# *Obox4* secures zygotic genome activation upon loss of *Dux*

Youjia Guo[#1], Tomohiro Kitano[#1], Kimiko Inoue[#2,3], Kensaku Murano[1], Michiko Hirose[2], Ten D. Li[1], Akihiko Sakashita[1], Hirotsugu Ishizu[1], Narumi Ogonuki[2,3], Shogo Matoba[2,3], Masayuki Sato[1], Atsuo Ogura[*2,3], Haruhiko Siomi[*1]

[1]Department of Molecular Biology, Keio University School of Medicine, Shinjuku, Tokyo 160-8582, Japan;
[2]Bioresource Engineering Division, Bioresource Center, RIKEN, Tsukuba, Ibaraki 305-0074, Japan;
[3]Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan

[#]These authors contributed equally: Youjia Guo, Tomohiro Kitano, and Kimiko Inoue.

[*]Corresponding author:
Haruhiko Siomi awa403@keio.jp
Atsuo Ogura ogura@rtc.riken.go.jp

## Abstract

Once fertilized, mouse zygotes rapidly proceed to zygotic genome activation (ZGA), during which long terminal repeats (LTRs) of murine endogenous retroviruses with leucine tRNA primer (MERVL) are potently activated by a conserved homeodomain-containing transcription factor, DUX. However, *Dux*-knockout embryos produce fertile mice, suggesting that ZGA is redundantly driven by an unknown factor(s). In the present study, we present multiple lines of evidence that the multicopy mouse homeobox gene, *Obox4*, encodes a transcription factor that is highly expressed in mouse 2-cell embryos and redundantly drives ZGA. Single depletion of *Obox4* or *Dux* is tolerated by embryogenesis, whereas *Obox4*/*Dux* double depletion completely compromises embryonic development. Genome-wide epigenetic profiling revealed that OBOX4 binds to MERVL LTRs as well as a subset of murine endogenous retroviruses with lysine tRNA primer (MERVK) LTRs, and mediates the deposition of active histone modifications. Our study identified OBOX4 as a transcription factor that provides genetic redundancy to pre-implantation development.

## Introduction

The mechanism by which zygotes acquire totipotency is a major question in Developmental Biology. Following fertilization, the zygote must build a *de novo* conceptus with a transcriptionally quiescent genome. The genome rapidly undergoes zygotic genome activation (ZGA), during which epigenetic reprogramming and expression of nascent transcripts enforce the replacement of parental infrastructures by their zygotic counterparts [1]. Fulfillment of metazoan ZGA results in blastomeres, a collection of cells that are totipotent enough to reflect their potential to individually produce both embryos and extraembryonic appendages [2,3].

Totipotency lasts until the morula stage, where the first cell fate decision takes place, following which it is remolded into pluripotency in cells destined for the embryonic lineage [4], to which embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) correspond. In placental mammals, ZGA is characterized by the massive reactivation of transposable elements (TEs) and epigenome remodeling, predominantly implemented by a collection of ZGA genes that have adapted long terminal repeats (LTRs) of endogenous retroviruses (ERVs) as stage-specific *cis*-regulatory elements [5].

Homeodomains are DNA-binding amino acid motifs encoded by a class of conserved genomic sequences termed homeoboxes [6-8]. Homeodomains are prevalent transcription factors that regulate development because of their ability to bind chromatin in a sequence-specific manner [9-11]. Among all homeobox-containing genes, Paired-like (PRD-like) homeobox class genes are particularly associated with ZGA [12]. Studies have shown that the PRD-like homeobox gene, double homeobox (*Dux*), plays a critical role in pre-implantation embryogenesis [13,14]. *Dux* is conserved throughout placentalia and is highly expressed during ZGA. In humans and mice, DUX specifically binds and activates endogenous retroviruses with leucine tRNA primer (ERVL) LTR-derived promoters, resulting in the expression of downstream ZGA genes [15-17]. *Dux* expression induces a 2-cell-embryo-like (2C-like) state in mouse embryonic stem cells (mESCs) [15-17], whereas *Dux*-knockout (KO) undermines mouse blastocyst formation [15]. These observations have led to the assumption that *Dux* resides at the top of the eutherian ZGA regulatory hierarchy [16]. However, subsequent studies have revealed that mouse embryogenesis is compatible with the loss of *Dux*, suggesting a redundant pathway of ZGA that is controlled by uncharacterized transcription factor(s) [18-21].

In this study, we sought to identify the redundant transcription factor that drives ZGA in the absence of *Dux*. We discovered that the multicopy homeobox gene, oocyte-specific homeobox 4 (*Obox4*), is essential for *Dux*-less ZGA. *Obox4* is abundantly expressed in mouse 2C-embryos, 2C-like mESCs, and totipotent blastomere-like cells (TBLCs). *Obox4* and *Dux* are collectively, but not respectively, essential for ZGA. Mechanistically, OBOX4 promotes ZGA by binding to the LTRs of murine endogenous retroviruses with leucine tRNA primer (MERVL) and murine endogenous retroviruses with lysine tRNA primer (MERVK), and thereby affecting the deposition of active epigenetic modifications. Taken together, our findings substantiate a pre-implantation development model, in which the fulfillment of ZGA is redundantly secured by *Dux* and *Obox4*.

## Results

### Expression profiling identifies transcription factor candidates

Genetic redundancy is commonly observed among genes that share sequence homology [22]. Because *Dux* is a homeobox gene, we speculated that its redundant factors are homeobox genes. We examined the expression profiles of all mouse homeobox genes during pre-implantation embryogenesis using published single-cell RNA-seq (scRNA-seq) data [23]. Clustering analysis defined a collection of 45 homeobox genes, whose transcript levels reached a maximum before the late 2C stage (Fig. 1a). We then sought to identify *bona fide* ZGA factors from this collection by looking for genes whose transcripts were of zygotic origin, instead of ones that were parentally inherited. TBLCs have recently been established *as in vitro* counterparts of 2C-embryos, which are

75 derived from splicing-inhibited mESCs, and hence, are free of parentally inherited transcripts [24]. Analysis of
76 published RNA sequencing (RNA-seq) data showed that the expression of four homeobox genes, *Duxf3*, *Emx2*,
77 *Hoxd13*, and *Obox4*, was ten times higher in TBLCs than in ESCs (Fig. 1b). We examined whether the
78 expression of these genes was affected in the *Dux* knockout embryos. Using published RNA-seq data [18,20], we
79 found that the expression of all single-copy candidate genes and a subset of the multicopy gene *Obox4* was
80 unaffected in *Dux* knockout embryos (Fig. 1c).
81
82 **_Obox4_ potently activates 2C-genes and TEs in mESC**
83 Mouse ZGA is characterized by the surging activity of genes that specifically express at 2C stage (2C-gene),
84 many of which have co-opted MERVL LTRs as stage-specific promoters [25]. We then sought to examined
85 whether the candidates had potential to promote ZGA by activating MERVL and 2C-genes (Supplementary Fig.
86 1a and Supplementary Table 1-2). We constructed a 2C::tdTomato reporter mESC line with transgenic
87 tdTomato red fluorescence protein driven by MERVL 5'-LTR. [26]. Candidate genes were cloned and
88 overexpressed in the reporter cell line (Fig. 1d and Supplementary Fig. 1b). Interestingly, ectopic expression of
89 *Obox4* markedly induced tdTomato expression, similar to that induced by *Dux* (Fig. 1e and Supplementary Fig.
90 1c-d). To characterize the impact of *Obox4* expression on 2C-genes, we established an ESC stable line bearing
91 an inducible *Obox4* transgene (Supplementary Fig. 1e). Differentially expressed gene analysis revealed that
92 *Obox4* induction for 12 hours and 48 hours led to drastic transcriptome changes in mESCs, characterized by
93 upregulation of 2C-genes and TEs that were highly expressed in early-to-middle 2C-embryos [27], naturally
94 occurred 2C-like mESCs [28], and *Dux* induced 2C-like mESCs [16] (Fig. 1f-j and Supplementary Fig. 1f-g ). These
95 results strongly suggest that *Obox4* is a potent inducer of 2C-like genes.
96
97 **Simultaneous depletion of _Obox4/Dux_ compromises pre-implantation development**
98 We then examined whether *Obox4* was crucial for ZGA, particularly in the absence of *Dux*. First, we
99 confirmed the presence of endogenous OBOX4 protein, as the *Obox4* loci are marked as pseudogenes in the
100 current genome annotation GRCm39 [29]. We generated mouse monoclonal antibodies that recognized OBOX4.
101 Immunofluorescence staining with anti-OBOX4 monoclonal antibodies confirmed that endogenous OBOX4 was
102 expressed in zygotes and highly abundant in 2C-embryos and 2C-like mESCs, which is consistent with
103 translatome profile captured by Ribo-seq that detects OBOX4 translation at PN5 stage [30] (Fig. 2a-c). To
104 determine whether *Obox4* is required for pre-implantation development, microinjection of antisense
105 oligonucleotide (ASO) into male pronuclei of zygotes was performed to knockdown *Obox4* and *Dux* (Fig. 2d).
106 At the transcript level, both *Obox4* and *Dux* knockdown reduced *Obox4* expression in 2C-embryos. At the
107 protein level, ASO targets were depleted in knockdown 2C-embryos, whereas OBOX4 was retained in *Dux*
108 knockdown 2C-embryos (Supplementary Fig. 2a-c). Monitoring development until 4.5 days post cotium (dpc)
109 showed that *Obox4* single knockdown resulted in moderate developmental retardation, with almost 50% of
110 embryos reaching the blastocyst stage. Similarly, blastocyst formation was preserved in nearly 25% of *Dux*
111 single knockdown embryos, which is consistent with previously reported *Dux* knockdown/knockout
112 experiments [15,18-21]. Intriguingly, the *Obox4*/*Dux* double knockdown completely compromised blastocyst
113 formation, resulted in more than 80% embryos degenerated or arrested before the 4C stage, and less than 20%

114 manifested morula-like morphology (Fig. 2e-f). The double knockdown embryos with morula-like morphology
115 manifest dysregulated transcriptome characterized by failure of expressing morula specific genes and
116 activation of apoptosis related pathways, suggesting that the development was dysfunctional instead of
117 delayed (Supplementary Fig. 2d-g).
118 Transient depletion of DUX has previously confused the field in that it caused an embryonic phototype
119 while subsequent *Dux* knockout females were shown to be fertile [15,18-21]. Therefore, it is of critical importance
120 to examine the functional requirement of OBOX4 and DUX in ZGA using a *Obox4*/*Dux* genetic double knockout
121 model. To address this concern, we generated *Obox4* and *Dux* single and double knockout mESC lines using
122 CRISPR-Cas9 (Supplementary Fig. 3a-b). While mono-allelic *Obox4* deletion was detected in 49.1% (335/682)
123 of the CRISPR-Cas9 edited clones, bi-allelic deletion was only detected in 0.44% (3/682) of the population,
124 suggesting that removal of *Obox4* allele was associated with severe genetic toxicity. In contrast, 36.7%
125 (47/128) mono-allelic and 4.7% (6/128) bi-allelic deletions were detected in *Dux* knockout clones
126 (Supplementary Fig. 3c-e). Somatic cell nuclear transfer (SCNT) is a technique to create embryos by
127 transferring nuclei of somatic cells into enucleated oocytes, which recapitulates ZGA [31]. The knockout mESCs
128 were subjected to SCNT as nucleus donors to demonstrate role of *Obox4* and *Dux* during ZGA (Fig. 2g).
129 Consistent with ASO-mediated knockdown experiments, wildtype and single knockout mESCs produced
130 normal blastocysts at 4 days post nuclear transfer (dpt), while double knockout blastocysts were of low
131 quality, judging from severe morphological abnormality (Fig. 2h). As control, 54.4% of WT mESC derived
132 embryos developed to blastocyst stage. *Obox4* and *Dux* single knockout led to reduced blastocyst formation
133 rate at 38.5% and 39.7% respectively, whereas double knockout resulted in additive effect that led to
134 significantly lowered blastocyst rate of 29.3% (Fig. 2i-j). Most of them are morphologically abnormal (Fig. 2h).
135 The consistency between ASO-mediated transient depletion and CRISPR-Cas9 mediated constitutive knockout
136 demonstrated that the expression of *Obox4* and *Dux* is collectively essential for pre-implantation, and that
137 *Obox4* is capable of securing ZGA in a *Dux*-independent manner.
138
139 **OBOX4 activates 2C-gene in the absence of DUX**
140 As *Obox4* has been shown to activate 2C-genes and TEs in mESCs, we questioned whether embryonic
141 depletion of OBOX4 impairs 2C-genes and TEs activation, and whether the impairment was ameliorated by
142 restoration of OBOX4. First, we determined whether the presence of OBOX4 can rescue the developmental
143 arrest of *Obox4*/*Dux* double knockdown embryos. Codon-optimized *Obox4* mRNA without the ASO target
144 motif was produced using *in vitro* transcription. The rescue experiment was performed by means of
145 co-microinjection of *Obox4* mRNA with *Obox4*/*Dux* double knockdown ASOs (Fig. 3a). Restoration of OBOX4
146 in *Obox4* mRNA-microinjected embryos was confirmed using immunofluorescence staining at the 2C stage
147 (Fig. 3b). As expected, development monitoring showed that *Obox4*/*Dux* double knockdown embryos failed
148 blastocyst formation at 4.5 dpc, whereas among *Obox4* mRNA-rescued embryos, blastocyst formation was
149 retained at a similar level to that observed in the *Dux* single knockdown experiment (Fig. 2e-f and Fig. 3c-d).
150 RNA-seq of 2C-embryos revealed exacerbated dysregulation of 2C-genes and TEs by *Obox4*/*Dux* double
151 knockdown comparing to single knockdown, and the dysregulated transcriptome of double knockdown was
152 rescued by resupplying OBOX4, revealed by number of up and down-regulated genes (Fig. 3e-f and

