Obox4 secures zygotic genome activation upon loss of *Dux*

Youjia Guo[#], Tomohiro Kitano[#], Kensaku Murano, Ten D. Li, Akihiko Sakashita, Hirotsugu Ishizu, Masayuki Sato, Haruhiko Siomi*

Department of Molecular Biology, Keio University School of Medicine, Tokyo 1608582, Japan.
[#]These authors contributed equally: Youjia Guo and Tomohiro Kitano.

*Corresponding author:
Haruhiko Siomi
awa403@keio.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 pioneer factor, DUX. However, *Dux*-knockout embryos produce fertile mice, suggesting that ZGA is redundantly driven by unknown factor(s). In the present study, we report that the multicopy mouse homeobox gene, *Obox4*, encodes a transcription factor that redundantly drives ZGA. OBOX4 is highly expressed in mouse 2-cell embryos. Single knockdown of *Obox4* or *Dux* is tolerated by embryogenesis, whereas *Obox4*/*Dux* double knockdown completely compromises embryonic development. Genome-wide epigenetic profiling revealed that OBOX4 binds to MERVL LTRs as well as murine endogenous retroviruses with lysine tRNA primer (MERVK) LTRs, and mediates the deposition of active histone modifications. Our study identified OBOX4 as a pioneer 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 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 pioneer factor(s) [18-21].

In this study, we sought to identify the redundant pioneer 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 guiding the deposition of active epigenetic modifications. Taken together, our findings substantiate a pre-implantation development model, in which the initiation of ZGA is redundantly secured by *Dux* and *Obox4*.

Results

Expression profiling identifies pioneer 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 (Figure 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 derived from splicing-inhibited mESCs, and hence, are free of parentally inherited transcripts [24]. Analysis of published RNA sequencing (RNA-seq) data showed that the expression of four homeobox genes, *Duxf3*, *Emx2*, *Hoxd13*, and *Obox4*, was ten times higher in TBLCs than in ESCs (Figure 1B). We examined whether the expression of these genes was affected in the *Dux* knockout embryos. Using published RNA-seq data [18,20], we

74 found that the expression of all single-copy candidate genes and a subset of the multicopy gene *Obox4* was
75 unaffected in *Dux* knockout embryos (Figure 1C).
76
77 ***Obox4* potently activates 2C-genes and TEs**
78 Next, we examined whether these candidate genes promoted ZGA gene expression. Mouse ZGA is
79 characterized by the surging activity of genes that have co-opted MERVL LTRs as stage-specific promoters [25].
80 We constructed a 2C::tdTomato reporter mESC line with transgenic tdTomato red fluorescence protein driven
81 by MERVL 5'-LTR. [26]. Candidate genes were cloned and overexpressed in the reporter cell line (Figures 1D and
82 S1A). Interestingly, ectopic expression of *Obox4* markedly induced tdTomato expression, similar to that
83 induced by *Dux* (Figures 1E, 1F, and S1B). To characterize the impact of *Obox4* expression on the mESC
84 transcriptome, we established an ESC stable line bearing an inducible *Obox4* transgene and performed
85 RNA-seq analysis with or without induced *Obox4* expression (Figure S1C and S1D). Differentially expressed
86 gene analysis revealed that *Obox4* expression is associated with drastic transcriptome changes in mESCs
87 (Figure 1G). Specifically, *Obox4* expression resulted in the upregulation of genes that were specifically
88 expressed in early-to-middle 2C-embryos (Figure 1H). Typical 2C-genes including *Zscan4*, *Zfp352*, *Dppa3*, and
89 *Sp110*, as well as 2C-specific TEs, MERVL and MERVK, were highly activated upon *Obox4* expression (Figure 1I
90 and 1J). These results indicated that *Obox4* is a potent inducer of the 2C-like state in mESCs.
91
92 ***Obox4*/*Dux* double knockdown compromises pre-implantation development**
93 We then examined whether *Obox4* was crucial for ZGA, particularly in the absence of *Dux*. First, we
94 confirmed the presence of endogenous OBOX4 protein, as the *Obox4* loci are marked as pseudogenes in the
95 current genome annotation GRCm39 [27]. We generated mouse monoclonal antibodies that recognized OBOX4
96 (Figure S2A). The sequence of OBOX4 was significantly divergent from those of other OBOX members (Figure
97 S2B), and monoclonal antibodies were negative for cross-reactivity with OBOX2 (Figure S2C).
98 Immunofluorescence staining with anti-OBOX4 monoclonal antibodies confirmed that endogenous OBOX4 was
99 highly abundant in 2C-embryos (Figure 2A) and 2C-like mESCs (Figure 2B). To determine whether *Obox4* is
100 essential for pre-implantation development, microinjection of antisense oligonucleotide (ASO) into male
101 pronuclei of zygotes was performed to knockdown *Obox4* and *Dux* (Figure 2C). At the transcript level, both
102 *Obox4* and *Dux* knockdown reduced *Obox4* expression in 2C-embryos (Figure 2D). At the protein level, ASO
103 targets were depleted in knockdown 2C-embryos (Figure 2E and 2F), whereas OBOX4 was retained in *Dux*
104 knockdown 2C-embryos (Figure 2G). Monitoring development until 4.5 days post cotium (dpc) showed that
105 *Obox4* single knockdown resulted in moderate developmental retardation, with almost 50% of embryos
106 reaching the blastocyst stage. Similarly, blastocyst formation was preserved in nearly 25% of *Dux* single
107 knockdown embryos, which is consistent with previously reported *Dux* knockdown/knockout experiments
108 [15,18-21]. Intriguingly, the *Obox4*/*Dux* double knockdown completely disrupted pre-implantation development,
109 with most embryos arrested at the 2C stage (Figure 2H and 2I). These observations showed that the
110 expression of *Obox4* and *Dux* is collectively required for pre-implantation development and strongly suggest
111 that *Obox4* is capable of securing ZGA in a *Dux*-independent manner.
112

113 **OBOX4 rescues *Obox4/Dux* double knockdown in pre-implantation embryos**

114 We then sought to verify whether OBOX4 is essential for pre-implantation development in the absence of
115 DUX, by determining whether the presence of OBOX4 can rescue the developmental arrest of *Obox4*/*Dux*
116 double knockdown embryos. First, codon-optimized *Obox4* mRNA without the ASO target motif was produced
117 using *in vitro* transcription. The rescue experiment was performed by means of co-microinjection of *Obox4*
118 mRNA with *Obox4*/*Dux* double knockdown ASOs (Figure 3A). Restoration of OBOX4 in *Obox4*
119 mRNA-microinjected embryos was confirmed using immunofluorescence staining at the 2C stage (Figure 3B).
120 As expected, development monitoring showed that *Obox4*/*Dux* double knockdown embryos failed blastocyst
121 formation at 4.5 dpc, whereas among *Obox4* mRNA-rescued embryos, blastocyst formation was retained at a
122 similar level to that observed in the *Dux* single knockdown experiment (Figures 3C, 3D, 2H, and 2I). These
123 results suggested that OBOX4 is responsible for securing development in the absence of DUX. As *Obox4* has
124 been shown to activate 2C-genes and TEs in mESCs, we questioned whether embryonic depletion of OBOX4
125 impairs ZGA. Consistent with the observation of embryonic development, transcriptome analysis revealed that
126 the ZGA pattern of *Dux*/*Obox4* double knockdown 2C-embryos was markedly different from that of control
127 embryos, while single knockdown embryos were less affected (Figure 3E). The dysregulated transcriptomes
128 were characterized by extensive hypoactivation of MERVL/MERVK elements (Figure 3F) and 2C-genes (Figure
129 3G), suggesting that the compromised pre-implantation development was due to ZGA failure.

130

131 ***Obox4* and *Dux* are essential for ESC-to-TBLC transition**

132 In 2021, it was reported that administration of the splicing inhibitor pladienolide B (PlaB) to ESC culture
133 drives pluripotent-to-totipotent transition, which results in TBLCs that are functionally and molecularly
134 comparable with totipotent blastomeres [24]. Since it remains unclear whether the ESC-to-TBLC transition relies
135 on pioneer factor activity, we attempted to knockout *Obox4* and *Dux* in mESCs, to examine whether they
136 contribute to *in vitro* totipotency acquisition in TBLCs. Following a previously published protocol [18], plasmids
137 encoding Cas9 and single guide RNA (sgRNA) flanking the *Dux* cluster were introduced into wild-type (WT)
138 mESCs, to knockout the *Dux* loci (Figure S3A). Plasmids encoding Cas9 and sgRNAs flanking the *Obox4* cluster,
139 and sgRNAs targeting the *Obox4* internal sequence were then introduced into *Dux* knockout mESCs and WT
140 mESCs, to knockout the *Obox4* loci (Figure S3B). Bi-allelic deletions of the *Dux* cluster and *Dux* internal
141 sequences were confirmed in the monoclonized knockout cell lines (Figure S3C and S3D). Deletion of the
142 *Obox4* cluster was confirmed in the monoclonized cell lines (Figure S3E). Knockout mESCs were treated with
143 PlaB to enforce TBLC transition and then examined for the depletion of OBOX4 at the protein level.
144 Immunostaining and western blotting revealed OBOX4 depletion in *Obox4/Dux* double knockout and *Obox4*
145 single knockout cell lines (Figure S3F and S3G). The cell lines were subjected to transcriptome profiling for
146 further validation (Figure 4A). RNA-seq analysis showed that both *Obox4* and *Dux* were highly expressed in
147 WT TBLC, while a single knockout of either gene resulted in inactivation of the other (Figure 4B), which was
148 not observed in embryo knockdown experiments (Figure 2D), suggesting that *Obox4* and *Dux* are individually
149 required for TBLC transition. Indeed, transcriptome analysis showed that *Dux* and *Obox4* single knockout was
150 sufficient to hinder the activation of TBLC-specific genes and TEs, while double knockout led to the
151 dysregulation of the TBLC transcriptome to a greater extent (Figure 4C and 4D). Consistent with the

152 observation in embryos, the expression of critical 2C-genes was abolished in PlaB-treated *Dux*/*Obox4* double
153 knockout mESCs (Figure 4E).
154
155 **Obox4 binds to 2C-gene promoters and LTR elements**
156 *Obox4* contains a homeobox, activates 2C-genes and TEs, and is essential for *Dux*-less pre-implantation
157 embryogenesis, which prompted us to examine whether OBOX4 is a pioneer factor that activates
158 ZGA-associated loci through direct binding with *cis*-regulatory elements. Cleavage under targets and release
159 using nuclease sequencing (CUT&RUN-seq) is an alternative method to chromatin immunoprecipitation
160 sequencing (ChIP-seq) that profiles protein-DNA interactions. This technique combines antibody-targeted
161 cleavage of proximal DNA with high-throughput sequencing, to identify the binding sites of DNA-associated
162 proteins [28]. To profile genome-wide OBOX4 binding, we performed CUT&RUN-seq against triple FLAG-tagged
163 OBOX4 (3×FLAG-OBOX4) expressed in mESCs, using a high-affinity anti-FLAG antibody [29] (Figure 5A).
164 CUT&RUN-seq yielded ~11,000 peaks that were shared among biological triplicates over the control samples
165 (q<0.05). Peak annotation revealed that OBOX4 was enriched in the promoter region and LTR elements
166 (Figure 5B). As expected, we found that most genes (3304 of 5172, 63.9%) whose promoters were occupied
167 by OBOX4 were highly expressed in 2C-embryos. Unlike DUX, which strictly binds to a conserved motif
168 exclusively found in MERVL LTR [17], OBOX4 displayed a broad binding spectrum that was strongly biased
169 towards both LTRs of MERVK and MERVL (Figure 5C). Consistently, *de novo* motif discovery failed to identify
170 a consensus OBOX4 binding motif shared among the MERVK and MERVL elements, while clear patterns were
171 discovered in each LTR element family (Figure 5D). Hence, we hypothesized that OBOX4 binds to the MERVL
172 and MERVK elements, to promote downstream gene expression. To test this hypothesis, we constructed
173 tdTomato reporters driven by LTRs of MERVL or MERVK and co-transfected them with *Dux* and *Obox4* into
174 *Dux*/*Obox4* double knockout ESCs (Figure 5E). Fluorescence-activated cell sorting (FACS) analysis of the
175 tdTomato-positive population showed that *Obox4* activated both MERVK and MERVL LTRs, whereas *Dux* only
176 activated MERVL LTR (Figure 5F). As OBOX4 and DUX share the MERVL LTR as binding targets, we reasoned
177 that OBOX4 and DUX are mutually redundant factors that activate core totipotency genes by activating their
178 MERVL LTR promoters. Indeed, by comparing OBOX4 CUT&RUN-seq and previously published DUX ChIP-seq
179 data [16], we found that OBOX4 targets more than half (1,089 out of 2,068) of the DUX target genes (Figure 5G),
180 among which are critical totipotency genes, including *Dppa3*, *Sp110*, and *Zscan4d* (Figure 5H).
181
182 **OBOX4 guides deposition of active histone marks**
183 Pioneer factors have been proposed to modulate the epigenetic landscape at their binding loci [30]. We
184 analyzed previously published ChIP-seq data [31,32] and found that OBOX4 target loci were enriched with the
185 active histone marks H3K4me3 and H3K27ac (Figure 6A) in 2C-embryos. To examine whether OBOX4
186 mediates the deposition of active histone modifications, we performed CUT&RUN-seq in WT and *Obox4*
187 knockout mESCs and TBLCs (Figure 6B). Differential analysis of histone modification revealed that upon PlaB
188 treatment, *Obox4* knockout mESCs manifested extensive loss of H3K4me3, but not H3K27me3, at the OBOX4
189 target MERVL and MERVK loci (Figure 6C and 6D). We then investigated whether the MERVL and MERVK loci
190 were discriminated in terms of histone modifications. Interestingly, most of the strongest OBOX4-associated

