Obox4 promotes 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[*1,4]

[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
[4]Human Biology Microbiome Quantum Research Center (WPI-Bio2Q), Keio University, Shinjuku, Tokyo 160-8545, Japan;

[*]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 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). Here we present multiple lines of evidence that the multicopy homeobox gene, *Obox4*, encodes a transcription factor that is highly expressed in mouse 2-cell embryos and redundantly drives ZGA. Genome-wide profiling revealed that OBOX4 specifically binds and activates MERVL LTRs as well as a subset of murine endogenous retroviruses with lysine tRNA primer (MERVK) LTRs. Depletion of *Obox4* is tolerated by embryogenesis, whereas concomitant *Obox4*/*Dux* depletion markedly compromises embryonic development. 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]. Completion 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 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 an important 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]. 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*), promotes *Dux*-less ZGA. *Obox4* is abundantly expressed in mouse 2C-embryos, 2C-like mESCs, and totipotent blastomere-like cells (TBLCs) [22]. 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. Concomitant, but not respective, depletion of *Obox4* and *Dux* severely compromises ZGA and preimplantation development. Taken together, our findings substantiate a pre-implantation development model, in which the ZGA is redundantly promoted by *Dux* and *Obox4*.

Results

Expression profiling identifies transcription factor candidates

Genetic redundancy is commonly observed among genes that share sequence homology [23]. 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 [24]. 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 an *in vitro* model of 2C-embryos, which are derived from splicing-inhibited mESCs, and hence, are free of parentally inherited transcripts [22]. 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 mESCs (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 found that only the expression of the multicopy gene *Obox4* was partially affected in *Dux* knockout embryos (Figure 1C).

Obox4 activates 2C-genes and TEs in mESCs

Before examining whether the candidates were crucial for embryogenesis, we first confirmed the presence of endogenous OBOX4 protein, as the *Obox4* loci are marked as pseudogenes in the current genome annotation GRCm39 [25]. We generated mouse anti-OBOX4 monoclonal antibodies (Figure 1 - figure supplement 1). Immunofluorescence staining using these antibodies confirmed that endogenous OBOX4 was expressed in zygotes and highly abundant in 2C-embryos and 2C-like mESCs (Figure 1D,E), which is consistent with translatome profiles captured by Ribo-seq that detects OBOX4 translation at the 2C stage [26] (Figure 1F). Mouse ZGA is characterized by the surging activity of genes that specifically express at the 2C stage (2C-gene), many of which have co-opted MT2_Mm, the LTR of MERVL, as stage-specific promoters [27]. We then sought to examine whether the candidates had potential to promote ZGA by activating MERVL and 2C-genes (Figure 2-figure supplement 1A; Supplementary File 1). We constructed a 2C::tdTomato reporter mESC line with transgenic tdTomato red fluorescence protein driven by the MERVL 5'-LTR [28] (Figure 2A). Candidate genes were cloned and overexpressed in the reporter cell line (Figure 2-figure supplement 2B,C). Interestingly, ectopic expression of *Obox4* markedly induced tdTomato expression, similar to that induced by *Dux* (Figure 2B; Figure 2-figure supplement 2A,B). To characterize the impact of *Obox4* expression on 2C-genes, we established an mESC stable line bearing an inducible *Obox4* transgene (Figure 2C). Differentially expressed gene analysis revealed that *Obox4* induction led to transcriptome changes in mESCs, characterized by upregulation of 2C-genes and TEs that were highly expressed in early-to-middle 2C-embryos [29] (Figure 2D,E), naturally occurring 2C-like mESCs [30], and *Dux* induced 2C-like mESCs [16] (Figure 2F; Supplementary File 2). A substantial fraction (159/164) of *Obox4* induced 2C-genes were also induced by *Dux* (Figure 2G). While the induction effect of *Obox4* is milder compared with *Dux* when considering the number and fold-upregulation of 2C-genes, transcriptomic perturbances introduced by *Dux* and *Obox4* are highly correlated (Figure 2H). Interestingly, *Dux* and *Obox4* were mutually inductive, where one promoted expression of the other. These results suggest that *Obox4* is an inducer of 2C-like genes and potentially a redundant factor of *Dux*.

Obox4 binds to 2C-gene promoters and LTR elements in mESCs

Obox4 contains a homeobox, activates 2C-genes and TEs, and upregulates *Dux*-induced genes, which prompted us to examine whether OBOX4 is a transcription factor that activates 2C-gene associated loci through direct binding with *cis*-regulatory elements. Cleavage under targets and release using nuclease (CUT&RUN) leverages antibody-targeted cleavage of proximal DNA to identify the binding sites of DNA-associated proteins [31]. To examine this technique, we performed CUT&RUN against triple FLAG-tagged DUX (3×FLAG-DUX) expressed in mESCs using a high-affinity anti-FLAG antibody [32], and confirmed that the

DUX binding pattern revealed by CUT&RUN recapitulated the published HA-tagged DUX ChIP data [16] (Figure 3-figure supplement 1A,B). We then proceeded with characterizing the genomic footprint of OBOX4 by performing CUT&RUN against 3×FLAG-OBOX4 expressed in mESCs, and discovered ~24,000 peaks, among which 39.5% were located in the gene promoter regions that covered 26.8% (273/1,019) of the 2C-genes (Figure 3A). *De novo* motif discovery using the top 500 CUT&RUN signal peaks predicted CTGGGATYWRMR as the top OBOX4 binding motif, which is enriched in the promoter regions of 2C-genes (Figure 3B). Collectively, OBOX4 and DUX targeted 48.1% (490/1,019) of the 2C-genes with a considerable overlap (36.3% for OBOX4 and 31.3% for DUX), which covered many important ZGA genes including *Dppa2*, *Sp110*, and *Zscan4d* (Figure 3C,D; Supplementary File 3). The overlap was substantiated by the observation that MERVL LTR MT2_Mm was among the top 10 LTR targets of both OBOX4 and DUX, based on loci coverage (Figure 3E,F; Supplementary File 4). Notably, while DUX strongly prefers MT2_Mm, OBOX4 binding was biased toward MERVK LTRs, namely RLTR9 and RLTR13 elements (Figure 3G), which also demonstrated a 2C-specific expression profile (Figure 3-figure supplement 1C,D). To examine whether these bindings were functional in the absence of *Dux*, we generated Obox4/Dux single and double knockout mESC lines, in which 14 of 15 protein-coding *Obox4* copies were removed from the genome (Figure 3-figure supplement 2,3). tdTomato reporters driven by MT2_Mm and RLTR13B2 were co-transfected with *Dux* and *Obox4* into *Obox4*/*Dux* double knockout mESCs (Figure 3H). Analysis of the tdTomato-positive population showed that *Obox4* activated both RLTR13B2 and MT2_Mm, whereas *Dux* only activated MT2_Mm (Figure 3I). These observations demonstrated that OBOX4 binds and activates a subset of DUX targets in mESCs and redundantly drives their expression in the absence of DUX.

Concomitant loss of *Obox4* and *Dux* impairs preimplantation development

We then sought to determine whether *Obox4* is functionally required for pre-implantation development, particularly in the absence of *Dux*. Transient depletion of DUX has previously confused the field in that it caused an embryonic phenotype while subsequent *Dux* knockout females were shown to be fertile [15,18-21]. Therefore, it is critical to examine the functional requirement of OBOX4 and DUX in genetic knockout models. Somatic cell nuclear transfer (SCNT) is a technique to create embryos by transferring nuclei of somatic cells into enucleated oocytes (Figure 4A), which recapitulates ZGA [33]. When the knockout mESCs were subjected to SCNT as nuclear donors, 54.4% of WT mESC derived embryos developed to the blastocyst stage 4 days post nuclear transfer (dpt). *Obox4* and *Dux* single knockout donors led to reduced blastocyst formation rate at 38.5% and 39.7% respectively, whereas double knockout resulted in an additive effect that led to a significantly lower blastocyst rate of 29.3% (Figure 4B). Notably, all blastocysts derived from double knockout mESCs were of low quality, judging from severe morphological abnormalities (Figure 4C).

As an alternative genetic knockout model, we generated mice bearing *Dux* and *Obox4* knockout alleles by direct CRISPR-Cas9 editing in embryos (Figure 4D; Figure 3-figure supplement 2A,B). Among 39 *Obox4* copies, there are 15 copies that maintain an intact open reading frame (ORF) for full-length OBOX4 (Figure 3-figure supplement 2C). While 14 of the 15 intact ORFs form a tightly packed cluster (*Obox4* cluster), a solo ORF (*Obox4-ps33*) remains distant from the *Obox4* cluster and is interspersed with *Obox1*/*2*/*3*, a collection of

Obox family members that are critical for ZGA [34]. To minimize collateral genetic toxicity and interference to the experimental result caused by the removal of other *Obox* members, only the *Obox4* cluster was knocked out and the solo ORF was retained (Figure 3-figure supplement 2D; Figure 4-figure supplement 1). The observed *Obox4*^{KO} frequency (7/29) in the progenies of *Obox4*^{Het} × *Obox4*^{Het} crossing is consistent with the expected 25% Mendelian ratio (Figure 4E; Figure 4-figure supplement 2; and Supplementary File 5). That *Obox4*^{KO} mice developed to adulthood without discernable abnormalities and are fertile when intercrossed suggest that development is compatible with loss of *Obox4* (Figure 4F,G). We next attempted to produce *Dux*/*Obox4* double knockout (*Dux*^{KO}/*Obox4*^{KO}) mice to examine whether concomitant loss of *Dux* and *Obox4* compromises embryogenesis. We genotyped 54 pups from three litters of *Dux*^{KO}/*Obox4*^{Het} × *Dux*^{KO}/*Obox4*^{Het} and six litters of *Dux*^{KO}/*Obox4*^{Het} × *Dux*^{Het}/*Obox4*^{Het} mating pairs (Figure 4-figure supplement 3; Supplementary File 5). One *Obox4*^{KO}/*Dux*^{KO} pup was present, and the frequency of *Obox4*^{KO}/*Dux*^{KO} and *Obox4*^{KO}/*Dux*^{Het} were significantly lower than the expected 25% Mendelian ratio (Figure 4H). Development monitoring and genotyping of embryos produced by *Dux*^{KO}/*Obox4*^{Het} × *Dux*^{Het}/*Obox4*^{Het} mating pairs at 4.5 days post coitum (dpc) revealed that *Dux*^{KO}/*Obox4*^{KO} was under-represented in blastocysts while over-represented in 2-cell arrest and degenerated embryos (Figure 4I, Figure 4-figure supplement 4A-C; Supplementary File 6). Despite repeated attempts, various crossing strategies failed to produce *Obox4*^{KO}/*Dux*^{KO} mating pairs that could be used to produce large number of *Obox4*^{KO}/*Dux*^{KO} embryos required for transcriptome analysis. As an alternative, we performed single blastomere genotyping and RNA-seq of 2C embryos produced by *Dux*^{KO}/*Obox4*^{Het} crossings (Figure 4-figure supplement 4D,E) which showed dysregulation of 2C-genes and TEs targeted by OBOX4 (Figure 4J; Supplementary File 7). The impaired development and transcriptome of *Dux*^{KO}/*Obox4*^{KO} embryos at the 2-cell stage showed that ZGA was defective in these embryos.

As large numbers of *Obox4*^{KO}/*Dux*^{KO} embryos are inaccessible due to technical limitations, we adopted an alternative approach to capture high quality transcriptomes upon *Dux*/*Obox4* depletion. Microinjection of antisense oligonucleotide (ASO) into male pronuclei of zygotes was performed to knockdown *Obox4* and *Dux* (Figure 4K; Figure 4-figure supplement 5A-C). Monitoring development until 4.5 dpc showed that *Obox4* single knockdown resulted in moderate developmental defects, with almost 50% of embryos reaching the blastocyst stage. Similarly, blastocyst formation was preserved in nearly 25% of *Dux* single knockdown embryos, which is consistent with previously reported *Dux* knockdown/knockout experiments [15,18-21]. The *Obox4*/*Dux* double knockdown markedly compromised blastocyst formation, resulting in more than 80% of embryos degenerated or arrested before the 4C stage, and less than 20% manifested morula-like morphology (Figure 4K,L). The double knockdown embryos with morula-like morphology manifest a dysregulated transcriptome characterized by a failure of expressing morula specific genes and activation of apoptosis related pathways, suggesting that the development was dysfunctional instead of delayed (Figure 4-figure supplement 5D-G).

The consistency among constitutive knockout in SCNT embryos, living mice and ASO-mediated transient depletion demonstrated that the expression of *Obox4* and *Dux* is collectively important for pre-implantation development, and that *Obox4* is capable of promoting ZGA in a *Dux*-independent manner. Collectively, these data show that *Obox4* promotes mouse preimplantation development in the absence of *Dux*.

OBOX4 promotes 2C-gene expression upon depletion of DUX

As *Obox4* has been shown to activate 2C-genes and TEs in mESCs, and embryonic depletion of OBOX4 impairs 2C-genes and TEs activation, we asked whether the impairment can be ameliorated by restoration of OBOX4. First, we determined whether the presence of OBOX4 can rescue the developmental arrest of *Obox4*/*Dux* double knockdown embryos. Codon-optimized *Obox4* mRNA without the ASO target motif was produced using *in vitro* transcription. The rescue experiment was performed by means of co-microinjection of *Obox4* mRNA with *Obox4*/*Dux* double knockdown ASOs (Figure 5A). Restoration of OBOX4 in *Obox4* mRNA-microinjected embryos was confirmed using immunofluorescence staining at the 2C stage (Figure 5B). As expected, development monitoring showed that *Obox4*/*Dux* double knockdown embryos failed blastocyst formation at 4.5 dpc, whereas among *Obox4* mRNA-rescued embryos, blastocyst formation was retained at a similar level to that observed in the *Dux* single knockdown experiment (Figure 5C,D; Figure 4K,L). RNA-seq of 2C-embryos revealed exacerbated dysregulation of 2C-genes and TEs by *Obox4*/*Dux* double knockdown compared to single knockdowns, and the dysregulated transcriptome of double knockdown was rescued by resupplying OBOX4, revealed by the number of up and down-regulated genes (Figure 5E,F). Differential expression analysis revealed that 58% (102/175) of the down-regulated 2C-genes in double knockdown embryos were rescued, including 2C stage markers and important ZGA factors (Figure 5G,H). These results showed that OBOX4 redundantly activates 2C genes to promote ZGA under DUX depletion.

Obox4/*Dux* redundancy is distinct among *Obox* family

Recently, Ji *et al.* demonstrated that OBOX family proteins are important regulators of ZGA in a DUX independent manner [34]. Interestingly, they showed that maternal (*Obox1*/*2*/*5*/*7*) and zygotic (*Obox3*/*4*) *Obox* members seem to be redundant, as loss of either is compatible with embryogenesis. This prompted us to ask whether *Obox4* targets are distinctive from that of other *Obox* family members. Comparative analysis revealed distinctive patterns of *Dux*/*Obox3*/*Obox4*/*Obox5* induced transcriptomes in mESCs (Figure 6A). While 2C-genes induced by *Obox3* and *Obox5* overlap substantially, the majority of them were not induced by *Obox4* (Figure 6B). The expression changes of *Obox3*/*Obox5* induced 2C-genes are highly correlated, characterized by coefficient of determination R^2 (Figure 6C), whereas the *Obox4* induced transcriptome poorly correlated with those of *Obox3* and *Obox5* (Figure 6D,E; Supplementary File 8). Except for MERVL, transposable elements induced by *Obox3* and *Obox5* were not induced by *Obox4* (Figure 6D,E; Supplementary File 8). Similarly, transcriptomes are poorly correlated between *Dux*/*Obox4* double knockdown (DKD) and *Obox* maternal/zygotic knockout (mzKO) 2C embryos, with down-regulation of distinct groups of 2C-genes (Figure 6F,G). The transcriptome of *Obox4* knockdown embryos showed a higher correlation with *Dux*/*Obox4* DKD embryos than *Obox* mzKO embryos (Figure 6G,H; Supplementary File 9), suggesting that the functional redundancy to *Dux* is a distinct feature of *Obox4* among *Obox* family.