153 Supplementary Fig. 4a). Differential expression analysis revealed that 58% (102/175) of the down-regulated
154 2C-genes in double knockdown embryos were rescued, including 2C stage markers and critical ZGA factors
155 (Fig. g-h). These results showed that that *Obox4* is capable of activating 2C-genes in the absence of *Dux*.
156 It was reported that splicing inhibitor pladienolide B (PlaB) drives mESC to TBLC transition by activating
157 2C-genes [24]. As a complementary experiment, knockout mESCs were treated with PlaB to examine roles of
158 *Obox4* and *Dux* on 2C-gene activation. Depletion of *Obox4* and *Dux* was confirmed at transcript and protein
159 levels (Supplementary Fig. 4b-c). In contrast to embryonic knockdown, transcriptome analysis showed that
160 *Dux* and *Obox4* single knockout was sufficient to hinder the activation of 2C-genes and TEs, while double
161 knockout led to the dysregulation of the transcriptome to a slightly greater extent (Fig. 3i-j and Supplementary
162 Fig. 4d-e). These observations demonstrated that *Obox4* and *Dux* are individually essential for TBLC
163 reprogramming, suggesting that TBLCs utilize a unique transcription regulatory network to activate 2C-genes.
164
165 **Obox4 binds to 2C-gene promoters and LTR elements**
166 *Obox4* contains a homeobox, activates 2C-genes and TEs, and is essential for *Dux*-less pre-implantation
167 embryogenesis, which prompted us to examine whether OBOX4 is a transcription factor that activates 2C-gene
168 associated loci through direct binding with *cis*-regulatory elements. Cleavage under targets and release using
169 nuclease (CUT&RUN) leverages antibody-targeted cleavage of proximal DNA to identify the binding sites of
170 DNA-associated proteins [32]. CUT&RUN against triple FLAG-tagged DUX (3×FLAG-DUX) expressed in mESCs
171 using a high-affinity anti-FLAG antibody [33] confirmed that the DUX binding pattern revealed by CUT&RUN
172 recapitulated the published HA-tagged DUX ChIP data [16] (Supplementary Fig. 5a-b). CUT&RUN against
173 3×FLAG-OBOX4 expressed in mESCs discovered ~24,000 peaks over non-expressing control, among which
174 39.5% located in the gene promoter regions that covered 26.8% (273/1,019) of the 2C-genes (Fig. 4a). *De*
175 *novo* motif discovery using the top 500 CUT&RUN signal peaks predicted CTGGGATYWRMR as top OBOX4
176 binding motif, which is enriched in the promoter regions of 2C-genes (Fig. 4b). Collectively, OBOX4 and DUX
177 targeted 48.1% (490/1,019) of the 2C-genes with a considerable overlap (36.3% for OBOX4 and 31.3% for
178 DUX), which covered many critical totipotency genes including *Dppa2*, *Sp110*, and *Zscan4d* (Fig. 4c-d). The
179 overlap was substantiated by the observation that MERVL LTR MT2_Mm was among the top 10 LTR targets of
180 both OBOX4 and DUX, based on loci coverage (Fig. 4e-f). Notably, while DUX strongly prefers MT2_Mm, OBOX4
181 binding was biased toward MERVK LTRs, namely RLTR9 and RLTR13 elements, which also demonstrated a
182 2C-specific expression profile (Fig. 4g and Supplementary Fig. 5e-f). To examine whether these bindings were
183 functional, tdTomato reporters driven by MT2_Mm and RLTR13B2 were co-transfected with *Dux* and *Obox4*
184 into *Obox4*/*Dux* double knockout ESCs (Supplementary Fig. 5g). Analysis of the tdTomato-positive population
185 showed that *Obox4* activated both RLTR13B2 and MT2_Mm, whereas *Dux* only activated MT2_Mm (Fig. 4h).
186 Based on these results, we reasoned that OBOX4 and DUX are mutually redundant factors that activate core
187 2C-genes by sharing MT2_Mm as target, while preserving biases toward distinctive cis-regulatory elements.
188
189 **OBOX4 is associated with deposition of active histone marks**
190 Gene activation is frequently associated with change of epigenetic landscape at gene bodies as well as
191 transcription factor binding loci [34]. We analyzed previously published ChIP-seq data and found that during

192 ZGA, OBOX4 binding loci were enriched with the active histone marks H3K4me3 and H3K27ac, while OBOX4
193 target gene bodies were depleted of the repressive histone mark H3K27me3 [35,36] (Fig. 5a). To examine
194 whether OBOX4 is correlated with the deposition of active histone modifications at its LTR targets, we treated
195 WT and *Obox4* knockout mESCs with PlaB and performed CUT&RUN against H3K4me3, H3K27ac, and
196 H3K27me3 (Fig. 5b). Upon PlaB treatment, overall H3K4me3 and H3K27ac level increased at MT2_Mm loci in
197 WT mESCs, while the trends were severely abolished in *Obox4* knockout mESCs, as revealed by quantitative
198 comparison with non-treated mESCs. RLTR9 and RLTR13 loci underwent divergent change in histone
199 modification level, with a slight but significant reduction of overall levels in *Obox4* knockout mESCs (Fig. 5c-d
200 and Supplementary Fig. 6a). No significant change of overall H3K27me3 level between WT and *Obox4*
201 knockout mESCs was observed at MT2_Mm, RLTR9, and RLTR13 and loci (Fig. 5e-f and Supplementary Fig. 6b).
202 Together with our finding that *Obox4* activates MERVK LTRs and is associated with active histone mark
203 deposition, we speculated that despite being functionally redundant to *Dux*, *Obox4* also acts as an activator of
204 2C-specific MERVK elements, which warrants further investigations.
205

206 ## Discussion

207 Starting with a transcriptionally inert genome, the eutherian ZGA engages in massive yet coordinated
208 expression of genes driven by LTR TEs. In mice, it has been reported that the transcription factor DUX accesses
209 and opens condensed chromatin of the MERVL LTR loci, which results in the activation of downstream ZGA
210 genes [15-17]. However, the establishment of fertile *Dux* knockout mice suggests that ZGA is redundantly fulfilled
211 by unknown transcription factor(s) [18-21]. In the present study, we found that *Obox4*, a cryptic multi-copy
212 cluster gene, encodes a homeodomain-containing protein that potently induces ZGA gene expression.
213 Remarkably, *Obox4*/*Dux* double knockdown completely compromised pre-implantation development, which
214 tolerated single knockdown of either gene. Double knockdown embryos exhibited severely dysregulated
215 transcriptomes, characterized by MERVL and ZGA gene activation failure. We then focused on characterizing
216 the molecular mechanisms underlying the biological significance of *Obox4* as well as its relevance to *Dux*.
217 Multi-omics analysis revealed that OBOX4 directly binds to MERVL (the target of DUX) and MERVK loci and is
218 associated with active histone mark deposition. We demonstrated that *Obox4* activates specific MERVL and
219 MERVK elements in a *Dux*-independent manner and that the *Obox4*-targeted MERVL elements are enriched
220 with active histone marks in PlaB treated mESCs. In summary, our results highlight that OBOX4 is a
221 transcription factor that is functionally redundant to DUX during ZGA.
222 *Obox* transcripts were first discovered in gonads [37], and were later found to be highly abundant in mouse
223 2C-embryos [38]. The *Obox* family has 67 members clustered in the sub-telomeric region of mouse chromosome
224 7 and is further divided into six subfamilies: *Obox1*, *Obox2*, *Obox3*, *Obox4*, *Obox5*, and *Obox6*, with *Obox4*
225 constituting nearly 60% (39 of 67) of the family population [39,40]. Although *Obox4* dominates the *Obox* family
226 population, members of the entire subfamily have been annotated as pseudogenes and have largely been
227 neglected from investigation. Despite being abundantly transcribed in the gonads, the biological significance of
228 *Obox* genes is poorly understood. The role of *Obox4* is particularly mysterious, as it remains to be determined
229 whether *Obox4* is a functional protein-coding gene. Transcription of *Obox4* was originally reported to be

230 testis-specific and contradicted by the detection of *Obox4* transcripts in the ovaries and oocytes [41]. *Obox4* is
231 involved in oocyte maturation and ESC differentiation [42-46]. However, these observations are supported by
232 limited evidence, with no evidence at the protein level. By detecting OBOX4 in various totipotent cell types
233 using high-quality monoclonal antibodies, we provided decisive evidence that *Obox4* is a functional
234 protein-coding gene. Our data demonstrated that OBOX4 binds to genomic loci of MERVL-derived promoters
235 and MERVK-derived enhancers and mediates their activation during ZGA. However, whether and how other
236 *Obox* family members contribute to ZGA remains unclear.

237 Unlike *Dux*, which is structurally and functionally conserved throughout placentalia [47], the ancestral locus
238 of the *Obox* family appears to have undergone mouse-specific duplication and generated a gene cluster that is
239 collectively syntenic to the *Tprx2* locus in other mammals [48]. Despite the distal homology between *Obox4* and
240 *Tprx2*, it has been recently reported that human *TPRX2* is expressed in 8-cell embryos [14]. Interestingly, while
241 our work was being reviewed, TPRX2 was shown to cause defective ZGA upon embryonic depletion and bind
242 critical ZGA genes in hESCs [49], suggesting that the *Tprx2* locus has undergone functionally convergent
243 evolution despite its divergent genetic context. However, whether *TPRX2* plays a similar role in humans to
244 *Obox4* in mice regarding the redundancy to *Dux* (*DUX4* in humans) remains to be elucidated.

245 Genetic redundancy is widespread in higher organisms and typically arises in signaling networks, in
246 which multiple functionally overlapping factors commit to a shared teleological objective, to counteract
247 sporadic mutations or defects [50,51]. This principle is of particular importance to embryogenesis, a process that
248 gives rise to a reproducing organism. Indeed, genetic redundancy in pluripotency has been evidenced by the
249 observation that naïve pluripotency can be elicited and maintained through various "independent inputs that
250 operate through both unique and convergent targets" [52]. Although the molecular pathway that governs
251 totipotency has been largely restricted to the scope that centers on *Dux*, it is reasonable to assume that the
252 acquisition of totipotency also complies with the rule of redundancy. This assumption was validated by the
253 generation of fertile *Dux* knockout mice; however, the panorama of ZGA redundancy remained hidden. Our
254 work on *Obox4* sheds light on this topic by discovering that *Obox4* serves as a redundant gene to *Dux* during
255 ZGA and is a bivalent activator of MERVL and MERVK elements. Although the function of MERVL has been
256 extensively investigated, particularly during early embryogenesis, the role of MERVK remains unclear.
257 Interestingly, recent studies have suggested that MERVK serves as a meiosis-specific enhancer of
258 spermatogenesis [53]. Mammalian Y chromosomes have undergone rapid evolution and have been intensively
259 inserted by TEs [54]. As *Obox4* is highly expressed in the testes, it is worth investigating whether *Obox4* plays a
260 regulatory role during spermatogenesis, by activating MERVK-derived enhancers.

261 Pioneer factors possess the ability to specifically recognize and access DNA motifs that are inaccessible to
262 other transcription factors [55]. In the scope of pluripotency-to-totipotency transition, OBOX4 and DUX are
263 qualified pioneer factor for they are capable of accessing and activating silenced 2C-genes in ESCs. However, it
264 is not likely that such pioneer activity demonstrated *in vitro* is accounted for ZGA initiation *in vivo*, considering
265 that genome-wide demethylation precede the timing of OBOX4 and DUX translation [56], and neither individual
266 nor collective depletion of OBOX4 and DUX caused strict 2C arrest. A recent study has identified DNA damage
267 induced p53 activation is important but dispensable for *Dux* activation in mouse zygotes [57], suggesting another

layer of genetic redundancy at ZGA initiation level. Hence, how *Dux* and *Obox4* are activated after fertilization and whether they share the same activation mechanism would be good questions to interrogate.

## Figure Legends

**Figure 1. *Obox4* promotes transcription of 2C-genes and transposable elements.**

a) Mouse homeobox genes that are specifically expressed during ZGA. The genes were identified by means of statistically determined *k*-means clustering based on their expression in pre-implantation embryos. *Dux* is shown in a bold italic font.

b) Scatterplot showing per-gene normalized read counts in mouse totipotent TBLCs *versus* ESCs. Genes with more than 10-fold normalized read counts (FC=10) have been highlighted. *Dux* is shown in a bold italic font.

c) MA plot displaying gene expression in *Dux* knockout 2C-embryos *versus* WT 2C-embryos. Loss of *Dux* partially deactivates the multicopy homeobox gene *Obox4*.

d) Diagram of the 2C::tdTomato reporter assay. ESCs bearing the tdTomato expression cassette under the control of the MERVL LTR promoter showed red fluorescence upon entering the 2C-like state. The expression of the transcription factor increased the 2C-like population, as detected using FACS. An EGFP expression plasmid was co-transfected with a gene of interest, to normalize the transfection efficiency.

e) Boxplot showing normalized 2C-like cell percentage in 2C::tdTomato reporter ESCs overexpressing candidate pioneer factors. *Dux* and *Obox4* potently induced a 2C-like state.

f) Left panel: volcano plot of DEGs in mESCs with *Obox4* induction for 48 hours. Representative 2C-genes are labeled with gene symbols. Right panel: expression profile of genes up-regulated by *Obox4* during embryogenesis.

g) Left panel: volcano plot of differentially expressed transposable elements in mESCs with *Obox4* induction for 48 hours. MERVL and MERVK elements were highlighted. Right panel: expression profile of MERVL and MERVK elements during pre-implantation embryogenesis.

h) Heatmaps of the expression of 2C-genes in preimplantation embryos, naturally occurred 2C-like mESCs, and induced 2C-like mESCs.

i) Scatterplot showing per-gene expression changes in *Dux* induced mESCs *versus* *Obox4* induced mESCs. 2C-genes are highlighted.

j) Gene set enrichment analysis (GSEA) of 2C-genes and *Dux* induced 2C-genes in *Obox4* induced mESCs.

**Figure 2. OBOX4 is necessary for DUX-less preimplantation development.**

a) Immunofluorescence staining of OBOX4 and MERVL GAG at different pre-implantation embryo stages.

b) Immunofluorescence staining of DUX, OBOX4, and OCT4 in 2C-like ESCs.

c) Translation profile of DUX and OBOX4 during preimplantation characterized by Ribo-seq signal.

d) Schematic representation of *Dux* and *Obox4* knockdown experiments in pre-implantation embryos. Male pronuclei of zygotes were microinjected with ASO targeting *Dux* or *Obox4* transcripts.

