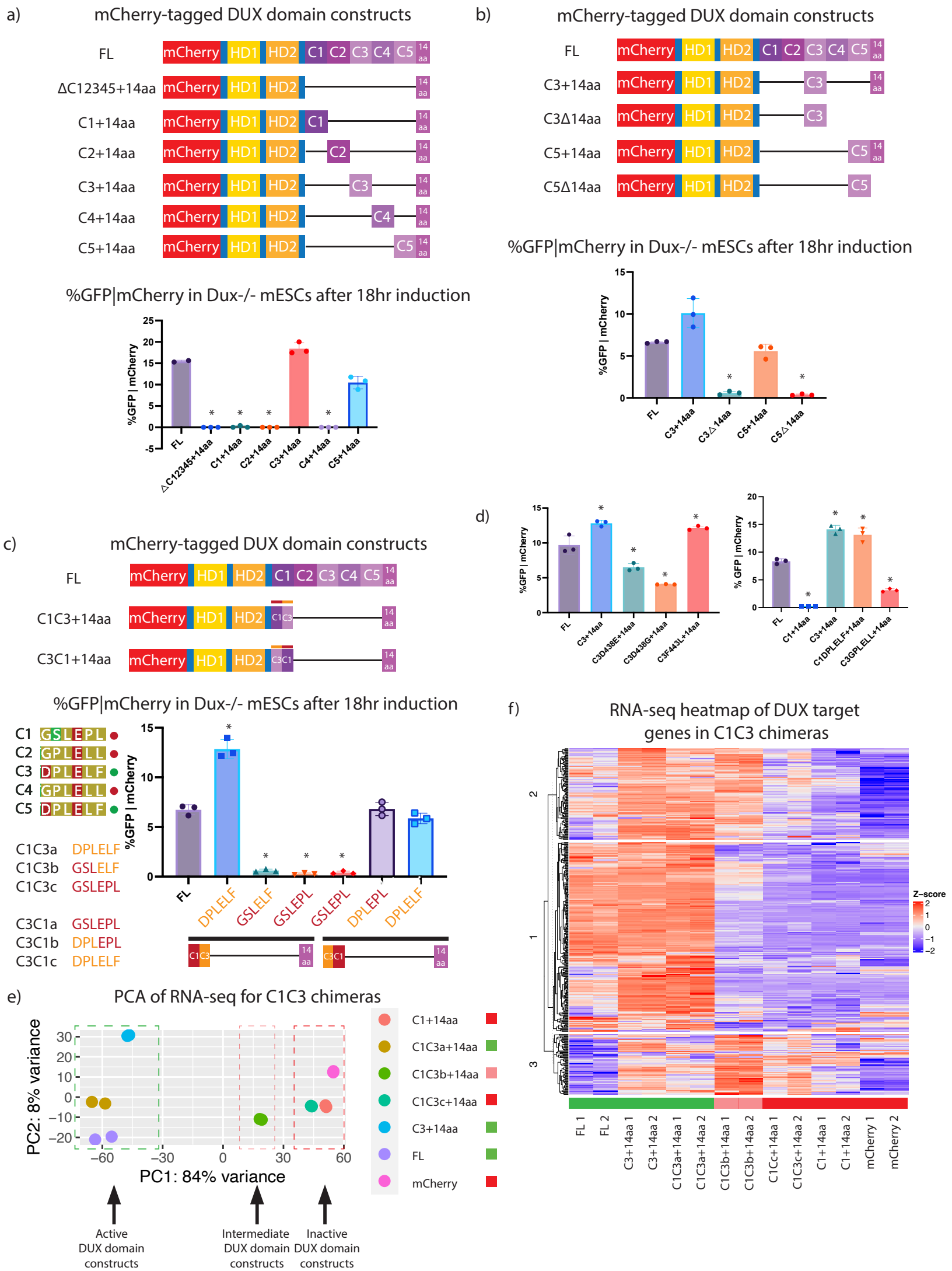
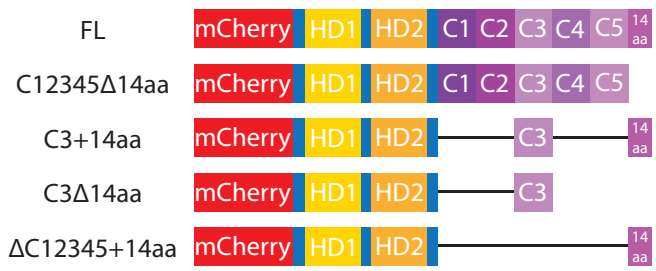
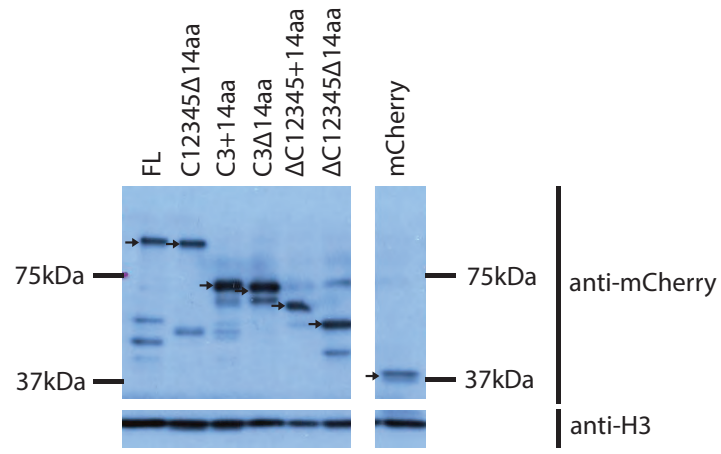

a)
mCherry-tagged DUX domain constructs
FL
ΔC12345+14aa
C1+14aa
C2+14aa
C3+14aa
C4+14aa
C5+14aa
mCherry
HD1
HD2
C1
C2
C3
C4
C5
14 aa
%GFP|mCherry in Dux-/- mESCs after 18hr induction
%GFP | mCherry
b)
mCherry-tagged DUX domain constructs
C3+14aa
C3Δ14aa
C5+14aa
C5Δ14aa
%GFP|mCherry in Dux-/- mESCs after 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

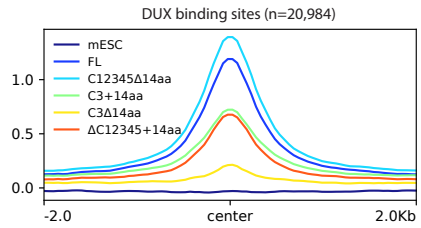
a)