Discussion

Starting with a transcriptionally inert genome, the eutherian ZGA engages in massive yet coordinated expression of genes driven by TE LTRs. In mice, it has been reported that the transcription factor DUX accesses

and opens condensed chromatin of the MERVL LTR loci, which results in the activation of downstream ZGA genes [15-17]. However, the establishment of fertile *Dux* knockout mice suggests that ZGA is redundantly driven by other transcription factor(s) [18-21]. *Obox* transcripts were first discovered in gonads [35], and were later found to be highly abundant in mouse 2C-embryos [36]. Despite being important regulators of ZGA [34], the biological significance of distinct *Obox* genes remain elusive. The *Obox* family has 67 members clustered in the sub-telomeric region of mouse chromosome 7 and is further divided into six subfamilies: *Obox1*, *Obox2*, *Obox3*, *Obox4*, *Obox5*, and *Obox6*, with *Obox4* constituting nearly 60% (39 of 67) of the family population [37,38]. The role of *Obox4* is particularly mysterious, as it remains to be determined whether *Obox4* is a functional protein-coding gene. In the present study, we found that *Obox4* is a *bona fide* multicopy protein-coding gene that encodes a homeodomain-containing protein that potently induces ZGA gene expression. Using multiple genetic knockout models, we showed that concomitant depletion of *Dux* and *Obox4* was hardly compatible with embryogenesis, as *Dux*/*Obox4* double knockout mESCs subjected to SCNT failed to produce pups, and *Dux*/*Obox4* double knockout mouse was produced at a sub-Mendelian ratio. Consistently, *Obox4*/*Dux* double knockdown markedly compromised pre-implantation development, but single knockdown of either gene was tolerated. Double knockdown embryos exhibited severely dysregulated transcriptomes, characterized by MERVL and ZGA gene activation defects. We also characterized the molecular mechanisms underlying the biological significance of *Obox4* as well as its relevance to *Dux*. OBOX4 directly binds to MERVL (the target of DUX) and MERVK loci and activates specific MERVL and MERVK elements in a *Dux*-independent manner. In summary, our results highlight that OBOX4 is a transcription factor that is functionally redundant to DUX during ZGA.

Despite the narrow time window, ZGA takes place as a stepped program. A wave of mild and promiscuous transcription of "minor ZGA" is commenced first, followed by "major ZGA" characterized by transcriptional bursting of a more defined set of genes [39]. While major ZGA in large mammals usually launches multiple cell cycles after minor ZGA, it is worth noting that mouse ZGA adopts an unique "early genome activator" strategy where the timing of minor and major ZGA converges at the early 2-cell stage [40]. Overlapping minor and major ZGA right after fertilization allows prompt initiation of the zygotic program and accelerated preimplantation development – a potentially favorable reproductive advantage in rodents. Multiple shared features of *Dux* and *Obox4* imply that these genes have adapted to this particular developmental scheme to serve redundantly as key ZGA factors. Qualitatively, the functional redundancy between DUX and OBOX4 buffers sporadic mutations or defects that could arise during ZGA. Their lack of introns enables the earliest possible translation under ZGA associated splicing deficiency [22]. Quantitively, both genes are tandemly arrayed in clusters, which assures rapid accumulation of their transcripts. Taken together, it seems that mouse ZGA may represent an evolutionary innovation where the first embryonic program shows surprising plasticity, likely supported by expression of multiple *Obox* family members throughout peri-ZGA stages (Figure 6I). Indeed, concomitant loss of *Dux* and nearly all *Obox4* copies only partially compromises the ZGA transcriptome, demonstrating that mouse ZGA is so strongly canalized that even such a potent perturbance in genes make very little difference to the phenotype.

Unlike *Dux*, which is structurally and functionally conserved throughout placentalia [41], the ancestral locus of the *Obox* family appears to have undergone mouse-specific duplication and generated a gene cluster that is collectively syntenic to the *Tprx2* locus in other mammals [42]. Despite the distal homology between *Obox4* and *Tprx2*, it has been recently reported that human *TPRX2* is expressed in 8-cell embryos [14]. Interestingly, TPRX2 was shown to cause defective ZGA upon embryonic depletion and bind important ZGA genes in hESCs [43], suggesting that the *Tprx2* locus might have undergone functionally convergent evolution despite its divergent genetic context. However, due to the distinction of minor and major ZGA timing across species, whether *TPRX2* plays a similar role in humans to *Obox4* in mice regarding the redundancy to *Dux* (*DUX4* in humans) warrants careful examination. Nevertheless, considering that LTR domestication is functionally convergent in preimplantation development [44], it might be helpful to speculate that homeobox gene mediated LTR activation is a common principle of ZGA. Meanwhile, because the repressive epigenetic landscape is barely established during minor ZGA, this may allow transcription factors like DUX and OBOX to potently shape the zygotic transcriptome by activating LTR elements. However, ZGA gene induction and epigenetic reprogramming can only partially recapitulate the 2C transcriptome in mESCs [45,46], suggesting that ZGA is a collective outcome of specific parental, zygotic, and epigenetic factors that may or may not be present in pluripotent stem cells. Indeed, OBOX4 is equally potent as DUX at activating a 2C::tdTomato reporter but less potent at inducing 2C-genes in mESCs, which may suggest that the repressive epigenetic landscape is less permissive to OBOX4 than to DUX. Thus, attempts to model ZGA *in vitro* need to be carefully formulated to address these fundamental distinctions between embryos and cultured cells.

Legends

Figure 1. OBOX4 is expressed during ZGA.

(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* mESCs. 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 downregulates the multicopy homeobox gene *Obox4*. **(D)** Immunofluorescence staining of OBOX4 and MERVL GAG at different pre-implantation embryo stages. **(E)** Immunofluorescence staining of DUX, OBOX4, and OCT4 in 2C-like mESCs. **(F)** Translation profile of DUX and OBOX4 during preimplantation characterized by Ribo-seq signal.

Figure 2. *Obox4* and *Dux* induce 2C-gene expression in mESCs.

(A) Diagram of the 2C::tdTomato reporter assay. mESCs 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. **(B)** Boxplot showing normalized 2C-like cell percentage in 2C::tdTomato reporter mESCs overexpressing candidate

pioneer factors. *Dux* and *Obox4* potently induced a 2C-like state. n=3 biological replicates. **(C)** Upper panel: schematic of *Obox4*-inducible cell line construction. Lower panel: western blot showing OBOX4 level upon induction by different concentrations of doxycycline. Expression of OBOX4 was carried out in a dose-dependent manner. **(D)** 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. n=3 biological replicates. **(E)** 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. n=3 biological replicates. **(F)** Heatmaps of the expression of 2C-genes in preimplantation embryos, naturally occurred 2C-like mESCs, and induced 2C-like mESCs. **(G)** Venn diagram showing overlap of 2C-genes with genes induced by ectopic expression of *Dux* and *Obox4* in mESCs. **(H)** Scatterplot showing per-gene expression changes in *Dux* induced *versus Obox4* induced mESCs. 2C-genes are highlighted in red.

Figure 2-Source Data Original and uncropped blot for Figure 2C.

Figure 3. OBOX4 binds and activates 2C-specific LTR elements.

(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: histograph showing the distribution of the predicted OBOX4 binding motif near 2C and random gene promoters. **(C)** Venn diagram showing the distinct and overlapping 2C-genes targeted by DUX and OBOX4. **(D)** Representative genomic track showing DUX and OBOX4 binding sites at the *Dppa2*, *Sp110,* and *Zscan4d* loci and their expression levels in *Obox4* and *Dux* induced mESCs. Read counts were CPM normalized. The OBOX4 binding sites overlapped with those of DUX. *Dppa2, Sp110*, and *Zscan4d* expression was upregulated upon *Obox4* and *Dux* induction. **(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. 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)** 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. **(I)** 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 4. Concomitant loss of OBOX4 and DUX severely hinders ZGA.

(A) Schematic representation of the somatic cell nuclear transfer (SCNT) experiment. The nuclei of knockout mESCs were transferred into enucleated oocytes to generate zygotes with knockout genotype. **(B)** Upper panel: percent SCNT embryos developed to different stages at 4 dpt. Lower panel: *p*-value of non-parametric analysis of variance (ANOVA) among different genotypes and stages comparing to wildtype SCNT embryos. Three independent experiments were conducted, with 150-200 embryos per condition. **(C)** Representative picture of SCNT embryos at 4 dpt. Morphologically abnormal blastocysts are highlighted. Blastocysts generated by double knockout mESCs were severely defective. **(D)** Schematic representation of CRISPR-Cas9 mediated *Dux* and *Obox4* knockout mouse production. *In vitro* fertilized mouse zygotes were electroporated with pre-assembled CRISPR-Cas9 complex targeting *Dux* and *Obox4* loci. **(E)** Bar plot showing genotype percentage of the pups delivered by $Obox4^{Het}$ intercrosses. 4 litters delivered 29 pups, litter size 7.25 ± 1.26. ns *p*-value = 0.9146, chi-square goodness of fit test. **(F)** Representative photos of $Obox4^{KO}$ and WT adult mice analyzed in **E**. **(G)** Photo of $Obox4^{KO}$ intercross litter with live pups. **(H)** Bar plot showing genotype of *Obox4* allele in the pups delivered by crossing of $Dux^{KO}/Obox4^{Het} \times Dux^{KO}/Obox4^{Het}$ or $Dux^{KO}/Obox4^{Het} \times Dux^{Het}/Obox4^{Het}$. Nine litters delivered 54 pups, *Dux* heterozygous and knockout allele were present in 31 and 23 pups, respectively. ** *p*-value = 0.001306; * *p*-value = 0.02218; chi-square goodness of fit test. **(I)** Bar plot showing observed percentages of different preimplantation stage embryos bearing *Obox4* KO allele with *Dux* heterozygous or knockout allele at 4.5 dpc. Among the total 94 embryos assessed, two degenerated, ten 2-cell arrest, two 4-8 cell arrest, four morula arrest embryos were observed at 4.5 dpc, whereas 76 developed to blastocyst. *** *p*-value = 3.564×10^{-5}; for 2-cell * *p*-value = 0.04348; for blastocyst * *p*-value = 0.02549; chi-square goodness of fit test. **(J)** Scatterplot showing expression log2 fold changes of genes (left panel) and TEs (right panel) in $Dux^{KO}/Obox4^{KO}$ and $Dux^{KO}/Obox4^{Het}$ *versus* $Dux^{KO}/Obox4^{WT}$ 2C embryos. 2C-genes and ERVK elements targeted by OBOX4 are highlighted in blue. MT2_Mm and MERVL-int are labeled. 79 and 5 2C-genes were down-regulated in $Dux^{KO}/Obox4^{KO}$ and $Dux^{KO}/Obox4^{Het}$ 2C embryos, respectively. **(K)** Upper panel: 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. Lower panel: the percentages of embryonic stages observed at 1.5 dpc, 2.5 dpc, 3.5 dpc, and 4.5 dpc. **(L)** Representative picture of KD embryos at 4.5 dpc. No blastocyst was observed among double ASO knockdown embryos.

Figure 5. OBOX4 promotes ZGA in the absence of DUX.

(A) Schematic of the double knockdown rescue experiment. Male pronuclei of zygotes were injected with ASO targeting *Obox4* and *Dux* transcripts as well as *in vitro*-transcribed codon-optimized *Obox4* mRNA. **(B)** Immunofluorescence staining of OBOX4 in early 2-cell embryos microinjected with scrambled ASO, double ASO (*Obox4* and *Dux*), or double ASO with codon-optimized *Obox4* mRNA. **(C)** Representative picture of knockdown and rescue embryos at 4.5 dpc. Codon-optimized *Obox4* mRNA injection rescued blastocyst formation in ASO knockdown embryos. **(D)** The percentages of embryonic stages observed at 1.5 dpc, 2.5 dpc, 3.5 dpc, and 4.5 dpc. The plot represents the sum of three independent experiments, with 80-100 embryos per condition. **(E)** Upper panel: volcano plot showing the results of DEG analysis of 2C-genes in knockdown and

rescue embryos, as compared to that in scramble ASO-injected embryos. Standard deviations of log_2(fold-change) were used to represent the degree of transcriptome dysregulation. Stdev, standard deviation. Lower panel: MA plot of differential expression of TEs in knockdown and recue embryos, as compared to that in scramble ASO-injected embryos. MERVK elements and MERVL are highlighted. n=3 biological replicates. **(F)** Heatmaps of the expression of 2C-genes in preimplantation embryos, knockdown 2C-embryos, and rescue 2C-embryos. **(G)** Rain plot displaying the expression change distribution of rescued 2C-genes in double knockdown and recue 2C-embryos. **(H)** Bar plots showing the expression of representative 2C-genes in knockdown and recue embryos; n=3 biological replicates.

Figure 6. OBOX4 mediated DUX redundancy is distinct among OBOX family.

(A) Heatmaps of the expression of 2C-genes in preimplantation embryos, non-induced mESCs, and *Dux*/*Obox4*/*Obox3*/*Obox5* induced mESCs. **(B)** Venn diagram showing overlap of 2C-genes induced by ectopic expression of *Obox3*, *Obox4*, and *Obox5* in mESCs. **(C)** Scatterplot showing per-gene expression changes in Obox3 versus Obox5 induced mESCs. 2C-genes are highlighted in red. **(D)** Scatterplot showing per gene (left panel) and per transposable element (right panel) expression changes in *Obox3* and induced *versus Obox4* induced mESCs. 2C-genes and *Obox3* induced TEs are highlighted in red. **(E)** Scatterplot showing per gene (left panel) and per transposable element (right panel) expression changes in *Obox5* and induced *versus Obox4* induced mESCs. 2C-genes and *Obox3* induced TEs are highlighted in red. **(F)** Venn diagram showing overlap of 2C-genes down-regulated upon *Obox4* knockdown, *Dux*/*Obox4* double knockdown, and *Obox* maternal-zygotic knockout. **(G)** Scatterplot showing per-gene expression changes in *Obox* maternal-zygotic knockout versus *Dux*/*Obox4* double knockdown 2C embryos. 2C-genes are highlighted in red. **(H)** Scatterplot showing per gene expression changes in *Obox4* knockdown versus *Dux*/*Obox4* double knockdown (left panel) and *Obox* maternal-zygotic knockout (right panel) 2C embryos. 2C-genes are highlighted in red. **(I)** Schematic model of different ZGA strategy employed by human and mouse. *Obox* family members with redundant functions are expressed in high dose at peri-ZGA stages to ensure rapid ZGA. *Obox4* evolved as a divergent *Obox* family member that provides functional redundancy to *Dux*.

Figure 1-figure supplement 1

(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.

Figure 2-figure supplement 1

(A) Expression profiles of *k*-means clustered gene sets during preimplantation development based on *z*-scored expression of scRNA-seq and bulk RNA-seq. Genes appeared in both scRNA-seq cluster 1 and bulk RNA-seq cluster 4 were designated as 2C-genes. **(B)** Western blot showing protein products of the candidate genes ectopically expressed in mESCs at 18 hours post transfection. **(C)** Bar plot showing MERVL transcript level in mESCs at 18 hours post *Dux* or *Obox4* transfection, detected by reverse transcription quantitative real-time PCR (RT-qPCR).

Figure 2-figure supplement 1-Source Data Original and uncropped blot for Figure 2-figure supplement 1B.

Figure 2-figure supplement 2

(A) FACS analysis of 2C::tdTomato cells co-transfected with candidate gene expression plasmids and EGFP. **(B)** Fluorescence microscopy of 2C::tdTomato mESCs transfected with empty plasmid, plasmid encoding DUX, or plasmid encoding OBOX4. Plasmids encoding EGFP were co-transfected to normalize the transfection efficiency in FACS analysis.

Figure 3-figure supplement 1

(A) Schematic diagram of the CUT&RUN workflow using mESCs with induced expression of FLAG-tagged DUX. mESCs bearing a doxycycline-inducible 3×FLAG-DUX expression cassette were induced with doxycycline for 24 h. DUX-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 5 OBOX4 covered LTR elements during embryogenesis.

Figure 3-figure supplement 2

(A) Schematic diagram of producing *Obox4* knockout mice using CRISPR-Cas9 (not drawn to scale). sgRNAs flanking the *Obox4* cluster as well as the internal sequence of individual *Obox4* were introduced into mouse zygotes 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 mice using CRISPR-Cas9 (not drawn to scale). Previously reported sgRNAs flanking the *Dux* cluster were introduced into mouse zygotes 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)** Multiple sequence alignment of DNA sequence among the *Obox4* copies. Full length *Obox4* with internal stop codons are marked by asterisks. **(D)** Schematic diagram of *Obox4* cluster removal strategy and 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.

Figure 3-figure supplement 3

(A) Agarose gel image of PCR-based genotyping of the *Obox4* knockout allele in founder mESC lines. The red arrows indicate cell lines bearing bi-allelic deletions of the *Obox4* cluster. **(B)** Agarose gel image of PCR-based genotyping of the *Dux* knockout allele in founder mESC lines. The red arrows indicate cell lines without the genomic *Dux* sequences. **(C)** Bar plot showing copy number of *Obox4* in the founder mESC lines analyzed in **A**,**B**, as detected by qPCR using purified genomic DNA.

Figure 3-figure supplement 3-Source Data 1 Original and uncropped gels for Figure 3-figure supplement 3A.

Figure 3-figure supplement 3-Source Data 2 Original and uncropped gels for Figure 3-figure supplement 3B.

Figure 4-figure supplement 1

(A) Agarose gel image of PCR-based genotyping of the *Obox4* knockout allele in the F_1 mice. **(B)** Agarose gel image of PCR-based genotyping of the *Dux* knockout allele in the F_1 mice. **(C)** Agarose gel image of PCR-based genotyping of the *Obox4* cluster 5' cutting site allele in the F_1 mice. **(D)** Agarose gel image of PCR-based genotyping of the *Dux* cluster 5' cutting site allele in the F_1 mice. **(E)** Agarose gel image of PCR-based genotyping of the *Dux* internal sequence in the F_1 mice. **(F)** Bar plot showing copy number of *Obox4* in the F_1 mice analyzed in **A-E**, as detected by qPCR using purified genomic DNA.