305 e) Representative picture of KD embryos at 1.5 dpc and 4.5 dpc. No blastocyst was observed among double
306 ASO knockdown embryos.
307 f) The percentages of embryonic stages observed at 1.5 dpc, 2.5 dpc, 3.5 dpc, and 4.5 dpc. The plot
308 represents the sum of three independent experiments, with 150-200 embryos per condition.
309 g) Schematic representation of somatic cell nuclear transfer (SCNT) experiment. The nuclei of knockout
310 mESCs were transferred to enucleated oocytes to generate zygotes with knockout genotype.
311 h) Representative picture of SCNT embryos at 4 dpt. Morphologically abnormal blastocysts are highlighted.
312 Blastocysts generated by double knockout mESCs were severely defective.
313 i) Upper panel: percent SCNT embryos developed to different stages at 4 dpt. Lower panel: *p*-value of
314 non-parametric analysis of variance (ANOVA) among different genotypes and stages comparing to
315 wildtype SCNT embryos.
316 j) The percentages of embryonic stages observed at 4 days post SCNT. The plot represents the sum of 3 to 7
317 independent experiments, with 70-170 embryos per condition.
318
319 **Figure 3. *Obox4/Dux* double knockdown embryos manifest a dysregulated transcriptome.**
320 a) Schematic of the double knockdown rescue experiment. Male pronuclei of zygotes were injected with ASO
321 targeting *Obox4* and *Dux* transcripts as well as *in vitro*-transcribed codon-optimized *Obox4* mRNA.
322 b) Immunofluorescence staining of OBOX4 in embryos microinjected with scrambled ASO, double ASO
323 (*Obox4* and *Dux*), or double ASO with codon-optimized *Obox4* mRNA.
324 c) Representative picture of knockdown and rescue embryos at 4.5 dpc.
325 d) The percentages of embryonic stages observed at 1.5 dpc, 2.5 dpc, 3.5 dpc, and 4.5 dpc. The plot
326 represents the sum of three independent experiments, with 80-100 embryos per condition.
327 e) Upper panel: volcano plot showing the results of DEG analysis of 2C-genes in knockdown embryos, as
328 compared to that in scramble ASO-injected embryos. Standard deviations of $\log_2$(fold-change) were used
329 to represent the degree of transcriptome dysregulation. Stdev, standard deviation. Lower panel: MA plot
330 of differential expression of TEs in knockdown embryos, as compared to that in scramble ASO-injected
331 embryos. MERVK elements and MERVL are highlighted.
332 f) Heatmaps of the expression of 2C-genes in preimplantation embryos, knockdown 2C-embryos, and rescue
333 2C-embryos.
334 g) Rain plot displaying the expression change distribution of rescued 2C-genes in double knockdown and
335 recue 2C-embryos.
336 h) Bar plots showing the expression of representative 2C-genes in knockdown and recue embryos; n=3
337 biological replicates.
338 i) Upper panel: schematic of the design of *in vitro* totipotency acquisition assay. *Obox4* and *Dux* single and
339 double knockout mESCs were subjected to PlaB treatment to enforce 2C-gene activation. Lower panel:
340 heatmaps of the expression of 2C-genes in PlaB treated mESCs.
341 j) Scatterplots showing per-2C-gene expression changes in knockdown 2C-emrbyos *versus* PlaB treated
342 knockout mESCs, as compared to scramble ASO-injected embryos and PlaB treated WT mESCs
343 respectively. 2C-genes that were significantly down-regulated ($\log_2$FC ≤ -1, q ≤ 0.01) in PlaB treated

knockout mESCs are highlighted, and linear-fitted to represent the degree of agreement between experimental settings.

**Figure 4. OBOX4 binds to LTR elements and activates 2C-genes.**

a) Left panel: pie-chart displaying proportions of annotated genomic regions of the OBOX4 binding sites. Right panel: heatmap showing the OBOX4 binding site distribution near 2C-gene promoters.

b) Left panel: the predicted OBOX4 binding motif using the top 500 CUT&RUN peaks. Right panel: Histocurve showing the distribution of predicted OBOX4 binding motif near 2C and random gene promoters.

c) Left panel: venn diagram showing the distinct and overlap 2C-genes targeted by DUX and OBOX4. Right panel: rain plots of expression of distinct and overlap 2C-genes targeted by DUX and OBOX4 in knockdown 2C-embryos, as compared to that in scramble ASO-injected embryos.

d) Representative genomic track showing DUX and OBOX4 binding sites at the *Dppa2*, *Sp110,* and *Zscan4d* loci and their expression levels in PlaB treated knockout mESCs. Read counts were CPM normalized. The OBOX4 binding sites overlapped with those of DUX. *Dppa2, Sp110*, and *Zscan4d* expression was hindered upon *Obox4* knockout.

e) Left panel: pie-chart displaying proportions of annotated TEs of the OBOX4 binding sites. Right panel: Bar plot showing the top 10 OBOX4 covered LTR elements.

f) Left panel: pie-chart displaying proportions of annotated TEs of the DUX binding sites. Right panel: Bar plot showing the top 10 DUX covered LTR elements.

g) Representative genomic track showing DUX and OBOX4 binding sites at MT2_Mm and RLTR13B2 and their expression levels in PlaB treated knockout mESCs. Read counts were CPM normalized. The OBOX4 binding sites overlapped with those of DUX at MT2_Mm loci, while RLTR13B2 was uniquely bound by OBOX4.

h) Bar plots showing the percentage of red fluorescence-positive mESCs upon expression of *Dux* or *Obox4*. Both *Dux* and *Obox4* activated MT2_Mm, whereas only *Obox4* activated RLTR13B2.

**Figure 5. OBOX4 guides deposition of active histone marks at its binding loci.**

a) Left panel: heatmaps showing change of H3K4me3 and H3K27Ac distributions near the OBOX4 binding sites during ZGA. Right panel: heatmaps showing change of H3K27me3 distribution at OBOX4 target genes during ZGA.

b) Schematic diagram showing the strategy for profiling histone modifications in PlaB treated mESCs. PlaB treated or untreated WT and *Obox4* knockout mESCs were subjected to CUT&RUN, with antibodies against H3K4me3, H3K27Ac, and H3K27me3.

c) Left panel: scatterplots showing per-locus H3K4me3 intensity changes at MT2_Mm and RLTR9/RLTR13 loci in PlaB treated mESCs, as compared to that in PlaB untreated mESCs. Right panel: box plots comparing distribution of intensity changes between WT and *Obox4* KO condition.

d) Left panel: scatterplots showing per-locus H3K27ac intensity at MT2_Mm and RLTR9/RLTR13 loci in PlaB treated mESCs, as compared to that in PlaB untreated mESCs. Right panel: box plots comparing distribution of intensity changes between WT and *Obox4* KO condition.

383 e) Left panel: scatterplots showing per-locus H3K27me3 intensity at MT2_Mm and RLTR9/RLTR13 loci in
384 PlaB treated mESCs, as compared to that in PlaB untreated mESCs. Right panel: box plots comparing
385 distribution of intensity changes between WT and *Obox4* KO condition.
386 f) Representative genomic track showing histone modification levels at *Zscan4d*, *Sp110*, MT2_Mm, and
387 RLTR9D loci in oocytes, 2C-embryos, and PlaB treated mESCs. Read counts were CPM normalized, set to
388 same scale among different samples, and hidden for better visualization. OBOX4 knockout is associated
389 with the reduction of active histone modifications H3K4me3 and H3K27Ac, but not with the repressive
390 histone modification H3K27me3.
391
392 **Supplemental figure legends**
393
394 **Supplementary Figure 1**
395 a) Expression profiles of *k*-means clustered gene sets during preimplantation development based on
396 *z*-scored expression of scRNA-seq and bulk RNA-seq. Genes appeared in both scRNA-seq cluster 1 and
397 bulk RNA-seq cluster 4 were designated as 2C-genes.
398 b) Western blot showing protein products of the candidate genes ectopically expressed in mESCs at 18 hours
399 post transfection. Notably, only *Dux* expression induced MERVL GAG protein production at this time
400 point.
401 c) FACS analysis of 2C::tdTomato cells co-transfected with candidate gene expression plasmids and EGFP.
402 d) Fluorescence microscopy of 2C::tdTomato ESCs transfected with empty plasmid, plasmid encoding DUX,
403 or plasmid encoding OBOX4. Plasmids encoding EGFP were co-transfected to normalize the transfection
404 efficiency in FACS analysis.
405 e) Upper panel: schematic of *Obox4*-inducible cell line construction. Middle panel: western blot showing
406 OBOX4 level upon induction by different concentrations of doxycycline. Expression of OBOX4 was carried
407 out in a dose-dependent manner. Lower panel: western blot showing OBOX4 and GAG level along time
408 course of *Obox4* induction. Prolonged *Obox4* expression induced mildly elevated GAG protein production.
409 f) Left panel: volcano plot of DEGs in mESCs with *Obox4* induction for 12 hours. Right panel: expression
410 profile of genes up-regulated by *Obox4* during embryogenesis.
411 g) Venn diagram showing overlap of 2C-genes with genes induced by ectopic expression of *Dux* and *Obox4* in
412 mESCs.
413
414 **Supplementary Figure 2**
415 a) Transcript levels of *Dux* and *Obox4* in ASO-injected embryos, as measured using RNA-seq. CPM, counts
416 per million.
417 b) Representative images of immunofluorescence co-staining of OBOX4 and DUX in 2C embryos subjected to
418 ASO-mediated knockdown.
419 c) Dot plots showing the per-nucleus intensity of OBOX4 and DUX normalized to dapi.

d) Heatmaps of the expression of morula-genes in preimplantation embryos and knockdown morulae. Genes appeared in both scRNA-seq cluster 6 and bulk RNA-seq cluster 3 in Supplementary Fig. 1a were designated as morula-genes.

e) Transcriptome-based PCA of knockdown morulae and preimplantation embryos.

f) Volcano plot of DEGs in double knockdown morulae comparing with scramble knockdown. Selected apoptosis associated genes are highlighted.

g) The enrichement of gene ontology (GO) terms of up-regulated genes in double knockdown morulae.

**Supplementary Figure 3**

a) Schematic diagram of producing *Obox4* knockout mESCs using CRISPR-Cas9 (not drawn to scale). sgRNAs flanking the *Obox4* cluster as well as the internal sequence of individual *Obox4* were introduced into mESCs with SpCas9, to knockout *Obox4*. Primers flanking Cas9 cutting sites (A'-D') were designed to validate removal of target allele. Primers inside of *Obox4* (E'-F') were designed to quantify *Obox4* copy number by qPCR.

b) Schematic diagram of producing *Dux* knockout mESCs using CRISPR-Cas9 (not drawn to scale). Previously reported sgRNAs flanking the *Dux* cluster were introduced into mESCs with SpCas9, to knockout *Dux*. Primers flanking Cas9 cutting sites (A-D) were used to validate removal of target allele. Primers inside of *Dux* (E-F) were designed to detect translocation of *Dux* cluster.

c) Agarose gel image of PCR-based genotyping of the *Obox4* knockout allele in founder cell lines. The red arrow indicates a cell line bearing a bi-allelic deletion of the *Obox4* cluster.

d) Agarose gel image of PCR-based genotyping of the *Dux* knockout allele in founder cell lines. The red arrow indicates a cell line without the genomic *Dux* sequence.

e) Upper panel: schematic diagram of *Obox4* copy number quantification. Primer E' and F' targets all 15 protein-coding as well as 3 non-protein-coding *Obox4* loci, among which 14 protein-coding and 1 non-protein-coding copies were subjected to knockout. Lower panel: bar plot showing copy number of *Obox4* in the founder cell lines, as detected by qPCR using purified genomic DNA.

**Supplementary Figure 4**

a) Venn diagram showing the distinct and overlap 2C-genes down-regulated in *Obox4*/*Dux* single and double knockdown 2C-embryos.

b) Bar charts showing transcript levels of *Obox4* and *Dux* in PlaB treated mESCs. WT, wild-type; mean ± SD (n=3 biological replicates).

c) Immunofluorescence staining of OBOX4 in *Obox4* knockout, *Dux* knockout, and double knockout mESCs subjected dto PlaB treatment.

d) Upper panel: volcano plot showing the results of DEG analysis of 2C-genes in PlaB treated knockout mESCs, as compared to wildtype. Lower panel: MA plot of differential expression of TEs in PlaB treated knockout mESCs, as compared to wildtype. MERVK elements and MERVL are highlighted.

e) Bar plots showing the expression of representative 2C-genes in wildtype and knockout mESCs subjected to PlaB treatment; n=3 biological replicates.

**Supplementary Figure 5**

a) Schematic diagram of the CUT&RUN workflow using mESCs with induced expression of FLAG-tagged DUX. mESCs bearing a doxycycline-inducible 3×FLAG-OBOX4 expression cassette were induced with doxycycline for 24 h. OBOX4-associated DNA was pulled down in the CUT&RUN assay, using a high-affinity anti-FLAG antibody.

b) Left panel: comparison of DUX binding motif predicted with published ChIP data and CUT&RUN result from this study. Right panel: overlap of DUX binding peaks discovered with the two data sets.

c) Heatmaps showing the OBOX4 binding site distribution near top covered LTR elements.

d) Expression profile of top OBOX4 covered LTR elements during embryogenesis.

e) Schematic design of the LTR::tdTomato reporter assay. Plasmids bearing a tdTomato ORF downstream of MT2_Mm or RLTR13B2 were co-transfected with *Dux* or *Obox4* expression plasmids. Activation of LTR elements resulted in an increased red fluorescence-positive mESC population, which was measured using FACS. EGFP expression plasmid was co-transfected into the culture, following which the green fluorescence-positive population was measured, to normalize the transfection efficiency.

**Supplementary Figure 6**

a) Heatmaps showing the H3K4me3 (left panel) and H3K27ac (right panel) distribution near top OBOX4 covered LTR elements in wildtype and *Obox4* knockout mESCs with and without PlaB treatment.

b) heatmaps showing change of H3K27me3 distribution at OBOX4 target genes in wildtype and *Obox4* knockout mESCs with and without PlaB treatment.