DUX domain constructs in stable clonal Dux-/- mESCs

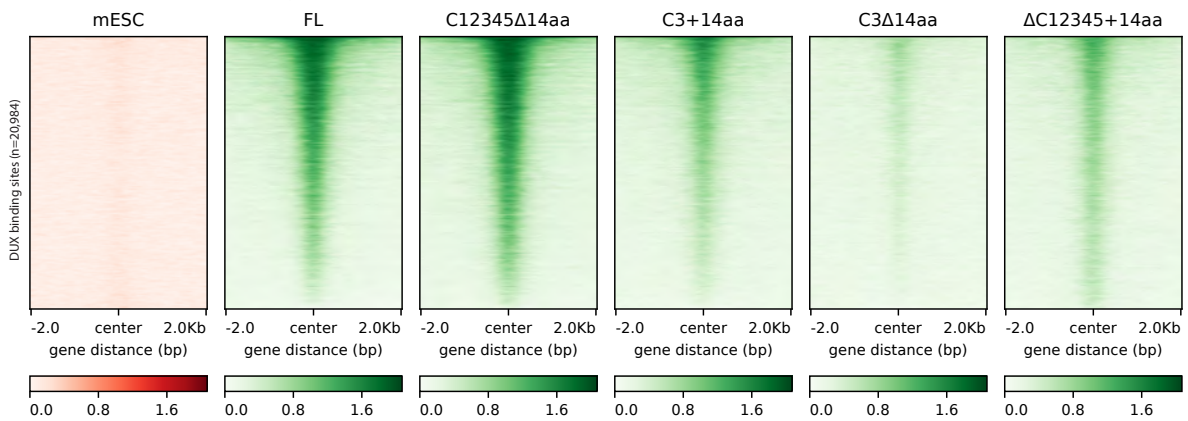




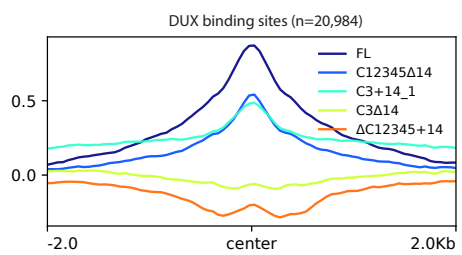
b) mCherry CUT&Tag of Dux domain constructs



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

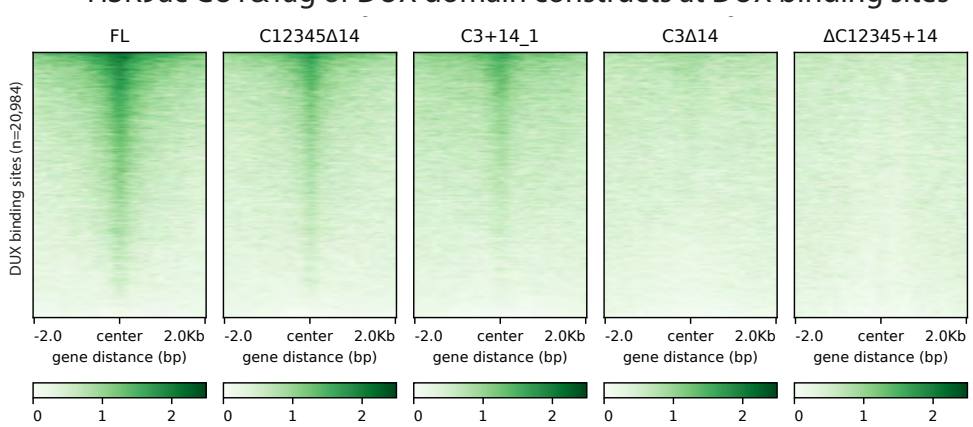
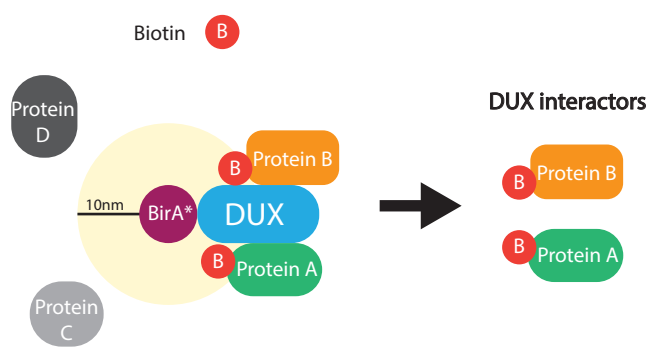
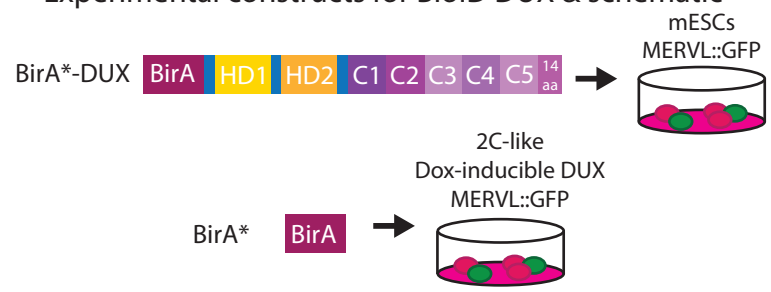