Figure 4-figure supplement 1-Source Data 1 Original and uncropped gels of F_1 mice 108 – 152 for Figure 4-figure supplement 1A-E.

Figure 4-figure supplement 1-Source Data 2 Original and uncropped gels of F_1 mice 200 – 204 for Figure 4-figure supplement 1A-E.

Figure 4-figure supplement 2

(A) Agarose gel image of PCR-based genotyping of the *Obox4* knockout and *Obox4* cluster 5' cutting site wild type allele in the F_2 mice produced by intercrossing $Obox4^{Het}$ F_1 mice. **(B)** Bar plot showing copy number of *Obox4* in the F_2 mice analyzed in **a**, as detected by qPCR using purified genomic DNA.

Figure 4-figure supplement 2-Source Data Original and uncropped gels of Figure 4-figure supplement 2A.

Figure 4-figure supplement 3

(A) Bar and dot plot showing litter sizes of intercross used to produce *Dux*/*Obox4* double knockout mice. Each dot represents individual litter size. six and three litters were analyzed for $Dux^{KO}/Obox4^{Het} \times Dux^{Het}/Obox4^{Het}$ and $Dux^{KO}/Obox4^{Het} \times Dux^{KO}/Obox4^{Het}$ mating, respectively. Crossing of $Dux^{KO}/Obox4^{Het} \times Dux^{KO}/Obox4^{Het}$ delivered reduced litter size than those of $Dux^{KO}/Obox4^{Het} \times Dux^{Het}/Obox4^{Het}$. * p-value = 0.01104, two-tailed Student's t test. Error bars indicate standard deviations. **(B)** Agarose gel image of PCR-based genotyping of the *Obox4* knockout, *Obox4* cluster 5' cutting site allele, and *Dux* internal sequence in the F_2 mice produced by intercrossing $Dux^{KO}/Obox4^{Het}$ and/or $Dux^{Het}/Obox4^{Het}$ F_1 mice. **(C)** Bar plot showing copy number of *Obox4* in the F_2 mice analyzed in **a**, as detected by qPCR using purified genomic DNA.

Figure 4-figure supplement 3-Source Data 1 Original and uncropped gels of F_2 mice 187s – 201s for Figure 4-figure supplement 3B.

Figure 4-figure supplement 3-Source Data 2 Original and uncropped gels of F_2 mice 202s – 206s for Figure 4-figure supplement 3B.

Figure 4-figure supplement 3-Source Data 3 Original and uncropped gels of F_2 mice 207s – 215s for Figure 4-figure supplement 3B.

Figure 4-figure supplement 3-Source Data 4 Original and uncropped gels of F_2 mice 216s – 240s for Figure 4-figure supplement 3B.

Figure 4-figure supplement 4

(A) Schematic diagram of knockout embryo production for development monitoring. *Dux*^{Het}/*Obox4*^{Het} female mice were mated with *Dux*^{KO}/*Obox4*^{Het} male mice. Embryos were collected at 0.5 dpc and cultured *in vitro* until 4.5 dpc, followed by developmental assessment and genotyping. **(B)** Agarose gel image of multiplex PCR-based genotyping of *Obox4* cluster 5' cutting site allele, *Dux* internal sequence, and *Tardbp* internal sequence in the embryos produced by intercrossing *Dux*^{KO}/*Obox4*^{Het} and *Dux*^{Het}/*Obox4*^{Het} F_1 mice. Note that multiplexing three sets of PCR primers produced artifact band in *Dux*^{KO}/*Obox4*^{KO} embryos. **(C)** Representative pictures of *Dux*^{KO}/*Obox4*^{KO} embryos. **(D)** Schematic diagram of knockout embryo production for RNA-seq. Embryos were collected at 0.5 dpc from *Dux*^{KO}/*Obox4*^{Het} mating pairs and cultured *in vitro* until 2-cell stage at 1.5 dpc, blastomeres of each embryo were separated and subject to genotyping and RNA-seq respectively. **(E)** Agarose gel image of genotyping of *Obox4* cluster 5' cutting site allele in the embryos produced by intercrossing *Dux*^{KO}/*Obox4*^{He} mice.

Figure 4-figure supplement 4-Source Data 1 Original and uncropped gels for Figure 4-figure supplement 4B.

Figure 4-figure supplement 4-Source Data 2 Original and uncropped gels for Figure 4-figure supplement 4E.

Figure 4-figure supplement 5

(A) Transcript levels of *Dux* and *Obox4* in ASO-injected embryos, as measured using RNA-seq. CPM, counts per million. **(B)** Representative images of immunofluorescence co-staining of OBOX4 and DUX in 2C embryos subjected to ASO-mediated knockdown. **(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 enrichment of gene ontology (GO) terms of up-regulated genes in double knockdown morulae.

Supplementary file 1

List of Z-score and k-means clustering of genes based on preimplantation expression profiling data by Wu *et al*. (GSE66390) and Deng *et al* (GSE45719).

Supplementary file 2

Deferential expression analysis results of genes and transposable elements in *Obox4* and *Dux* overexpressing mESCs.

Supplementary file 3

List OBOX4 and DUX bond genes.

Supplementary file 4

Total numbers of annotated loci in each transposable element family and numbers of loci bond by OBOX4.

Supplementary file 5

List of mating pairs and genotyping results of *Dux*/*Obox4* knockout mice.

Supplementary file 6

Genotyping and development monitoring results of preimplantation embryos produced by $Dux^{KO}/Obox4^{Het}$ × $Dux^{Het}/Obox4^{Het}$ mating pair.

Supplementary file 7

Deferential expression analysis results of genes and transposable elements in $Dux^{KO}/Obox4^{Het}$ and $Dux^{KO}/Obox4^{KO}$ versus $Dux^{KO}/Obox4^{WT}$ 2C embryos.

Supplementary file 8

Combined deferential expression analysis results of genes and transposable elements in *Obox3*/*4*/*5* overexpressing mESCs.

Supplementary file 9

Combined deferential expression analysis results of genes and transposable elements in *Dux*/*Obox4* knockdown and *Obox* family knockout 2C embryos.

Supplementary file 10

Primer sequences used in this study.

Supplementary file 11

Antisense oligonucleotides sequences used in this study.

Supplementary file 12

sgRNA sequences used in this study.

Supplementary file 13

Key resources table.

Materials and Methods

Key resources

Sequences of primers, ASOs, and sgRNAs used in this study can be found in Supplementary File 11,12. A complete list of key reagents and biological samples can be found in Supplementary File 13.

Mice

Male/female BDF1 and female BALB/c mice were purchased from Japan SLC Inc. Mice were fed regular 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 the Animal Experimentation Committee at the RIKEN Tsukuba Institute and conducted in compliance with the Keio University Code of Research Ethics (License #11045-4) and the RIKEN's guiding principles (T2023-Jitsu015).

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 incubation at room temperature, the samples were centrifuged at 12,000 × *g* for 15 min at 4°C. The supernatant was then discarded, following which the RNA pellets were washed with 70% ethanol (catalog no. 057-00456, Wako). The RNA pellets were then air-dried and dissolved in RNase-free water. The RNA solution was subjected to DNase treatment using TURBO DNA-*free*™ kit (catalog no. AM1907, Thermo Fisher), according to the manufacturer's instructions, to remove genomic DNA carryover. RNA (1 μg) was reverse transcribed to cDNA using the Transcriptor First Strand cDNA Synthesis Kit (catalog no. 04379012001, Roche), according to the manufacturer's instructions. PrimeSTAR® Max DNA Polymerase (catalog no. R045A, TaKaRa) and ProFlex™ PCR System (catalog no. 4484073, Thermo Fisher) were used to amplify the sequence of interest from the cDNA. NEBuilder® HiFi DNA Assembly Kit (catalog no. E2621L, NEB) was used to clone the sequence of interest into plasmid backbone.

Immunofluorescence

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

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

Western blot

Freshly harvested cells were re-suspended in phosphate-buffered saline (PBS) (catalog no. 14249-95, Nacalai Tesque) and lysed by means of sonication. Whole cell lysates were mixed with sample buffer containing reducing reagent (catalog no. 09499-14, Nacalai Tesque) and boiled at 95°C for 5 min. Protein samples were loaded on a 10% tris-glycine gel, run in AllView PAGE Buffer® (catalog no. DS520, BioDynamics Laboratory), and then transferred to nitrocellulose membranes (catalog no. 10600003, Cytiva) using a Power Blotter System (catalog no. PB0012, Thermo Fisher). The membranes were blocked with PBST containing 2% skim milk (catalog no. 4902720131292, Morinaga), at room temperature for 15 min. Membranes were incubated with primary antibodies in PBST containing 2% skim milk, at room temperature for 30 min. After three washes with PBST, the membrane was incubated with secondary antibodies in PBST containing 2% skim milk, at room temperature for 15 min, with shaking. After three washes with PBST, the membrane was incubated with ECL reagents (catalog no. RPN2232, Cytiva) and exposed to an X-ray film (catalog no. 28906839, Cytiva) or detected using a digital chemiluminescence imager (catalog no. 17001402JA, Bio-Rad).

qPCR

qPCR was performed using TB Green Fast qPCR Mix (catalog no. RR430A, TaKaRa) according to the manufacturer's instructions. The qPCR reactions were carried out and the signals were detected using a real time PCR system (catalog no. TP950, TaKaRa). For RT-qPCR, first strand cDNA of mESCs total RNA was produced as described in methods **cDNA synthesis and cloning** section. Primers targeting MERVL Gag consensus sequence were used to detect MERVL transcript. Primers targeting beta actin mRNA were used as internal control in all qPCR assays performed in this study.

smFISH

smFISH probes targeting MERVL were designed and synthesized by LGC Biosearch Technologies as previously described [47]. smFISH followed by immunofluorescence staining was performed according to the manufacturer's instructions. Briefly, Quasar 570 labeled probes were hybridized against MERVL RNA at 37°C for 16 h. Followed by immunofluorescence staining described in methods **Immunofluorescence** section, without blocking step to prevent RNase contamination. The slides were mounted with ProLong™ Glass

Antifade Mountant (catalog no. P36982, Thermo Fisher). Fluorescence images were taken using a confocal laser scanning microscope (catalog no. FV3000, Olympus).

Flow cytometry

The cells were digested with 0.25% trypsin-EDTA (catalog no. 25200072, Gibco), at 37°C for 5 min, and then re-suspended in FluoroBrite™ DMEM (catalog no. A1896701, Gibco). The cell suspension was filtered through a 35 μM cell strainer (catalog no. 352235, Falcon). The samples were analyzed using a cell sorter (catalog no. SH800Z, Sony). Data analysis was performed using the Sony SH800Z cell sorter software.

Generation of transgenic mESC lines

The 2C::tdTomato reporter cell line was generated by transfection of EB3 mESCs (catalog no. AES0139, RIKEN BRC Cell Bank) with a linearized 2C::tdTomato reporter plasmid (catalog no. 40281, Addgene). After 48 h, the cells were subjected to 500 $\mu g.mL^{-1}$ hygromycin (catalog no. 08906151, Wako) for 7 d. The selected cells were then seeded at a density of 2×10^2 cm^{-2} in culture medium containing 250 $\mu g.mL^{-1}$ hygromycin. Single-cell clones were picked and expanded after 7 d.

A tetracycycline-controlled transcription activation system and a piggyBac transposon system were used as previously described [48], to generate an *Obox4*-inducible cell line. Briefly, pPB-TRE-3×FLAG-Obox4, pPB-CAG-rtTA3G-IRES-Hygro, and pCMV-HyPBase-PGK-Puro plasmids were co-transfected into EB3 mESCs. After 48 h, the cells were subjected to selection with 500 $\mu g.mL^{-1}$ hygromycin and 500 $\mu g.mL^{-1}$ G418 (catalog no. 07805961, Wako), for 7 d. The selected cells were then seeded at a density of 2×10^2 cm^{-2} in a culture medium containing 250 $\mu g.mL^{-1}$ hygromycin and 250 $\mu g.mL^{-1}$ G418. Single-cell clones were picked and expanded after 7 d.

Generation of monoclonal antibodies

The anti-OBOX4, anti-DUX, and anti-GAG monoclonal antibodies were produced as previously described [49] (Supplementary Fig. 11a). Briefly, BALB/c mice were immunized by means of intraperitoneal injection of glutathione S-transferase-fused protein-of-interest (PoI). The mice were routinely immunized until their sera tested positive for the PoI-fused maltose binding protein (MPB), after which two boosting immunizations were performed. Following boosting, splenocytes from immunized mice were fused with SP2/O myeloma and cultured in hypoxanthine-aminopterin-thymidine medium for 10 d, to select for hybridomas. Hybridomas were subsequently screened using enzyme-linked immunosorbent assay against PoI-fused MBP. The positive polyclonal hybridomas were monoclonalized and expanded. For anti-OBOX4 hybridomas, additional screening using OBOX2-fused MBP was conducted. The sequence of OBOX4 was significantly divergent from those of other OBOX members, and the monoclonal antibodies were negative for cross-reactivity with OBOX2 (Supplementary Fig. 11b-c). The N-terminal 100 amino acids of OBOX4 and OBOX2 were used for immunization and cross-reactivity examination respectively (Supplementary Fig. 11d). The culture supernatants of monoclonal hybridomas were sterilized by passing them through a 0.22 μM pore size filter (catalog no. 430769, Corning) and used directly as an antibody solution in other assays.

719

720 **Generation of knockout mouse lines**

721 Single guide RNAs (sgRNAs) designed on flanking and inside of the *Obox4* cluster were synthesized using
722 a Precision gRNA Synthesis Kit (catalog no. A29377, Thermo Fisher) according to the manufacturer's
723 instructions. 150 ng.mL^{-1} sgRNA and 125 ng/μL Alt-R S.p. Cas9 Nuclease (catalog no. 1081058, Integrated DNA
724 Technologies) diluted with Hepes-buffered KSOM were electroporated into BDF1 × B6 *in vitro* fertilized
725 embryos with an electroporator (catalog no. NEPA21, NEPA GENE). Treated embryos were cultured in KSOM
726 for 18 h and embryos that reached the two-cell stage were transferred into the oviducts of pseudopregnant
727 ICR female mice on day 0.5. Pups were retrieved on day 19.5 and genotyped with specific PCR primers.

728

729 **Generation of knockout mESC lines**

730 Plasmids encoding Cas9 and single guide RNA (sgRNA) flanking and inside of the *Obox4* cluster were
731 constructed to perform knockout experiment. The *Obox4* knockout cell line was generated by means of
732 co-transfection of EB3 mESCs with the pX330-Puro-Obox4KO-3, pX330-Puro-Obox4KO-4,
733 pX330-Puro-Obox4KO-6, and pX330-Puro-Obox4KO-7 plasmids. After 18 h, the cells were subjected to
734 selection with 1 μg.mL^{-1} puromycin (catalog no. 16023151, Wako) for 54 h. The selected cells were then
735 seeded at a very low density in culture medium without antibiotics. Single-cell clones were picked and
736 expanded after 7 d, and then subjected to genotyping.

737 Following a previously published protocol [18], plasmids encoding Cas9 and sgRNAs flanking the *Dux*
738 cluster were constructed to perform knockout experiment. The *Dux* knockout cell line and the *Obox4*/*Dux*
739 double knockout cell line were generated by means of co-transfection of EB3 mESCs and *Obox4* knockout cell
740 lines with pX330-DuxKO-A and pX330-DuxKO-B plasmids. After 48 h, the cells were seeded at a very low
741 density. Single-cell clones were picked and expanded after 7 d and then subjected to genotyping.

742 While mono-allelic *Obox4* deletion was detected in 49.1% (335/682) of the CRISPR-Cas9 edited clones,
743 bi-allelic deletion was only detected in 0.44% (3/682) of the population, suggesting that removal of *Obox4*
744 allele was associated with severe genetic toxicity. In contrast, 36.7% (47/128) mono-allelic and 4.7% (6/128)
745 bi-allelic deletions were detected in *Dux* knockout clones.

746

747 **Genotyping and copy number examination**

748 For mouse and culture cell, mouse right hindlimb toe or tail tip, or 1×10^6 culture cells were incubated in
749 400 μL ProK solution containing 20 mM Tris-HCl pH = 7.9, 1 mM EDTA, 1% w/v sodium dodecyl sulfate (SDS),
750 150 mM NaCl, 20 mM trisodium citrate, and 80 μg recombinant proteinase K (catalog no. 161-28701, Wako) at
751 55°C for 2 h with vigorous shaking. The solution was extracted with 2 volumes of phenol/chloroform/isoamyl
752 alcohol (25:24:1) (catalog no. 311-90151, Nippon Gene) twice. DNA in the aqueous phase was precipitated by
753 adding 0.1 volume of 3 M sodium acetate and 2 volumes of 99.5% ethanol (catalog no. 057-00456, Wako),
754 snap freeze in liquid nitrogen, and centrifuged at 15,000 × *g* for 12 min at 4°C. After discarding supernatant,
755 the DNA pellets were washed by 200 μL 70% ethanol, air dried at room temperature, and dissolved in 220 μL
756 TE solution (Tris-HCl pH = 7.9, 1 mM EDTA) supplemented with 1 μg RNase A (catalog no. 131-01461, Nippon

Gene). The solution was incubated at 37°C for 1 h, then phenol/chloroform/isoamyl alcohol extracted and ethanol precipitated again as described above, to obtain highly purified genomic DNA. The DNA was dissolved in 200 μL TE solution.