191 H3K27ac loci were deposited at the ERVK, but not ERVL loci (Figure 6E). As H3K27ac is highly associated with
192 enhancer activity, together with our finding that *Obox4* activates MERVK LTR and is associated with intense
193 H3K27ac deposition, we speculated that despite being functionally redundant to *Dux*, *Obox4* also acts as an
194 activator of ZGA-specific ERVK enhancers. This hypothesis is supported by the observation that in addition to
195 H3K4me3 enrichment at promoter regions, H3K27ac is also deposited at OBOX4 target ERVKs that are
196 proximal to critical ZGA-genes (Figure 6F).
197

198 ## Discussion

199 Starting with a transcriptionally inert genome, the eutherian ZGA engages in massive yet coordinated
200 expression of genes driven by LTR TEs. In mice, it has been reported that the pioneer factor DUX accesses and
201 opens condensed chromatin of the MERVL LTR loci, which results in the activation of downstream ZGA genes
202 [15-17]. However, the establishment of fertile *Dux* knockout mice suggests that ZGA is redundantly fulfilled by
203 unknown pioneer factor(s) [18-21]. In the present study, we found that *Obox4*, a cryptic multi-copy cluster gene,
204 encodes a homeodomain-containing protein that potently induces ZGA gene expression. Remarkably,
205 *Dux*/*Obox4* double knockdown completely compromised pre-implantation development, which tolerated
206 single knockdown of either gene. Double knockdown embryos exhibited severely dysregulated transcriptomes,
207 characterized by MERVL and ZGA gene activation failure. We then focused on characterizing the molecular
208 mechanisms underlying the biological significance of *Obox4* as well as its relevance to *Dux*. Multi-omics
209 analysis revealed that OBOX4 directly binds to MERVL (the target of DUX) and MERVK loci, to mediate active
210 histone mark deposition. We demonstrated that *Obox4* activates MERVL and MERVK in a *Dux*-independent
211 manner and that the *Obox4*-targeted MERVK elements are enriched with the active enhancer mark H3K27Ac
212 in TBLCs. In summary, our results highlight that OBOX4 is a pioneer factor that is functionally redundant to
213 DUX during ZGA.

214 The human PRD-like homeobox gene, TPRX2, is highly expressed during ZGA [14]. Ectopic expression of
215 *TPRX2* in human embryonic stem cells (hESCs) and fibroblasts alters the expression of genes related to
216 totipotency and pluripotency [14,33]. *Obox* is a divergent ortholog of *TPRX2* in mice [33]. Obox transcripts were first
217 discovered in gonads [34], and were later found to be highly abundant in mouse 2C-embryos [35]. The *Obox* family
218 has 67 members clustered in the sub-telomeric region of mouse chromosome 7 and is further divided into six
219 subfamilies: *Obox1*, *Obox2*, *Obox3*, *Obox4*, *Obox5*, and *Obox6*, with *Obox4* constituting nearly 60% (39 of 67)
220 of the family population [36,37]. Although *Obox4* dominates the *Obox* family population, members of the entire
221 subfamily have been annotated as pseudogenes and have largely been neglected from investigation. Despite
222 being abundantly transcribed in the gonads, the biological significance of *Obox* genes is poorly understood.
223 The role of *Obox4* is particularly mysterious, as it remains to be determined whether *Obox4* is a functional
224 protein-coding gene. Transcription of *Obox4* was originally reported to be testis-specific and contradicted by
225 the detection of *Obox4* transcripts in the ovaries and oocytes [38]. *Obox4* is involved in oocyte maturation and
226 ESC differentiation [39-43]. However, these observations are supported by limited evidence, with no evidence at
227 the protein level. By detecting OBOX4 in various totipotent cell types using high-quality monoclonal antibodies,
228 we provided decisive evidence that *Obox4* is a functional protein-coding gene. Our data demonstrated that

229 OBOX4 binds to genomic loci of MERVL-derived promoters and MERVK-derived enhancers and mediates their
230 activation during ZGA. However, whether and how other *Obox* family members contribute to ZGA remains
231 unclear.

232 Genetic redundancy is widespread in higher organisms and typically arises in signaling networks, in
233 which multiple functionally overlapping factors commit to a shared teleological objective, to counteract
234 sporadic mutations or defects [44,45]. This principle is of particular importance to embryogenesis, a process that
235 gives rise to a reproducing organism. Indeed, genetic redundancy in pluripotency has been evidenced by the
236 observation that naïve pluripotency can be elicited and maintained through various "independent inputs that
237 operate through both unique and convergent targets" [46]. Although the molecular pathway that governs
238 totipotency has been largely restricted to the scope that centers on *Dux*, it is reasonable to assume that the
239 acquisition of totipotency also complies with the rule of redundancy. This assumption was validated by the
240 generation of fertile *Dux* knockout mice; however, the panorama of ZGA redundancy remained hidden. Our
241 work on *Obox4* sheds light on this topic by discovering that *Obox4* serves as a redundant gene to *Dux* during
242 ZGA and is a bivalent activator of MERVL and MERVK elements. Although the function of MERVL has been
243 extensively investigated, particularly during early embryogenesis, the role of MERVK remains unclear.
244 Interestingly, recent studies have suggested that MERVK serves as a meiosis-specific enhancer of
245 spermatogenesis [47]. Mammalian Y chromosomes have undergone rapid evolution and have been intensively
246 inserted by TEs [48]. As *Obox4* is highly expressed in the testes, it is worth investigating whether *Obox4* plays a
247 regulatory role during spermatogenesis, by activating MERVK-derived enhancers.

248 Unlike *Dux*, which is structurally and functionally conserved throughout placentalia [49], the ancestral locus
249 of the *Obox* family appears to have undergone mouse-specific duplication and generated a gene cluster that is
250 collectively syntenic to the *Tprx2* locus in other mammals [33]. Despite the distal homology between *Obox4* and
251 *Tprx2*, it has been reported that human *TPRX2* is expressed in 8-cell embryos [14] and that ectopic expression of
252 *TPRX2* in hESCs affects the expression of totipotency- and pluripotency-related genes [14,33]. Whether *TPRX2*
253 plays a similar role in humans to *Obox4* in mice remains to be elucidated. In addition, *Obox4* and *Dux* are
254 zygotically expressed, denying their role as ZGA initiators. How they are activated after fertilization and
255 whether they share the same activation mechanism remain elusive. Both *Dux* and *Obox4* are tandem repeat
256 genes that cluster in either sub-telomeric or centromeric regions and form heterochromatin in most cell
257 lineages. These observations have led to the hypothesis that pioneer factors, such as *Obox4* and *Dux*, are
258 passively activated by biased epigenetic remodeling towards their loci.
259

260 Figure Legends

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

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

265 (B) Dot plot showing per-gene normalized read counts in mouse totipotent TBLCs *versus* ESCs. Genes with
266 more than 10-fold normalized read counts (FC=10) have been highlighted.

267 (C) Bar graph displaying gene expression in *Dux* knockout 2C-embryos *versus* WT 2C-embryos. Loss of *Dux*
268 partially deactivates the multicopy homeobox gene *Obox4*.
269 (D) Diagram of the 2C::tdTomato reporter assay. ESCs bearing the tdTomato expression cassette under the
270 control of the MERVL promoter showed red fluorescence upon entering the 2C-like state. The expression
271 of the pioneer factor increased the 2C-like population, as detected using FACS. An EGFP expression
272 plasmid was co-transfected with a gene of interest, to normalize the transfection efficiency.
273 (E) Boxplot showing normalized 2C-like cell percentage in 2C::tdTomato reporter ESCs overexpressing
274 candidate pioneer factors. *Dux* and *Obox4* potently induced a 2C-like state.
275 (F) Fluorescence microscope view of 2C::tdTomato ESCs transfected with empty plasmid, plasmid encoding
276 DUX, or plasmid encoding OBOX4. Plasmids encoding EGFP were co-transfected to normalize the
277 transfection efficiency in FACS analysis.
278 (G) Volcano plot of DEGs in *Obox4*-overexpressing mESCs. Representative ZGA genes are labeled with gene
279 symbols.
280 (H) Expression profile of genes up-regulated by *Obox4* during embryogenesis.
281 (I) Volcano plot of differentially expressed transposable elements in *Obox4*-overexpressing ESCs. ERVL and
282 ERVK were activated upon *Obox4* induction.
283 (J) Expression profile of ERVL and ERVK elements during pre-implantation embryogenesis.
284
285 **Figure 2. OBOX4 is necessary for DUX-less preimplantation development.**
286 (A) Immunofluorescence staining of histone H3, OBOX4, and MERVL GAG at different pre-implantation
287 embryo stages.
288 (B) Immunofluorescence staining of DUX, OBOX4, and OCT4 in 2C-like ESCs.
289 (C) Schematic representation of *Dux* and *Obox4* knockdown experiments in pre-implantation embryos. Male
290 pronuclei of zygotes were microinjected with ASO targeting *Dux* or *Obox4* transcripts. The embryos were
291 monitored for development until 4.5 dpc, or subjected to RNA-seq at the middle 2C stage.
292 (D) Transcript levels of *Dux* and *Obox4* in ASO-injected embryos, as measured using RNA-seq. CPM, counts per
293 million
294 (E) Immunofluorescence staining of OBOX4 in embryos microinjected with *Obox4* ASO.
295 (F) Immunofluorescence staining of DUX in embryos microinjected with *Dux* ASO.
296 (G) Immunofluorescence staining of OBOX4 in embryos microinjected with *Dux* ASO.
297 (H) Development monitoring of KD embryos at 1.5 dpc and 4.5 dpc.
298 (I) The percentages of embryonic stages observed at 1.5 dpc, 2.5 dpc, 3.5 dpc, and 4.5 dpc. The plot represents
299 the sum of three independent experiments, with 150–200 embryos per condition.
300
301 **Figure 3. *Obox4*/*Dux* double knockdown embryos manifest a dysregulated transcriptome.**
302 (A) Schematic of the double knockdown rescue experiment. Male pronuclei of zygotes were injected with ASO
303 targeting *Obox4* and *Dux* transcripts as well as *in vitro*-transcribed codon-optimized *Obox4* mRNA.
304 Embryos were monitored for development until 4.5 dpc.

305 (B) Immunofluorescence staining of OBOX4 in embryos microinjected with scrambled ASO, double ASO
306 (*Obox4* and *Dux*), or double ASO with codon-optimized *Obox4* mRNA.
307 (C) Development monitoring of knockdown and rescue embryos at 1.5 dpc and 4.5 dpc.
308 (D) The percentages of embryonic stages observed at 1.5 dpc, 2.5 dpc, 3.5 dpc, and 4.5 dpc. The plot represents
309 the sum of three independent experiments, with 80–100 embryos per condition.
310 (E) Heatmap of expression level of 2C-genes in ASO-injected embryos.
311 (F) Volcano plot showing the results of DEG analysis of 2C-genes and TEs in knockdown embryos, as
312 compared to that in scramble ASO-injected embryos. Standard deviations of $\log_2$(fold-change) were used
313 to represent the degree of transcriptome dysregulation. Stdev, standard deviation
314 (G) Bar chart showing the differential expression of representative 2C-genes in ASO-injected embryos; n=3
315 biological replicates.
316
317 **Figure 4. *Obox4* is necessary for TBLC formation.**
318 (A) Schematic diagram of *Obox4* and *Dux* knockout using CRISPR-Cas9 in mESCs. Previously reported sgRNAs
319 flanking the *Dux* cluster were introduced into mESCs with SpCas9, to knockout *Dux*. sgRNAs flanking the
320 *Obox4* cluster as well as the homeobox sequence of individual *Obox4* were introduced into mESCs with
321 SpCas9, to knockout *Obox4*. WT and knockout mESCs were treated with PlaB, to induce TBLC formation.
322 RNA-seq was performed on untreated and PlaB-treated mESCs, to profile transcriptome changes.
323 (B) Bar charts showing transcript levels of *Obox4* and *Dux* in mESCs and TBLCs are described in (A). WT,
324 wild-type; mean ± SD (n=3 biological replicates).
325 (C) Heatmap of TBLC gene expression levels in normally cultured mESCs and PlaB-treated mESCs.
326 (D) Volcano plot showing differential expression of TBLC genes and TEs in PlaB-treated knockout mESCs, as
327 compared to that in WT TBLCs. Standard deviations of all $\log_2$ values (fold change) were used to represent
328 the degree of transcriptome dysregulation. Stdev, standard deviation
329 (E) Bar chart showing differential expression of representative 2C-genes (upper panel) and pluripotency
330 genes (lower panel) in PlaB-treated knockout mESCs, as compared to that in WT TBLCs; n=3 biological
331 replicates.
332
333 **Figure 5. OBOX4 binds to LTR elements and activates 2C-genes.**
334 (A) Schematic diagram of the CUT&RUN-seq workflow using ESCs with induced expression of FLAG-tagged
335 OBOX4. ESCs bearing a doxycycline-inducible 3×FLAG-OBOX4 expression cassette were induced with
336 doxycycline for 24 h. OBOX4-associated DNA was pulled down in the CUT&RUN assay, using a high-affinity
337 anti-FLAG antibody.
338 (B) Pie-chart displaying proportions of annotated features in the OBOX4 binding sites.
339 (C) Heatmap showing the OBOX4 binding site distribution near 2C-gene promoters and LTR elements. OBOX4
340 binding is highly condensed in the 2C-gene TSS.
341 (D) The OBOX4 binding motif was enriched *de novo* with the MEME-ChIP algorithm, using the top 500
342 CUT&RUN-seq peaks discovered at the MERVK loci or all peaks discovered at the MERVL loci. The stacked
343 letter suggested divergence of the motif, which matched 339 out of 500 peaks at the MERVK and 126 out of