**Supplementary Figure 7**

a) Schematic of the production of monoclonal antibodies. Mice were immunized with the GST-OBOX4 fusion protein. After fusing B-cells with myelomas, the antibodies produced by hybridomas were examined against MBP-OBOX4 fusion protein, followed by MBP-OBOX2 fusion protein, for identification of OBOX4-specific antibodies.

b) Analysis of protein sequence similarity among the OBOX family members. The Smith–Waterman local alignment method was used to align the amino acid sequences, and the divergence among different sequences was represented in the form of a block substitution matrix.

c) ELISA signal of anti-OBOX4 monoclonal antibodies against the MBP-OBOX4 and MBP-OBOX2 fusion proteins. Anti-OBOX4 monoclonal antibodies were not reactive to MBP-OBOX2.

d) Multiple sequence alignment of protein sequence among the OBOX family members. The homeodomains are highlighted by magenta box. The peptide sequences used for antibody production are highlighted by black box.

**Supplementary Figure 8**

Multiple sequence alignment of DNA sequence among the *Obox4* copies. Full length *Obox4* with internal top codons are marked by asterisks.

498

499 # Methods

500 **Mice**

501 Male/female BDF1 and female BALB/c mice were purchased from Japan SLC Inc. Mice were fed regular
502 chow and housed in a controlled room under a 14 h/10 h light/dark cycle, at 22°C. All animal experiments
503 were approved by the Animal Care and Use Committee of Keio University and the Animal Experimentation
504 Committee at the RIKEN Tsukuba Institute and conducted in compliance with the Keio University Code of
505 Research Ethics and the RIKEN's guiding principles.

506

507 **Cell culture**

508 All mouse mESC lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) (catalog no. 08488-55,
509 Nacalai Tesque) supplemented with 10% fetal bovine serum (lot no. S10581S1820, Biowest), 1× GlutaMAX™
510 (catalog no. 35050061, Gibco), 1× penicillin-streptomycin (catalog no. 15140163, Gibco), 1× non-essential
511 amino acids (catalog no. 11140050, Gibco), 1× sodium pyruvate (catalog no. S8636, Merck), 50 μM
512 2-mercaptoethanol (catalog no. 21985023, Gibco), 1 μM CHIR99021 (catalog no. 034-23103, Wako), 3 μM
513 PD0325901 (catalog no. 168-25293, Wako), and mouse leukemia inhibitory factor produced in-house. The
514 culture medium was supplemented with iMatrix-511 silk (recombinant human laminin-511 E8 fragment)
515 (catalog no. 892021, Matrixome) at a concentration of 125 ng.cm$^{-2}$ culture vessel area, before seeding mESCs
516 into it. The cells were cultured in a humidified atmosphere containing 5% $CO_2$ at 37°C, with the media changed
517 every 48 h.

518 TBLCs were generated by culturing mESCs in an ordinary medium supplemented with 2.5 nM PLaB
519 (catalog no. 16538, Cayman Chemicals). Cells were sub-cultured every 2–4 d, at a seeding density of 1×10$^5$
520 cm$^{-2}$ culture vessel area. After five rounds of sub-culture, the cells were deemed TBLCs.

521 SP2/O-Ag14 myeloma and primary clones of hybridomas were cultured in GIT medium (catalog no.
522 63725715, Wako) supplemented with recombinant human interleukin-6 (IL-6) (catalog no. 20006,
523 PeproTech). Cells were cultured in a humidified atmosphere containing 5% $CO_2$ at 37°C and sub-cultured
524 every day, at a seeding density of 3×10$^5$ mL$^{-1}$.

525 For monoclonal antibody production, hybridomas were cultured in hybridoma serum-free medium
526 supplemented with IL-6. The cells were cultured in a humidified atmosphere containing 5% $CO_2$ at 37°C, until
527 they were over-confluent.

528

529 **Cell transfection**

530 Plasmid transfections were performed using jetOPTIMUS® DNA Transfection Reagent (catalog no.
531 101000025, Polyplus), according to the manufacturer's instructions. Briefly, mESCs were seeded at a cell
532 density of 8×10$^4$ cm$^{-2}$ culture vessel area. After 20 min, the jetOPTIMUS® reagent and plasmids were diluted in
533 jetOPTIMUS® buffer, incubated for 10 min at room temperature, and then applied to the cell culture. After 48 h,
534 the cells were collected for downstream experiments.

535

536 **cDNA synthesis and cloning**

537 Total RNA was extracted from TBLCs using ISOGEN (catalog no. 311-02501, Nippon Gene), according to
538 the manufacturer's instructions. Briefly, 500 μL ISOGEN was added to $1.5\times10^6$ freshly harvested TBLCs. After
539 5 min of incubation at room temperature, 100 μL of chloroform (catalog no. 03802606, Wako) was added to
540 the cells and mixed by means of vigorous shaking. After 2 min of incubation at room temperature, the samples
541 were centrifuged at 12,000 × *g* for 15 min at 4°C. The upper aqueous phase was then transferred to a new tube
542 and mixed with 240 μL isopropanol (catalog no. 15-2320, Merck), to precipitate the RNA. After 5 min of
543 incubation at room temperature, the samples were centrifuged at 12,000 × *g* for 15 min at 4°C. The
544 supernatant was then discarded, following which the RNA pellets were washed with 70% ethanol (catalog no.
545 057-00456, Wako). The RNA pellets were then air-dried and dissolved in RNase-free water. The RNA solution
546 was subjected to DNase treatment using TURBO DNA-*free*™ kit (catalog no. AM1907, Thermo Fisher),
547 according to the manufacturer's instructions, to remove genomic DNA carryover. RNA (1 μg) was reverse
548 transcribed to cDNA using the Transcriptor First Strand cDNA Synthesis Kit (catalog no. 04379012001, Roche),
549 according to the manufacturer's instructions. PrimeSTAR® Max DNA Polymerase (catalog no. R045A, TaKaRa)
550 and ProFlex™ PCR System (catalog no. 4484073, Thermo Fisher) were used to amplify the sequence of interest
551 from the cDNA. NEBuilder® HiFi DNA Assembly Kit (catalog no. E2621L, NEB) was used to clone the sequence
552 of interest into plasmid backbone.

553

554 **Generation of transgenic mESC lines**

555 The 2C::tdTomato reporter cell line was generated by transfection of EB3 mESCs (catalog no. AES0139,
556 RIKEN BRC Cell Bank) with a linearized 2C::tdTomato reporter plasmid (catalog no. 40281, Addgene). After 48
557 h, the cells were subjected to 500 $\mu g.mL^{-1}$ hygromycin (catalog no. 08906151, Wako) for 7 d. The selected cells
558 were then seeded at a density of $2\times10^2$ $cm^{-2}$ in culture medium containing 250 $\mu g.mL^{-1}$ hygromycin. Single-cell
559 clones were picked and expanded after 7 d.

560 A tetracycycline-controlled transcription activation system and a piggyBac transposon system were used
561 as previously described [58], to generate an *Obox4*-inducible cell line. Briefly, pPB-TRE-3×FLAG-Obox4,
562 pPB-CAG-rtTA3G-IRES-Hygro, and pCMV-HyPBase-PGK-Puro plasmids were co-transfected into EB3 mESCs.
563 After 48 h, the cells were subjected to selection with 500 $\mu g.mL^{-1}$ hygromycin and 500 $\mu g.mL^{-1}$ G418 (catalog
564 no. 07805961, Wako), for 7 d. The selected cells were then seeded at a density of $2\times10^2$ $cm^{-2}$ in a culture
565 medium containing 250 $\mu g.mL^{-1}$ hygromycin and 250 $\mu g.mL^{-1}$ G418. Single-cell clones were picked and
566 expanded after 7 d.

567

568 **Generation of knockout mESC lines**

569 Among 39 *Obox4* copies, there are 15 copies that maintain intact open reading frame (ORF) for full-length
570 OBOX4 (Supplementary Fig. 8). While 14 of the 15 intact ORFs form tightly packed cluster (*Obox4* cluster), a
571 solo ORF (*Obox4-ps33*) remains distant from the *Obox4* cluster and is interspersed with other *Obox1/2/3*.
572 Since *Obox* family locates at the chromosome 7 subtelomeric region, only *Obox4* cluster was knocked out and
573 the solo ORF was retained to minimize collateral genetic toxicity and interference to experiment result caused
574 by removal of other Obox members (Supplementary Fig. 3a). Plasmids encoding Cas9 and single guide RNA

575 (sgRNA) flanking and inside of the *Obox4* cluster were constructed to perform knockout experiment. The
576 *Obox4* knockout cell line was generated by means of co-transfection of EB3 mESCs with the
577 pX330-Puro-Obox4KO-3, pX330-Puro-Obox4KO-4, pX330-Puro-Obox4KO-6, and pX330-Puro-Obox4KO-7
578 plasmids. After 18 h, the cells were subjected to selection with 1 $\mu g.mL^{-1}$ puromycin (catalog no. 16023151,
579 Wako) for 54 h. The selected cells were then seeded at a very low density in culture medium without
580 antibiotics. Single-cell clones were picked and expanded after 7 d, and then subjected to genotyping.
581 Following a previously published protocol [18], plasmids encoding Cas9 and sgRNAs flanking the *Dux*
582 cluster were constructed to perform knockout experiment. The *Dux* knockout cell line and the *Obox4*/*Dux*
583 double knockout cell line were generated by means of co-transfection of EB3 mESCs and *Obox4* knockout cell
584 lines with pX330-DuxKO-A and pX330-DuxKO-B plasmids. After 48 h, the cells were seeded at a very low
585 density. Single-cell clones were picked and expanded after 7 d and then subjected to genotyping.
586
587 **Flow cytometry**
588 The cells were digested with 0.25% trypsin-EDTA (catalog no. 25200072, Gibco), at 37°C for 5 min, and
589 then re-suspended in FluoroBrite™ DMEM (catalog no. A1896701, Gibco). The cell suspension was filtered
590 through a 35 μM cell strainer (catalog no. 352235, Falcon). The samples were analyzed using a cell sorter
591 (catalog no. SH800Z, Sony). Data analysis was performed using the Sony SH800Z cell sorter software.
592
593 **Western blot**
594 Freshly harvested cells were re-suspended in phosphate-buffered saline (PBS) (catalog no. 14249-95,
595 Nacalai Tesque) and lysed by means of sonication. Whole cell lysates were mixed with sample buffer
596 containing reducing reagent (catalog no. 09499-14, Nacalai Tesque) and boiled at 95°C for 5 min. Protein
597 samples were loaded on a 10% tris-glycine gel, run in AllView PAGE Buffer® (catalog no. DS520, BioDynamics
598 Laboratory), and then transferred to nitrocellulose membranes (catalog no. 10600003, Cytiva) using a Power
599 Blotter System (catalog no. PB0012, Thermo Fisher). The membranes were blocked with PBST containing 2%
600 skim milk (catalog no. 4902720131292, Morinaga), at room temperature for 15 min. Membranes were
601 incubated with primary antibodies in PBST containing 2% skim milk, at room temperature for 30 min. After
602 three washes with PBST, the membrane was incubated with secondary antibodies in PBST containing 2% skim
603 milk, at room temperature for 15 min, with shaking. After three washes with PBST, the membrane was
604 incubated with ECL reagents (catalog no. RPN2232, Cytiva) and exposed to an X-ray film (catalog no.
605 28906839, Cytiva) or detected using a digital chemiluminescence imager (catalog no. 17001402JA, Bio-Rad).
606
607 **Generation of monoclonal antibodies**
608 The anti-OBOX4, anti-DUX, and anti-GAG monoclonal antibodies were produced as previously described [59]
609 (Supplementary Fig. 7a). Briefly, BALB/c mice were immunized by means of intraperitoneal injection of
610 glutathione S-transferase-fused protein-of-interest (PoI). The mice were routinely immunized until their sera
611 tested positive for the PoI-fused maltose binding protein (MPB), after which two boosting immunizations were
612 performed. Following boosting, splenocytes from immunized mice were fused with SP2/O myeloma and
613 cultured in hypoxanthine-aminopterin-thymidine medium for 10 d, to select for hybridomas. Hybridomas

614 were subsequently screened using enzyme-linked immunosorbent assay against PoI-fused MBP. The positive
615 polyclonal hybridomas were monoclonalized and expanded. For anti-OBOX4 hybridomas, additional screening
616 using OBOX2-fused MBP was conducted. The sequence of OBOX4 was significantly divergent from those of
617 other OBOX members, and the monoclonal antibodies were negative for cross-reactivity with OBOX2
618 (Supplementary Fig. 7b-c). The N-terminal 100 amino acids of OBOX4 and OBOX2 were used for immunization
619 and cross-reactivity examination respectively (Supplementary Fig. 7d). The culture supernatants of
620 monoclonal hybridomas were sterilized by passing them through a 0.22 μM pore size filter (catalog no.
621 430769, Corning) and used directly as an antibody solution in other assays.
622
623 ***In vitro* transcription**
624 The *Obox4* coding sequence was codon-optimized using GeneArt Instant Designer (Thermo Fisher), to
625 remove the ASO target motif and improve translation efficiency. Codon-optimized DNA was synthesized using
626 Prime Gene Synthesis Services (Thermo Fisher). Codon-optimized *Obox4* mRNA was transcribed *in vitro* using
627 the mMESSAGE mMACHINE™ T7 Transcription Kit (catalog no. AM1344, Thermo Fisher) and then
628 polyadenylated using a poly(A) tailing kit (catalog no. AM1350, Thermo Fisher), according to the
629 manufacturer's instructions. The polyadenylated mRNA was purified using the RNeasy Mini Kit (catalog no.
630 74004, Qiagen), according to the manufacturer's instructions, and dissolved in RNase-free water, at a
631 concentration of 100 $ng.\mu L^{-1}$.
632
633 **Pronuclear injection of mouse embryos**
634 Eight-week-old BDF1 female mice were injected with 150 μL of CARD HyperOva (catalog no. KYD-010-EX,
635 Kyudo Co., Ltd.). After 48 hours, the female mice were injected with 5IU human chorionic gonadotropin (hCG)
636 (catalog no. GONATROPIN, ASKA Animal Health Co. Ltd.) and mated with BDF1 male mice. After 22 h, the PN3
637 zygotes were collected. Cumulus cells were removed by briefly culturing the zygotes in potassium simplex
638 optimization medium (KSOM) (catalog no. MR-101-D, Merck) supplemented with 0.3 $\mu g.\mu L^{-1}$ hyaluronidase
639 (catalog no. H4272, Merck). Embryos were cultured in KSOM medium drops covered with liquid paraffin
640 (catalog no. 26137-85, Nacalai Tesque), in a humidified atmosphere with 5% $CO_2$, at 37°C. After 4 h, PBS
641 containing either 40 μM scramble ASO, 20 μM anti-*Dux* ASO, 20 μM anti-*Obox4* ASO, 40 μM equimolar
642 anti-*Dux*/anti-*Obox4* ASO mixture, or 40 μM equimolar anti-*Dux*/anti-*Obox4* ASO mixture with 100 $ng.\mu L^{-1}$
643 codon-optimized *Obox4* mRNA was microinjected into the male pronuclei of zygotes using a microinjector
644 (catalog no. 5252000013, Eppendorf). For developmental monitoring, embryos were cultured for 4 d after
645 microinjection and assessed for developmental stage at 18 h, 42 h, 66 h, and 90 h after microinjection, which
646 corresponded to 1.5 dpc, 2.5 dpc, 3.5 dpc, and 4.5 dpc, respectively.
647
648 **Somatic cell nuclear transfer (SCNT)**
649 SCNT was performed as described previously [60] using wild-type and knockout mESC lines as nuclear
650 donor cells. They were cultured at a high density until they reached confluency for about four to six days
651 before SCNT. BDF1 female mice were superovulated by the injection of 7.5 IU of equine chorionic
652 gonadotropin (eCG, ZENOAQ) and 7.5 IU of hCG (ASKA Pharmaceutical Co., Ltd.) at a 48 h interval. At 15 h after