For single blastomeres and abnormal embryos, whole genome amplifications were performed on individual embryos by using REPLI-g Advanced DNA Single Cell Kit (catalog no. 150363, Qiagen) according to the manufacturer's instructions. For blastocysts, crude genomic DNA from single blastocysts was prepared according to the method described by Sakurai *et al.*[50] with some modifications. Briefly, single blastocysts in 0.5 μL KSOM medium were transferred to the bottom of 0.1 mL PCR tubes, followed by addition of 10 μL blastocyst lysis buffer containing of 120 μg/mL recombinant proteinase K, 100mM Tris-HCl pH = 7.9, 100mM KCl, 0.45% NP-40, and 10 μg/mL yeast tRNA (catalog no. AM7119, Thermo Fisher). After brief vortex and pulse spin, the tubes were incubated at 55°C for 10 min followed by 95°C for 10 min.

For PCR based genotyping, 10 ng purified genomic DNA, 10ng whole genome amplified DNA, or 2μL of blastocyst lysate were used as template and amplified with target specific primers by KOD One® PCR Master Mix (catalog no. KMM-101, TOYOBO) according to the manufacturer's instructions. Primers targeting mouse *Tardbp* were used as internal control when whole genome amplification product or blastocyst lysate were used as multiplex PCR template. For copy number examination, 5 ng genomic DNA template was amplified as described in methods **qPCR** section.

In vitro transcription

The *Obox4* coding sequence was codon-optimized using GeneArt Instant Designer (Thermo Fisher), to remove the ASO target motif and improve translation efficiency. Codon-optimized DNA was synthesized using Prime Gene Synthesis Services (Thermo Fisher). Codon-optimized *Obox4* mRNA was transcribed *in vitro* using the mMESSAGE mMACHINE™ T7 Transcription Kit (catalog no. AM1344, Thermo Fisher) and then polyadenylated using a poly(A) tailing kit (catalog no. AM1350, Thermo Fisher), according to the manufacturer's instructions. The polyadenylated mRNA was purified using the RNeasy Mini Kit (catalog no. 74004, Qiagen), according to the manufacturer's instructions, and dissolved in RNase-free water, at a concentration of 100 ng.μL^{-1}.

Pronuclear injection of mouse embryos

Eight-week-old BDF1 female mice were injected with 150 μL of CARD HyperOva (catalog no. KYD-010-EX, Kyudo Co., Ltd.). After 48 hours, the female mice were injected with 5 IU human chorionic gonadotropin (hCG) (catalog no. GONATROPIN, ASKA Animal Health Co. Ltd.) and mated with BDF1 male mice. After 22 h, the PN3 zygotes were collected. Cumulus cells were removed by briefly culturing the zygotes in potassium simplex optimization medium (KSOM) (catalog no. MR-101-D, Merck) supplemented with 0.3 μg.μL^{-1} hyaluronidase (catalog no. H4272, Merck). Embryos were cultured in KSOM medium drops covered with liquid paraffin (catalog no. 26137-85, Nacalai Tesque), in a humidified atmosphere with 5% CO_2, at 37°C. After 4 h, PBS containing either 40 μM scramble ASO, 20 μM anti-*Dux* ASO, 20 μM anti-*Obox4* ASO, 40 μM equimolar anti-*Dux*/anti-*Obox4* ASO mixture, or 40 μM equimolar anti-*Dux*/anti-*Obox4* ASO mixture with 100 ng.μL^{-1}

codon-optimized *Obox4* mRNA was microinjected into the male pronuclei of zygotes using a microinjector (catalog no. 5252000013, Eppendorf). For developmental monitoring, embryos were cultured for 4 d after microinjection and assessed for developmental stage at 18 h, 42 h, 66 h, and 90 h after microinjection, which corresponded to 1.5 dpc, 2.5 dpc, 3.5 dpc, and 4.5 dpc, respectively.

Somatic cell nuclear transfer (SCNT)

SCNT was performed as described previously [51] using wild-type and knockout mESC lines as nuclear donor cells. They were cultured at a high density until they reached confluency for about four to six days before SCNT. BDF1 female mice were superovulated by the injection of 7.5 IU of equine chorionic gonadotropin (eCG, ZENOAQ) and 7.5 IU of hCG (ASKA Pharmaceutical Co., Ltd.) at a 48 h interval. At 15 h after hCG injection, cumulus–oocyte complexes were collected from the oviducts. After removal of cumulus cells by 0.1% bovine testicular hyaluronidase (catlog no. 385931, Calbiochem), oocytes were enucleated in Hepes-buffered KSOM containing 7.5 μg/mL cytochalasin B. After culture in KSOM for at least 1 h, the enucleated oocytes were injected with donor mESCs using a Piezo-driven micromanipulator (catalog no. PMM-150FU, Primetech). After culture in KSOM for about 1 h, injected oocytes were cultured in Ca^{2+}-free KSOM containing 2.5 mM $SrCl_2$, 5 μM latrunculin A (LatA) (catalog no. L5163, Merck) with 50 nM trichostatin A (TSA) (catalog no. T8552, Merck) for 1 h. Then, they were cultured in KSOM containing 5 μM LatA and 50 nM 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. The embryos that reached 2-cells at 24 h after activation were considered successfully cloned with cell cycle-matched mESCs (G0/G1 phase).

scRNA-seq analysis

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

Ribo-seq analysis

Raw Ribo-seq data was obtained from the dataset of Xiong et al. (GSE165782). Reads were quality-controlled, aligned, and counted, as described above. Reads aligned to all protein coding *Obox4* loci were added up to represent translation level of OBOX4.

Bulk RNA-seq

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

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

The libraries were quantified with a 2100 Bioanalyzer (catalog no. G2939BA, Agilent) using a High Sensitivity DNA Kit (catalog no. 5067-4626, Agilent). Quantified libraries were pooled and sequenced using an Illumina NovaSeq 6000 System in 150 bp paired-end mode (Illumina). Base calling and de-multiplexing were performed using the bcl2fastq2 (Illumina) v2.20. De-multiplexed reads were quality-controlled, aligned, and counted, as described above. DESeq2 [58] v1.32.0 was used to perform differential expression analysis. Raw RNA-seq reads for TBLCs were downloaded from the dataset of Shen et al. (GSE168728). Raw RNA-seq reads of *Dux* knockout embryos were downloaded from the datasets of Chen & Zhang and De Iaco et al. (GSE121746 and GSE141321 respectively). Raw RNA-seq reads of preimplantation embryos were downloaded from the datasets of Wu et al. (GSE66390). Raw RNA-seq reads of 2C-like mESCs were downloaded from the datasets of Zhu et al. (GSE159623). The published data were analyzed using the same method.

CUT&RUN-seq

Cultured cells (1×10^5) freshly prepared by trypsin-EDTA digestion were used to perform CUT&RUN with the CUT&RUN Assay Kit (catalog no. 86652, Cell Signaling Technology), according to the manufacturer's instructions. Enriched DNA was subjected to library construction using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (catalog no. E7645L, NEB), with the adapter ligation step performed at 50°C instead of 65°C, to prevent the denaturation of small DNA inserts. The libraries were indexed using NEBNext® Multiplex 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 --normalizeUsing CPM --smoothLength 30), and then normalized by subtracting the signal from non-immune IgG and wildtype mESCs using the deepTools bamCompare function (--scaleFactorsMethod 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. Heatmaps were generated using deepTools computeMatrix and plotHeatmap functions. Visualization of genomic tracks was performed using trackplot [64]

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. For single hypothesis testing, significance is defined as a *p-value* less than 0.05 indicated with asterisk (* $p\text{-value} < 0.05$, ** $p\text{-value} < 0.01$, *** $p\text{-value} < 0.001$). Error bar represents the standard deviation (SD) of the mean of the replicates. Significant change in gene expression is denoted by greater than 1-fold difference and less than 0.01 false discovery rate.

Materials and Data availability

The materials underlying this article will be shared upon reasonable request to the corresponding author. 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), Katsuhiko Hayashi (Department of Genome Biology, Osaka University) and Therese Solberg (WPI-Bio2Q, Keio 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. and 19H05758 to A.O.), 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. and 22H02534 to K.I.), 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.

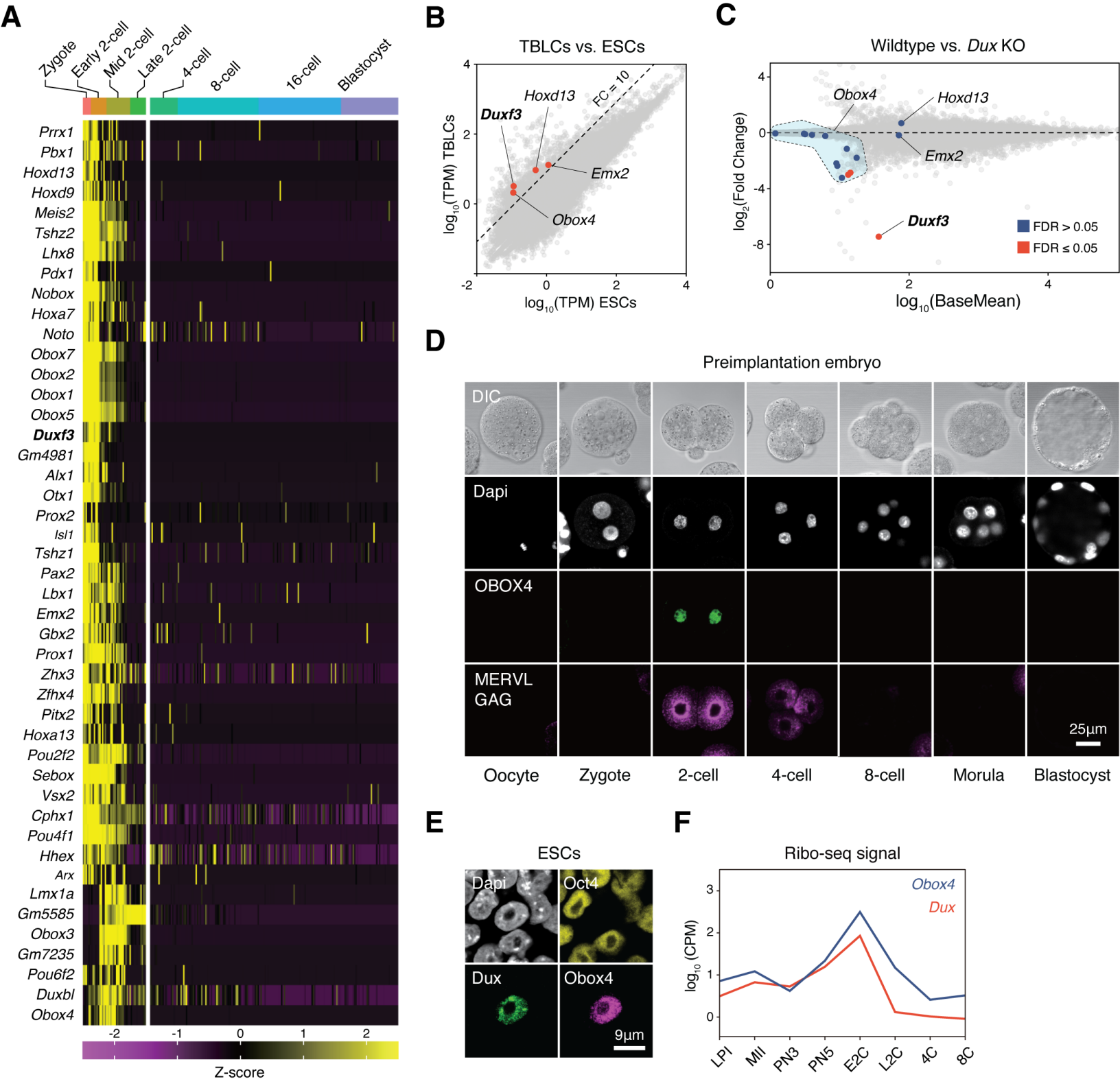
References

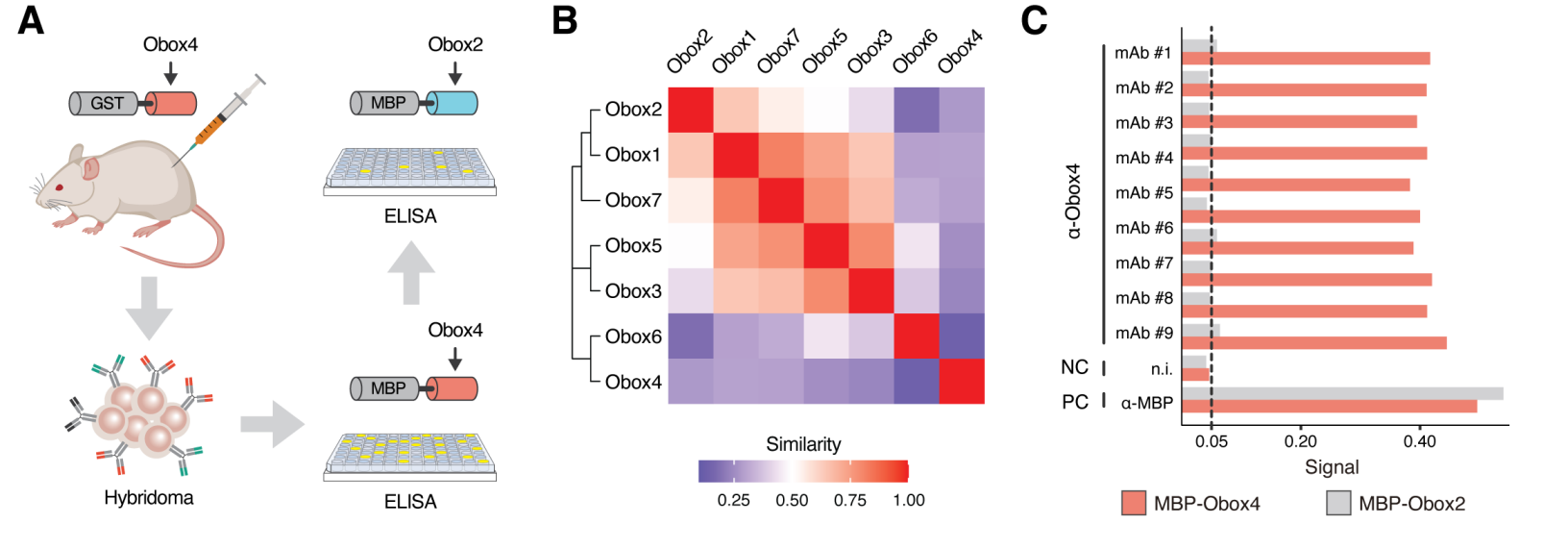
1. Lee, M.T., Bonneau, A.R. & Giraldez, A.J. Zygotic genome activation during the maternal-to-zygotic transition. *Annual review of cell and developmental biology* **30**, 581-613 (2014).
2. Tarkowski, A.K. Experiments on the development of isolated blastomeres of mouse eggs. *Nature* **184**, 1286-1287 (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).