344 138 peaks at the MERVL loci. The histogram given in the figure displays the distribution of conserved
345 motifs within the 500 nt windows flanking the peak summits.
346 (E) Schematic design of the reporter assay. Plasmids bearing a tdTomato ORF downstream of the ERVL or
347 ERVK LTR were co-transfected with *Dux* or *Obox4* expression plasmids. Activation of LTR elements
348 resulted in an increased red fluorescence-positive mESC population, which was measured using FACS.
349 EGFP expression plasmid was co-transfected into the culture, following which the green
350 fluorescence-positive population was measured, to normalize the transfection efficiency.
351 (F) Box plot showing the percentage of red fluorescence-positive mESCs upon expression of *Dux* or *Obox4*.
352 Both *Dux* and *Obox4* can activate ERVL, whereas only *Obox4* activates ERVK.
353 (G) Euler diagram showing the overlap of DUX and OBOX4 target genes. Selected 2C-genes are listed on the
354 right.
355 (H) Genomic track view showing DUX and OBOX4 binding sites at the *Sp110, Dppa3*, and *Zscan4d* loci and
356 their expression levels upon *Obox4* overexpression. Reads were converted to CPM for normalization of the
357 sequencing depth. The OBOX4 binding sites overlapped with those of DUX. *Sp110, Dppa3*, and *Zscan4d*
358 were highly expressed upon overexpression of *Obox4*.
359
360 **Figure 6. OBOX4 guides deposition of active histone marks at its binding loci.**
361 (A) Heatmap showing the distributions of H3K4me3, H3K27Ac, and H3K27me3 near the OBOX4 binding sites
362 in 2C-embryos.
363 (B) Schematic diagram showing the strategy for profiling histone modifications in PlaB-treated mESCs.
364 PlaB-treated or untreated WT and *Obox4* knockout mESCs were examined using CUT&RUN-seq, with
365 antibodies against H3K4me3, H3K27Ac, and H3K27me3. Loci subjected to TBLC and *Obox4*-associated
366 histone modifications were identified by comparing the signal intensities between PlaB-treated and
367 non-treated mESCs.
368 (C) Change in H3K4me3 modification intensity at OBOX4-binding ERV loci in PlaB-treated mESCs, as
369 compared to that in PlaB-untreated mESCs.
370 (D) Change in H3K27me3 modification intensity at OBOX4-binding ERV loci in PlaB-treated mESCs, as
371 compared to that in PlaB-untreated mESCs.
372 (E) Change in H3K27Ac modification intensity at OBOX4-binding ERV loci in PlaB-treated mESCs, as compared
373 to that in PlaB-untreated mESCs.
374 (F) Genomic track view of histone modification levels at *Dppa3*, *Zscan4c*, and selected ERVK loci in
375 PlaB-treated WT and *Obox4* knockout mESCs. Reads were converted to CPM, for normalization of the
376 sequencing depth. OBOX4 binding is associated with the deposition of active histone modifications
377 H3K4me3 and H3K27Ac, but not with the repressive histone modification H3K27me3.
378
379 **Supplemental figure legends**
380
381 **Figure S1**
382 (A) Western blot showing protein products of the candidate genes ectopically expressed in mESCs.

383 (B) FACS analysis of 2C::tdTomato cells co-transfected with candidate gene expression plasmids and EGFP.
384 (C) Schematic of *Obox4* inducible cell line construction.
385 (D) Western blot showing 3×FLAG-OBOX4 in *Obox4* inducible cell line, induced by different concentrations of
386 doxycycline. Expression of 3×FLAG-OBOX4 was carried out in a dose-dependent manner.
387
388 **Figure S2**
389 (A) Schematic of the production of monoclonal antibodies. Mice were immunized with the GST-OBOX4 fusion
390 protein. After fusing B-cells with myelomas, the antibodies produced by hybridomas were examined
391 against MBP-OBOX4 fusion protein, followed by MBP-OBOX2 fusion protein, for identification of
392 OBOX4-specific antibodies.
393 (B) Analysis of protein sequence similarity among the OBOX family members. The Smith–Waterman local
394 alignment method was used to align the amino acid sequences, and the divergence among different
395 sequences was represented in the form of a block substitution matrix.
396 (C) ELISA signal of anti-OBOX4 monoclonal antibodies against the MBP-OBOX4 and MBP-OBOX2 fusion
397 proteins. Anti-OBOX4 monoclonal antibodies were not reactive to MBP-OBOX2.
398
399 **Figure S3**
400 (A) Schematic of the strategy for knocking out the *Dux* cluster (not drawn to scale). *Dux* cluster was removed
401 by means of CRISPR-Cas9 targeting of the flanking regions. PCR-based genotyping, with primers targeting
402 the knockout allele and *Dux* internal sequence, was used to validate the genomic removal of *Dux*.
403 (B) Schematic of the strategy for knocking out the *Obox4* cluster (not drawn to scale). The *Obox4* cluster was
404 removed by means of CRISPR-Cas9 targeting of the flanking regions as well as the internal regions of
405 individual *Obox4* copies. PCR-based genotyping, with primers targeting the knockout allele, was used to
406 validate the genomic removal of the cluster.
407 (C) Agarose gel image of PCR-based genotyping of the *Dux* knockout allele in founder cell lines. The red arrow
408 indicates a cell line bearing a bi-allelic deletion of the *Dux* cluster.
409 (D) Agarose gel image of PCR-based genotyping of the *Dux* internal sequences in founder cell lines. The red
410 arrow indicates a cell line without the genomic *Dux* sequence.
411 (E) Agarose gel image of PCR-based genotyping of the *Obox4* knockout allele in founder cell lines. The red
412 arrow indicates a cell line bearing a bi-allelic deletion of the *Obox4* cluster.
413 (F) Immunofluorescence analysis of OBOX4 in *Obox4* knockout, *Dux* knockout, and double knockout cell lines.
414 (G) Western blot of OBOX4 and DUX in *Obox4* knockout, *Dux* knockout, and double knockout cell lines.
415

416 ## Methods

417 **Mice**

418 Male/female BDF1 and female BALB/c mice were purchased from Japan SLC Inc. Mice were fed regular
419 chow and housed in a controlled room under a 14 h/10 h light/dark cycle, at 22°C. All animal experiments

were approved by the Animal Care and Use Committee of Keio University and conducted in compliance with the Keio University Code of Research Ethics.

Cell culture

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

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

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

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

Cell transfection

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

cDNA synthesis and cloning

Total RNA was extracted from TBLCs using ISOGEN (catalog no. 311-02501, Nippon Gene), according to the manufacturer's instructions. Briefly, 500 μL ISOGEN was added to 1.5×10^6 freshly harvested TBLCs. After 5 min of incubation at room temperature, 100 μL of chloroform (catalog no. 03802606, Wako) was added to the cells and mixed by means of vigorous shaking. After 2 min of incubation at room temperature, the samples were centrifuged at 12,000 × *g* for 15 min at 4°C. The upper aqueous phase was then transferred to a new tube and mixed with 240 μL isopropanol (catalog no. 15-2320, Merck), to precipitate the RNA. After 5 min of

459 incubation at room temperature, the samples were centrifuged at 12,000 × *g* for 15 min at 4°C. The
460 supernatant was then discarded, following which the RNA pellets were washed with 70% ethanol (catalog no.
461 057-00456, Wako). The RNA pellets were then air-dried and dissolved in RNase-free water. The RNA solution
462 was subjected to DNase treatment using TURBO DNA-*free*™ kit (catalog no. AM1907, Thermo Fisher),
463 according to the manufacturer's instructions, to remove genomic DNA carryover. RNA (1 μg) was reverse
464 transcribed to cDNA using a Transcriptor First Strand cDNA Synthesis Kit (catalog no. 04379012001, Roche),
465 according to the manufacturer's instructions. PrimeSTAR® Max DNA Polymerase (catalog no. R045A, TaKaRa)
466 and ProFlex™ PCR System (catalog no. 4484073, Thermo Fisher) were used to amplify the sequence of interest
467 from the cDNA. NEBuilder® HiFi DNA Assembly Kit (catalog no. E2621L, NEB) was used to clone the sequence
468 of interest into plasmid backbone.
469
470 **Generation of transgenic mESC lines**
471 The 2C::tdTomato reporter cell line was generated by transfection of EB3 mESCs (catalog no. AES0139,
472 RIKEN BRC Cell Bank) with a linearized 2C::tdTomato reporter plasmid (catalog no. 40281, Addgene). After 48
473 h, the cells were subjected to 500 μg.mL^{-1} hygromycin (catalog no. 08906151, Wako) for 7 d. The selected cells
474 were then seeded at a density of 2×10^{2} cm^{-2} in culture medium containing 250 μg.mL^{-1} hygromycin. Single-cell
475 clones were picked and expanded after 7 d.
476 A tetracycycline-controlled transcription activation system and a piggyBac transposon system were used
477 as previously described [50], to generate an *Obox4*-inducible cell line. Briefly, pPB-TRE-3×FLAG-Obox4,
478 pPB-CAG-rtTA3G-IRES-Hygro, and pCMV-HyPBase-PGK-Puro plasmids were co-transfected into EB3 mESCs.
479 After 48 h, the cells were subjected to selection with 500 μg.mL^{-1} hygromycin and 500 μg.mL^{-1} G418 (catalog
480 no. 07805961, Wako), for 7 d. The selected cells were then seeded at a density of 2×10^{2} cm^{-2} in a culture
481 medium containing 250 μg.mL^{-1} hygromycin and 250 μg.mL^{-1} G418. Single-cell clones were picked and
482 expanded after 7 d.
483
484 **Generation of knockout mESC lines**
485 The *Dux* knockout cell line was generated by means of co-transfection of EB3 mESCs with
486 pX330-DuxKO-A and pX330-DuxKO-B plasmids. After 48 h, the cells were seeded at a very low density.
487 Single-cell clones were picked and expanded after 7 d and then subjected to genotyping.
488 The *Obox4* knockout cell line was generated by means of co-transfection of EB3 mESCs with the
489 pX330-Puro-Obox4KO-3, pX330-Puro-Obox4KO-4, pX330-Puro-Obox4KO-6, and pX330-Puro-Obox4KO-7
490 plasmids. After 18 h, the cells were subjected to selection with 1 μg.mL^{-1} puromycin (catalog no. 16023151,
491 Wako) for 54 h. The selected cells were then seeded at a very low density in culture medium without
492 antibiotics. Single-cell clones were picked and expanded after 7 d, and then subjected to genotyping.
493
494 **Flow cytometry**
495 The cells were digested with 0.25% trypsin-EDTA (catalog no. 25200072, Gibco), at 37°C for 5 min, and
496 then re-suspended in FluoroBrite™ DMEM (catalog no. A1896701, Gibco). The cell suspension was filtered

497 through a 35 μM cell strainer (catalog no. 352235, Falcon). The samples were analyzed using a cell sorter
498 (catalog no. SH800Z, Sony). Data analysis was performed using the Sony SH800Z cell sorter software.
499
500 **Western blot**
501 Freshly harvested cells were re-suspended in phosphate-buffered saline (PBS) (catalog no. 14249-95,
502 Nacalai Tesque) and lysed by means of sonication. Whole cell lysates were mixed with sample buffer
503 containing reducing reagent (catalog no. 09499-14, Nacalai Tesque) and boiled at 95°C for 5 min. Protein
504 samples were loaded on a 10% tris-glycine gel, run in AllView PAGE Buffer® (catalog no. DS520, BioDynamics
505 Laboratory), and then transferred to nitrocellulose membranes (catalog no. 10600003, Cytiva) using a Power
506 Blotter System (catalog no. PB0012, Thermo Fisher). The membranes were blocked with PBST containing 2%
507 skim milk (catalog no. 4902720131292, Morinaga), at room temperature for 15 min. Membranes were
508 incubated with primary antibodies in PBST containing 2% skim milk, at room temperature for 30 min. After
509 three washes with PBST, the membrane was incubated with secondary antibodies in PBST containing 2% skim
510 milk, at room temperature for 15 min, with shaking. After three washes with PBST, the membrane was
511 incubated with ECL reagents (catalog no. RPN2232, Cytiva) and exposed to an X-ray film (catalog no.
512 28906839, Cytiva) or detected using a digital chemiluminescence imager (catalog no. 17001402JA, Bio-Rad).
513
514 **Generation of monoclonal antibodies**
515 The anti-OBOX4, anti-DUX, and anti-GAG monoclonal antibodies were produced as previously described [51].
516 Briefly, BALB/c mice were immunized by means of intraperitoneal injection of glutathione S-transferase-fused
517 protein-of-interest (PoI). The mice were routinely immunized until their sera tested positive for the PoI-fused
518 maltose binding protein (MPB), after which two boosting immunizations were performed. Following boosting,
519 splenocytes from immunized mice were fused with SP2/O myeloma and cultured in
520 hypoxanthine-aminopterin-thymidine medium for 10 d, to select for hybridomas. Hybridomas were
521 subsequently screened using enzyme-linked immunosorbent assay against PoI-fused MBP. The positive
522 polyclonal hybridomas were monoclonalized and expanded. For anti-OBOX4 hybridomas, additional screening
523 using OBOX2-fused MBP was conducted. The culture supernatants of monoclonal hybridomas were sterilized
524 by passing them through a 0.22 μM pore size filter (catalog no. 430769, Corning) and used directly as an
525 antibody solution in other assays.
526
527 ***In vitro* transcription**
528 The *Obox4* coding sequence was codon-optimized using GeneArt Instant Designer (Thermo Fisher), to
529 remove the ASO target motif and improve translation efficiency. Codon-optimized DNA was synthesized using
530 Prime Gene Synthesis Services (Thermo Fisher). Codon-optimized *Obox4* mRNA was transcribed *in vitro* using
531 the mMESSAGE mMACHINE™ T7 Transcription Kit (catalog no. AM1344, Thermo Fisher) and then
532 polyadenylated using a poly(A) tailing kit (catalog no. AM1350, Thermo Fisher), according to the
533 manufacturer's instructions. The polyadenylated mRNA was purified using the RNeasy Mini Kit (catalog no.
534 74004, Qiagen), according to the manufacturer's instructions, and dissolved in RNase-free water, at a
535 concentration of 100 $ng.\mu L^{-1}$.