653 hCG injection, cumulus–oocyte complexes were collected from the oviducts. After removal of cumulus cells by
654 0.1% bovine testicular hyaluronidase (catlog no. 385931, Calbiochem), oocytes were enucleated in
655 Hepes-buffered KSOM containing 7.5 μg/ml cytochalasin B. After culture in KSOM for at least 1 h, the
656 enucleated oocytes were injected with donor mESCs using a Piezo-driven micromanipulator (catalog no.
657 PMM-150FU, Primetech). After culture in KSOM for about 1 h, injected oocytes were cultured in $Ca^{2+}$-free
658 KSOM containing 2.5 mM $SrCl_2$, 5 μM latrunculin A (LatA) (catalog no. L5163, Merck) with 50 nM trichostatin
659 A (TSA) (catalog no. T8552, Merck) for 1 h. Then, they were cultured in KSOM containing 5 μM LatA and 50 nM
660 TSA for 7 h. After washing, the SCNT embryos were cultured in KSOM under 5% $CO_2$ in air at 37 °C for 96 h.
661 The embryos that reached 2-cells at 24 h after activation were considered successfully cloned with cell
662 cycle-matched mESCs (G0/G1 phase).
663
664 **Immunofluorescence**
665 For embryos, middle-2C-embryos were collected at 1.5 dpc. The zona pellucida was removed by treating
666 the embryos with acidic Tyrode's solution (catalog no. MR-0040-D, Merck). The embryos were fixed in 4%
667 paraformaldehyde (catalog no. 09154-14, Nacalai Tesque) in PBS for 10 min. Following three washes in PBS,
668 fixed embryos were permeabilized with 0.1% Triton™ X-100 (catalog no. 1610407, Bio-Rad) in PBS, for 20 min
669 at room temperature. Following three washes in PBS, permeabilized embryos were blocked in 2% bovine
670 serum albumin (BSA; catalog no. 011-27055, Wako) in PBS, for 20 min at room temperature. The blocked
671 embryos were incubated with primary antibodies diluted in 2% BSA in PBS, at 4°C overnight. Following three
672 washes in PBS, the embryos were incubated with 2% BSA in PBS containing 1:200 diluted
673 4′,6-diamidino-2-phenylindole (DAPI) (catalog no. 19178-91, Nacalai Tesque) and 1:500 diluted Alexa
674 Fluor™-conjugated anti-mouse IgG1 secondary antibody (catalog no. A-21127, Thermo Fisher) or Alexa
675 Fluor™-conjugated anti-mouse IgG2a secondary antibody (catalog no. A-21131, Thermo Fisher) and, for 1 h at
676 room temperature. Following three washes in PBS, the embryos were transferred into liquid paraffin-covered
677 PBS drops on a glass-bottom dish (catalog no. D11130H, Matsunami).
678 For culturing, cells seeded on glass bottom chamber slides (catalog no. SCS-N02, Matsunami) were fixed
679 with 4% paraformaldehyde in PBS, for 10 min. After three washes in PBS, the fixed cells were permeabilized
680 with 0.1% Triton™ X-100 in PBS, for 15 min at room temperature. Following three washes in PBS,
681 permeabilized cells were incubated with primary antibodies diluted with 2% skim milk in PBS, for 30 min at
682 room temperature. Following three washes in PBS, the cells were incubated with 2% BSA in PBS containing
683 1:500 diluted Alexa Fluor™-conjugated secondary antibody and 1:200 diluted DAPI, for 1 h at room
684 temperature. Following three washes in PBS, the chamber was removed and the slides were mounted with
685 ProLong™ Glass Antifade Mountant (catalog no. P36982, Thermo Fisher). Fluorescence images were taken
686 using a confocal laser scanning microscope (catalog no. FV3000, Olympus).
687
688 **scRNA-seq analysis**
689 Raw scRNA-seq data was obtained from the dataset of Deng et al. (GSE45719). Quality control and
690 adapter trimming was done using fastp [61] v0.23.2. Quality-controlled reads were aligned to the GRCm38.p6
691 reference genome using STAR [62] v2.7.9a, with default arguments. Reads were counted against GRCm38.p6

692 comprehensive gene annotation [63] and mm10 repeats from the University of California, Santa Cruz (UCSC)
693 RepeatMasker using Subread [64] v2.0.1 featureCounts function, and multi-mapping reads were discarded for
694 non-TE features and counted fractionally for TEs. Seurat [65] v4.1.0 was used to process the read counts of
695 scRNA-seq. Cells with greater than 7.5% mitochondrial reads or less than 14,000 annotated features were
696 discarded. Expression levels were log-normalized. The expression profiles of homeobox genes listed in
697 HomeoDB2 [66] were clustered into 11 *k*-means after *z*-score transformation.
698
699 **Ribo-seq analysis**
700 Raw Ribo-seq data was obtained from the dataset of Xiong et al. (GSE165782). Reads were
701 quality-controlled, aligned, and counted, as described above. Reads aligned to all protein coding *Obox4* loci
702 were added up to represent translation level of OBOX4.
703
704 **Bulk RNA-seq**
705 For embryos, 30 middle-2C-embryos were collected 20 h after microinjection in each independent
706 biological replicate. The zona pellucida was removed by treating embryos with acidic Tyrode's solution.
707 Libraries were constructed using the SMART-Seq® Stranded Kit (catalog no. 634442, TaKaRa) and indexed
708 using the SMARTer® RNA Unique Dual Index Kit (catalog no. 634451, TaKaRa), according to the
709 manufacturer's instructions.
710 For cell culture, total RNA was prepared using the RNeasy Kit, according to the manufacturer's
711 instructions. The RNA solution was subjected to DNase treatment to remove genomic DNA carryover. Using 1
712 μg total RNA per sample, libraries were constructed with the NEBNext® Ultra™ II Directional RNA Library Prep
713 Kit for Illumina (catalog no. E7760L, NEB) and indexed with NEBNext® Multiplex Oligos for Illumina (catalog
714 no. E6440S, NEB), according to the manufacturer's instructions.
715 The libraries were quantified with a 2100 Bioanalyzer (catalog no. G2939BA, Agilent) using a High
716 Sensitivity DNA Kit (catalog no. 5067-4626, Agilent). Quantified libraries were pooled and sequenced using an
717 Illumina NovaSeq 6000 System in 150 bp paired-end mode (Illumina). Base calling and de-multiplexing were
718 performed using the bcl2fastq2 (Illumina) v2.20. De-multiplexed reads were quality-controlled, aligned, and
719 counted, as described above. DESeq2 [67] v1.32.0 was used to perform differential expression analysis. Raw
720 RNA-seq reads for TBLCs were downloaded from the dataset of Shen et al. (GSE168728). Raw RNA-seq reads
721 of *Dux* knockout embryos were downloaded from the datasets of Chen & Zhang and De Iaco et al. (GSE121746
722 and GSE141321 respectively). Raw RNA-seq reads of preimplantation embryos were downloaded from the
723 datasets of Wu et al. (GSE66390). Raw RNA-seq reads of 2C-like mESCs were downloaded from the datasets of
724 Zhu et al. (GSE159623). The published data were analyzed using the same method.
725
726 **CUT&RUN-seq**
727 Cultured cells ($1\times10^5$) freshly prepared by trypsin-EDTA digestion were used to perform CUT&RUN with
728 the CUT&RUN Assay Kit (catalog no. 86652, Cell Signaling Technology), according to the manufacturer's
729 instructions. Enriched DNA was subjected to library construction using the NEBNext® Ultra™ II DNA Library
730 Prep Kit for Illumina (catalog no. E7645L, NEB), with the adapter ligation step performed at 50°C instead of

731 65°C, to prevent the denaturation of small DNA inserts. The libraries were indexed using NEBNext® Multiplex
732 Oligos for Illumina, according to the manufacturer's instructions.
733 The libraries were quantified, pooled, sequenced, de-multiplexed, and quality controlled, as described
734 above. Processed reads were aligned to the GRCm38.p6 reference genome using Bowtie2 [68] v2.4.1, with default
735 settings. Peaks were called in each biological replicate using the MACS3 [69] v3.0.0a6 callpeak function (-f
736 BAMPE). Alignment tracks were first generated using deepTools [70] v3.5.1 bamCoverage function (--binSize 10
737 --normalizeUsing CPM --smoothLength 30), and then normalized by subtracting the signal from non-immune
738 IgG (for histone modifications) or wildtype mESCs (for OBOX4 and DUX) using the deepTools bamCompare
739 function (--scaleFactorsMethod None --operation subtract --binSize 10 --smoothLength 30). ChIPseeker [71]
740 v1.28.3 was used to annotate the peaks. The MEME suite [72] v5.4.1 was used to identify the binding motifs.
741 Quantitative histone modification analysis of replicated samples was performed using MAnorm2 [73] v1.2.0.
742 Heatmaps were generated using deepTools computeMatrix and plotHeatmap functions. Visualization of
743 genomic tracks was performed using trackplot [74] v1.3.10. Raw ChIP-seq reads of DUXs in mESCs were
744 downloaded from Hendrickson et al. (GSE85632). The published data were analyzed using the same method.
745
746 **Quantification and statistical analysis**
747 Descriptive and comparative statistics were employed in the manuscript as described in the figure
748 legends with the number of replicates indicated. Significance is defined as a *p-value* less than 0.05 indicated
749 with asterisk (* *p-value* < 0.05, ** *p-value* < 0.01, *** *p-value* < 0.001). Error bar represents the standard
750 deviation (SD) of the mean of the replicates.
751
752 **Data and code availability**
753 The RNA-seq and CUT&RUN-seq data generated in this study have been deposited at NCBI Gene Expression
754 Omnibus (GEO) database under the accession code GSE196671.
755
756

757 ## Acknowledgements

758 We thank all the members of the Siomi Laboratory for their discussions and comments on this work. We
759 thank Takehiko Yokomizo (Department of Biochemistry, Juntendo University) for providing the
760 anti-DYKDDDDK monoclonal antibody (clone 2H8). We also thank Daisuke Motooka (Research Institute for
761 Microbial Diseases, Osaka University) for generating the sequencing data. We are grateful to Azusa Inoue
762 (Center for Integrative Medical Sciences, RIKEN) and Katsuhiko Hayashi (Department of Genome Biology,
763 Osaka University) for their comments on this manuscript. This work was supported by the MEXT Grant-in-Aid
764 for Scientific Research in Innovative Areas (19H05753 to H.S. and 19H05758 to A.O.), AMED project to
765 elucidate and control mechanisms of aging and longevity (1005442 to H.S.), JSPS Grant-in-Aid for Scientific
766 Research KAKENHI (20K21507 to K.M. and 22H02534 to K.I.), Mochida Memorial Foundation Research Grant
767 to K.M., Sumitomo Foundation Research Grant to K.M., Keio University Doctorate Student Grant-in-Aid
768 Program to Y.G., and the JST Doctoral Program Student Support Fellowship to Y.G.

## Author contributions

H.S. and Y.G. conceived and designed the project; Y.G., T.K., K.M., T.D.L., A.S., H.I., and M.S. performed the experiments; Y.G. and T.K. analyzed the data; H.I. developed the ASO-mediated knockdown experiments; H.S. and K.M. supervised and coordinated the experiments; Y.G. and H.S. wrote the paper with input from all authors. All authors reviewed and approved the final version of the manuscript.

## Declaration of interests

The authors declare no competing interests.