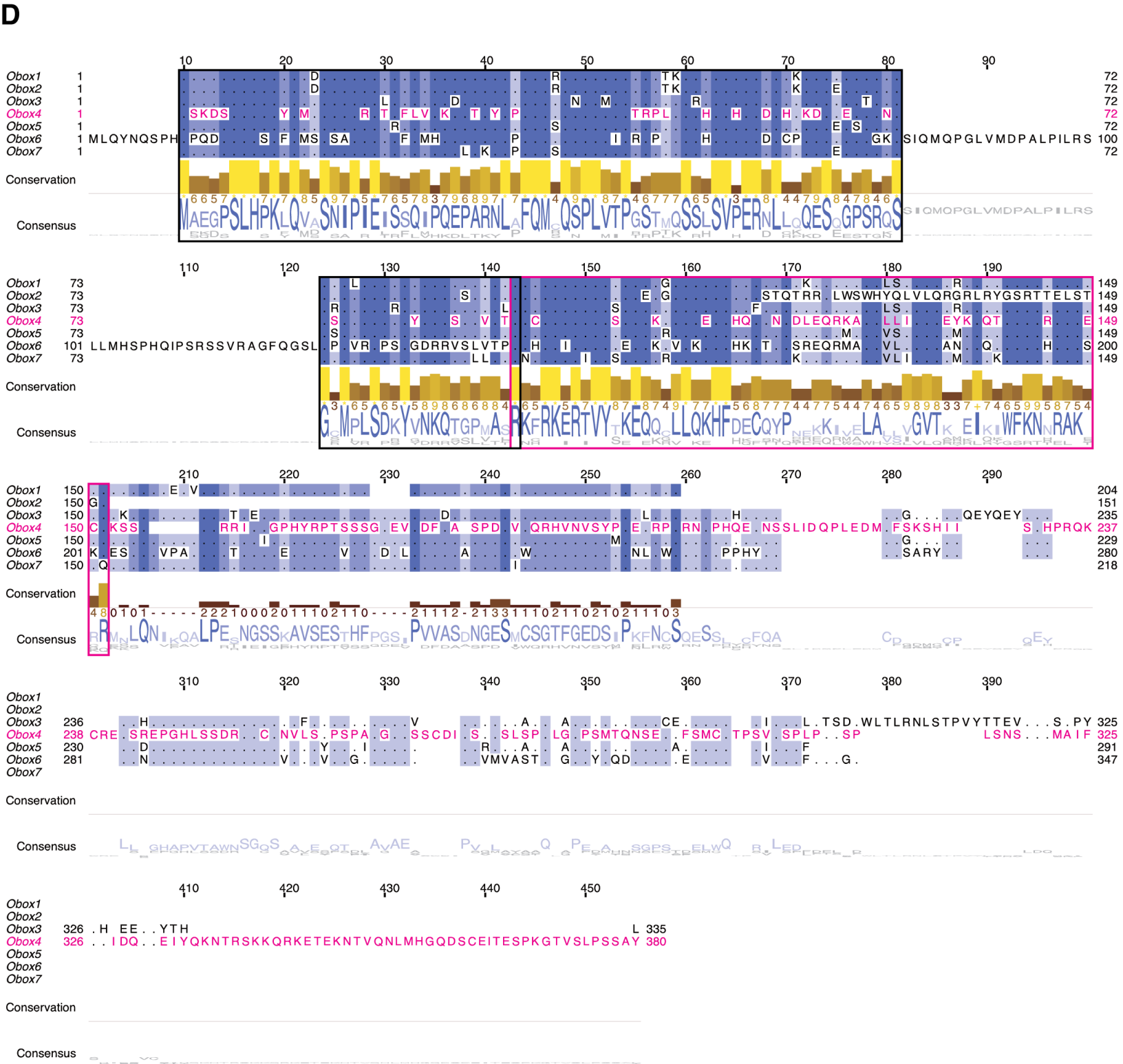
4. Suwińska, A., Czołowska, R., Ożdżeński, W. & Tarkowski, A.K. Blastomeres of the mouse embryo lose totipotency after the fifth cleavage division: expression of Cdx2 and Oct4 and developmental potential of inner and outer blastomeres of 16-and 32-cell embryos. *Developmental biology* **322**, 133-144 (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. Shen, H. *et al.* Mouse totipotent stem cells captured and maintained through spliceosomal repression. *Cell* **184**, 2843-2859. e20 (2021).
23. Wagner, A. Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. *Biological cybernetics* **74**, 557-567 (1996).

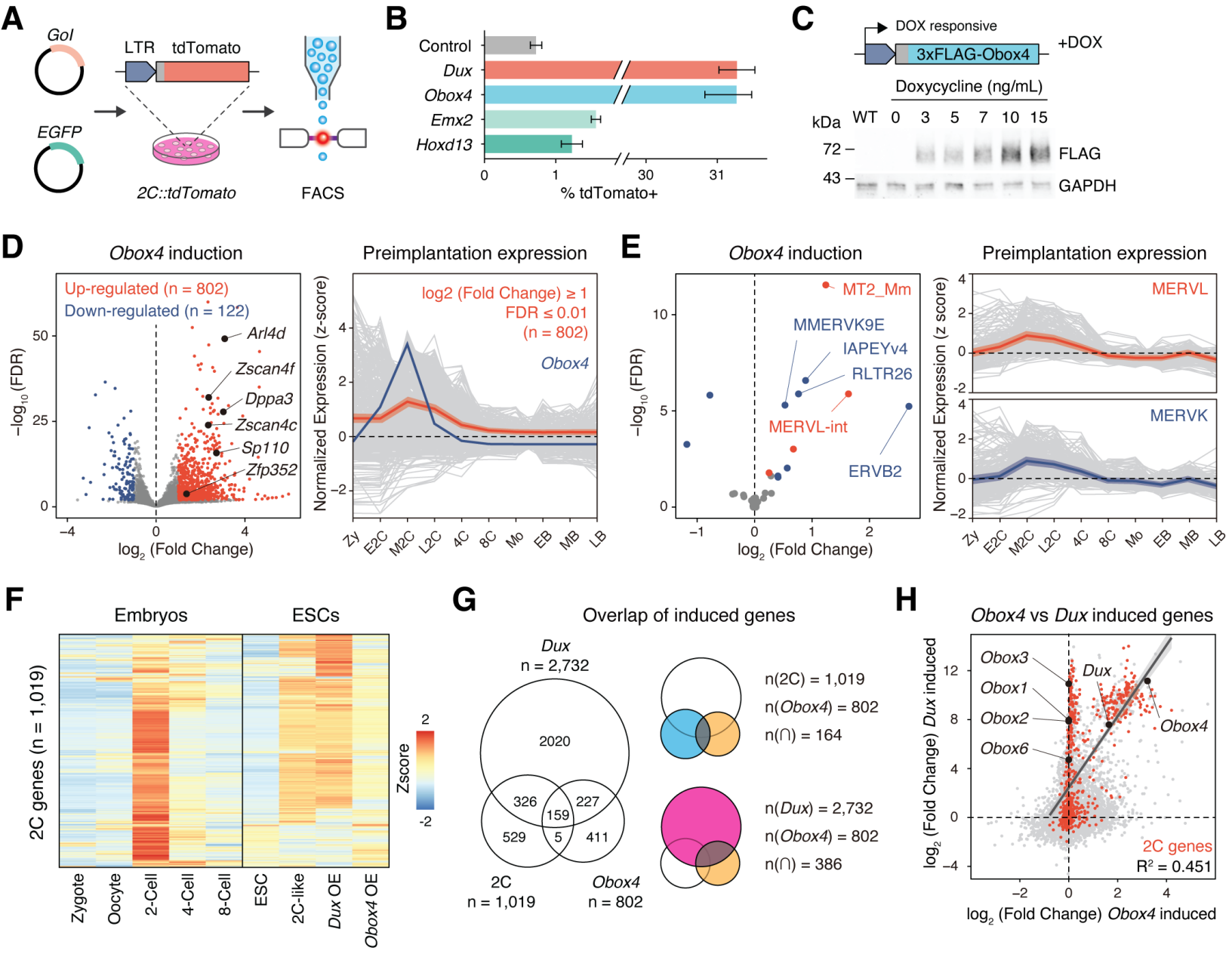
24. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193-196 (2014).
25. Howe, K.L. *et al.* Ensembl 2021. *Nucleic Acids Research* **49**, D884-D891 (2020).
26. Xiong, Z. *et al.* Ultrasensitive Ribo-seq reveals translational landscapes during mammalian oocyte-to-embryo transition and pre-implantation development. *Nature Cell Biology* **24**, 968-980 (2022).
27. Macfarlan, T.S. *et al.* Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes & development* **25**, 594-607 (2011).
28. Macfarlan, T.S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57-63 (2012).
29. Wu, J. *et al.* The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652-657 (2016).
30. 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).
31. Skene, P.J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, e21856 (2017).
32. 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).
33. Markoulaki, S., Meissner, A. & Jaenisch, R. Somatic cell nuclear transfer and derivation of embryonic stem cells in the mouse. *Methods* **45**, 101-114 (2008).
34. Ji, S. *et al.* OBOX regulates mouse zygotic genome activation and early development. *Nature* **620**, 1047-1053 (2023).
35. Rajkovic, A., Yan, C., Yan, W., Klysik, M. & Matzuk, M.M. Obox, a family of homeobox genes preferentially expressed in germ cells. *Genomics* **79**, 711-717 (2002).
36. Ge, S.X. Exploratory bioinformatics investigation reveals importance of "junk" DNA in early embryo development. *BMC genomics* **18**, 1-19 (2017).
37. Zhong, Y.-f. & Holland, P.W. The dynamics of vertebrate homeobox gene evolution: gain and loss of genes in mouse and human lineages. *BMC evolutionary biology* **11**, 1-13 (2011).
38. Wilming, L.G., Boychenko, V. & Harrow, J.L. Comprehensive comparative homeobox gene annotation in human and mouse. *Database* **2015**(2015).
39. Tadros, W. & Lipshitz, H.D. The maternal-to-zygotic transition: a play in two acts. *Development* **136**, 3033-3042 (2009).
40. Svoboda, P. Mammalian zygotic genome activation. in *Seminars in cell & developmental biology* Vol. 84 118-126 (Elsevier, 2018).
41. Leidenroth, A. & Hewitt, J.E. A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. *BMC evolutionary biology* **10**, 1-13 (2010).
42. Maeso, I. *et al.* Evolutionary origin and functional divergence of totipotent cell homeobox genes in eutherian mammals. *BMC biology* **14**, 1-14 (2016).
43. Zou, Z. *et al.* Translatome and transcriptome co-profiling reveals a role of TPRXs in human zygotic genome activation. *Science*, eabo7923 (2022).
44. Guo, Y., Li, T.D., Modzelewski, A.J. & Siomi, H. Retrotransposon renaissance in early embryos. *Trends in Genetics* **40**, 39-51 (2024).
45. Posfai, E. *et al.* Evaluating totipotency using criteria of increasing stringency. *Nature cell biology* **23**, 49-60 (2021).

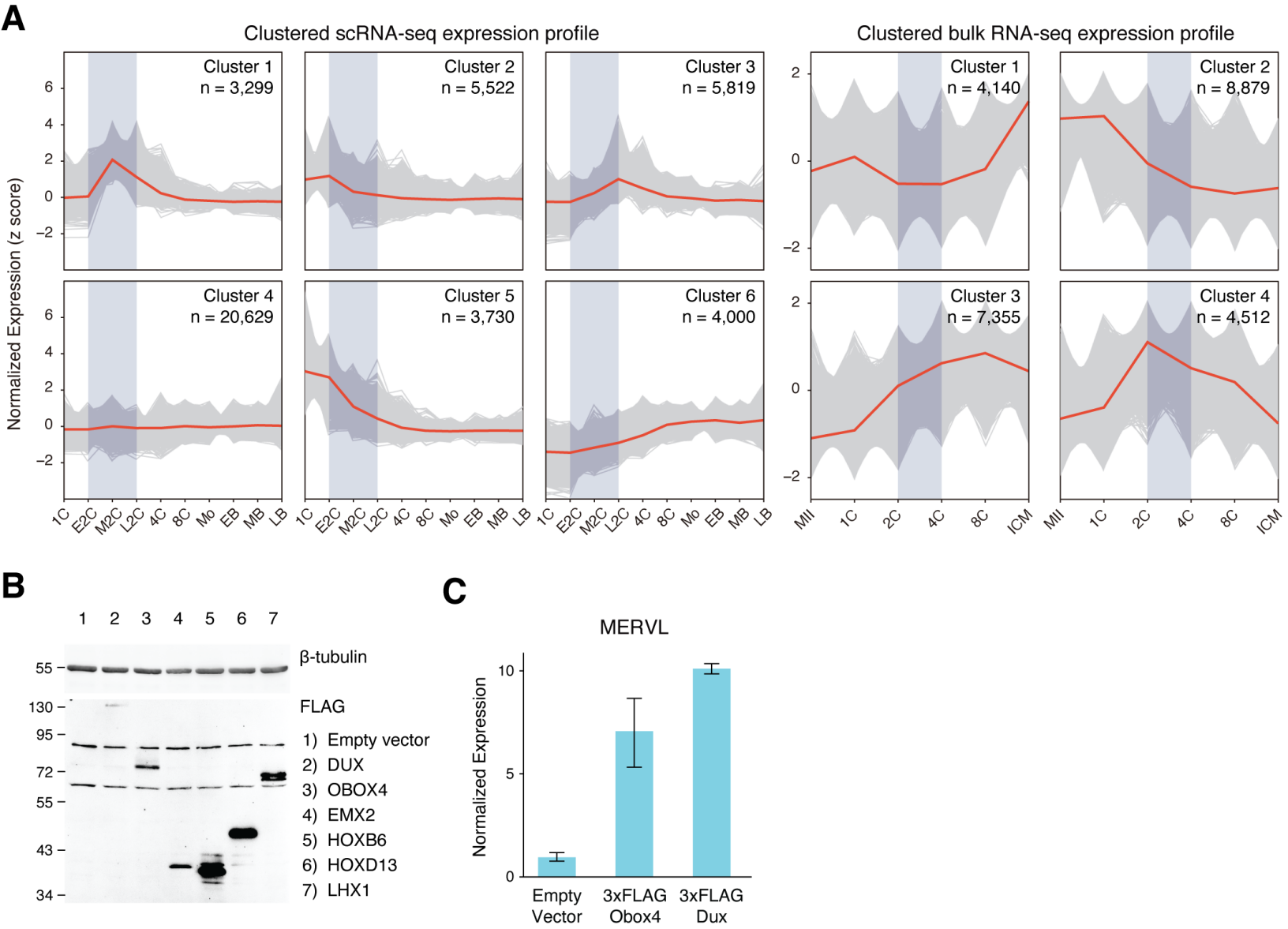
46. Hu, Y. *et al.* Induction of mouse totipotent stem cells by a defined chemical cocktail. *Nature* **617**, 792-797 (2023).
47. Sakashita, A. *et al.* Transcription of MERVL retrotransposons is required for preimplantation embryo development. *Nature genetics* **55**, 484-495 (2023).
48. Takeuchi, C., Murano, K., Ishikawa, M., Okano, H. & Iwasaki, Y.W. Generation of Stable Drosophila Ovarian Somatic Cell Lines Using the piggyBac System. in *piRNA: Methods and Protocols* 143-153 (Springer, 2022).
49. Guo, Y. *et al.* Potent mouse monoclonal antibodies that block SARS-CoV-2 infection. *Journal of Biological Chemistry* **296**(2021).
50. Sakurai, T., Watanabe, S., Kamiyoshi, A., Sato, M. & Shindo, T. A single blastocyst assay optimized for detecting CRISPR/Cas9 system-induced indel mutations in mice. *BMC biotechnology* **14**, 1-11 (2014).
51. Inoue, K. *et al.* Loss of H3K27me3 imprinting in the Sfmbt2 miRNA cluster causes enlargement of cloned mouse placentas. *Nature communications* **11**, 1-12 (2020).
52. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018).
53. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
54. Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Research* **49**, D916-D923 (2020).
55. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
56. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* (2021).
57. Zhong, Y.f. & Holland, P.W. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evolution & development* **13**, 567-568 (2011).
58. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 1-21 (2014).
59. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359 (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. Pohl, A. & Beato, M. bwtool: a tool for bigWig files. *Bioinformatics* **30**, 1618-1619 (2014).