Figure 4

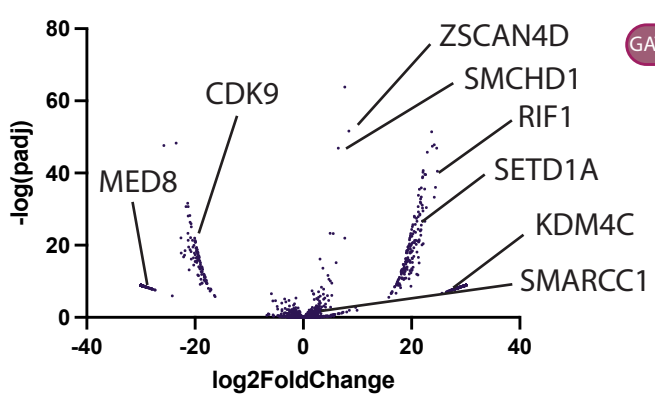
a) BioID schematic



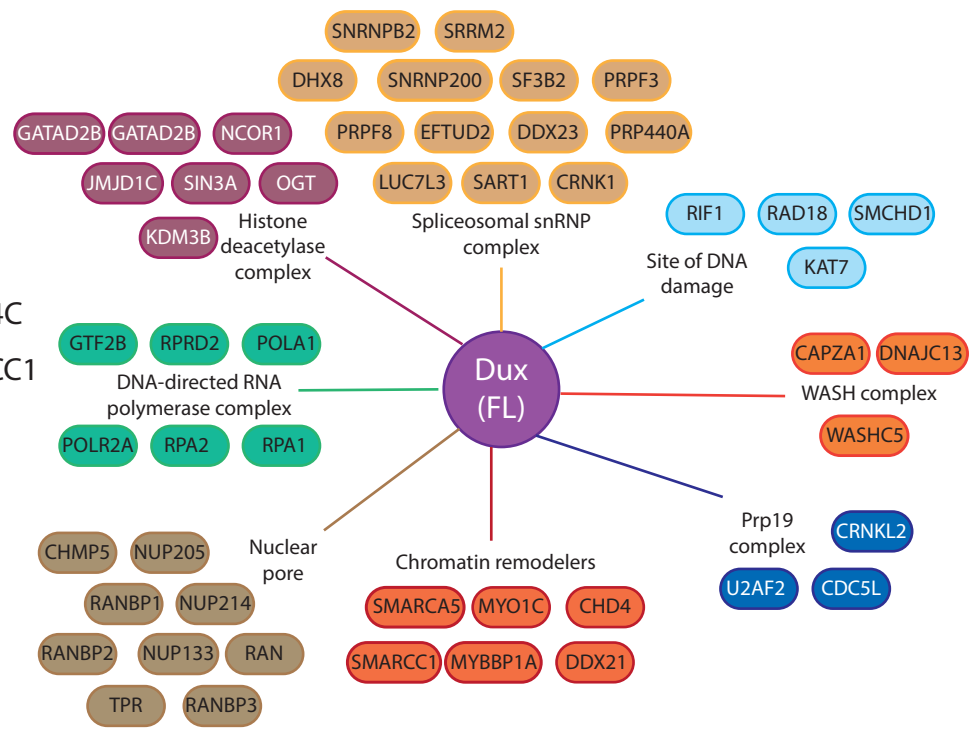
b) Experimental constructs for BioID DUX & schematic



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

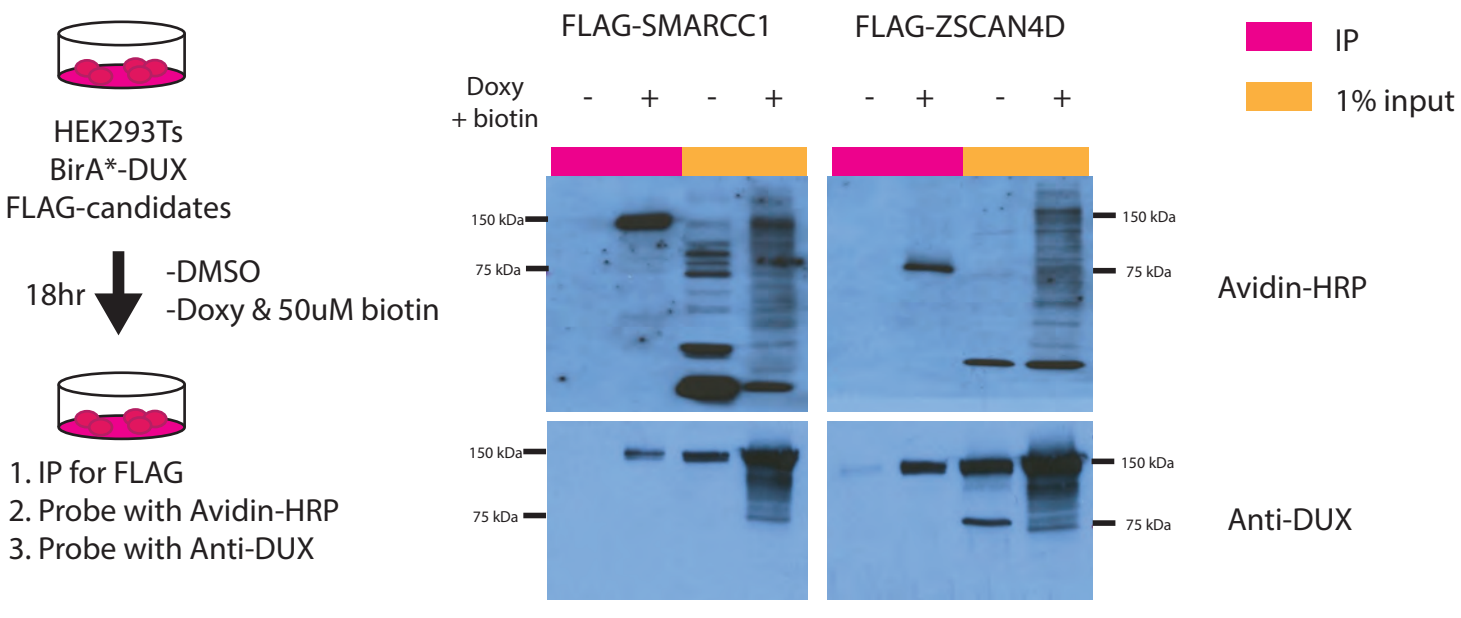
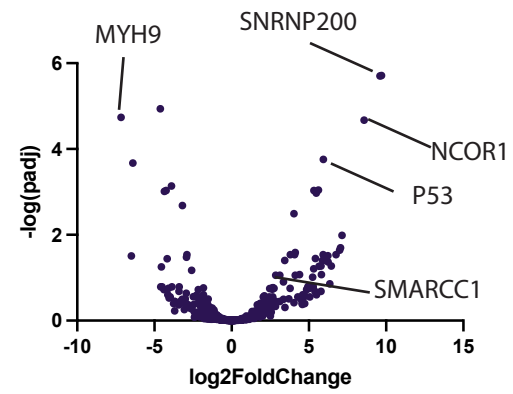