536

537 **Pronuclear injection of mouse embryos**

538 Eight-week-old BDF1 female mice were injected with 150 μL of CARD HyperOva (catalog no. KYD-010-EX,
539 Kyudo Co., Ltd.). After 48 hours, the female mice were injected with 5IU human chorionic gonadotropin (hCG)
540 (catalog no. GONATROPIN, ASKA Animal Health Co. Ltd.) and mated with BDF1 male mice. After 22 h, the PN3
541 zygotes were collected. Cumulus cells were removed by briefly culturing the zygotes in potassium simplex
542 optimization medium (KSOM) (catalog no. MR-101-D, Merck) supplemented with 0.3 $\mu g.\mu L^{-1}$ hyaluronidase
543 (catalog no. H4272, Merck). Embryos were cultured in KSOM medium drops covered with liquid paraffin
544 (catalog no. 26137-85, Nacalai Tesque), in a humidified atmosphere with 5% CO_2, at 37°C. After 4 h, PBS
545 containing either 40 μM scramble ASO, 20 μM anti-*Dux* ASO, 20 μM anti-*Obox4* ASO, 40 μM equimolar
546 anti-*Dux*/anti-*Obox4* ASO mixture, or 40 μM equimolar anti-*Dux*/anti-*Obox4* ASO mixture with 100 $ng.\mu L^{-1}$
547 codon-optimized *Obox4* mRNA was microinjected into the male pronuclei of zygotes using a microinjector
548 (catalog no. 5252000013, Eppendorf). For developmental monitoring, embryos were cultured for 4 d after
549 microinjection and assessed for developmental stage at 18 h, 42 h, 66 h, and 90 h after microinjection, which
550 corresponded to 1.5 dpc, 2.5 dpc, 3.5 dpc, and 4.5 dpc, respectively.

551

552 **Immunofluorescence**

553 For embryos, middle-2C-embryos were collected at 1.5 dpc. The zona pellucida was removed by treating
554 the embryos with acidic Tyrode's solution (catalog no. MR-0040-D, Merck). The embryos were fixed in 4%
555 paraformaldehyde (catalog no. 09154-14, Nacalai Tesque) in PBS for 10 min. Following three washes in PBS,
556 fixed embryos were permeabilized with 0.1% Triton™ X-100 (catalog no. 1610407, Bio-Rad) in PBS, for 20 min
557 at room temperature. Following three washes in PBS, permeabilized embryos were blocked in 2% bovine
558 serum albumin (BSA; catalog no. 011-27055, Wako) in PBS, for 20 min at room temperature. The blocked
559 embryos were incubated with primary antibodies diluted in 2% BSA in PBS, at 4°C overnight. Following three
560 washes in PBS, the embryos were incubated with 2% BSA in PBS containing 1:200 diluted
561 4′,6-diamidino-2-phenylindole (DAPI) (catalog no. 19178-91, Nacalai Tesque) and 1:500 diluted Alexa
562 Fluor™-conjugated anti-mouse IgG1 secondary antibody (catalog no. A-21127, Thermo Fisher) or Alexa
563 Fluor™-conjugated anti-mouse IgG2a secondary antibody (catalog no. A-21131, Thermo Fisher) and, for 1 h at
564 room temperature. Following three washes in PBS, the embryos were transferred into liquid paraffin-covered
565 PBS drops on a glass-bottom dish (catalog no. D11130H, Matsunami).

566 For culturing, cells seeded on glass bottom chamber slides (catalog no. SCS-N02, Matsunami) were fixed
567 with 4% paraformaldehyde in PBS, for 10 min. After three washes in PBS, the fixed cells were permeabilized
568 with 0.1% Triton™ X-100 in PBS, for 15 min at room temperature. Following three washes in PBS,
569 permeabilized cells were incubated with primary antibodies diluted with 2% skim milk in PBS, for 30 min at
570 room temperature. Following three washes in PBS, the cells were incubated with 2% BSA in PBS containing
571 1:500 diluted Alexa Fluor™-conjugated secondary antibody and 1:200 diluted DAPI, for 1 h at room
572 temperature. Following three washes in PBS, the chamber was removed and the slides were mounted with
573 ProLong™ Glass Antifade Mountant (catalog no. P36982, Thermo Fisher). Fluorescence images were taken
574 using a confocal laser scanning microscope (catalog no. FV3000, Olympus).

575

576 **scRNA-seq analysis**

577 Raw scRNA-seq data was obtained from the dataset of Deng et al. (GSE45719). Quality control and
578 adapter trimming was done using fastp [52] v0.23.2. Quality-controlled reads were aligned to the GRCm38.p6
579 reference genome using STAR [53] v2.7.9a, with default arguments. Reads were counted against GRCm38.p6
580 comprehensive gene annotation [54] and mm10 repeats from the University of California, Santa Cruz (UCSC)
581 RepeatMasker using Subread [55] v2.0.1 featureCounts function, and multi-mapping reads were discarded for
582 non-TE features and counted fractionally for TEs. Seurat [56] v4.1.0 was used to process the read counts of
583 scRNA-seq. Cells with greater than 7.5% mitochondrial reads or less than 14,000 annotated features were
584 discarded. Expression levels were log-normalized. The expression profiles of homeobox genes listed in
585 HomeoDB2 [57] were clustered into 11 *k*-means after *z*-score transformation.

586

587 **RNA-seq**

588 For embryos, 30 middle-2C-embryos were collected 20 h after microinjection in each independent
589 biological replicate. The zona pellucida was removed by treating embryos with acidic Tyrode's solution.
590 Libraries were constructed using the SMART-Seq® Stranded Kit (catalog no. 634442, TaKaRa) and indexed
591 using the SMARTer® RNA Unique Dual Index Kit (catalog no. 634451, TaKaRa), according to the
592 manufacturer's instructions.

593 For cell culture, total RNA was prepared using the RNeasy Kit, according to the manufacturer's
594 instructions. The RNA solution was subjected to DNase treatment to remove genomic DNA carryover. Using 1
595 μg total RNA per sample, libraries were constructed with the NEBNext® Ultra™ II Directional RNA Library Prep
596 Kit for Illumina (catalog no. E7760L, NEB) and indexed with NEBNext® Multiplex Oligos for Illumina (catalog
597 no. E6440S, NEB), according to the manufacturer's instructions.

598 The libraries were quantified with a 2100 Bioanalyzer (catalog no. G2939BA, Agilent) using a High
599 Sensitivity DNA Kit (catalog no. 5067-4626, Agilent). Quantified libraries were pooled and sequenced using an
600 Illumina NovaSeq 6000 System in 150 bp paired-end mode (Illumina). Base calling and de-multiplexing were
601 performed using the bcl2fastq2 (Illumina) v2.20. De-multiplexed reads were quality-controlled, aligned, and
602 counted, as described above. DESeq2 [58] v1.32.0 was used to perform differential expression analysis. Raw
603 RNA-seq reads for TBLCs were downloaded from the dataset of Shen et al. (GSE168728). Raw RNA-seq reads
604 of *Dux* knockout embryos were downloaded from the datasets of Chen & Zhang and De Iaco et al. (GSE121746
605 and GSE141321 respectively). The published data were analyzed using the same method.

606

607 **CUT&RUN-seq**

608 Cultured cells (1×10^5) freshly prepared by trypsin-EDTA digestion were used to perform CUT&RUN with
609 the CUT&RUN Assay Kit (catalog no. 86652, Cell Signaling Technology), according to the manufacturer's
610 instructions. Enriched DNA was subjected to library construction using the NEBNext® Ultra™ II DNA Library
611 Prep Kit for Illumina (catalog no. E7645L, NEB), with the adapter ligation step performed at 50°C instead of
612 65°C, to prevent the denaturation of small DNA inserts. The libraries were indexed using NEBNext® Multiplex
613 Oligos for Illumina, according to the manufacturer's instructions.

The libraries were quantified, pooled, sequenced, de-multiplexed, and quality controlled, as described above. Processed reads were aligned to the GRCm38.p6 reference genome using Bowtie2 [59] v2.4.1, with default settings. Peaks were called in each biological replicate using the MACS3 [60] v3.0.0a6 callpeak function (-f BAMPE). Alignment tracks were first generated using deepTools [61] v3.5.1 bamCoverage function (--binSize 10 --normalized using CPM --smoothLength 30), and then normalized by subtracting the signal from non-immune IgG using the deepTools bamCompare function (--scaleFactorssMethod None --operation subtract --binSize 10 --smoothLength 30). ChIPseeker [62] v1.28.3 was used to annotate the peaks. The MEME suite [63] v5.4.1 was used to identify the binding motifs. Quantitative histone modification analysis of replicated samples was performed using MAnorm2 [64] v1.2.0. Visualization of genomic tracks was performed using trackplot [65] v1.3.10. Raw ChIP-seq reads of DUXs in mESCs were downloaded from Hendrickson et al. (GSE85632). The published data were analyzed using the same method.

Quantification and statistical analysis

Descriptive and comparative statistics were employed in the manuscript as described in the figure legends with the number of replicates indicated. Significance is defined as a *p-value* less than 0.05 indicated with asterisk (* *p-value* < 0.05, ** *p-value* < 0.01, *** *p-value* < 0.001). Error bar represents the standard deviation (SD) of the mean of the replicates.

Data and code availability

The RNA-seq and CUT&RUN-seq data generated in this study have been deposited at NCBI Gene Expression Omnibus (GEO) database under the accession code GSE196671.

Acknowledgements

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

692 15. De Iaco, A. *et al.* DUX-family transcription factors regulate zygotic genome activation in placental mammals.
693 *Nature genetics* **49**, 941-945 (2017).
694 16. Hendrickson, P.G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes
695 and MERVL/HERVL retrotransposons. *Nature genetics* **49**, 925-934 (2017).
696 17. Whiddon, J.L., Langford, A.T., Wong, C.-J., Zhong, J.W. & Tapscott, S.J. Conservation and innovation in the
697 DUX4-family gene network. *Nature genetics* **49**, 935-940 (2017).
698 18. Chen, Z. & Zhang, Y. Loss of DUX causes minor defects in zygotic genome activation and is compatible with
699 mouse development. *Nature genetics* **51**, 947-951 (2019).
700 19. Guo, M. *et al.* Precise temporal regulation of Dux is important for embryo development. *Cell research* **29**,
701 956-959 (2019).
702 20. De Iaco, A., Verp, S., Offner, S., Grun, D. & Trono, D. DUX is a non-essential synchronizer of zygotic genome
703 activation. *Development* **147**, dev177725 (2020).
704 21. Bosnakovski, D., Gearhart, M.D., Ho Choi, S. & Kyba, M. Dux facilitates post-implantation development, but is
705 not essential for zygotic genome activation. *Biology of reproduction* **104**, 83-93 (2021).
706 22. Wagner, A. Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional
707 regulators. *Biological cybernetics* **74**, 557-567 (1996).
708 23. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic
709 gene expression in mammalian cells. *Science* **343**, 193-196 (2014).
710 24. Shen, H. *et al.* Mouse totipotent stem cells captured and maintained through spliceosomal repression. *Cell*
711 **184**, 2843-2859. e20 (2021).
712 25. Macfarlan, T.S. *et al.* Endogenous retroviruses and neighboring genes are coordinately repressed by
713 LSD1/KDM1A. *Genes & development* **25**, 594-607 (2011).
714 26. Macfarlan, T.S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*
715 **487**, 57-63 (2012).
716 27. Howe, K.L. *et al.* Ensembl 2021. *Nucleic Acids Research* **49**, D884-D891 (2020).
717 28. Skene, P.J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding
718 sites. *Elife* **6**, e21856 (2017).
719 29. Sasaki, F. *et al.* A high-affinity monoclonal antibody against the FLAG tag useful for G-protein-coupled
720 receptor study. *Analytical biochemistry* **425**, 157-165 (2012).
721 30. Mayran, A. & Drouin, J. Pioneer transcription factors shape the epigenetic landscape. *Journal of Biological*
722 *Chemistry* **293**, 13795-13804 (2018).
723 31. Liu, X. *et al.* Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos.
724 *Nature* **537**, 558-562 (2016).
725 32. Dahl, J.A. *et al.* Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition.
726 *Nature* **537**, 548-552 (2016).
727 33. Maeso, I. *et al.* Evolutionary origin and functional divergence of totipotent cell homeobox genes in eutherian
728 mammals. *BMC biology* **14**, 1-14 (2016).
729 34. Rajkovic, A., Yan, C., Yan, W., Klysik, M. & Matzuk, M.M. Obox, a family of homeobox genes preferentially
730 expressed in germ cells. *Genomics* **79**, 711-717 (2002).
731 35. Ge, S.X. Exploratory bioinformatics investigation reveals importance of "junk" DNA in early embryo
732 development. *BMC genomics* **18**, 1-19 (2017).
733 36. Zhong, Y.-f. & Holland, P.W. The dynamics of vertebrate homeobox gene evolution: gain and loss of genes in
734 mouse and human lineages. *BMC evolutionary biology* **11**, 1-13 (2011).