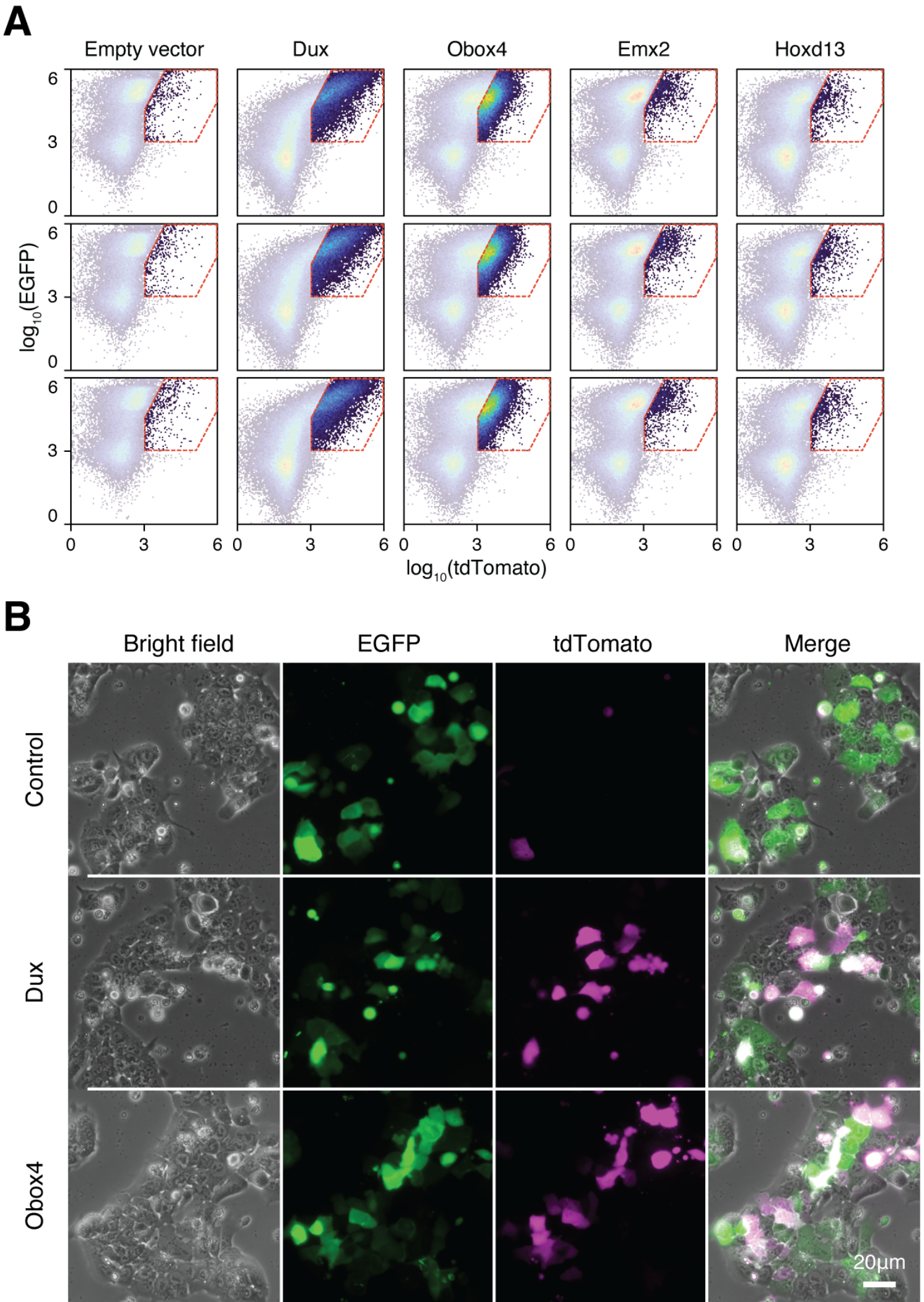

A
Zygote
Early 2-cell
Mid 2-cell
Late 2-cell
4-cell
8-cell
16-cell
Blastocyst
Prrx1
Pbx1
Hoxd13
Hoxd9
Meis2
Tshz2
Lhx8
Pdx1
Nobox
Hoxa7
Noto
Obox7
Obox2
Obox1
Obox5
Duxf3
Gm4981
Alx1
Otx1
Prox2
Isl1
Tshz1
Pax2
Lbx1
Emx2
Gbx2
Prox1
Zhx3
Zfhx4
Pitx2
Hoxa13
Pou2f2
Sebox
Vsx2
Cphx1
Pou4f1
Hhex
Arx
Lmx1a
Gm5585
Obox3
Gm7235
Pou6f2
Duxbl
Obox4
-2
-1
0
1
2
Z-score
B
TBLCs vs. ESCs
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(BaseMean)
D
Preimplantation embryo
DIC
Dapi
OBOX4
MERVL GAG
25μm
Oocyte
Zygote
2-cell
4-cell
8-cell
Morula
Blastocyst
E
ESCs
Dapi
Oct4
Dux
Obox4
9μm
F
Ribo-seq signal
Obox4
Dux
log10 (CPM)
LPI
MII
PN3
PN5
E2C
L2C
4C
8C

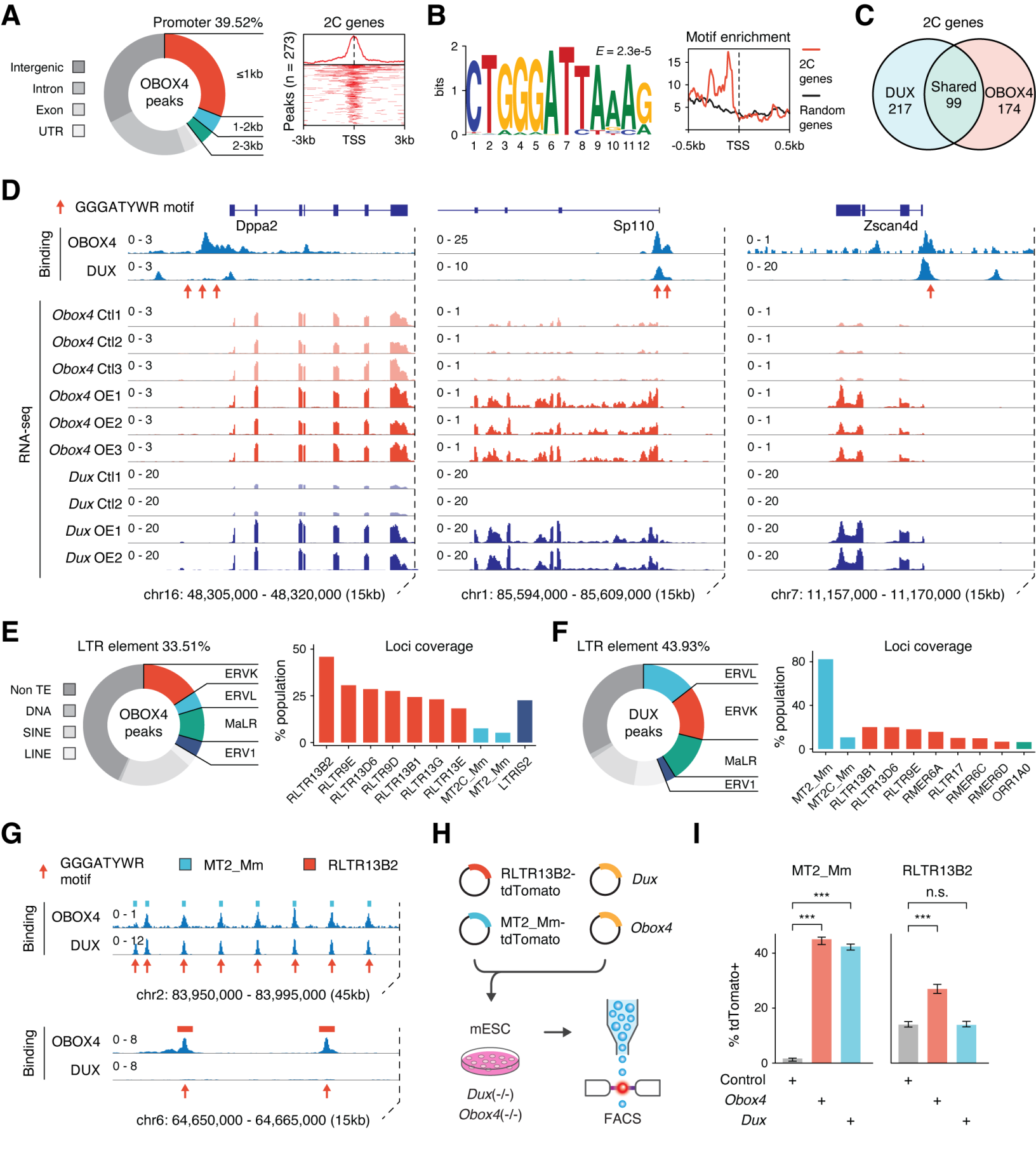

A
Obox4
GST
Obox2
MBP
ELISA
Hybridoma
Obox4
MBP
ELISA
B
Obox2
Obox1
Obox7
Obox5
Obox3
Obox6
Obox4
Similarity
0.25
0.50
0.75
1.00
C
α-Obox4
mAb #1
mAb #2
mAb #3
mAb #4
mAb #5
mAb #6
mAb #7
mAb #8
mAb #9
NC
n.i.
PC
α-MBP
0.05
0.20
0.40
Signal
MBP-Obox4
MBP-Obox2

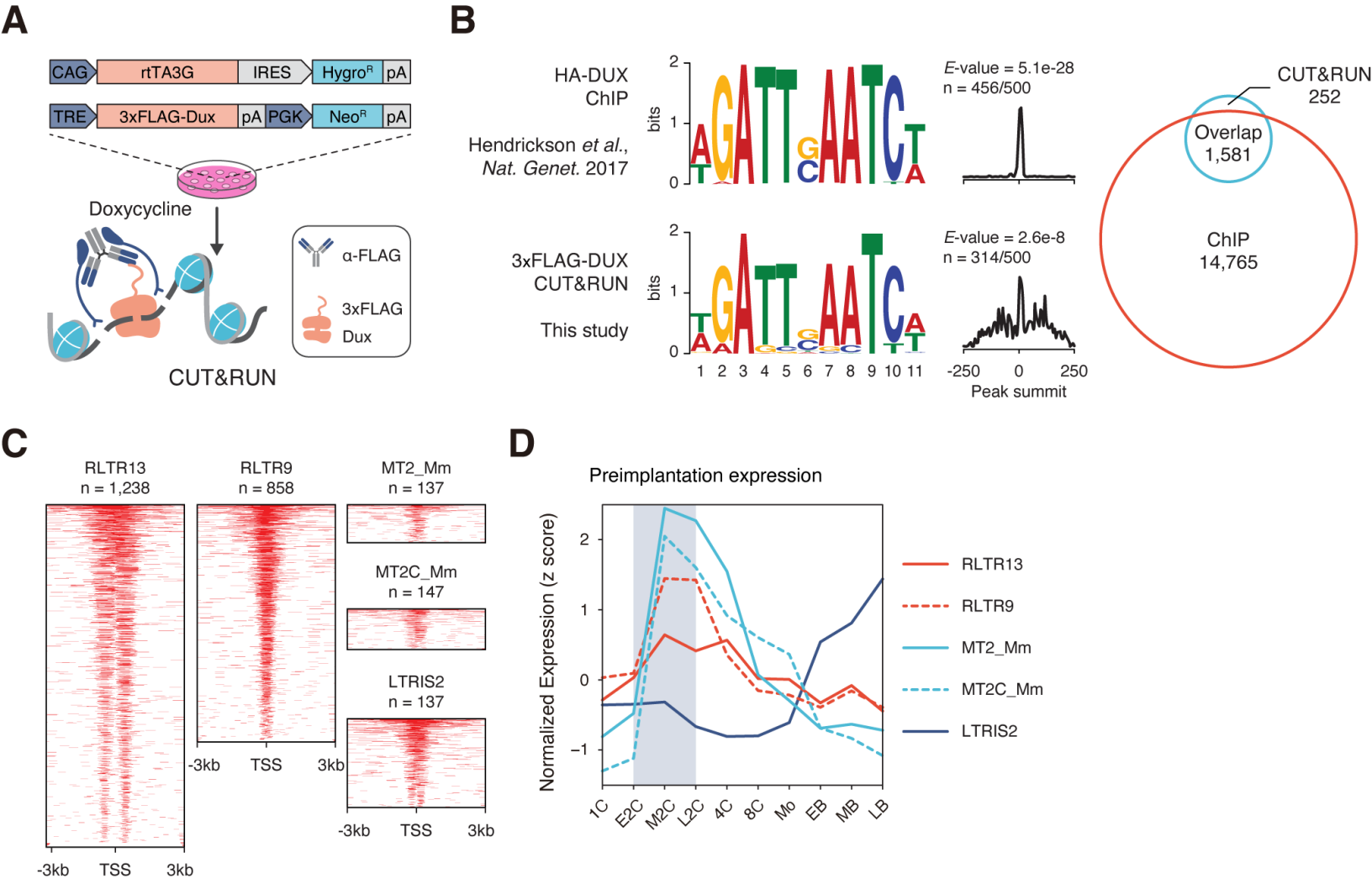

D
Obox1
Obox2
Obox3
Obox4
Obox5
Obox6
Obox7
Conservation
Consensus

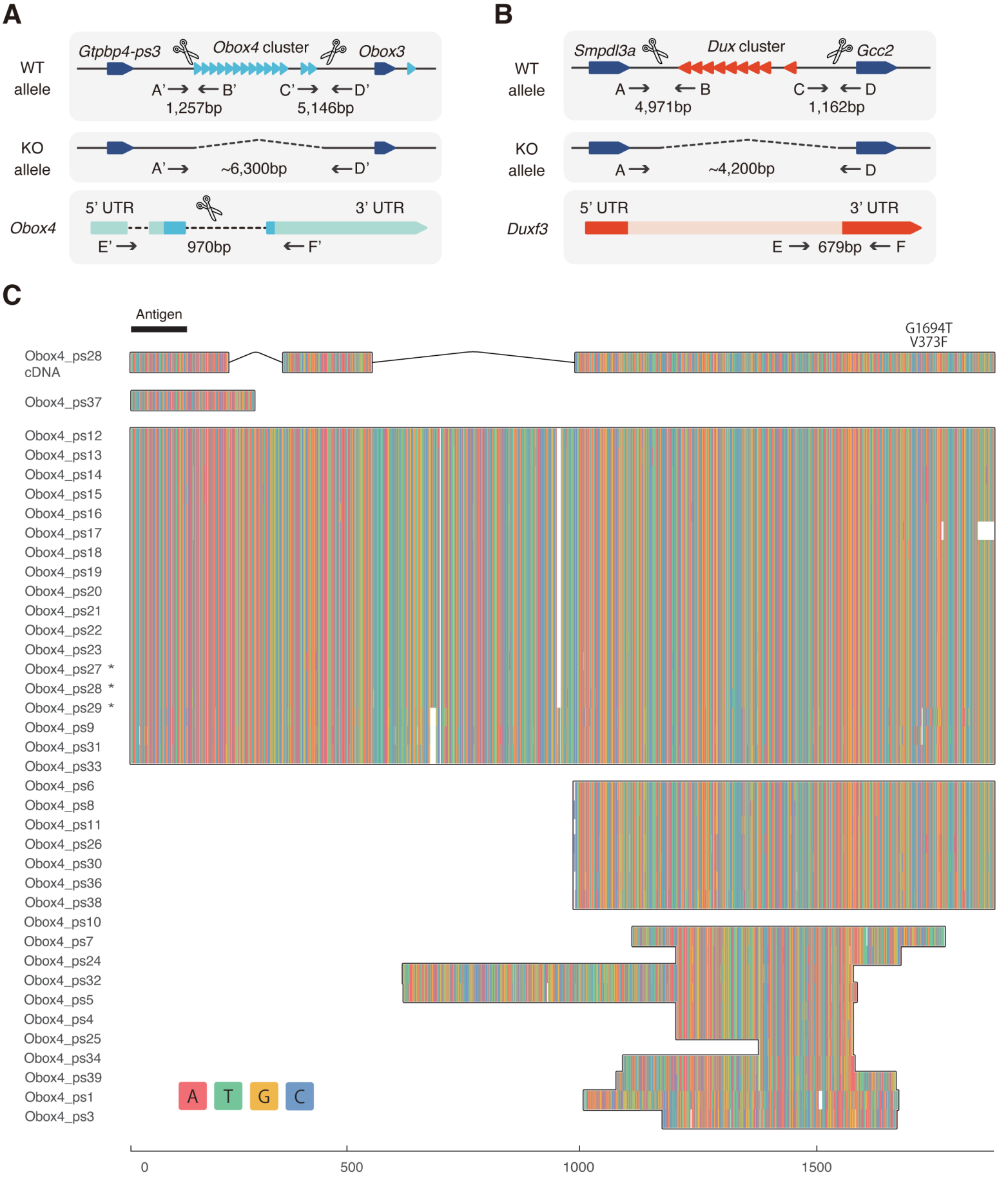

A
GoI
EGFP
LTR
tdTomato
2C::tdTomato
FACS
B
Control
Dux
Obox4
Emx2
Hoxd13
0
1
30
31
% tdTomato+
C
DOX responsive
3xFLAG-Obox4
+DOX
Doxycycline (ng/mL)
kDa
WT
0
3
5
7
10
15
72 –
43 –
FLAG
GAPDH
D
Obox4 induction
Up-regulated (n = 802)
Down-regulated (n = 122)
Arl4d
Zscan4f
Dppa3
Zscan4c
Sp110
Zfp352
−log10 (FDR)
log2 (Fold Change)
Preimplantation expression
log2 (Fold Change) ≥ 1
FDR ≤ 0.01
(n = 802)
Obox4
Normalized Expression (z-score)
Zy
E2C
M2C
L2C
4C
8C
Mo
EB
MB
LB
E
Obox4 induction
MT2_Mm
MMERVK9E
IAPEYv4
RLTR26
MERVL-int
ERVB2
−log10 (FDR)
log2 (Fold Change)
Preimplantation expression
MERVL
MERVK
Normalized Expression (z score)
F
Embryos
ESCs
2C genes (n = 1,019)
Zscore
Zygote
Oocyte
2-Cell
4-Cell
8-Cell
ESC
2C-like
Dux OE
Obox4 OE
G
Overlap of induced genes
Dux
n = 2,732
2020
326
159
227
529
5
411
2C
n = 1,019
Obox4
n = 802
n(2C) = 1,019
n(Obox4) = 802
n(∩) = 164
n(Dux) = 2,732
n(Obox4) = 802
n(∩) = 386
H
Obox4 vs Dux induced genes
Obox3
Obox1
Obox2
Obox6
Dux
Obox4
2C genes
R² = 0.451
log2 (Fold Change) Dux induced
log2 (Fold Change) Obox4 induced