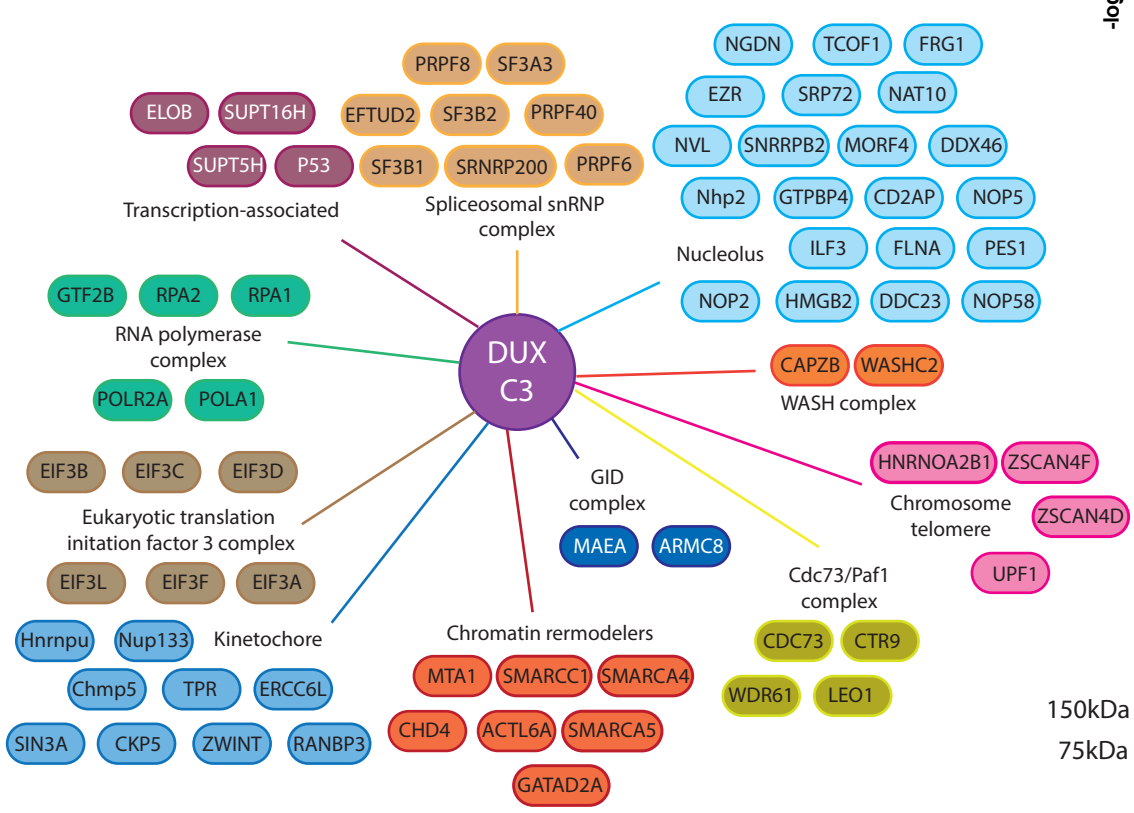
Figure 5

a) Experimental constructs for BioID C1C3

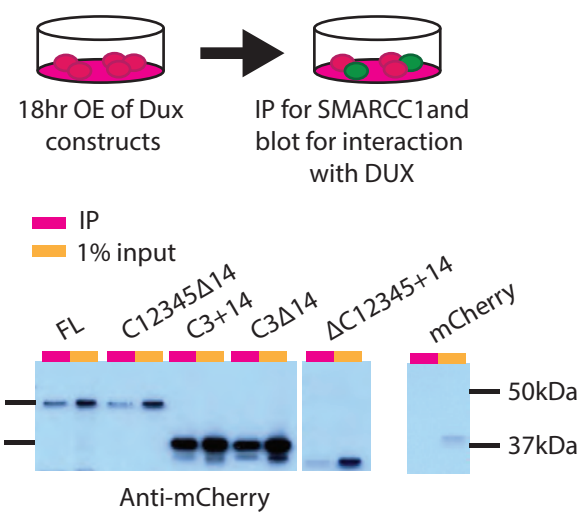
b) Volcano plot for BirA*-C3 vs. BirA*-C1



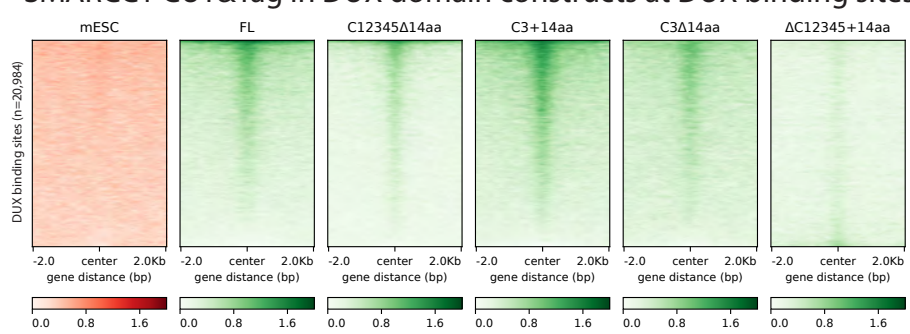
c) Interactome of BirA*-C3 in comparison to BirA*-C1



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

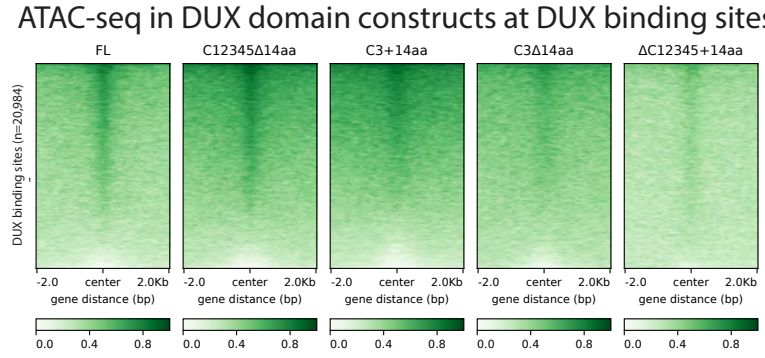


Figure 6

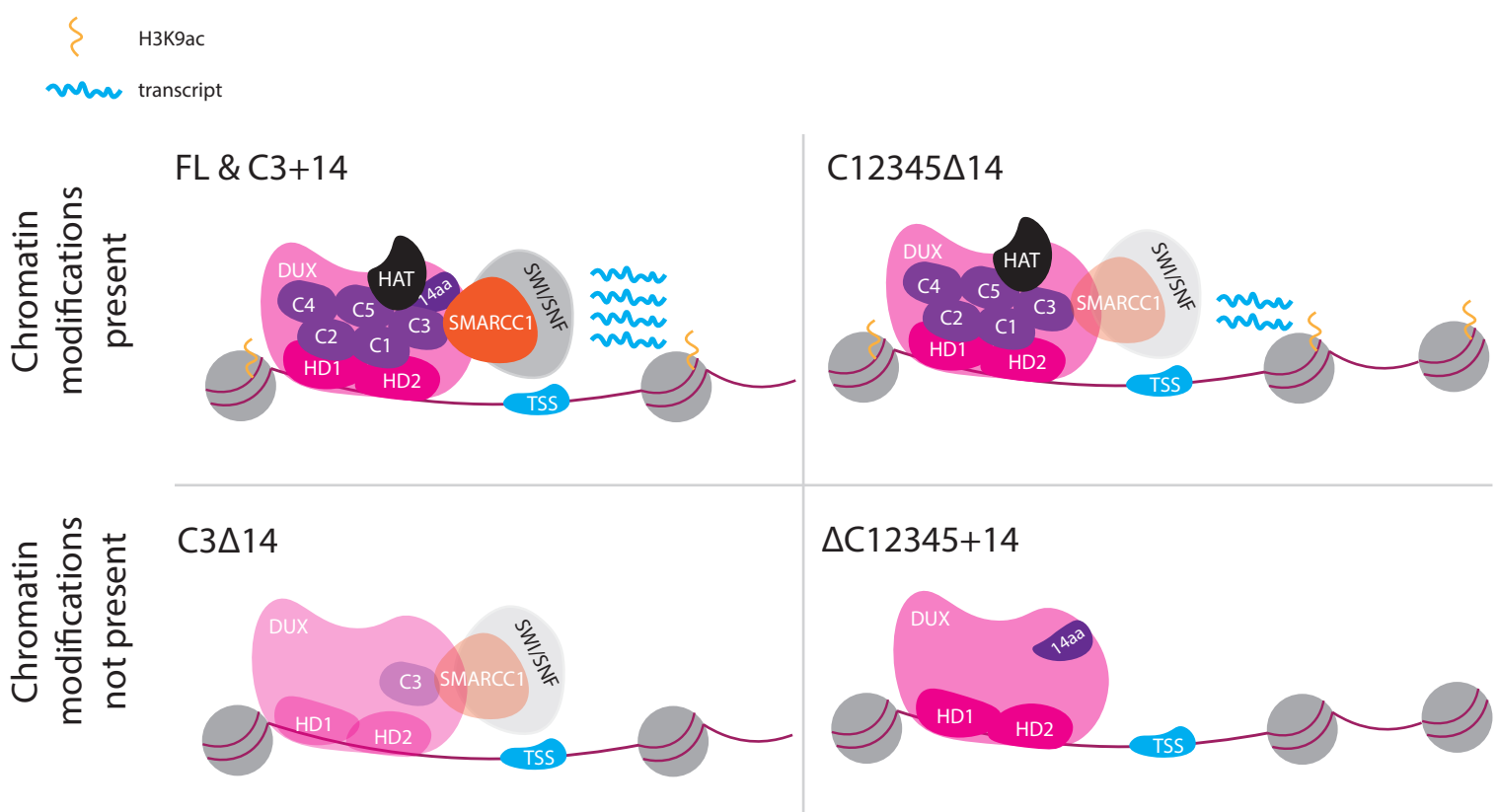
a)