735 37. Wilming, L.G., Boychenko, V. & Harrow, J.L. Comprehensive comparative homeobox gene annotation in
736 human and mouse. *Database* **2015**(2015).
737 38. Lee, H.-S. & Lee, K.-A. Characterization and functional analysis of Obox4 during oocyte maturation by RNA
738 interference. *Clinical and Experimental Reproductive Medicine* **34**, 293-303 (2007).
739 39. Lee, H.S. *et al.* Obox4 critically regulates cAMP-dependent meiotic arrest and MI-MII transition in oocytes.
740 *The FASEB Journal* **24**, 2314-2324 (2010).
741 40. Lee, K.-A., Lee, H.-S., Kim, E.-Y., Kim, K.-H. & Kim, Y. Oocyte-specific homeobox 4 (Obox4) is a key regulatory
742 component of the cAMP-dependent meiotic arrest during in vitro oocyte maturation. *Fertility and Sterility*
743 **90**, S108 (2008).
744 41. Lee, H.-S., Kim, E.-Y. & Lee, K.-A. Changes in gene expression associated with oocyte meiosis after Obox4
745 RNAi. *Clinical and experimental reproductive medicine* **38**, 68 (2011).
746 42. Lee, H.-S. *et al.* Obox4-silencing-activated STAT3 and MPF/MAPK signaling accelerate nuclear membrane
747 breakdown in mouse oocytes. *Reproduction* **151**, 369-378 (2016).
748 43. Kim, H.-M. *et al.* Obox4 regulates the expression of histone family genes and promotes differentiation of
749 mouse embryonic stem cells. *FEBS letters* **584**, 605-611 (2010).
750 44. Nowak, M.A., Boerlijst, M.C., Cooke, J. & Smith, J.M. Evolution of genetic redundancy. *Nature* **388**, 167-171
751 (1997).
752 45. Vavouri, T., Semple, J.I. & Lehner, B. Widespread conservation of genetic redundancy during a billion years
753 of eukaryotic evolution. *Trends in genetics* **24**, 485-488 (2008).
754 46. Hackett, J.A. & Surani, M.A. Regulatory principles of pluripotency: from the ground state up. *Cell stem cell* **15**,
755 416-430 (2014).
756 47. Sakashita, A. *et al.* Endogenous retroviruses drive species-specific germline transcriptomes in mammals.
757 *Nature Structural & Molecular Biology* **27**, 967-977 (2020).
758 48. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration.
759 *Nature Reviews Genetics* **14**, 113-124 (2013).
760 49. Leidenroth, A. & Hewitt, J.E. A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals
761 the ancestral DUX gene. *BMC evolutionary biology* **10**, 1-13 (2010).
762 50. Takeuchi, C., Murano, K., Ishikawa, M., Okano, H. & Iwasaki, W., Y. Generation of stable drosophila ovarian
763 somatic cell lines using piggyBac system. *Methods in Molecular Biology* **In press**(2022).
764 51. Guo, Y. *et al.* Potent mouse monoclonal antibodies that block SARS-CoV-2 infection. *Journal of Biological*
765 *Chemistry* **296**(2021).
766 52. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**,
767 i884-i890 (2018).
768 53. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
769 54. Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Research* **49**, D916-D923 (2020).
770 55. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence
771 reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
772 56. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* (2021).
773 57. Zhong, Y.f. & Holland, P.W. HomeoDB2: functional expansion of a comparative homeobox gene database for
774 evolutionary developmental biology. *Evolution & development* **13**, 567-568 (2011).
775 58. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data
776 with DESeq2. *Genome biology* **15**, 1-21 (2014).
777 59. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359
778 (2012).

60. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, 1-9 (2008).
61. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* **44**, W160-W165 (2016).
62. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383 (2015).
63. Bailey, T.L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* **37**, W202-W208 (2009).
64. Tu, S. *et al.* MAnorm2 for quantitatively comparing groups of ChIP-seq samples. *Genome Research* **31**, 131-145 (2021).
65. Pohl, A. & Beato, M. bwtool: a tool for bigWig files. *Bioinformatics* **30**, 1618-1619 (2014).

Figure 1

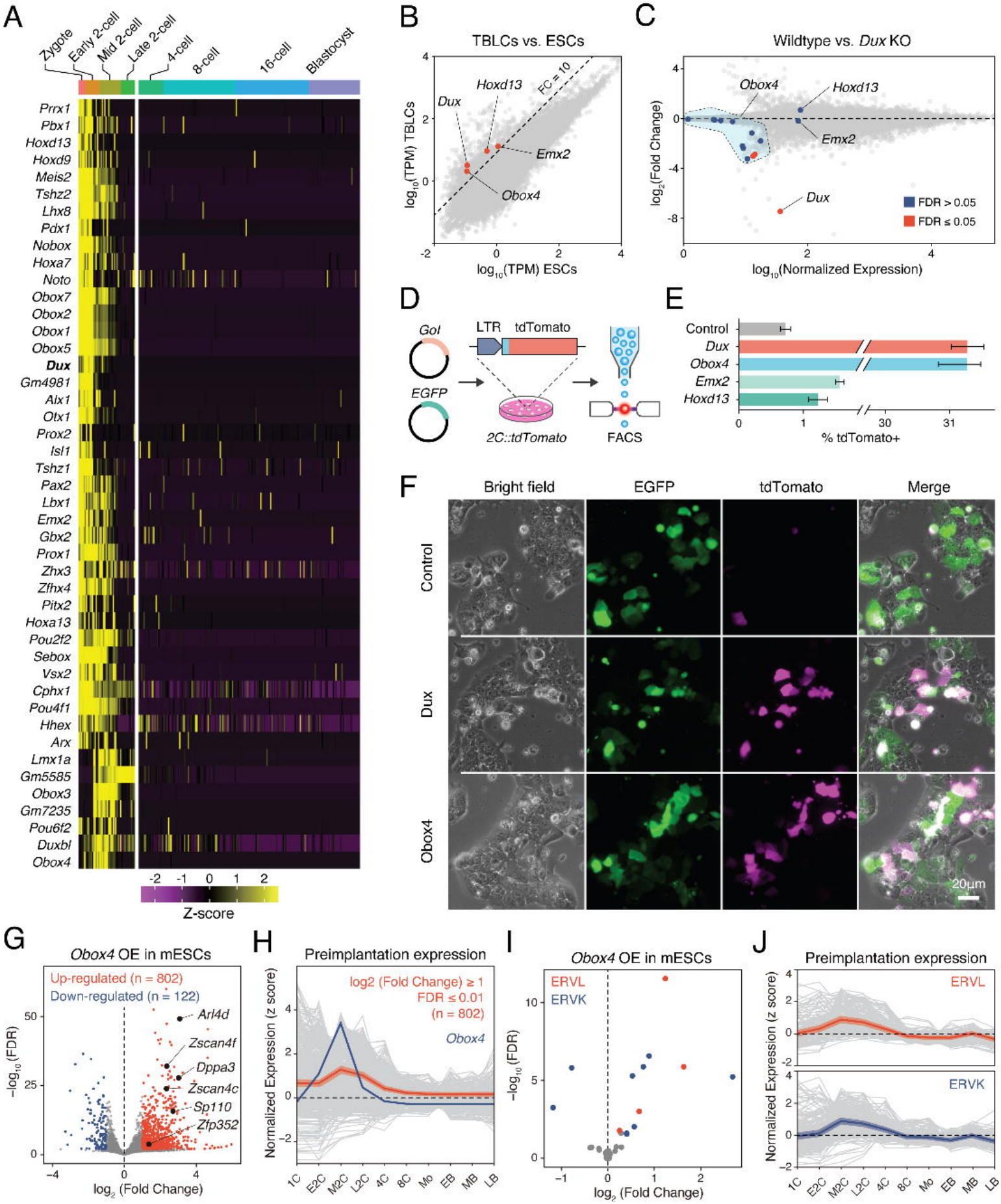

A
Zygote
Early 2-cell
Mid 2-cell
Late 2-cell
4-cell
8-cell
16-cell
Blastocyst
Prrx1
Pbx1
Hoxd13
Hoxd9
Meis2
Tshz2
Lhx8
Pdx1
Nobox
Hoxa7
Noto
Obox7
Obox2
Obox1
Obox5
Dux
Gm4981
Alx1
Otx1
Prox2
Isl1
Tshz1
Pax2
Lbx1
Emx2
Gbx2
Prox1
Zhx3
Zfhx4
Pitx2
Hoxa13
Pou2f2
Sebox
Vsx2
Cphx1
Pou4f1
Hhex
Arx
Lmx1a
Gm5585
Obox3
Gm7235
Pou6f2
Duxbl
Obox4
-2 -1 0 1 2
Z-score
B
TBLCs vs. ESCs
FC = 10
Dux
Hoxd13
Emx2
Obox4
log10(TPM) TBLCs
log10(TPM) ESCs
C
Wildtype vs. Dux KO
Obox4
Hoxd13
Emx2
Dux
FDR > 0.05
FDR ≤ 0.05
log2(Fold Change)
log10(Normalized Expression)
D
Gol
EGFP
LTR
tdTomato
2C::tdTomato
FACS
E
Control
Dux
Obox4
Emx2
Hoxd13
% tdTomato+
F
Bright field
EGFP
tdTomato
Merge
Control
Dux
Obox4
20µm
G
Obox4 OE in mESCs
Up-regulated (n = 802)
Down-regulated (n = 122)
Arl4d
Zscan4f
Dppa3
Zscan4c
Sp110
Zfp352
−log10(FDR)
log2 (Fold Change)
H
Preimplantation expression
log2 (Fold Change) ≥ 1
FDR ≤ 0.01
(n = 802)
Obox4
Normalized Expression (z score)
1C
E2C
M2C
L2C
4C
8C
Mo
EB
MB
LB
I
Obox4 OE in mESCs
ERVL
ERVK
−log10 (FDR)
log2 (Fold Change)
J
Preimplantation expression
ERVL
ERVK
Normalized Expression (z score)