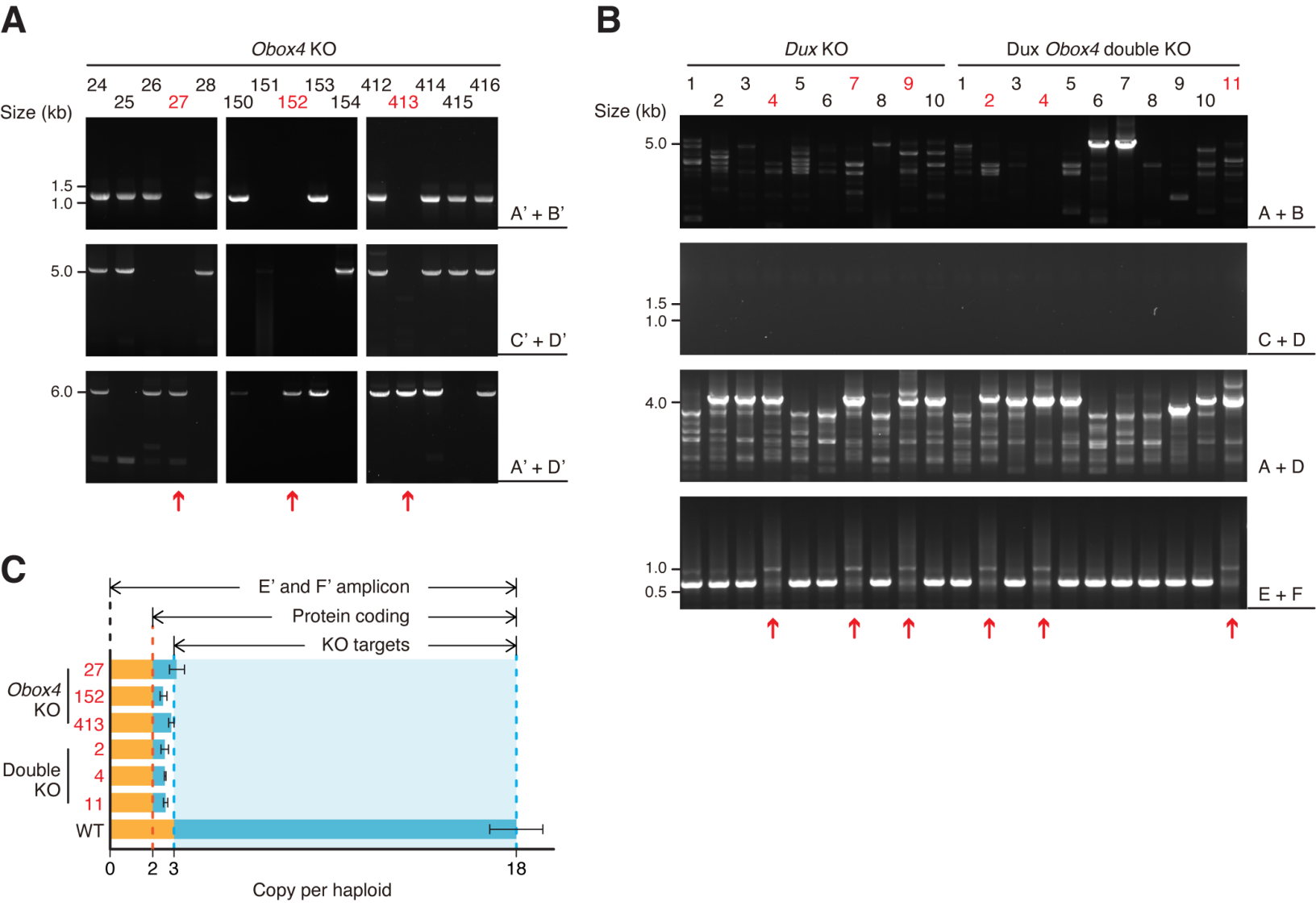

A
Clustered scRNA-seq expression profile
Clustered bulk RNA-seq expression profile
Normalized Expression (z score)
Cluster 1
n = 3,299
Cluster 2
n = 5,522
Cluster 3
n = 5,819
Cluster 4
n = 20,629
Cluster 5
n = 3,730
Cluster 6
n = 4,000
Cluster 1
n = 4,140
Cluster 2
n = 8,879
Cluster 3
n = 7,355
Cluster 4
n = 4,512
1C E2C M2C L2C 4C 8C Mo EB MB LB
MII 1C 2C 4C 8C ICM
B
1 2 3 4 5 6 7
β-tubulin
FLAG
1) Empty vector
2) DUX
3) OBOX4
4) EMX2
5) HOXB6
6) HOXD13
7) LHX1
55
130
95
72
55
43
34
C
MERVL
Normalized Expression
Empty Vector
3xFLAG Obox4
3xFLAG Dux

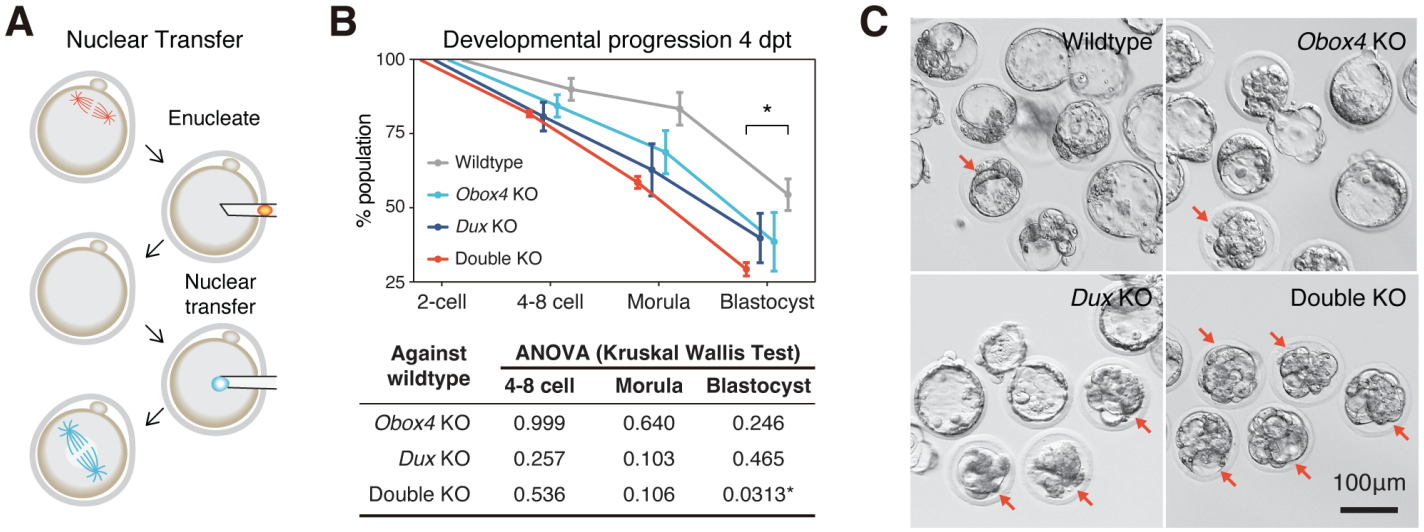

A
Empty vector
Dux
Obox4
Emx2
Hoxd13
log10(EGFP)
log10(tdTomato)
0
3
6
B
Bright field
EGFP
tdTomato
Merge
Control
Dux
Obox4
20μm


A
Promoter 39.52%
Intergenic
Intron
Exon
UTR
OBOX4 peaks
≤1kb
1-2kb
2-3kb
2C genes
Peaks (n = 273)
-3kb
TSS
3kb
B
E = 2.3e-5
bits
Motif enrichment
2C genes
Random genes
-0.5kb
TSS
0.5kb
C
2C genes
DUX 217
Shared 99
OBOX4 174
D
GGGATYWR motif
Binding
OBOX4
DUX
Dppa2
Sp110
Zscan4d
RNA-seq
Obox4 Ctl1
Obox4 Ctl2
Obox4 Ctl3
Obox4 OE1
Obox4 OE2
Obox4 OE3
Dux Ctl1
Dux Ctl2
Dux OE1
Dux OE2
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
RLTR13D6
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
RLTR13D6
RLTR9E
RMER6A
RLTR17
RMER6C
RMER6D
ORR1A0
G
GGGATYWR motif
MT2_Mm
RLTR13B2
Binding
OBOX4
DUX
chr2: 83,950,000 - 83,995,000 (45kb)
chr6: 64,650,000 - 64,665,000 (15kb)
H
RLTR13B2-tdTomato
MT2_Mm-tdTomato
Dux
Obox4
mESC
Dux(-/-)
Obox4(-/-)
FACS
I
MT2_Mm
RLTR13B2
n.s.
% tdTomato+
Control
Obox4
Dux

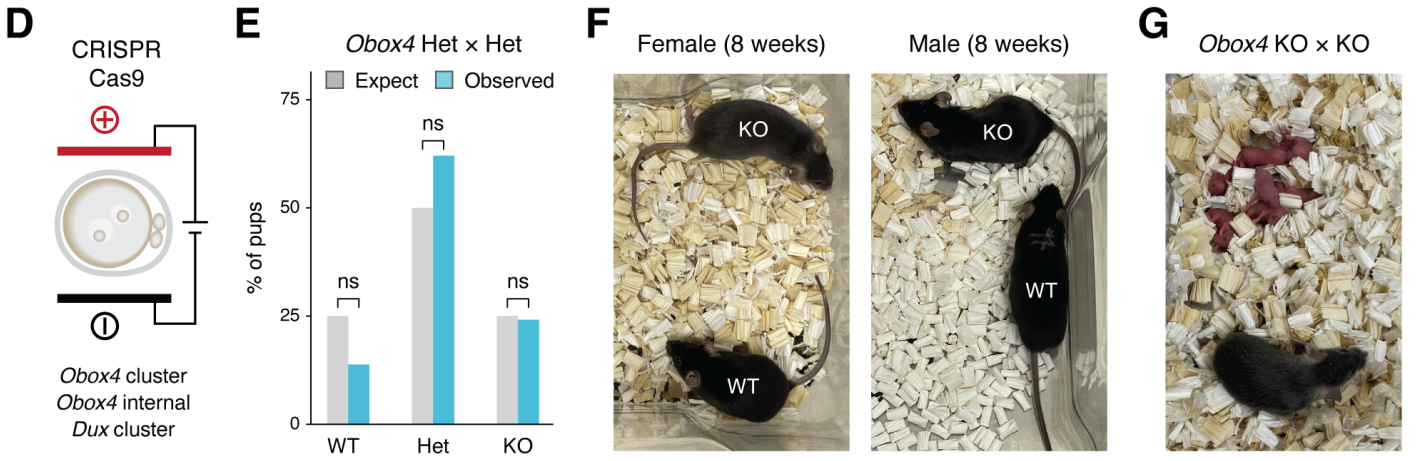

A
CAG
rtTA3G
IRES
Hygro^R
pA
TRE
3xFLAG-Dux
pA
PGK
Neo^R
pA
Doxycycline
α-FLAG
3xFLAG
Dux
CUT&RUN
B
HA-DUX
ChIP
Hendrickson et al.,
Nat. Genet. 2017
bits
E-value = 5.1e-28
n = 456/500
3xFLAG-DUX
CUT&RUN
This study
E-value = 2.6e-8
n = 314/500
1 2 3 4 5 6 7 8 9 10 11
-250 0 250
Peak summit
CUT&RUN
252
Overlap
1,581
ChIP
14,765
C
RLTR13
n = 1,238
RLTR9
n = 858
MT2_Mm
n = 137
MT2C_Mm
n = 147
LTRIS2
n = 137
-3kb TSS 3kb
D
Preimplantation expression
Normalized Expression (z score)
1C E2C M2C L2C 4C 8C Mo EB MB LB
RLTR13
RLTR9
MT2_Mm
MT2C_Mm
LTRIS2

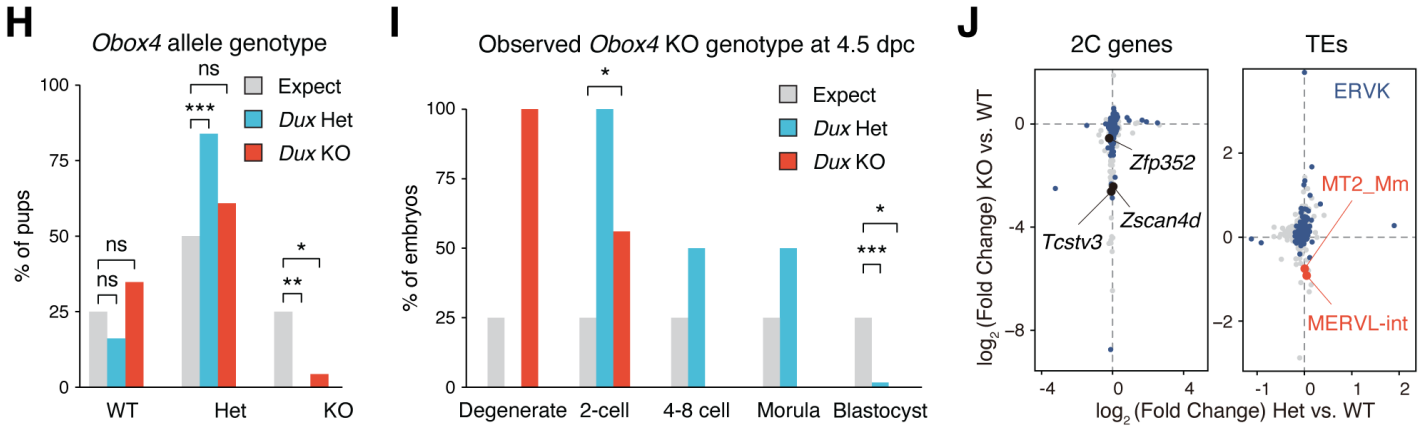

A
WT allele
Gtpbp4-ps3
Obox4 cluster
Obox3
A'→ ←B'
1,257bp
C'→ ←D'
5,146bp
KO allele
A'→
~6,300bp
←D'
Obox4
5' UTR
3' UTR
E'→
970bp
←F'
B
WT allele
Smpdl3a
Dux cluster
Gcc2
A→ ←B
4,971bp
C→ ←D
1,162bp
KO allele
A→
~4,200bp
←D
Duxf3
5' UTR
3' UTR
E→ 679bp ←F
C
Antigen
G1694T
V373F
Obox4_ps28 cDNA
Obox4_ps37
Obox4_ps12
Obox4_ps13
Obox4_ps14
Obox4_ps15
Obox4_ps16
Obox4_ps17
Obox4_ps18
Obox4_ps19
Obox4_ps20
Obox4_ps21
Obox4_ps22
Obox4_ps23
Obox4_ps27 *
Obox4_ps28 *
Obox4_ps29 *
Obox4_ps9
Obox4_ps31
Obox4_ps33
Obox4_ps6
Obox4_ps8
Obox4_ps11
Obox4_ps26
Obox4_ps30
Obox4_ps36
Obox4_ps38
Obox4_ps10
Obox4_ps7
Obox4_ps24
Obox4_ps32
Obox4_ps5
Obox4_ps4
Obox4_ps25
Obox4_ps34
Obox4_ps39
Obox4_ps1
Obox4_ps3
A T G C
0
500
1000
1500
D
WT allele
Gtpbp4-ps3
Obox4 cluster
Obox3
KO allele
Protein-coding
Non-protein-coding

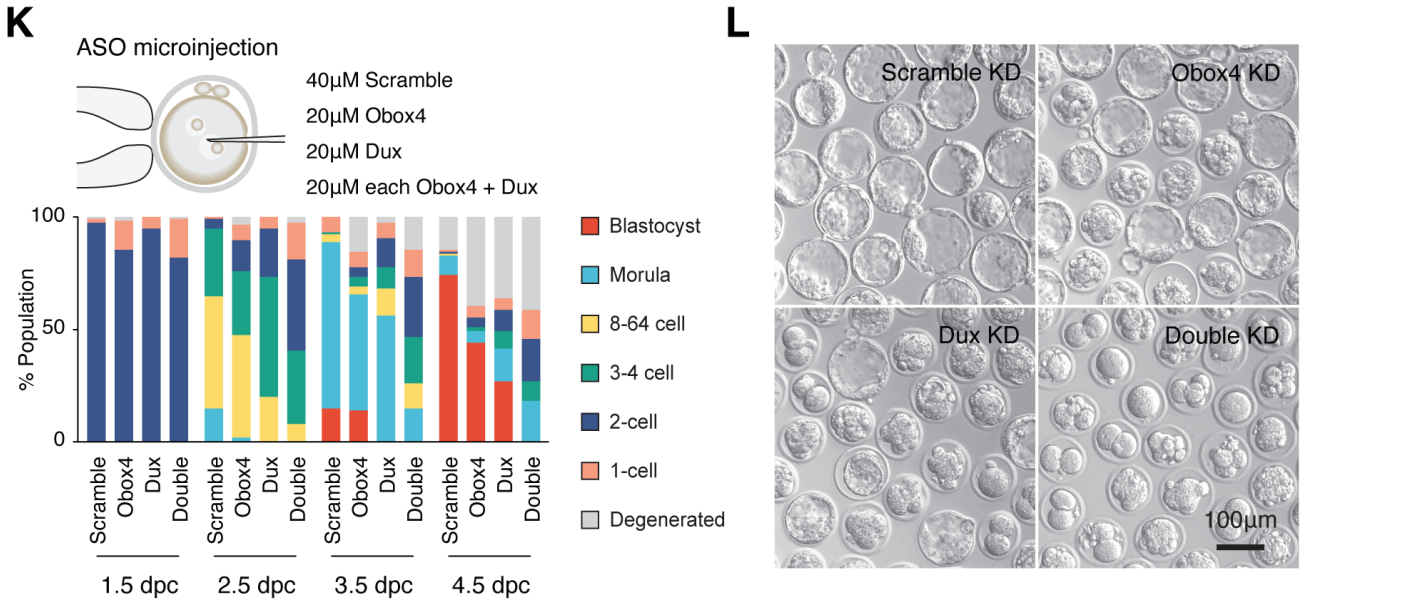

A
Obox4 KO
24 25 26 27 28
151 150 152 153 154
412 413 414 415 416
Size (kb)
1.5
1.0
A' + B'
5.0
C' + D'
6.0
A' + D'
B
Dux KO
Dux Obox4 double KO
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10 11
Size (kb)
5.0
A + B
1.5
1.0
C + D
4.0
A + D
1.0
0.5
E + F
C
E' and F' amplicon
Protein coding
KO targets
Obox4 KO
27
152
413
Double KO
2
4
11
WT
0
2
3
18
Copy per haploid