## References


1. Lee, M.T., Bonneau, A.R. & Giraldez, A.J. Zygotic genome activation during the maternal-to-zygotic transition. *Annual review of cell and developmental biology* **30**, 581-613 (2014).
2. Tarkowski, A.K. Experiments on the development of isolated blastomeres of mouse eggs. *Nature* **184**, 1286-1287 (1959).
3. Tarkowski, A.K. Experimental studies on regulation in the development of isolated blastomeres of mouse eggs; Badania eksperymentalne nad rozwojem izolowanych blastomerów jaj myszy. *Acta Theriologica* **3**, 191-267 (1959).
4. Suwińska, A., Czołowska, R., Ożdżeński, W. & Tarkowski, A.K. Blastomeres of the mouse embryo lose totipotency after the fifth cleavage division: expression of Cdx2 and Oct4 and developmental potential of inner and outer blastomeres of 16-and 32-cell embryos. *Developmental biology* **322**, 133-144 (2008).
5. Svoboda, P. *et al.* RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Developmental biology* **269**, 276-285 (2004).
6. McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A. & Gehring, W.J. A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. *Nature* **308**, 428-433 (1984).
7. Carrasco, A.E., McGinnis, W., Gehring, W.J. & De Robertis, E.M. Cloning of an X. laevis gene expressed during early embryogenesis coding for a peptide region homologous to Drosophila homeotic genes. *Cell* **37**, 409-414 (1984).
8. McGinnis, W., Hart, C.P., Gehring, W.J. & Ruddle, F.H. Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of drosophila. *Cell* **38**, 675-680 (1984).
9. Scott, M.P. & Weiner, A.J. Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. *Proceedings of the National Academy of Sciences* **81**, 4115-4119 (1984).
10. Gehring, W.J. & Hiromi, Y. Homeotic genes and the homeobox. *Annual review of genetics* **20**, 147-173 (1986).
11. Gehring, W.J. Homeo boxes in the study of development. *Science* **236**, 1245-1252 (1987).
12. Lewin, T.D., Royall, A.H. & Holland, P.W. Dynamic Molecular Evolution of Mammalian Homeobox Genes: Duplication, Loss, Divergence and Gene Conversion Sculpt PRD Class Repertoires. *Journal of Molecular Evolution*, 1-19 (2021).

13. Töhönen, V. *et al.* Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nature communications* **6**, 8207 (2015).
14. Madissoon, E. *et al.* Characterization and target genes of nine human PRD-like homeobox domain genes expressed exclusively in early embryos. *Scientific reports* **6**, 1-14 (2016).
15. De Iaco, A. *et al.* DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nature genetics* **49**, 941-945 (2017).
16. Hendrickson, P.G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nature genetics* **49**, 925-934 (2017).
17. Whiddon, J.L., Langford, A.T., Wong, C.-J., Zhong, J.W. & Tapscott, S.J. Conservation and innovation in the DUX4-family gene network. *Nature genetics* **49**, 935-940 (2017).
18. Chen, Z. & Zhang, Y. Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nature genetics* **51**, 947-951 (2019).
19. Guo, M. *et al.* Precise temporal regulation of Dux is important for embryo development. *Cell research* **29**, 956-959 (2019).
20. De Iaco, A., Verp, S., Offner, S., Grun, D. & Trono, D. DUX is a non-essential synchronizer of zygotic genome activation. *Development* **147**, dev177725 (2020).
21. Bosnakovski, D., Gearhart, M.D., Ho Choi, S. & Kyba, M. Dux facilitates post-implantation development, but is not essential for zygotic genome activation. *Biology of reproduction* **104**, 83-93 (2021).
22. Wagner, A. Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. *Biological cybernetics* **74**, 557-567 (1996).
23. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193-196 (2014).
24. Shen, H. *et al.* Mouse totipotent stem cells captured and maintained through spliceosomal repression. *Cell* **184**, 2843-2859. e20 (2021).
25. Macfarlan, T.S. *et al.* Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes & development* **25**, 594-607 (2011).
26. Macfarlan, T.S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57-63 (2012).
27. Wu, J. *et al.* The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652-657 (2016).
28. Zhu, Y. *et al.* Relaxed 3D genome conformation facilitates the pluripotent to totipotent-like state transition in embryonic stem cells. *Nucleic acids research* **49**, 12167-12177 (2021).
29. Howe, K.L. *et al.* Ensembl 2021. *Nucleic Acids Research* **49**, D884-D891 (2020).
30. Xiong, Z. *et al.* Ultrasensitive Ribo-seq reveals translational landscapes during mammalian oocyte-to-embryo transition and pre-implantation development. *Nature Cell Biology* **24**, 968-980 (2022).
31. Markoulaki, S., Meissner, A. & Jaenisch, R. Somatic cell nuclear transfer and derivation of embryonic stem cells in the mouse. *Methods* **45**, 101-114 (2008).
32. Skene, P.J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, e21856 (2017).
33. Sasaki, F. *et al.* A high-affinity monoclonal antibody against the FLAG tag useful for G-protein-coupled receptor study. *Analytical biochemistry* **425**, 157-165 (2012).
34. Mayran, A. & Drouin, J. Pioneer transcription factors shape the epigenetic landscape. *Journal of Biological Chemistry* **293**, 13795-13804 (2018).

851 35. Liu, X. *et al.* Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos.
852 *Nature* **537**, 558-562 (2016).
853 36. Dahl, J.A. *et al.* Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition.
854 *Nature* **537**, 548-552 (2016).
855 37. Rajkovic, A., Yan, C., Yan, W., Klysik, M. & Matzuk, M.M. Obox, a family of homeobox genes preferentially
856 expressed in germ cells. *Genomics* **79**, 711-717 (2002).
857 38. Ge, S.X. Exploratory bioinformatics investigation reveals importance of "junk" DNA in early embryo
858 development. *BMC genomics* **18**, 1-19 (2017).
859 39. Zhong, Y.-f. & Holland, P.W. The dynamics of vertebrate homeobox gene evolution: gain and loss of genes in
860 mouse and human lineages. *BMC evolutionary biology* **11**, 1-13 (2011).
861 40. Wilming, L.G., Boychenko, V. & Harrow, J.L. Comprehensive comparative homeobox gene annotation in
862 human and mouse. *Database* **2015**(2015).
863 41. Lee, H.-S. & Lee, K.-A. Characterization and functional analysis of Obox4 during oocyte maturation by RNA
864 interference. *Clinical and Experimental Reproductive Medicine* **34**, 293-303 (2007).
865 42. Lee, H.S. *et al.* Obox4 critically regulates cAMP-dependent meiotic arrest and MI-MII transition in oocytes.
866 *The FASEB Journal* **24**, 2314-2324 (2010).
867 43. Lee, K.-A., Lee, H.-S., Kim, E.-Y., Kim, K.-H. & Kim, Y. Oocyte-specific homeobox 4 (Obox4) is a key regulatory
868 component of the cAMP-dependent meiotic arrest during in vitro oocyte maturation. *Fertility and Sterility*
869 **90**, S108 (2008).
870 44. Lee, H.-S., Kim, E.-Y. & Lee, K.-A. Changes in gene expression associated with oocyte meiosis after Obox4
871 RNAi. *Clinical and experimental reproductive medicine* **38**, 68 (2011).
872 45. Lee, H.-S. *et al.* Obox4-silencing-activated STAT3 and MPF/MAPK signaling accelerate nuclear membrane
873 breakdown in mouse oocytes. *Reproduction* **151**, 369-378 (2016).
874 46. Kim, H.-M. *et al.* Obox4 regulates the expression of histone family genes and promotes differentiation of
875 mouse embryonic stem cells. *FEBS letters* **584**, 605-611 (2010).
876 47. Leidenroth, A. & Hewitt, J.E. A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals
877 the ancestral DUX gene. *BMC evolutionary biology* **10**, 1-13 (2010).
878 48. Maeso, I. *et al.* Evolutionary origin and functional divergence of totipotent cell homeobox genes in eutherian
879 mammals. *BMC biology* **14**, 1-14 (2016).
880 49. Zou, Z. *et al.* Translatome and transcriptome co-profiling reveals a role of TPRXs in human zygotic genome
881 activation. *Science*, eabo7923 (2022).
882 50. Nowak, M.A., Boerlijst, M.C., Cooke, J. & Smith, J.M. Evolution of genetic redundancy. *Nature* **388**, 167-171
883 (1997).
884 51. Vavouri, T., Semple, J.I. & Lehner, B. Widespread conservation of genetic redundancy during a billion years
885 of eukaryotic evolution. *Trends in genetics* **24**, 485-488 (2008).
886 52. Hackett, J.A. & Surani, M.A. Regulatory principles of pluripotency: from the ground state up. *Cell stem cell* **15**,
887 416-430 (2014).
888 53. Sakashita, A. *et al.* Endogenous retroviruses drive species-specific germline transcriptomes in mammals.
889 *Nature Structural & Molecular Biology* **27**, 967-977 (2020).
890 54. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration.
891 *Nature Reviews Genetics* **14**, 113-124 (2013).
892 55. Zaret, K.S. Pioneer transcription factors initiating gene network changes. *Annual review of genetics* **54**, 367
893 (2020).

894 56. Eckersley-Maslin, M.A., Alda-Catalinas, C. & Reik, W. Dynamics of the epigenetic landscape during the
895 maternal-to-zygotic transition. *Nature Reviews Molecular Cell Biology* **19**, 436-450 (2018).
896 57. Grow, E.J. *et al.* p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral
897 muscular dystrophy cell models. *Nature genetics* **53**, 1207-1220 (2021).
898 58. Takeuchi, C., Murano, K., Ishikawa, M., Okano, H. & Iwasaki, W., Y. Generation of stable drosophila ovarian
899 somatic cell lines using piggyBac system. *Methods in Molecular Biology* **In press**(2022).
900 59. Guo, Y. *et al.* Potent mouse monoclonal antibodies that block SARS-CoV-2 infection. *Journal of Biological*
901 *Chemistry* **296**(2021).
902 60. Inoue, K. *et al.* Loss of H3K27me3 imprinting in the Sfmbt2 miRNA cluster causes enlargement of cloned
903 mouse placentas. *Nature communications* **11**, 1-12 (2020).
904 61. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**,
905 i884-i890 (2018).
906 62. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
907 63. Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Research* **49**, D916-D923 (2020).
908 64. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence
909 reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
910 65. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* (2021).
911 66. Zhong, Y.f. & Holland, P.W. HomeoDB2: functional expansion of a comparative homeobox gene database for
912 evolutionary developmental biology. *Evolution & development* **13**, 567-568 (2011).
913 67. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data
914 with DESeq2. *Genome biology* **15**, 1-21 (2014).
915 68. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359
916 (2012).
917 69. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, 1-9 (2008).
918 70. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids*
919 *research* **44**, W160-W165 (2016).
920 71. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison
921 and visualization. *Bioinformatics* **31**, 2382-2383 (2015).
922 72. Bailey, T.L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* **37**,
923 W202-W208 (2009).
924 73. Tu, S. *et al.* MAnorm2 for quantitatively comparing groups of ChIP-seq samples. *Genome Research* **31**,
925 131-145 (2021).
926 74. Pohl, A. & Beato, M. bwtool: a tool for bigWig files. *Bioinformatics* **30**, 1618-1619 (2014).

927

Figure 1

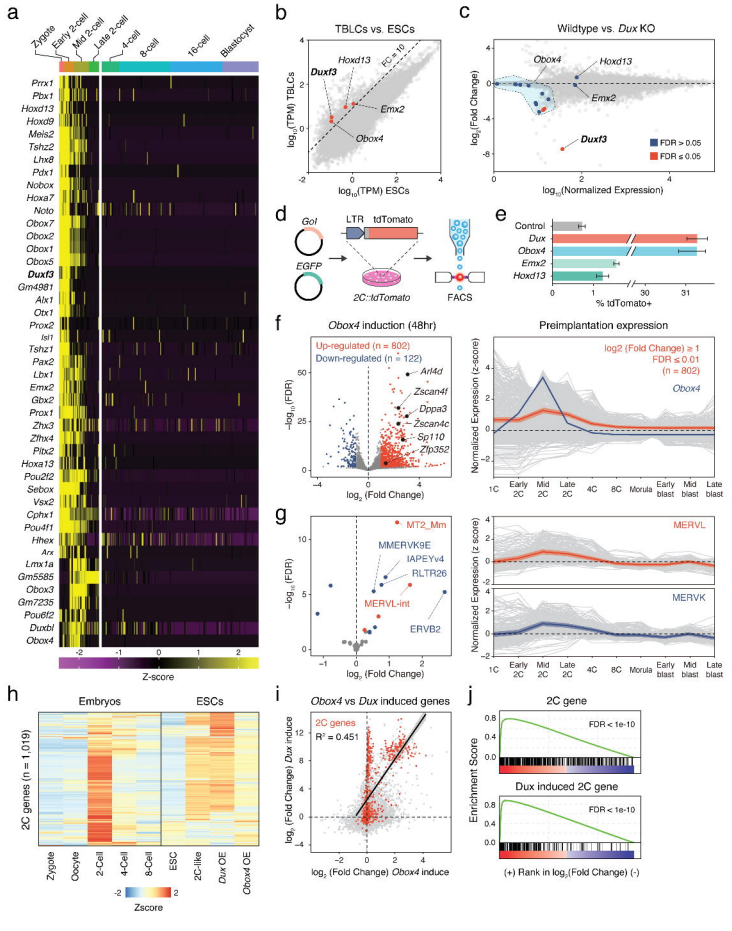
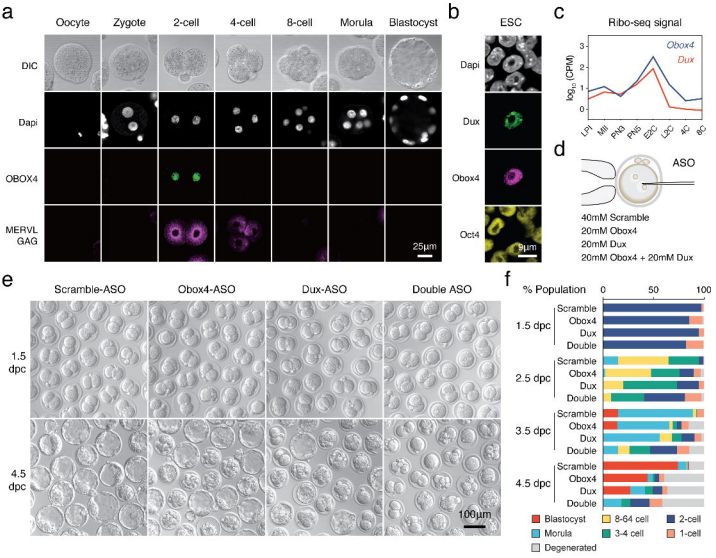