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	++	++	++	+	+

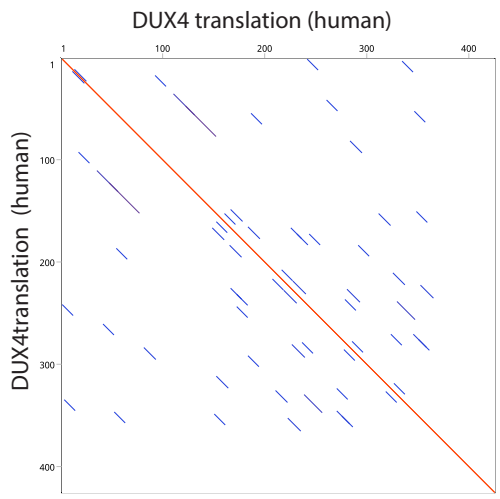
b)

Model for DUX mechanism of target gene activation

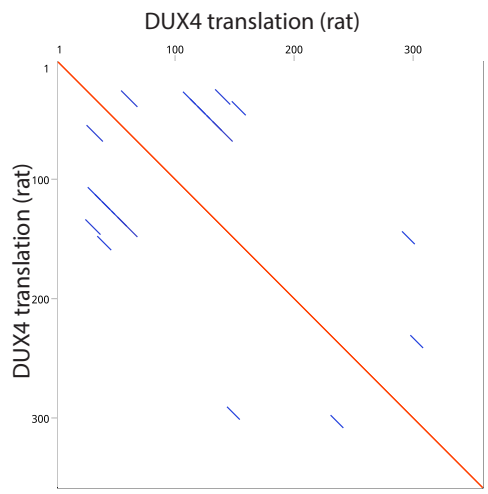


Supplementary Figure 1

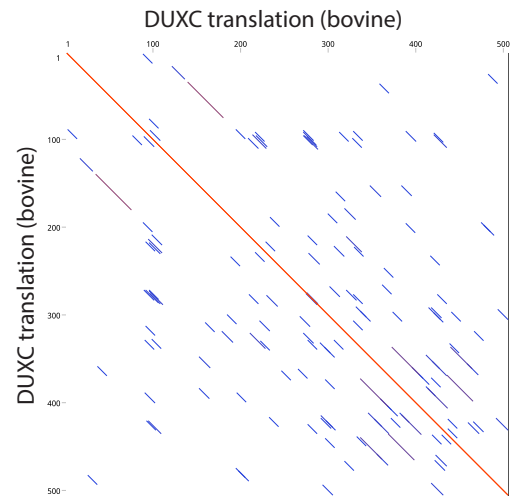
a)



b)



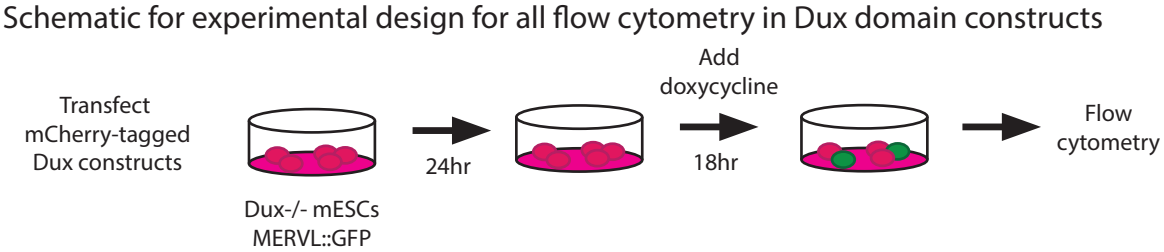
c)



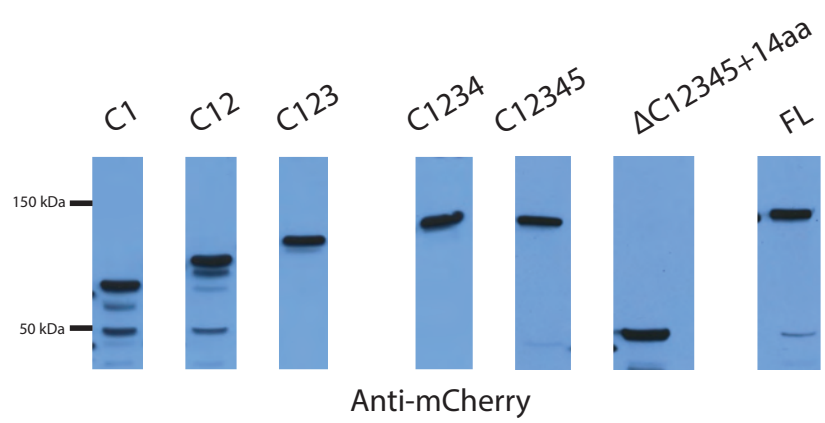
d) Table comparing percent identity for mouse C-terminal repeats and C-termini of DUX orthologs

	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		68.5	69.2	64.1	51.1	13.6	8.6	9.3
C-term repeat 2	68.5		62.0	91.3	48.4	14.6	8.2	10.0
C-term repeat 3	69.2	62.0		56.5	73.3	13.6	8.2	8.6
C-term repeat 4	64.1	91.3	56.5		44.6	13.5	7.8	11.3
C-term repeat 5	51.1	48.4	73.3	44.6		11.6	8.8	6.5
DUX4 C-term (H)	13.6	14.6	13.6	13.5	11.6		28.8	9.8
DUXC C-term (B)	8.6	8.2	8.2	7.8	8.8	28.8		8.9
DUX4 C-term (R)	9.3	10.0	8.6	11.3	6.5	9.8	8.9	