Figure 2

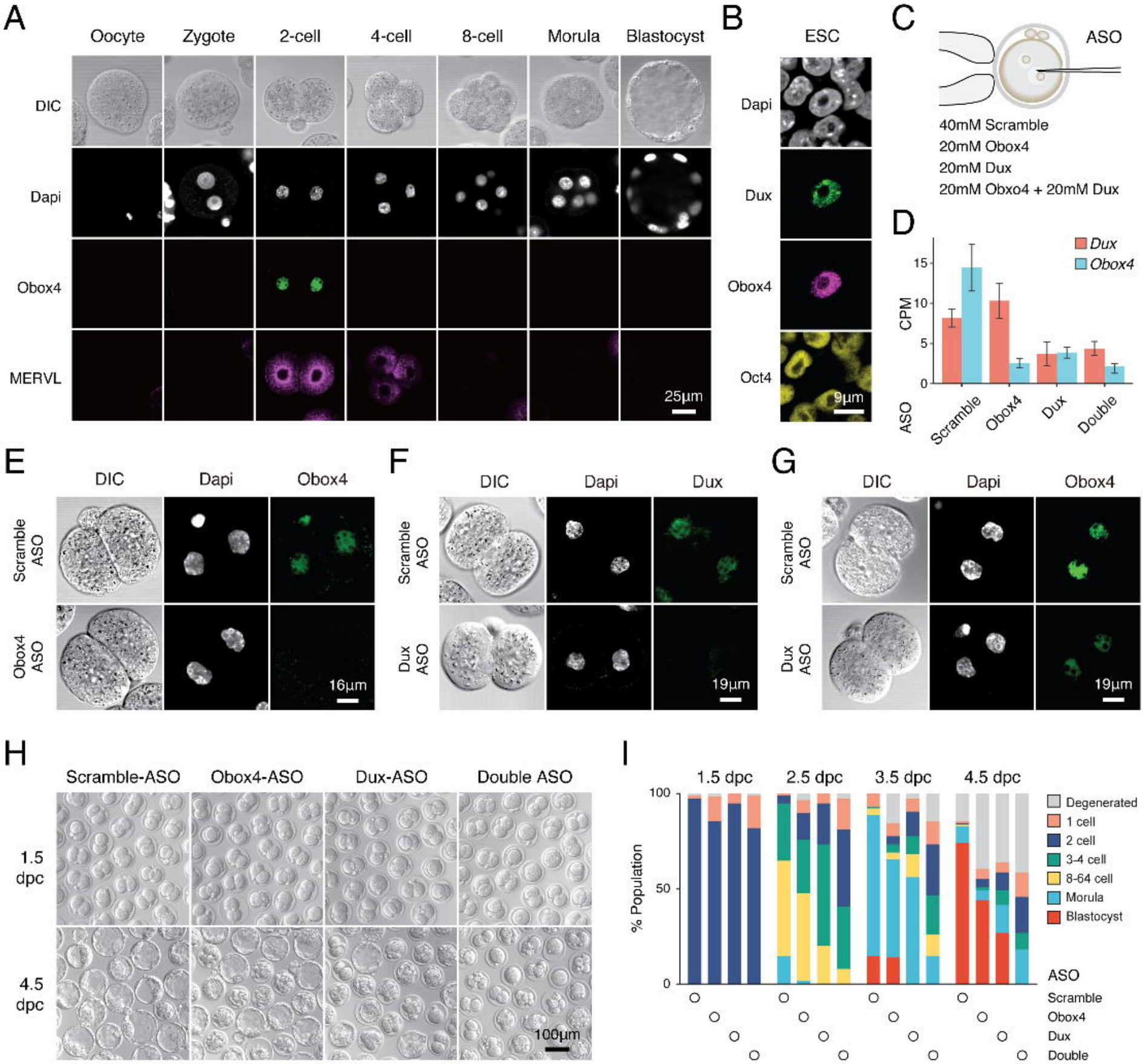

A
Oocyte
Zygote
2-cell
4-cell
8-cell
Morula
Blastocyst
DIC
Dapi
Obox4
MERVL
25µm
B
ESC
Dapi
Dux
Obox4
Oct4
9µm
C
ASO
40mM Scramble
20mM Obox4
20mM Dux
20mM Obxo4 + 20mM Dux
D
Dux
Obox4
CPM
15
10
5
0
ASO
Scramble
Obox4
Dux
Double
E
DIC
Dapi
Obox4
Scramble ASO
Obox4 ASO
16µm
F
DIC
Dapi
Dux
Scramble ASO
Dux ASO
19µm
G
DIC
Dapi
Obox4
Scramble ASO
Dux ASO
19µm
H
Scramble-ASO
Obox4-ASO
Dux-ASO
Double ASO
1.5 dpc
4.5 dpc
100µm
I
1.5 dpc
2.5 dpc
3.5 dpc
4.5 dpc
100
50
0
% Population
Degenerated
1 cell
2 cell
3-4 cell
8-64 cell
Morula
Blastocyst
ASO
Scramble
Obox4
Dux
Double

Figure 3

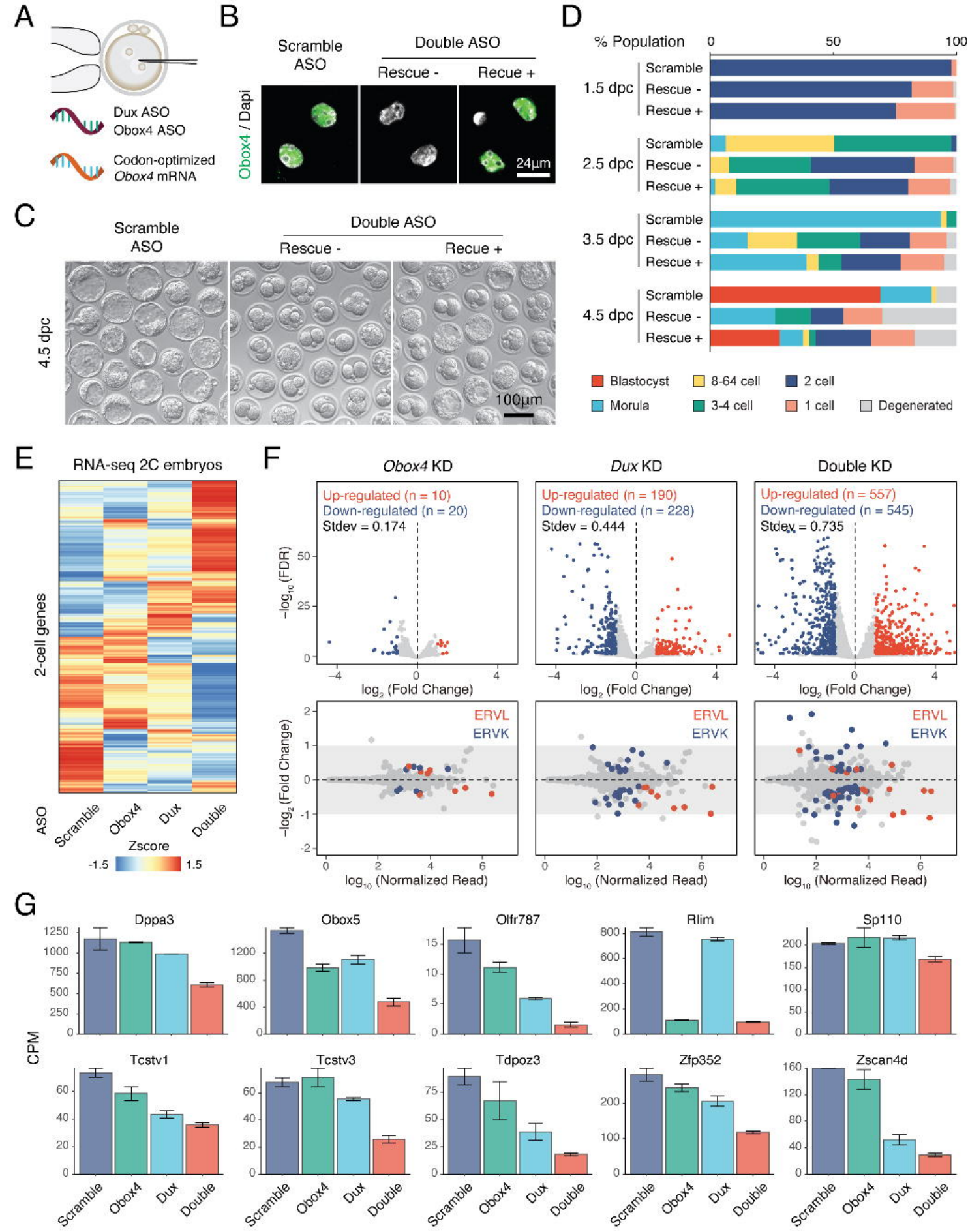

A
Dux ASO
Obox4 ASO
Codon-optimized Obox4 mRNA
B
Scramble ASO
Double ASO
Rescue -
Recue +
Obox4 / Dapi
24µm
C
Scramble ASO
Double ASO
Rescue -
Recue +
4.5 dpc
100µm
D
% Population
0
50
100
1.5 dpc
2.5 dpc
3.5 dpc
4.5 dpc
Scramble
Rescue -
Rescue +
Blastocyst
8-64 cell
2 cell
Morula
3-4 cell
1 cell
Degenerated
E
RNA-seq 2C embryos
2-cell genes
ASO
Scramble
Obox4
Dux
Double
Zscore
-1.5
1.5
F
Obox4 KD
Dux KD
Double KD
Up-regulated (n = 10)
Down-regulated (n = 20)
Stdev = 0.174
Up-regulated (n = 190)
Down-regulated (n = 228)
Stdev = 0.444
Up-regulated (n = 557)
Down-regulated (n = 545)
Stdev = 0.735
−log10 (FDR)
log2 (Fold Change)
ERVL
ERVK
−log2 (Fold Change)
log10 (Normalized Read)
G
Dppa3
Obox5
Olfr787
Rlim
Sp110
Tcstv1
Tcstv3
Tdpoz3
Zfp352
Zscan4d
CPM
Scramble
Obox4
Dux
Double

Figure 4

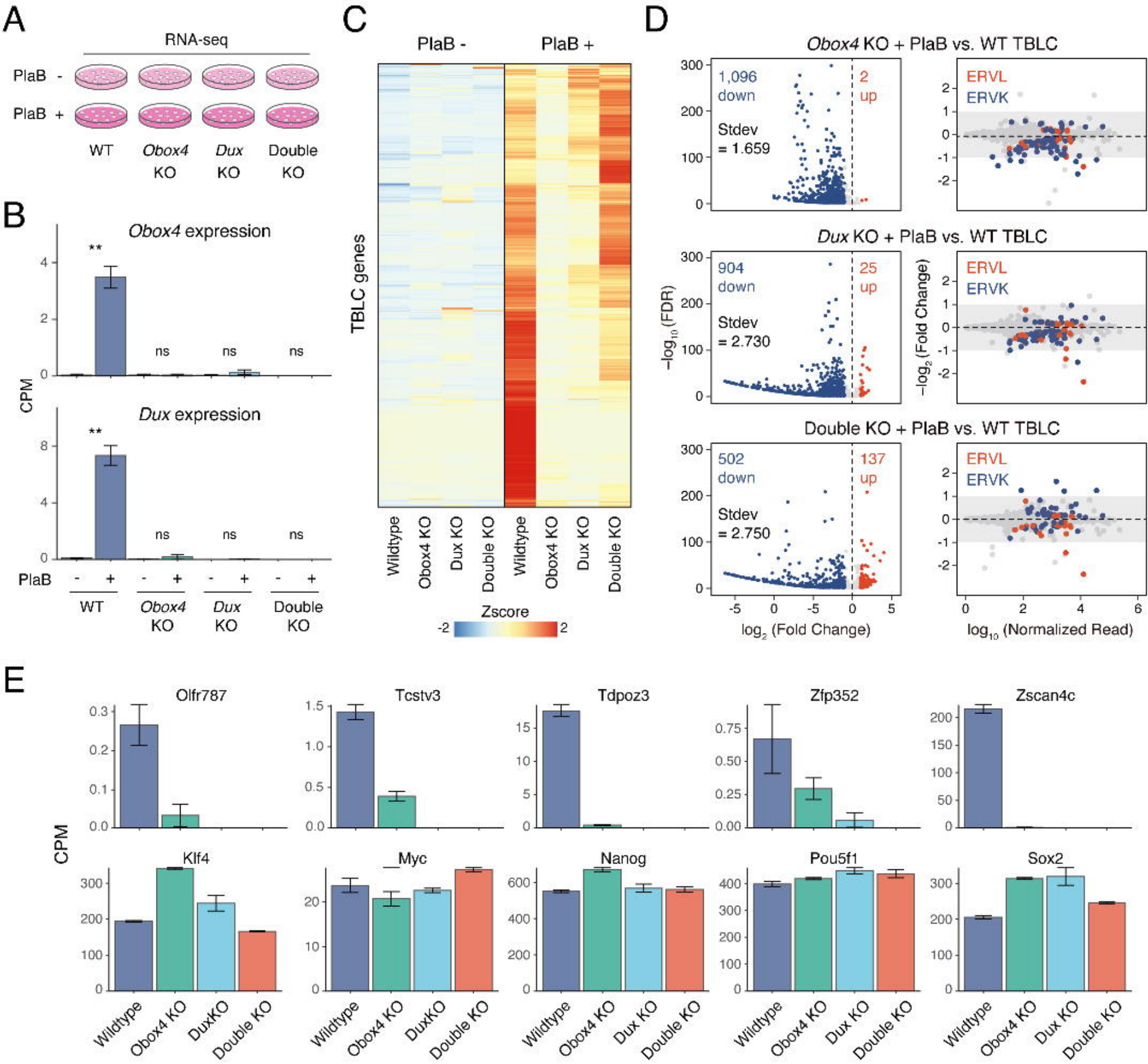

A
RNA-seq
PlaB -
PlaB +
WT
Obox4 KO
Dux KO
Double KO
B
Obox4 expression
Dux expression
CPM
PlaB
ns
C
PlaB -
PlaB +
TBLC genes
Wildtype
Obox4 KO
Dux KO
Double KO
Zscore
D
Obox4 KO + PlaB vs. WT TBLC
1,096 down
2 up
Stdev = 1.659
Dux KO + PlaB vs. WT TBLC
904 down
25 up
Stdev = 2.730
Double KO + PlaB vs. WT TBLC
502 down
137 up
Stdev = 2.750
ERVL
ERVK
−log10 (FDR)
−log2 (Fold Change)
log2 (Fold Change)
log10 (Normalized Read)
E
Olfr787
Tcstv3
Tdpoz3
Zfp352
Zscan4c
Klf4
Myc
Nanog
Pou5f1
Sox2
CPM
Wildtype
Obox4 KO
DuxKO
Double KO