a
Zygote
Early 2-cell
Mid 2-cell
Late 2-cell
4-cell
8-cell
16-cell
Blastocyst
Prrx1
Pbx1
Hoxd13
Hoxd9
Meis2
Tsh2
Lhx8
Pdx1
Nobox
Hoxa7
Noto
Obox7
Obox2
Obox1
Obox5
Duxf3
Gm4981
Alx1
Otx1
Prox2
Isl1
Tshz1
Pax2
Lbx1
Emx2
Gbx2
Prrx1
Zhx3
Zfhx4
Pitx2
Hoxa13
Pou2f2
Sebox
Vsx2
Cphx1
Pou4f1
Hhex
Arx
Lmx1a
Gm5585
Obox3
Gm7235
Pou6f2
Duxbl
Obox4
-2
-1
0
1
2
Z-score
b
TBLCs vs. ESCs
Hoxd13
FC = 10
Duxf3
Emx2
Obox4
log10(TPM) TBLCs
log10(TPM) ESCs
c
Wildtype vs. Dux KO
Obox4
Hoxd13
Emx2
Duxf3
FDR > 0.05
FDR ≤ 0.05
log2(Fold Change)
log10(Normalized Expression)
d
Gol
LTR
tdTomato
EGFP
2C::tdTomato
FACS
e
Control
Dux
Obox4
Emx2
Hoxd13
% tdTomato+
f
Obox4 induction (48hr)
Up-regulated (n = 802)
Down-regulated (n = 122)
Arl4d
Zscan4f
Dppa3
Zscan4c
Sp110
Zfp352
−log10(FDR)
log2 (Fold Change)
Preimplantation expression
log2 (Fold Change) ≥ 1
FDR ≤ 0.01
(n = 802)
Obox4
Normalized Expression (z-score)
1C
Early 2C
Mid 2C
Late 2C
4C
8C
Morula
Early blast
Mid blast
Late blast
g
MT2_Mm
MMERVK9E
IAPEYv4
RLTR26
MERVL-int
ERVB2
MERVL
MERVK
h
Embryos
ESCs
2C genes (n = 1,019)
Zygote
Oocyte
2-Cell
4-Cell
8-Cell
ESC
2C-like
Dux OE
Obox4 OE
Zscore
i
Obox4 vs Dux induced genes
2C genes
R² = 0.451
log2 (Fold Change) Dux induce
log2 (Fold Change) Obox4 induce
j
2C gene
FDR < 1e-10
Dux induced 2C gene
Enrichment Score
(+) Rank in log2(Fold Change) (-)

Figure 2




| Against wildtype | ANOVA (Kruskal Wallis Test) | | |
|---|---|---|---|
| | 4-8 cell | Morula | Blastocyst |
| *Obox4* KO | 0.999 | 0.640 | 0.246 |
| *Dux* KO | 0.257 | 0.103 | 0.465 |
| Double KO | 0.536 | 0.106 | 0.0313* |

Figure 3

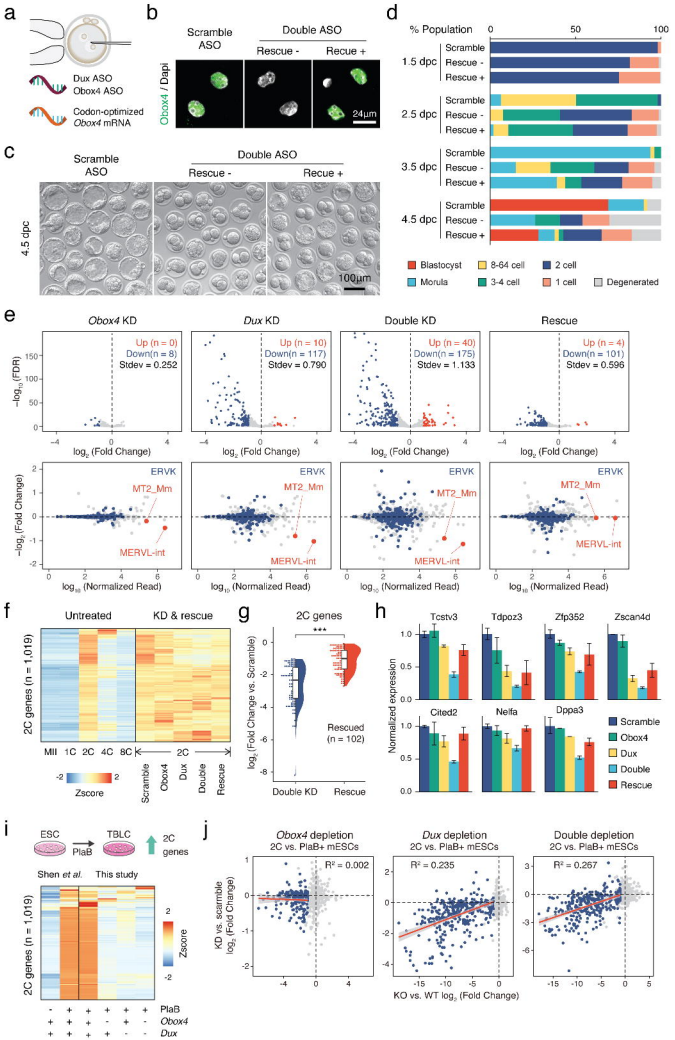
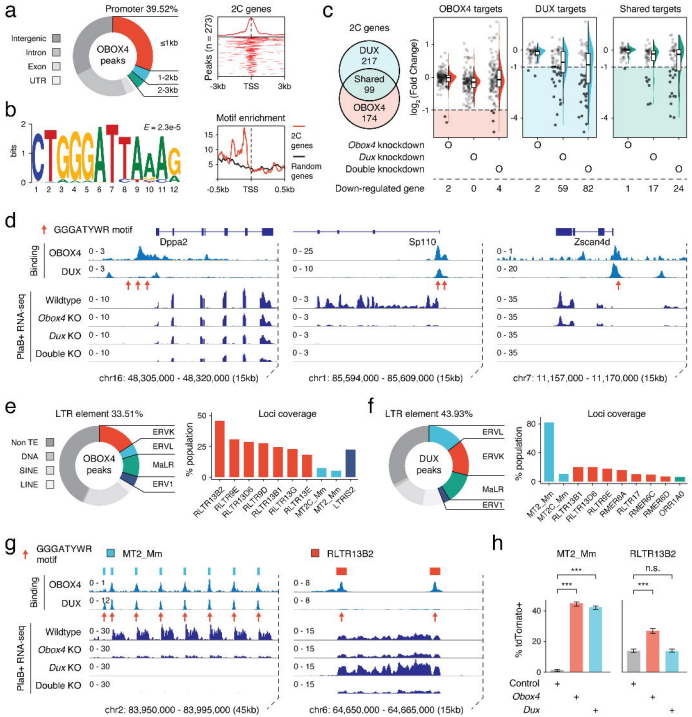
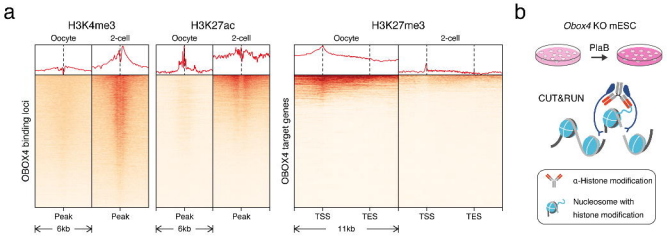

a
Dux ASO
Obox4 ASO
Codon-optimized Obox4 mRNA
b
Scramble ASO
Double ASO
Rescue -
Recue +
Obox4 / Dapi
24μm
c
Scramble ASO
Double ASO
Rescue -
Recue +
4.5 dpc
100μm
d
% Population
0
50
100
1.5 dpc
2.5 dpc
3.5 dpc
4.5 dpc
Scramble
Rescue -
Rescue +
Blastocyst
Morula
8-64 cell
3-4 cell
2 cell
1 cell
Degenerated
e
Obox4 KD
Dux KD
Double KD
Rescue
Up (n = 0)
Down(n = 8)
Stdev = 0.252
Up (n = 10)
Down(n = 117)
Stdev = 0.790
Up (n = 40)
Down(n = 175)
Stdev = 1.133
Up (n = 4)
Down(n = 101)
Stdev = 0.596
−log10 (FDR)
log2 (Fold Change)
ERVK
MT2_Mm
MERVL-int
log10 (Normalized Read)
f
Untreated
KD & rescue
2C genes (n = 1,019)
MII 1C 2C 4C 8C
Scramble
Obox4
Dux
Double
Rescue
Zscore
g
2C genes
***
log2 (Fold Change vs. Scramble)
Rescued (n = 102)
Double KD
Rescue
h
Tcstv3
Tdpoz3
Zfp352
Zscan4d
Cited2
Nelfa
Dppa3
Normalized expression
Scramble
Obox4
Dux
Double
Rescue
i
ESC
TBLC
PlaB
2C genes
Shen et al.
This study
2C genes (n = 1,019)
Zscore
PlaB
Obox4
Dux
j
Obox4 depletion
Dux depletion
Double depletion
2C vs. PlaB+ mESCs
R² = 0.002
R² = 0.235
R² = 0.267
KD vs. scramble log2 (Fold Change)
KO vs. WT log2 (Fold Change)

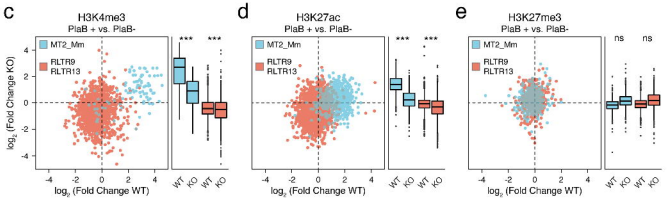
Figure 4

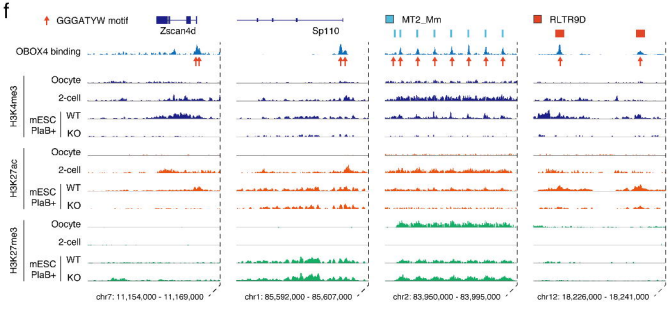


a
Promoter 39.52%
Intergenic
Intron
Exon
UTR
OBOX4 peaks
≤1kb
1-2kb
2-3kb
2C genes
Peaks (n = 273)
-3kb
TSS
3kb
b
bits
E = 2.3e-5
Motif enrichment
2C genes
Random genes
-0.5kb
TSS
0.5kb
c
2C genes
DUX 217
Shared 99
OBOX4 174
OBOX4 targets
DUX targets
Shared targets
log2 (Fold Change)
Obox4 knockdown
Dux knockdown
Double knockdown
Down-regulated gene
d
GGGATYWR motif
Dppa2
Sp110
Zscan4d
Binding
OBOX4
DUX
PlaB+ RNA-seq
Wildtype
Obox4 KO
Dux KO
Double KO
chr16: 48,305,000 - 48,320,000 (15kb)
chr1: 85,594,000 - 85,609,000 (15kb)
chr7: 11,157,000 - 11,170,000 (15kb)
e
LTR element 33.51%
Non TE
DNA
SINE
LINE
OBOX4 peaks
ERVK
ERVL
MaLR
ERV1
Loci coverage
% population
RLTR13B2
RLTR9E
RLTR13D5
RLTR9D
RLTR13B1
RLTR13G
RLTR13E
MT2C_Mm
MT2_Mm
LTRIS2
f
LTR element 43.93%
DUX peaks
ERVL
ERVK
MaLR
ERV1
Loci coverage
% population
MT2_Mm
MT2C_Mm
RLTR13B1
RLTR13D5
RLTR9E
RMER6A
RLTR17
RMER6C
RMER6D
ORR1A0
g
GGGATYWR motif
MT2_Mm
RLTR13B2
Binding
OBOX4
DUX
PlaB+ RNA-seq
Wildtype
Obox4 KO
Dux KO
Double KO
chr2: 83,950,000 - 83,995,000 (45kb)
chr6: 64,650,000 - 64,665,000 (15kb)
h
MT2_Mm
RLTR13B2
n.s.
% tdTomato+
Control
Obox4
Dux

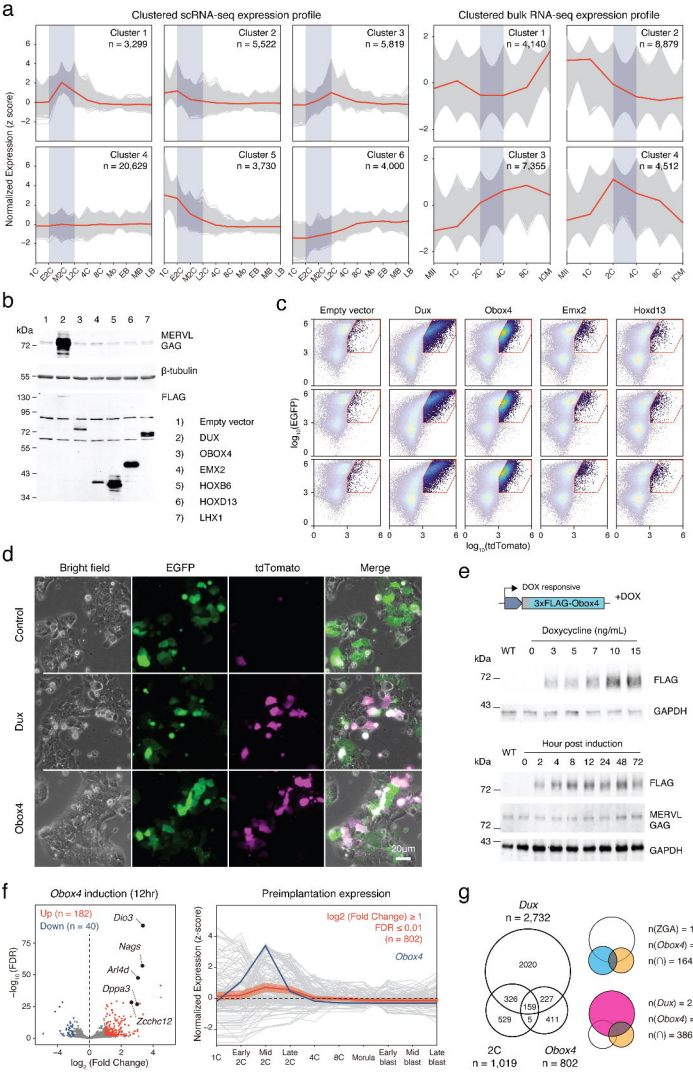
Figure 5


a
H3K4me3
H3K27ac
H3K27me3
Oocyte
2-cell
OBOX4 binding loci
OBOX4 target genes
Peak
TSS
TES
6kb
11kb
b
Oboxα KO mESC
PlaB
CUT&RUN
α-Histone modification
Nucleosome with histone modification