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

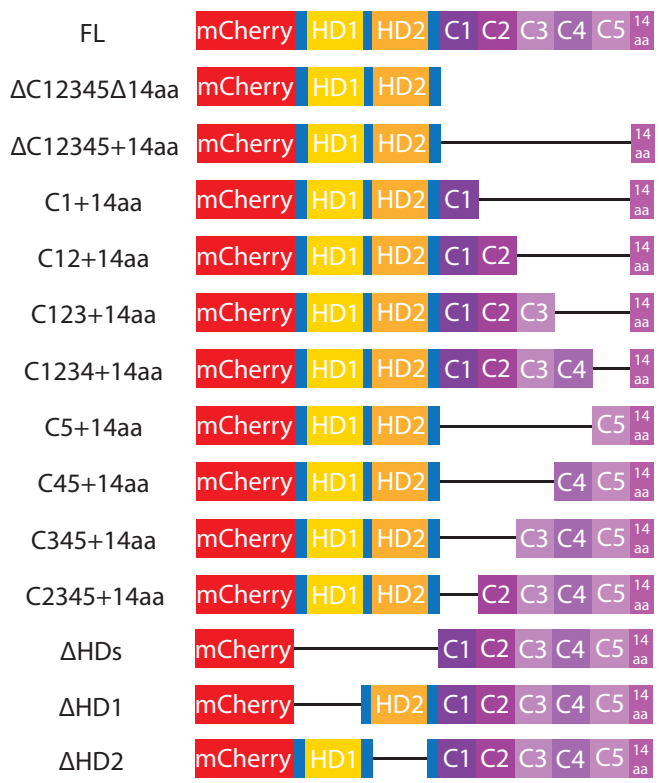


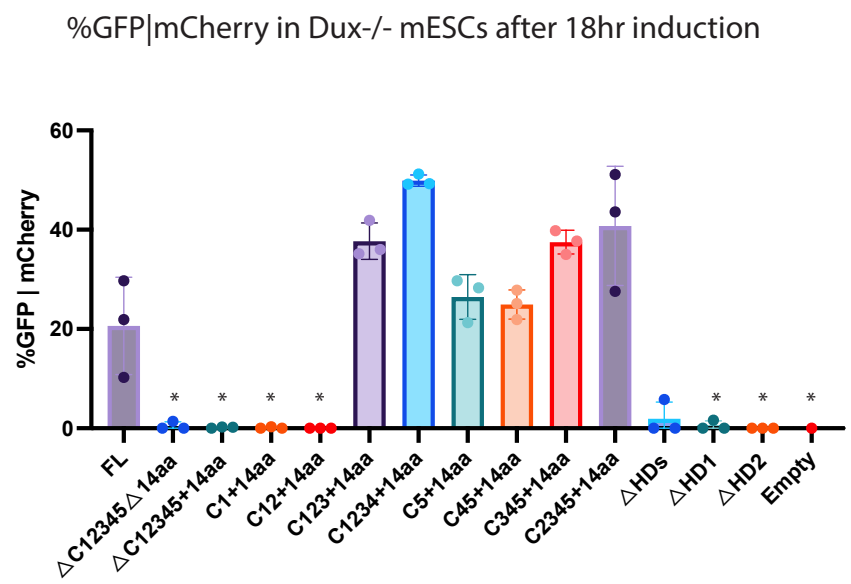
f) Western blot for mCherry-tagged Dux domain constructs