Figure 5

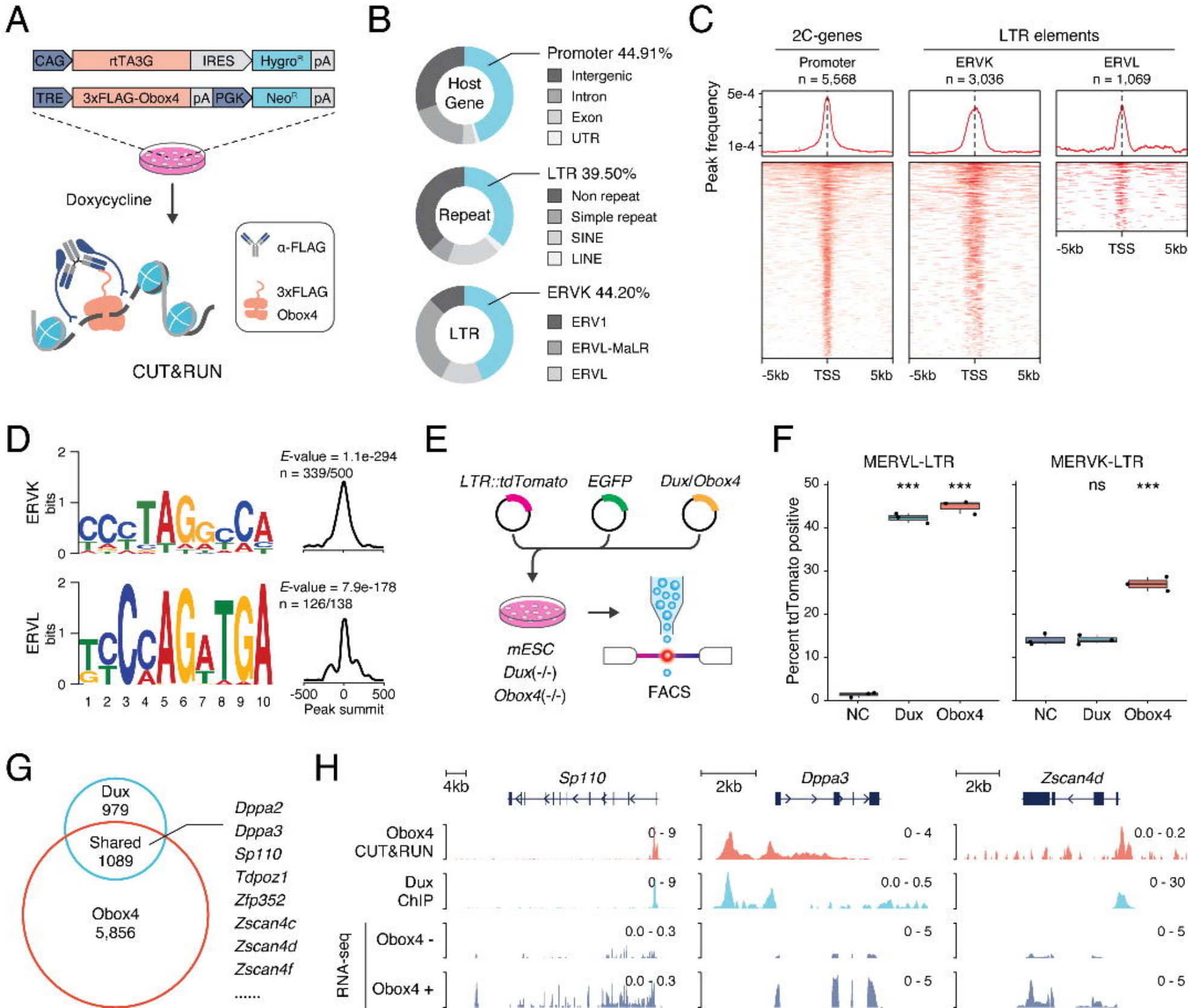

A
CAG
rtTA3G
IRES
Hygro^R
pA
TRE
3xFLAG-Obox4
pA
PGK
Neo^R
pA
Doxycycline
α-FLAG
3xFLAG
Obox4
CUT&RUN
B
Host Gene
Promoter 44.91%
Intergenic
Intron
Exon
UTR
Repeat
LTR 39.50%
Non repeat
Simple repeat
SINE
LINE
LTR
ERVK 44.20%
ERV1
ERVL-MaLR
ERVL
C
2C-genes
LTR elements
Promoter
n = 5,568
ERVK
n = 3,036
ERVL
n = 1,069
Peak frequency
5e-4
1e-4
-5kb
TSS
5kb
D
ERVK
bits
E-value = 1.1e-294
n = 339/500
ERVL
E-value = 7.9e-178
n = 126/138
-500
0
500
Peak summit
E
LTR::tdTomato
EGFP
Dux/Obox4
mESC
Dux(-/-)
Obox4(-/-)
FACS
F
MERVL-LTR
MERVK-LTR
Percent tdTomato positive
***
ns
NC
Dux
Obox4
G
Dux
979
Shared
1089
Obox4
5,856
Dppa2
Dppa3
Sp110
Tdpoz1
Zfp352
Zscan4c
Zscan4d
Zscan4f
.....
H
4kb
Sp110
2kb
Dppa3
2kb
Zscan4d
Obox4
CUT&RUN
Dux
ChIP
RNA-seq
Obox4 -
Obox4 +
0 - 9
0 - 4
0.0 - 0.2
0.0 - 0.5
0 - 30
0.0 - 0.3
0 - 5

Figure 6

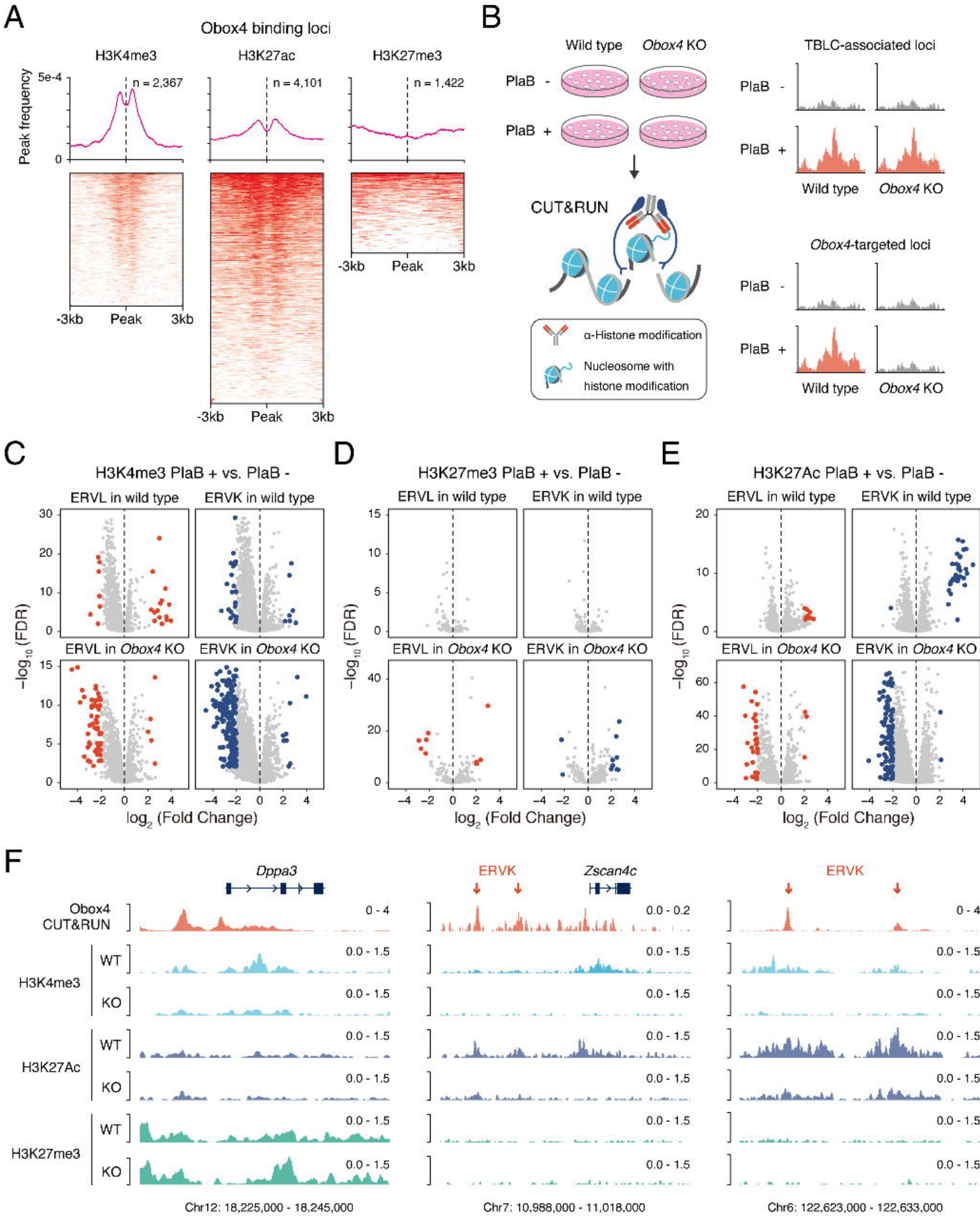
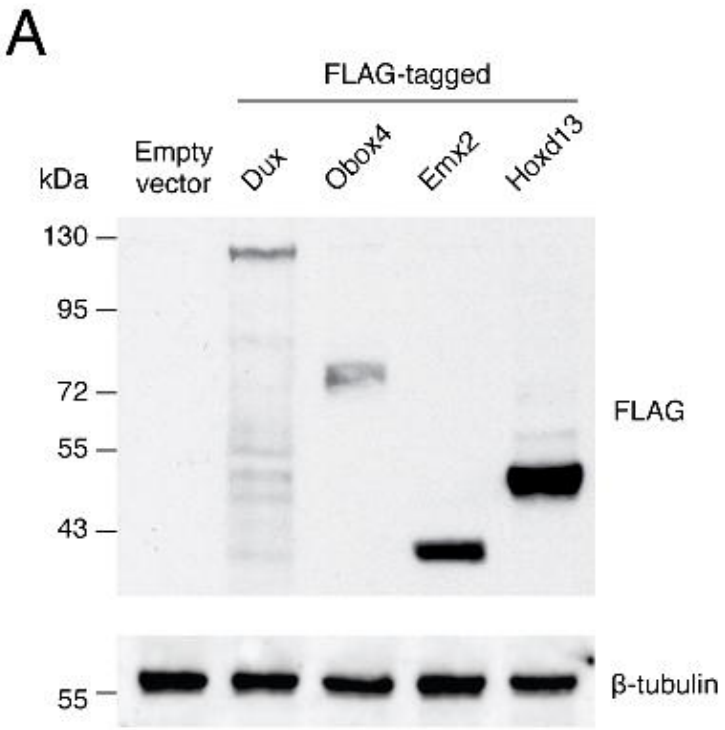
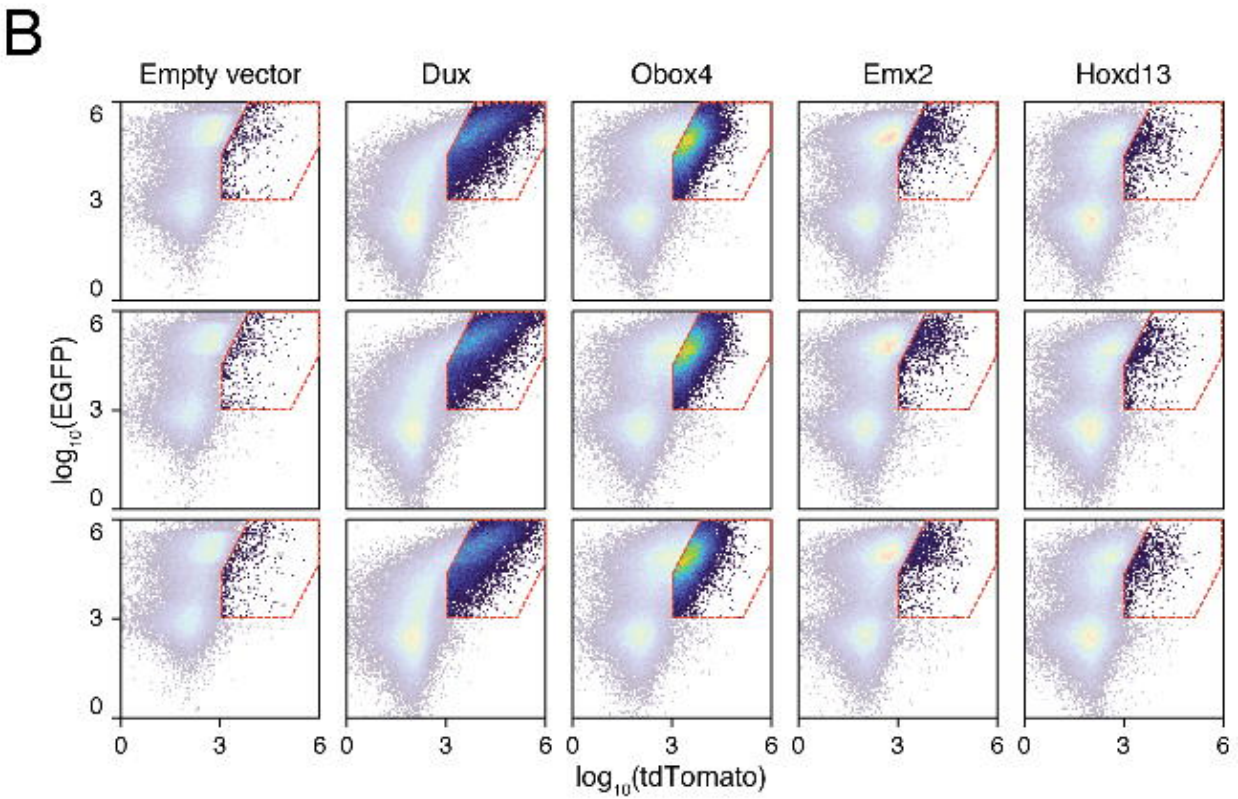
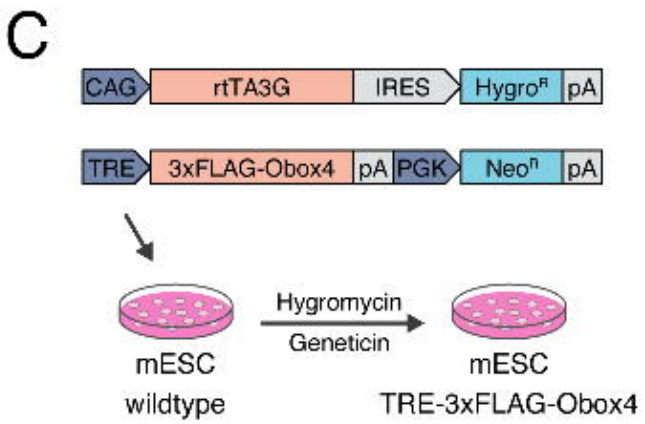