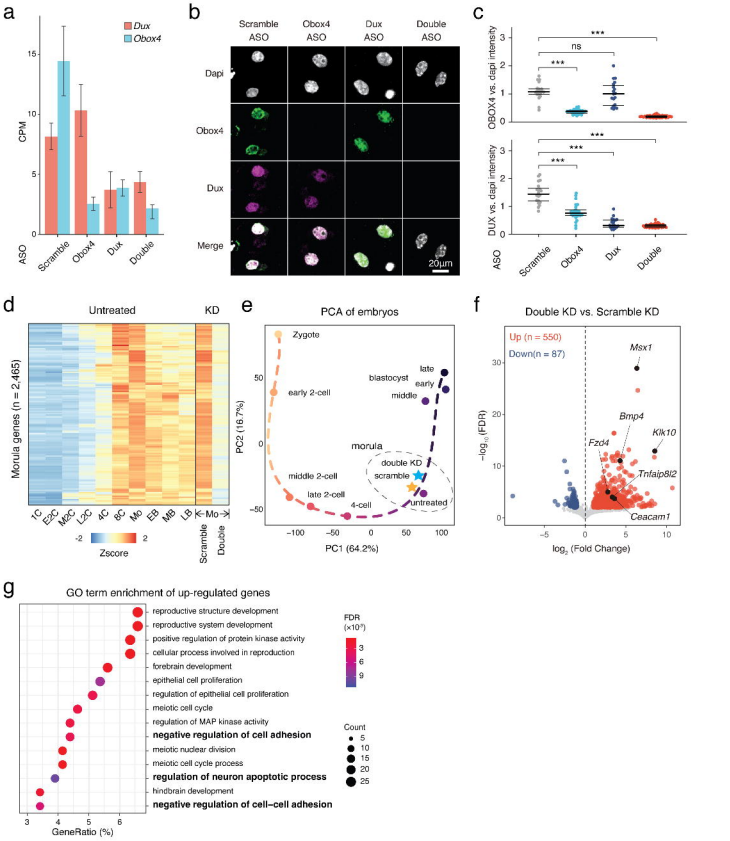

c
d
e
H3K4me3
H3K27ac
H3K27me3
PlaB + vs. PlaB-
MT2_Mm
RLTR9
RLTR13
log2 (Fold Change KO)
log2 (Fold Change WT)
***
ns
WT
KO


f
GGGATYW motif
Zscan4d
Sp110
MT2_Mm
RLTR9D
OBOX4 binding
H3K4me3
H3K27ac
H3K27me3
Oocyte
2-cell
mESC PlaB+
WT
KO
chr7: 11,154,000 - 11,169,000
chr1: 85,592,000 - 85,607,000
chr2: 83,950,000 - 83,995,000
chr12: 18,226,000 - 18,241,000

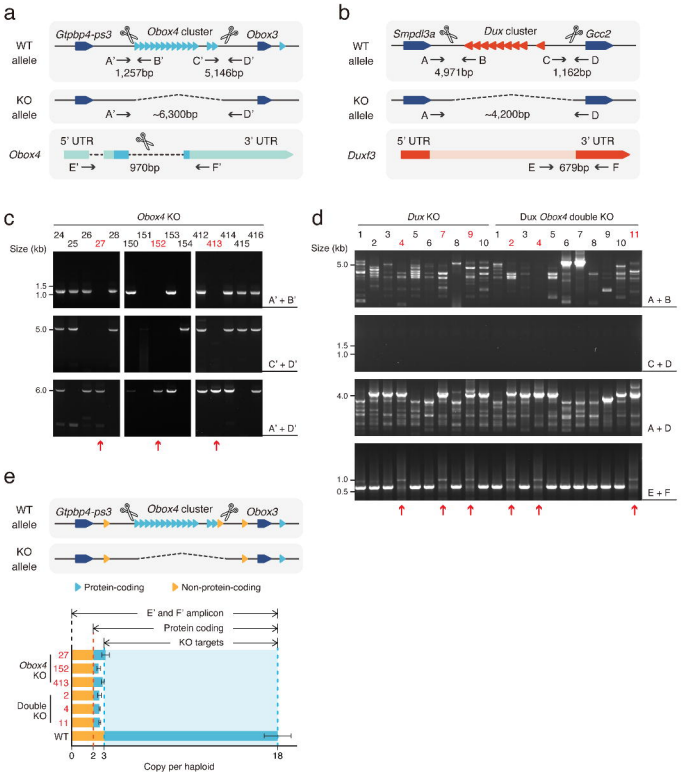
# Supplementary Figure 1



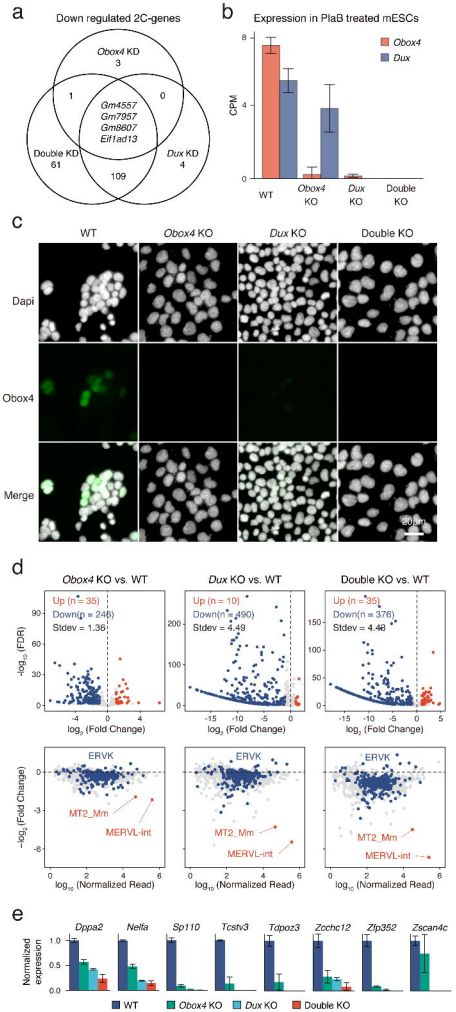
# Supplementary Figure 2



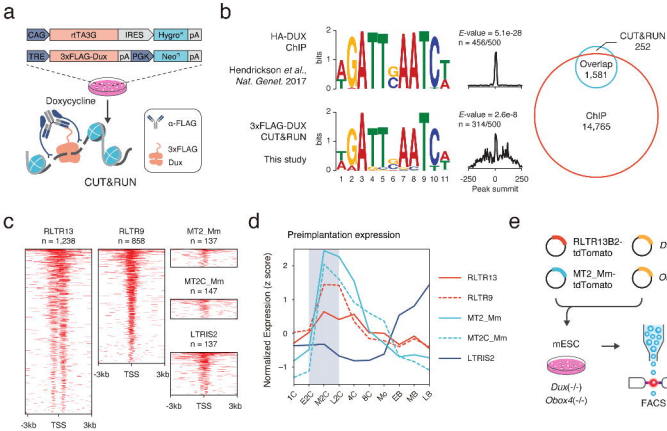
# Supplementary Figure 3



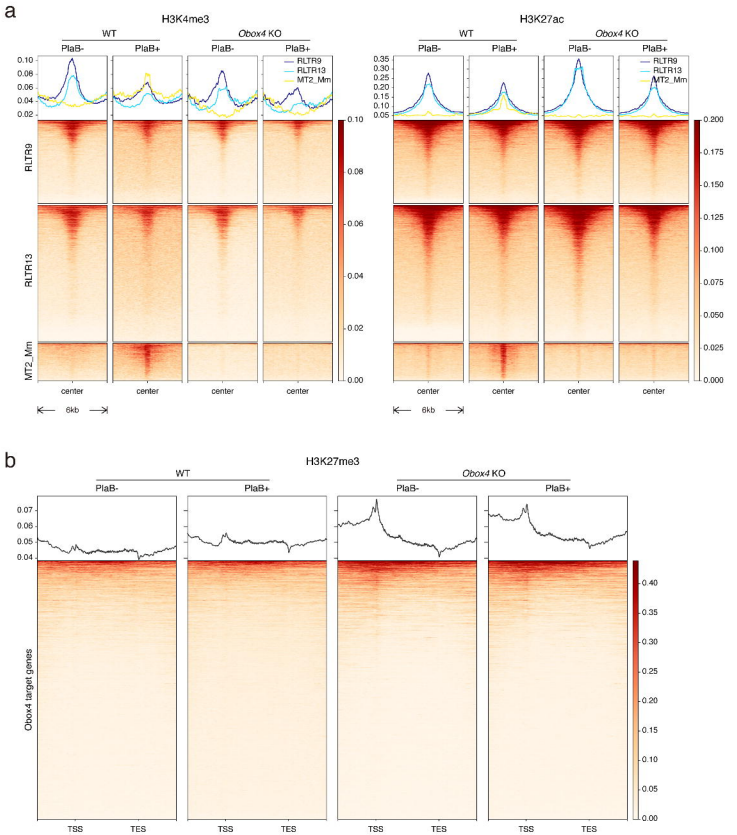
# Supplementary Figure 4



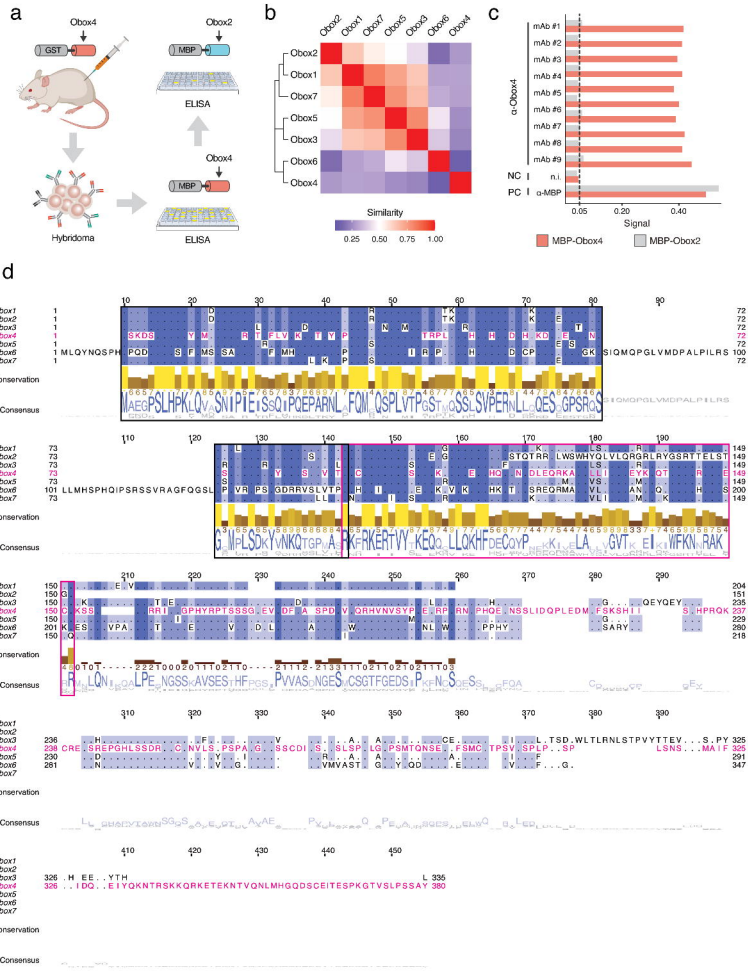
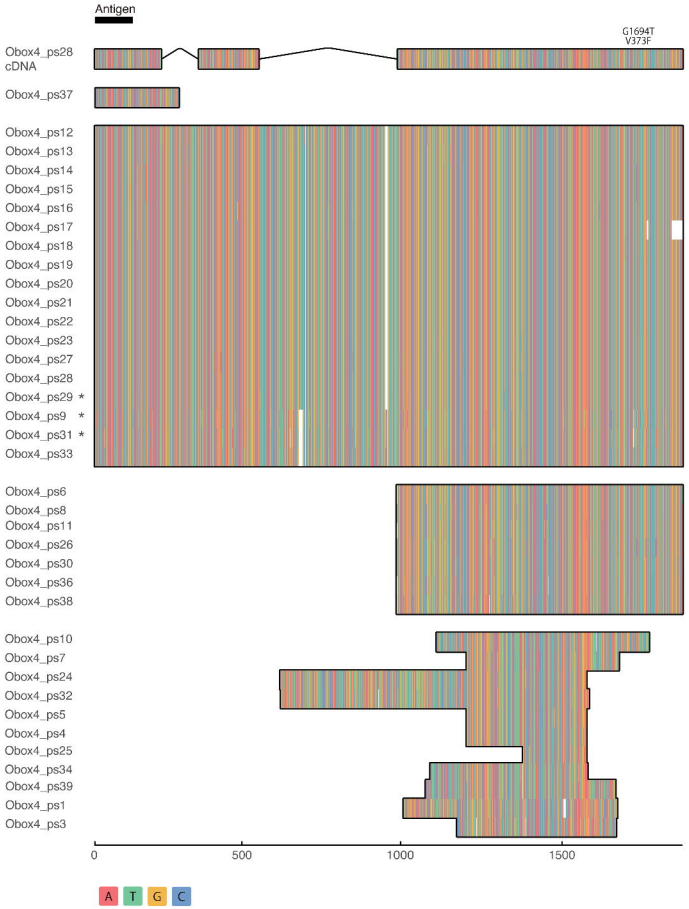
# Supplementary Figure 5



Supplementary Figure 6



# Supplementary Figure 7



Supplementary Figure 8