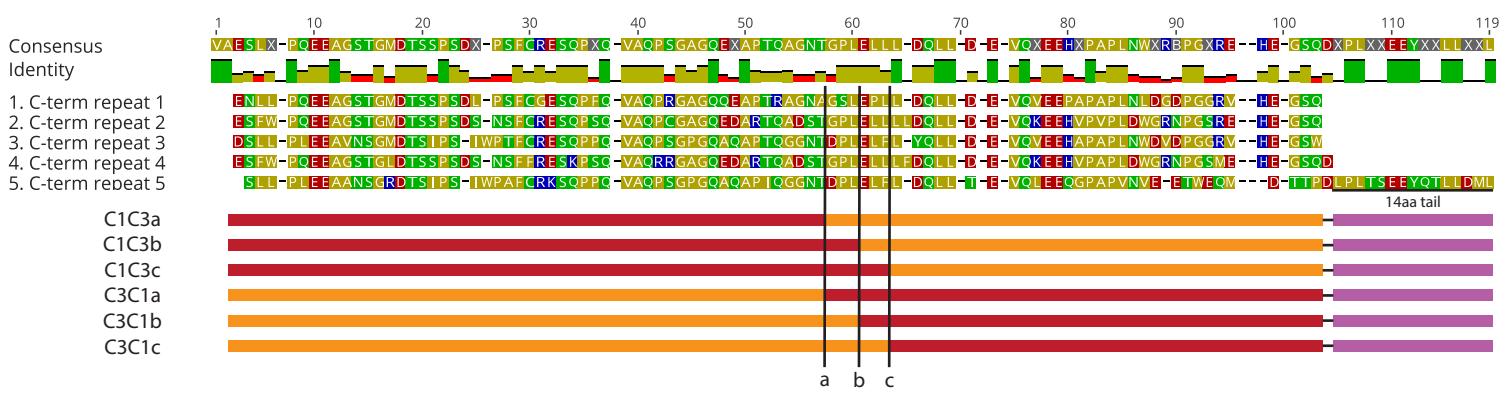
Supplementary Figure 2

a) mCherry-tagged DUX domain constructs

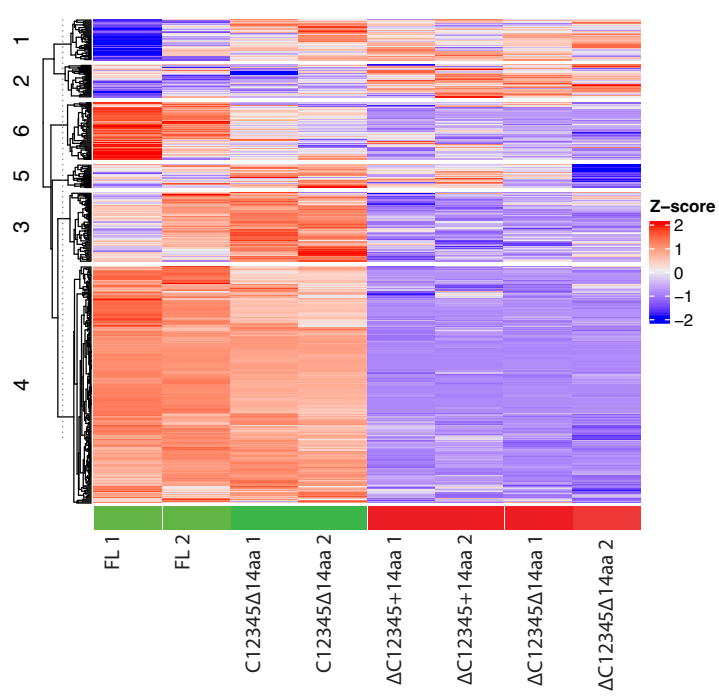




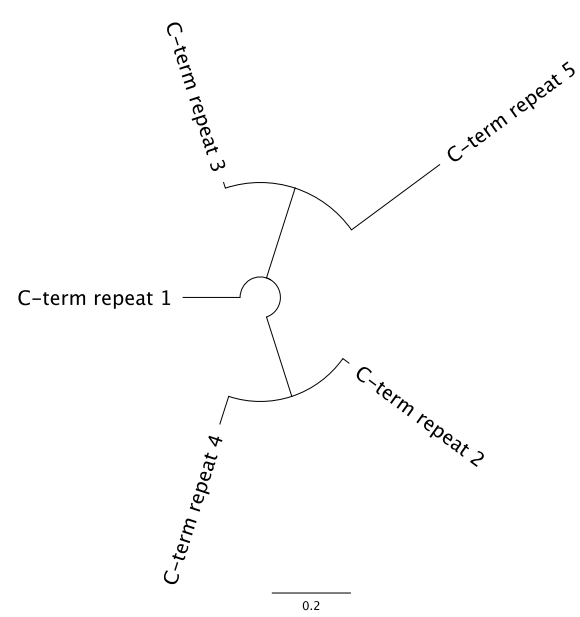
b) C1C3 chimera cutoff map aligned to mouse C-terminal repeat alignment



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

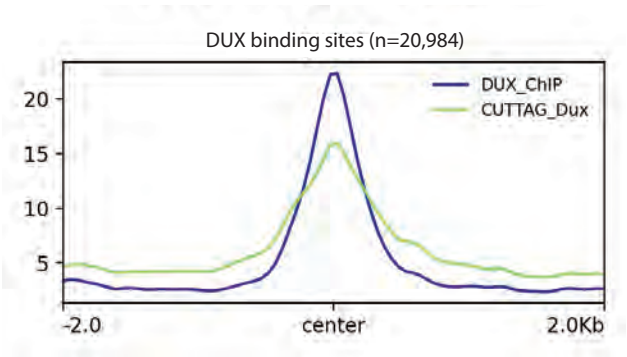


d) Phylogenetic analysis for mouse C-terminal repeats

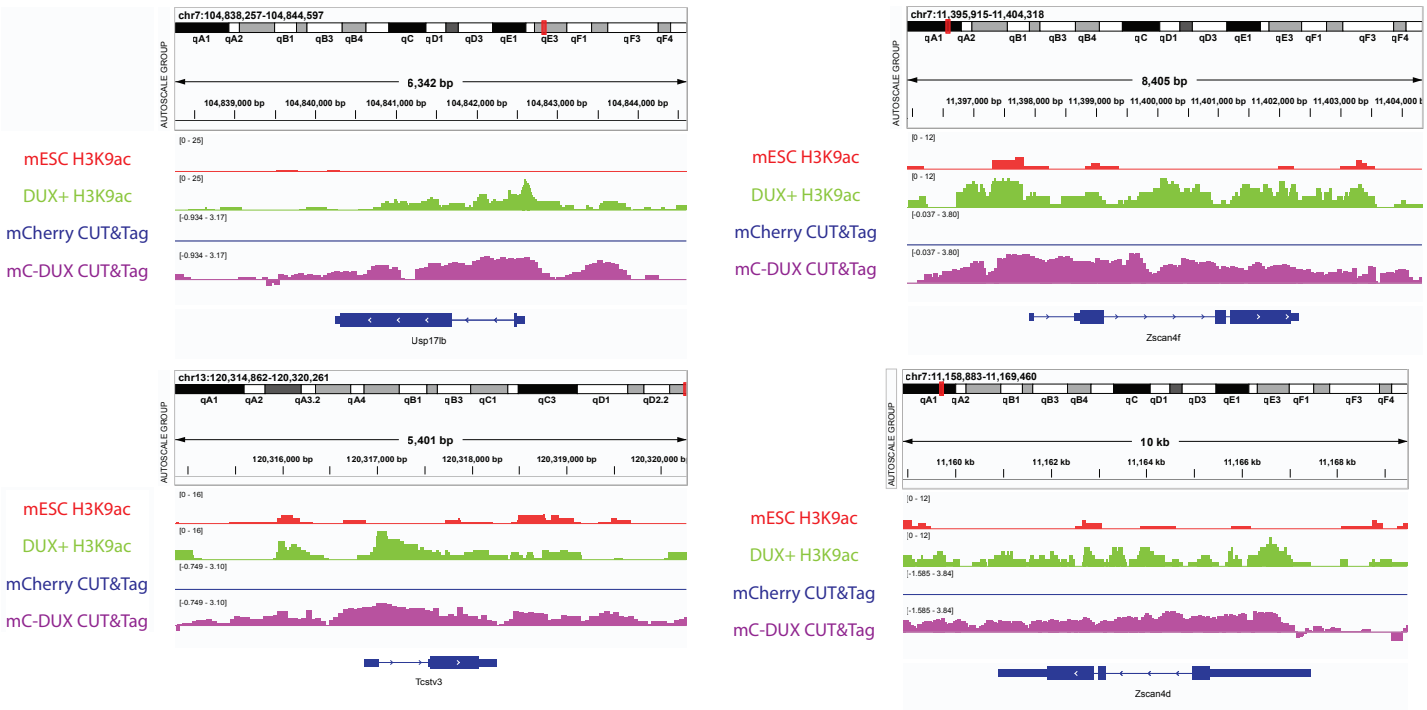


Supplementary Figure 3

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

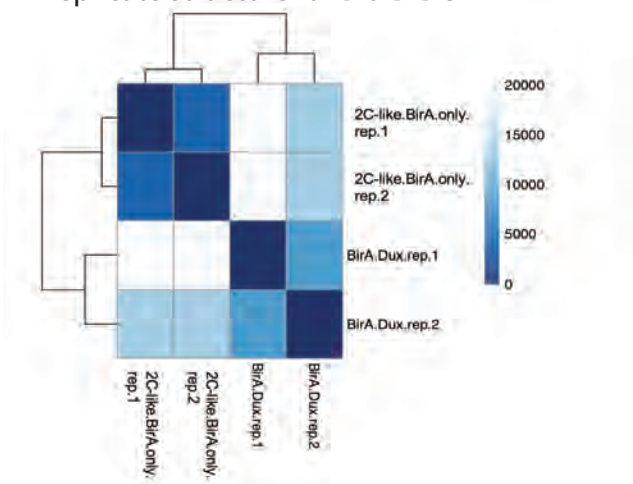


b) IGV screenshots of mESC H3K9ac and DUX-expressing cells H3K9ac at DUX target genes