A
Obox4 binding loci
H3K4me3
H3K27ac
H3K27me3
Peak frequency
5e-4
0
n = 2,367
n = 4,101
n = 1,422
-3kb
Peak
3kb
B
Wild type
Obox4 KO
PlaB -
PlaB +
CUT&RUN
α-Histone modification
Nucleosome with histone modification
TBLC-associated loci
Obox4-targeted loci
C
H3K4me3 PlaB + vs. PlaB -
D
H3K27me3 PlaB + vs. PlaB -
E
H3K27Ac PlaB + vs. PlaB -
ERVL in wild type
ERVK in wild type
ERVL in Obox4 KO
ERVK in Obox4 KO
−log10 (FDR)
log2 (Fold Change)
F
Dppa3
ERVK
Zscan4c
Obox4 CUT&RUN
H3K4me3
H3K27Ac
H3K27me3
WT
KO
0 - 4
0.0 - 0.2
0.0 - 1.5
Chr12: 18,225,000 - 18,245,000
Chr7: 10,988,000 - 11,018,000
Chr6: 122,623,000 - 122,633,000

Figure S1


A
FLAG-tagged
kDa
Empty vector
Dux
Obox4
Emx2
Hoxd13
130
95
72
55
43
FLAG
55
β-tubulin


B
Empty vector
Dux
Obox4
Emx2
Hoxd13
log10(EGFP)
log10(tdTomato)
0
3
6


C
CAG
rtTA3G
IRES
HygroR
pA
TRE
3xFLAG-Obox4
pA
PGK
NeoR
pA
Hygromycin
Geneticin
mESC
wildtype
mESC
TRE-3xFLAG-Obox4

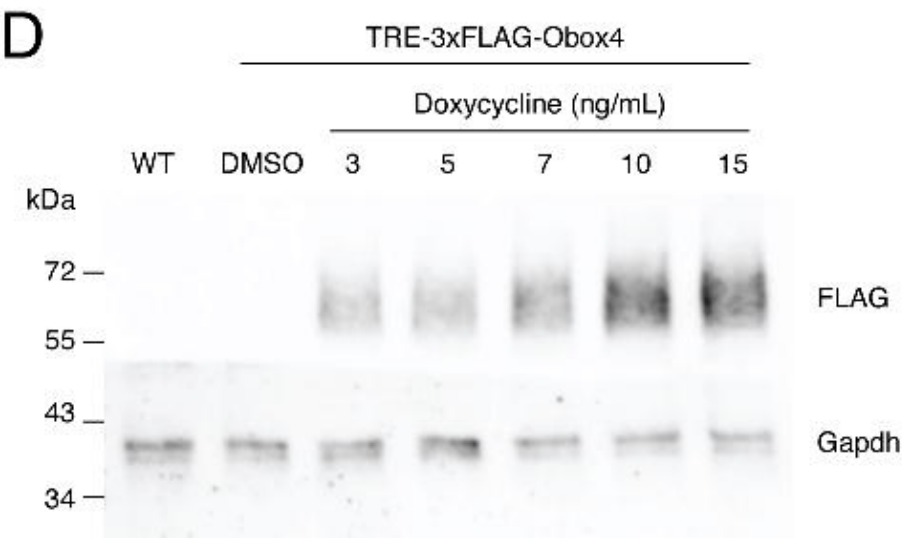
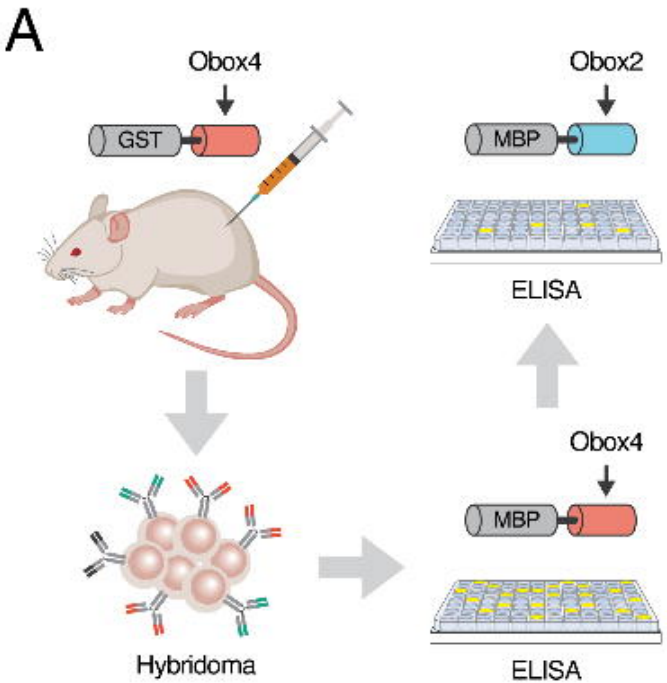
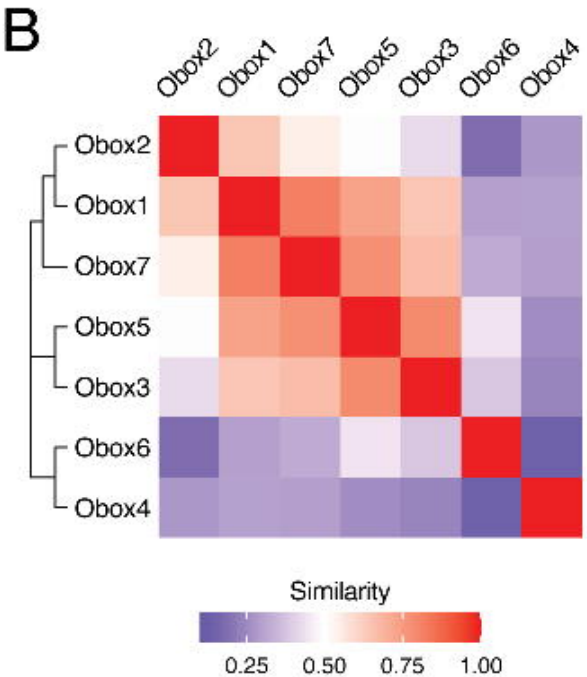

D
TRE-3xFLAG-Obox4
Doxycycline (ng/mL)
WT
DMSO
3
5
7
10
15
kDa
72
55
43
34
FLAG
Gapdh

Figure S2


A
Obox4
GST
Hybridoma
Obox2
MBP
ELISA
Obox4
MBP
ELISA


B
Obox2
Obox1
Obox7
Obox5
Obox3
Obox6
Obox4
Similarity
0.25
0.50
0.75
1.00

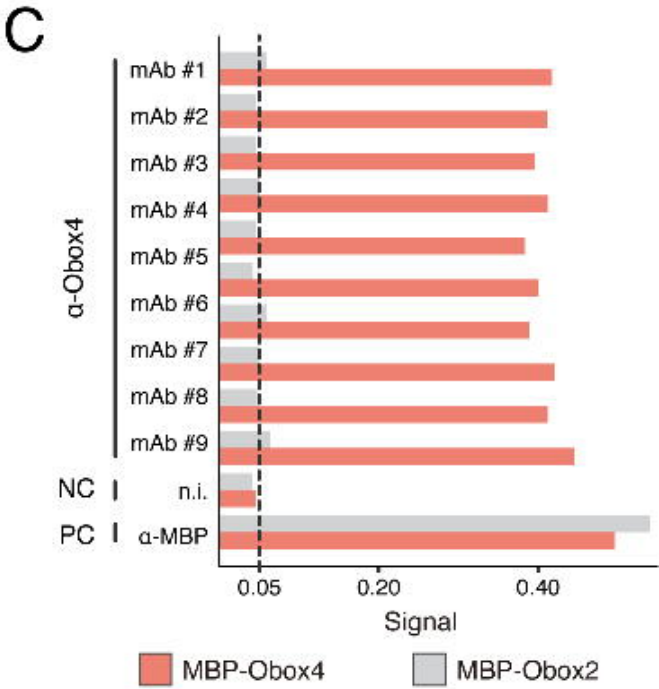
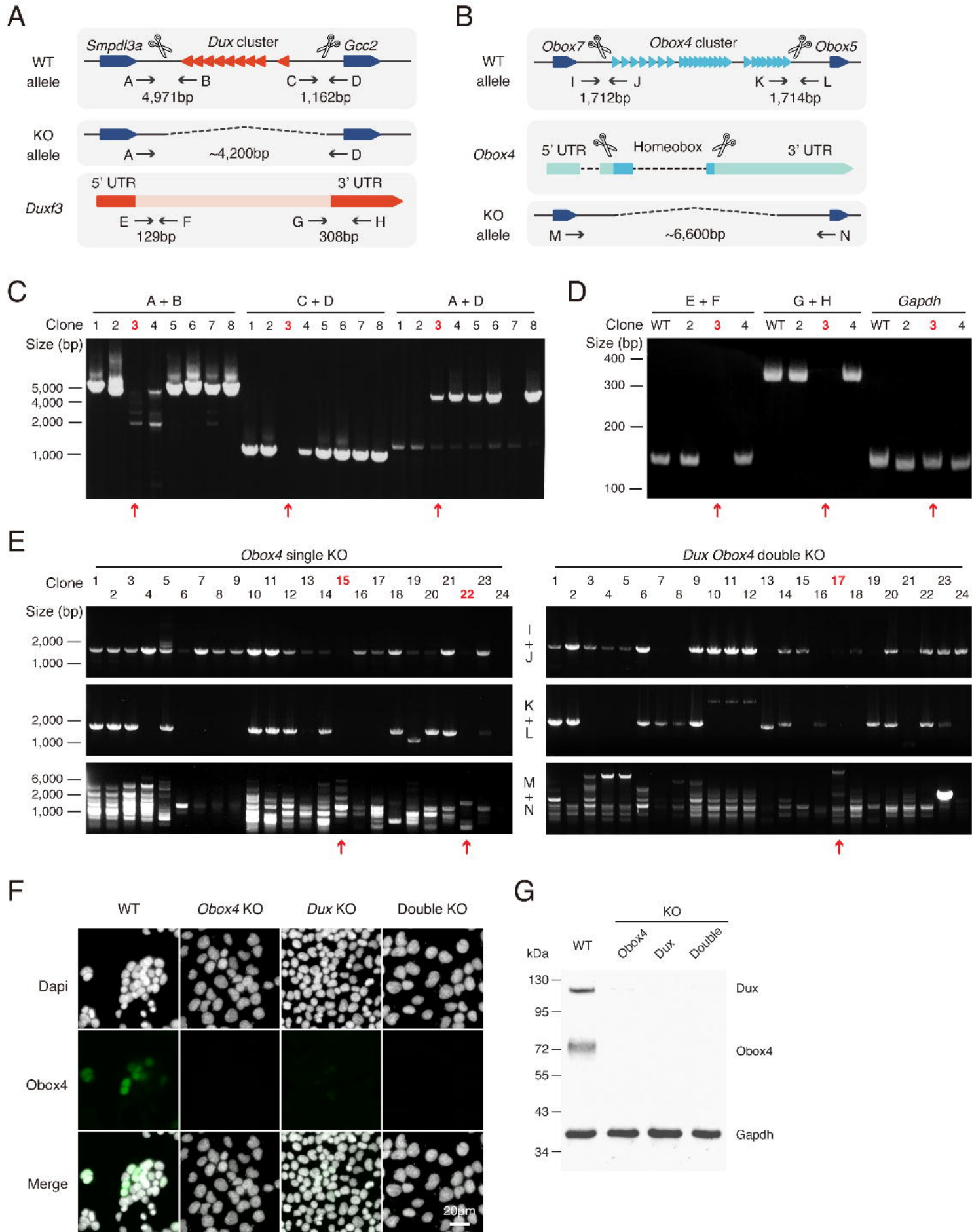

C
α-Obox4
mAb #1
mAb #2
mAb #3
mAb #4
mAb #5
mAb #6
mAb #7
mAb #8
mAb #9
NC
n.i.
PC
α-MBP
0.05
0.20
0.40
Signal
MBP-Obox4
MBP-Obox2

Figure S3


A
WT allele
Smpdl3a
Dux cluster
Gcc2
A → ← B
4,971bp
C → ← D
1,162bp
KO allele
A →
~4,200bp
← D
Duxf3
5' UTR
3' UTR
E → ← F
129bp
G → ← H
308bp
B
WT allele
Obox7
Obox4 cluster
Obox5
I → ← J
1,712bp
K → ← L
1,714bp
Obox4
5' UTR
Homeobox
3' UTR
KO allele
M →
~6,600bp
← N
C
A + B
C + D
A + D
Clone
Size (bp)
5,000
4,000
2,000
1,000
D
E + F
G + H
Gapdh
Clone
WT
Size (bp)
400
300
200
100
E
Obox4 single KO
Dux Obox4 double KO
Clone
Size (bp)
2,000
1,000
6,000
I + J
K + L
M + N
F
WT
Obox4 KO
Dux KO
Double KO
Dapi
Obox4
Merge
20μm
G
KO
WT
Obox4
Dux
Double
kDa
130
95
72
55
43
34
Dux
Obox4
Gapdh