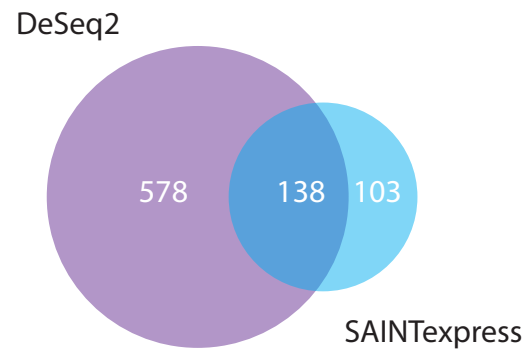


Supplementary Figure 4

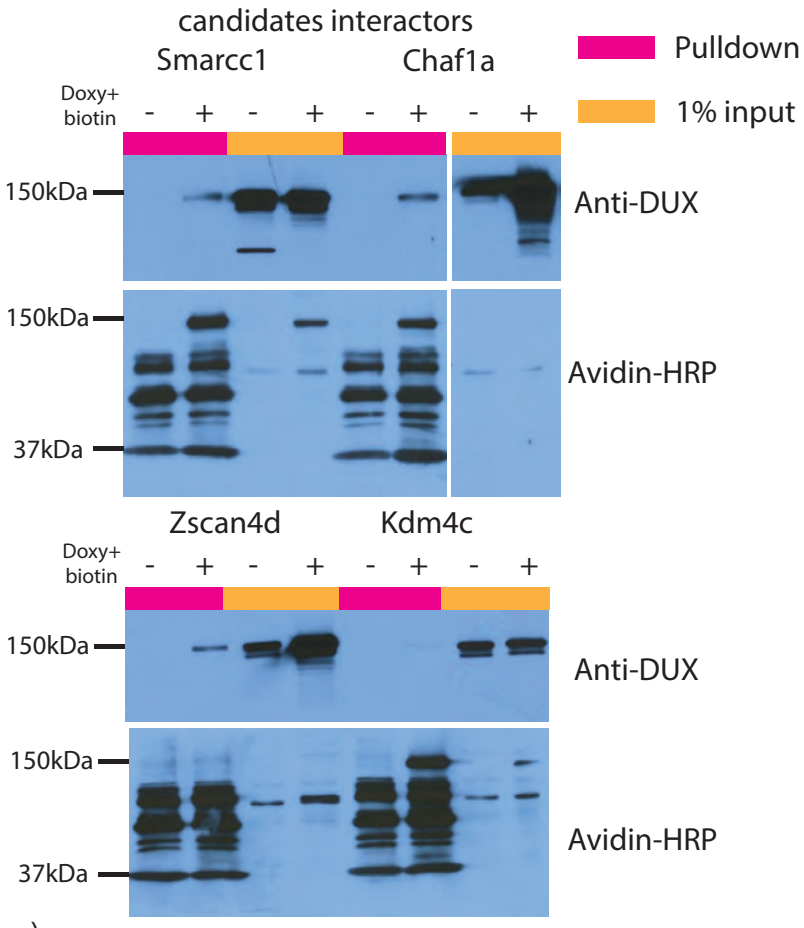
a) Replicate structure for BioID DUX



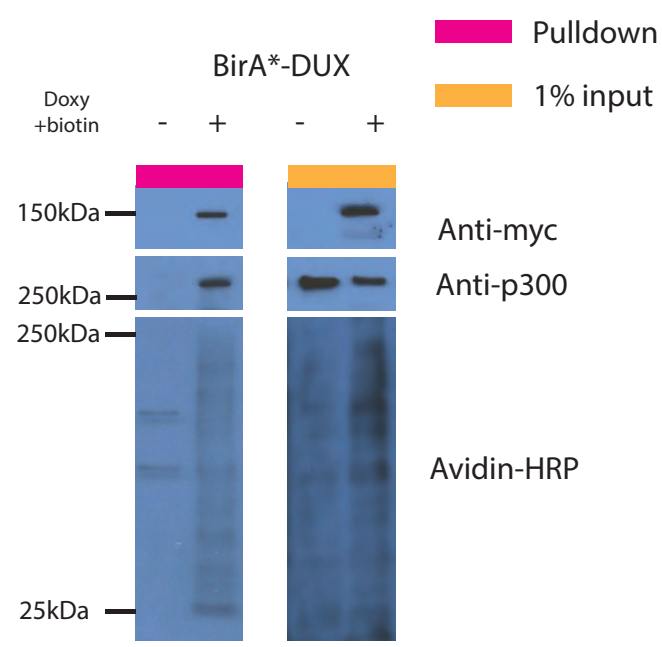
b) Number of statistically significant proteins in two different BioID DUX analysis packages



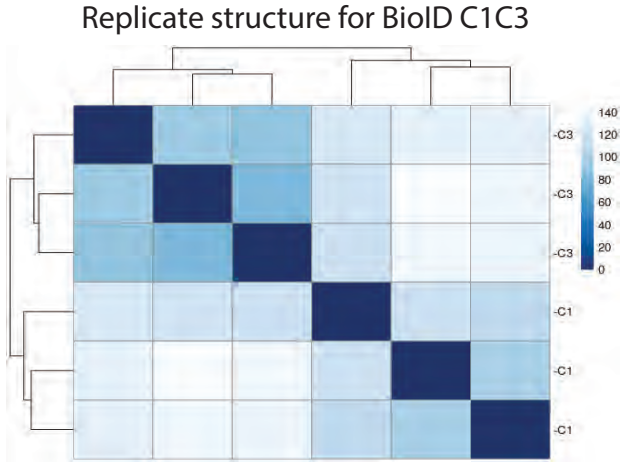
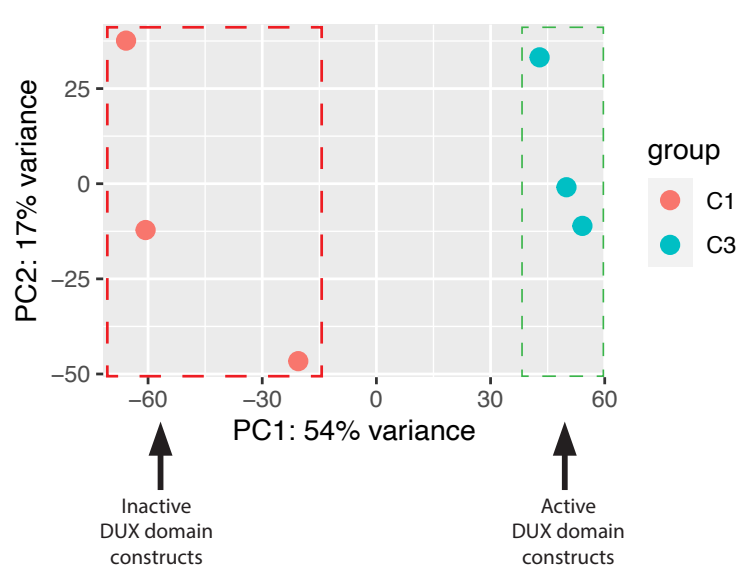
c) BioID experiment with FLAG-tagged candidates interactors



d) BioID Dux experiment with p300 Western blot



e) Replicate structure for BioID C1C3



f) Principle component analysis for BioID C1C3