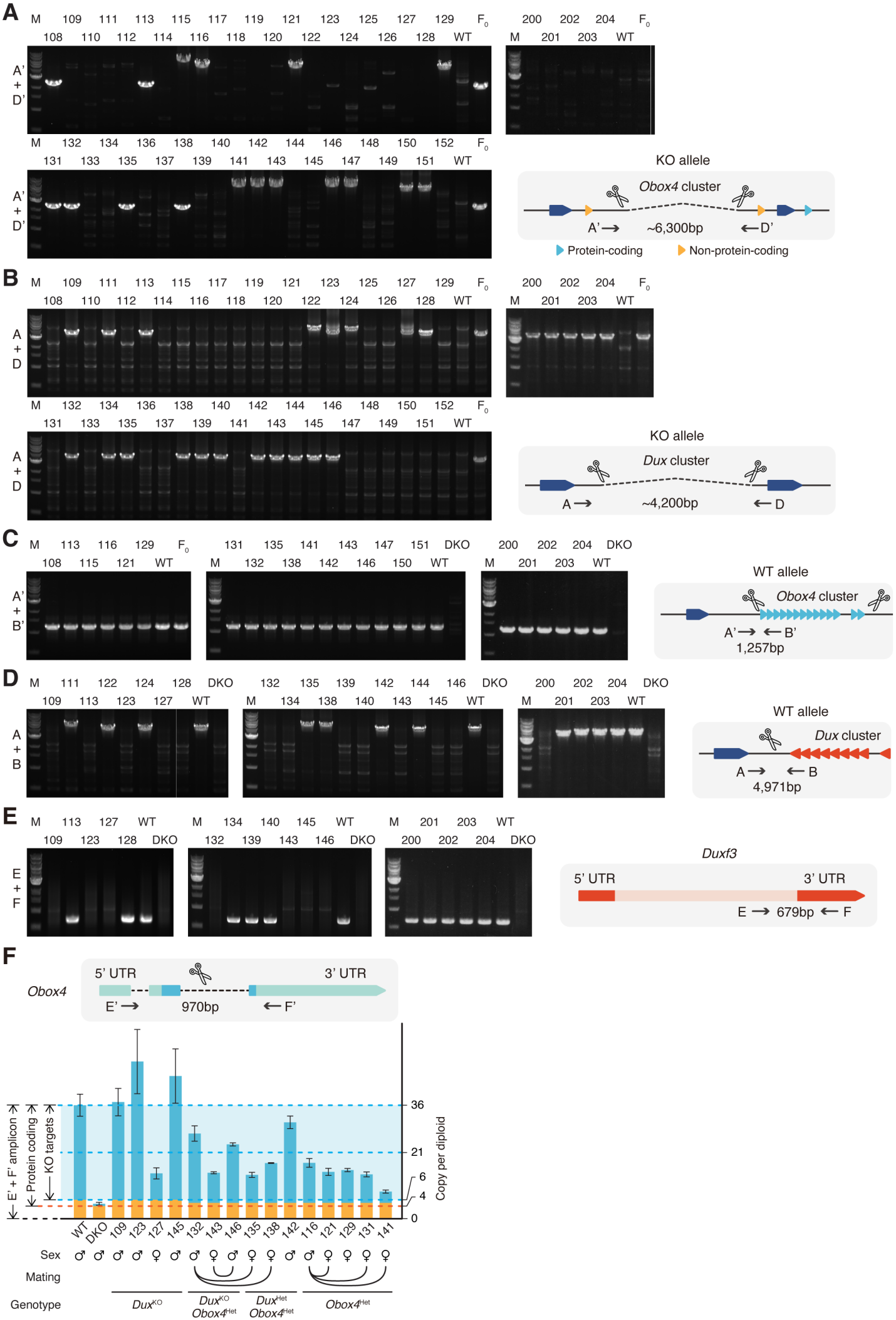


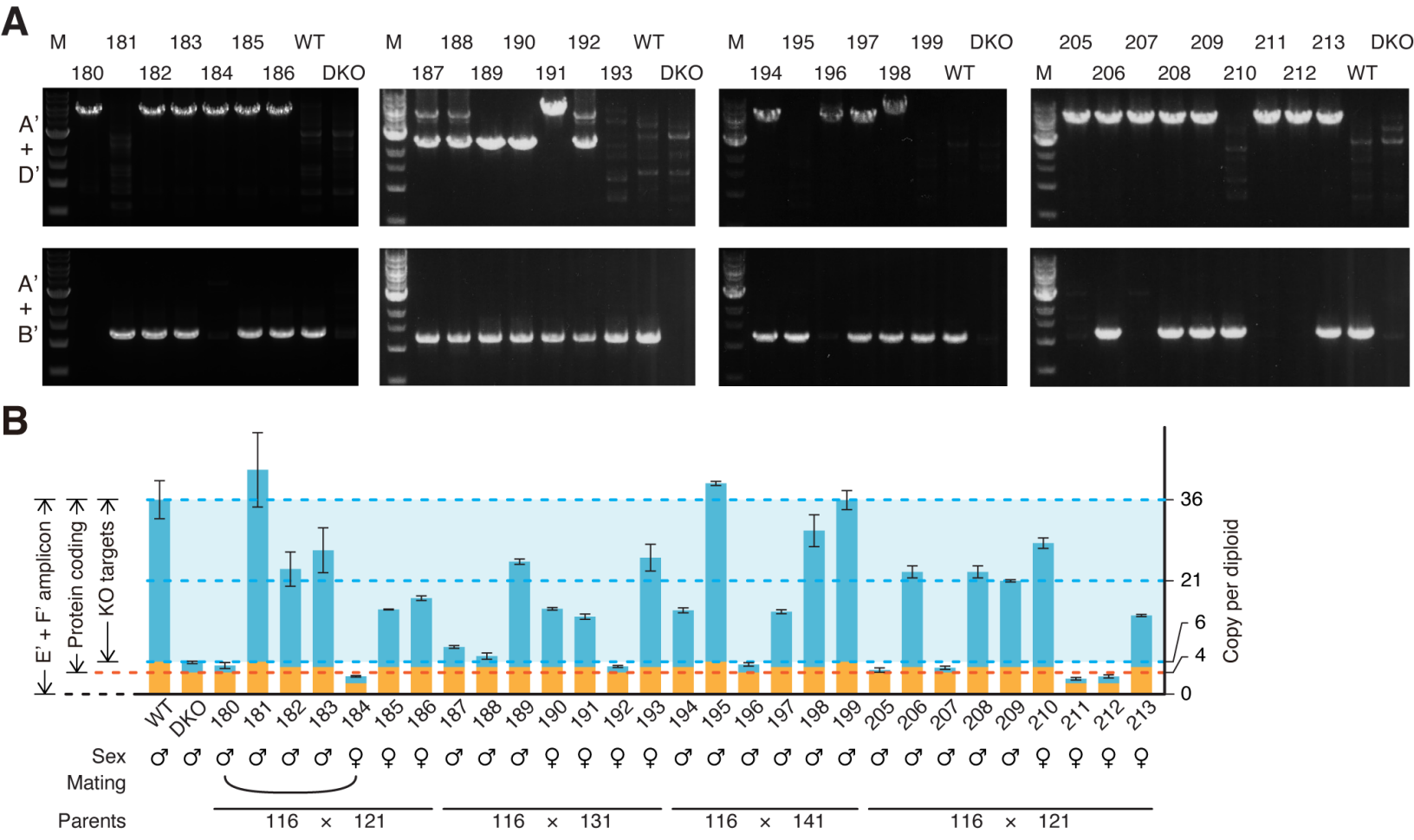
Against wildtype	ANOVA (Kruskal Wallis Test)		
	4-8 cell	Morula	Blastocyst
Obox4 KO	0.999	0.640	0.246
Dux KO	0.257	0.103	0.465
Double KO	0.536	0.106	0.0313*

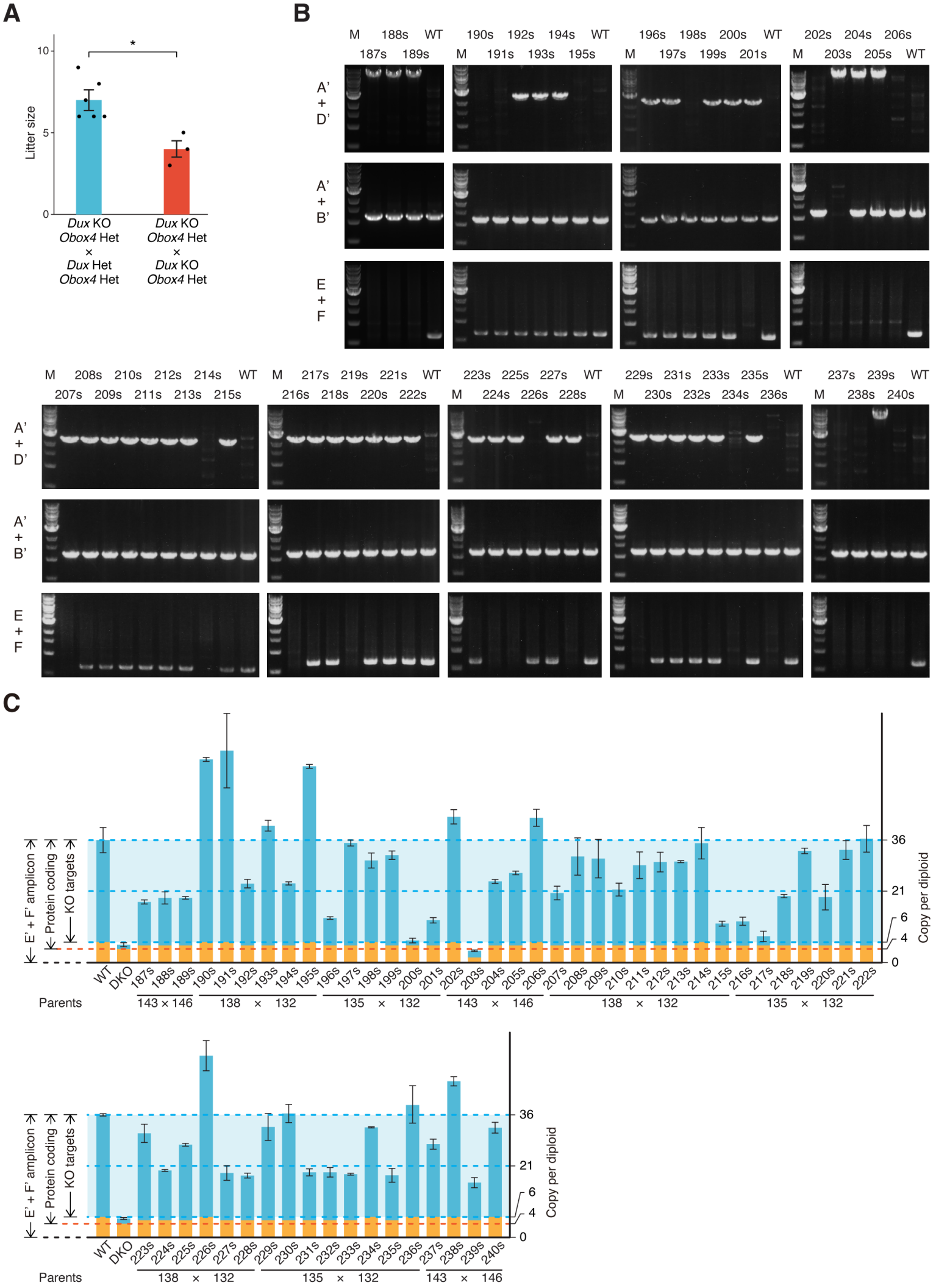


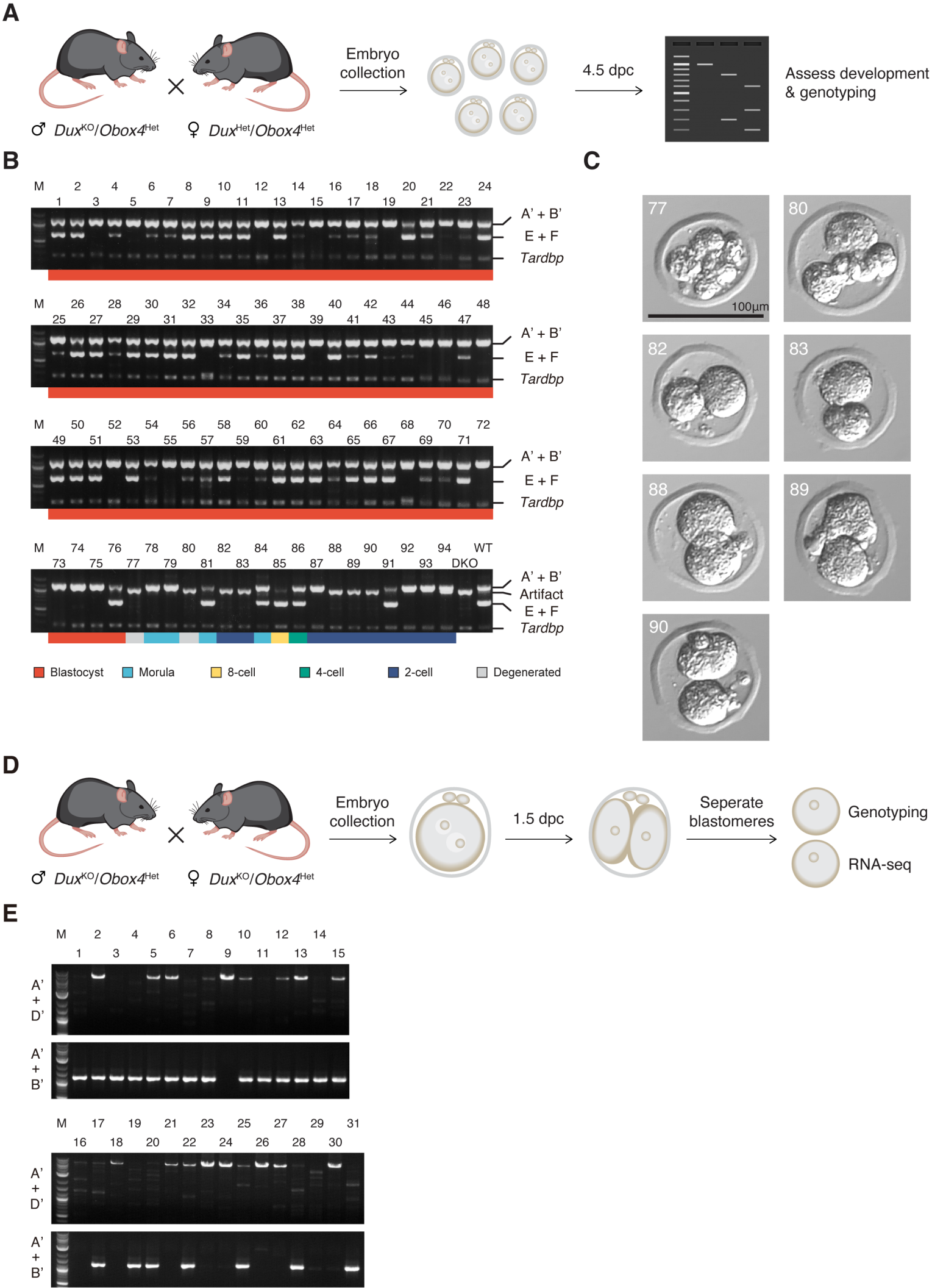


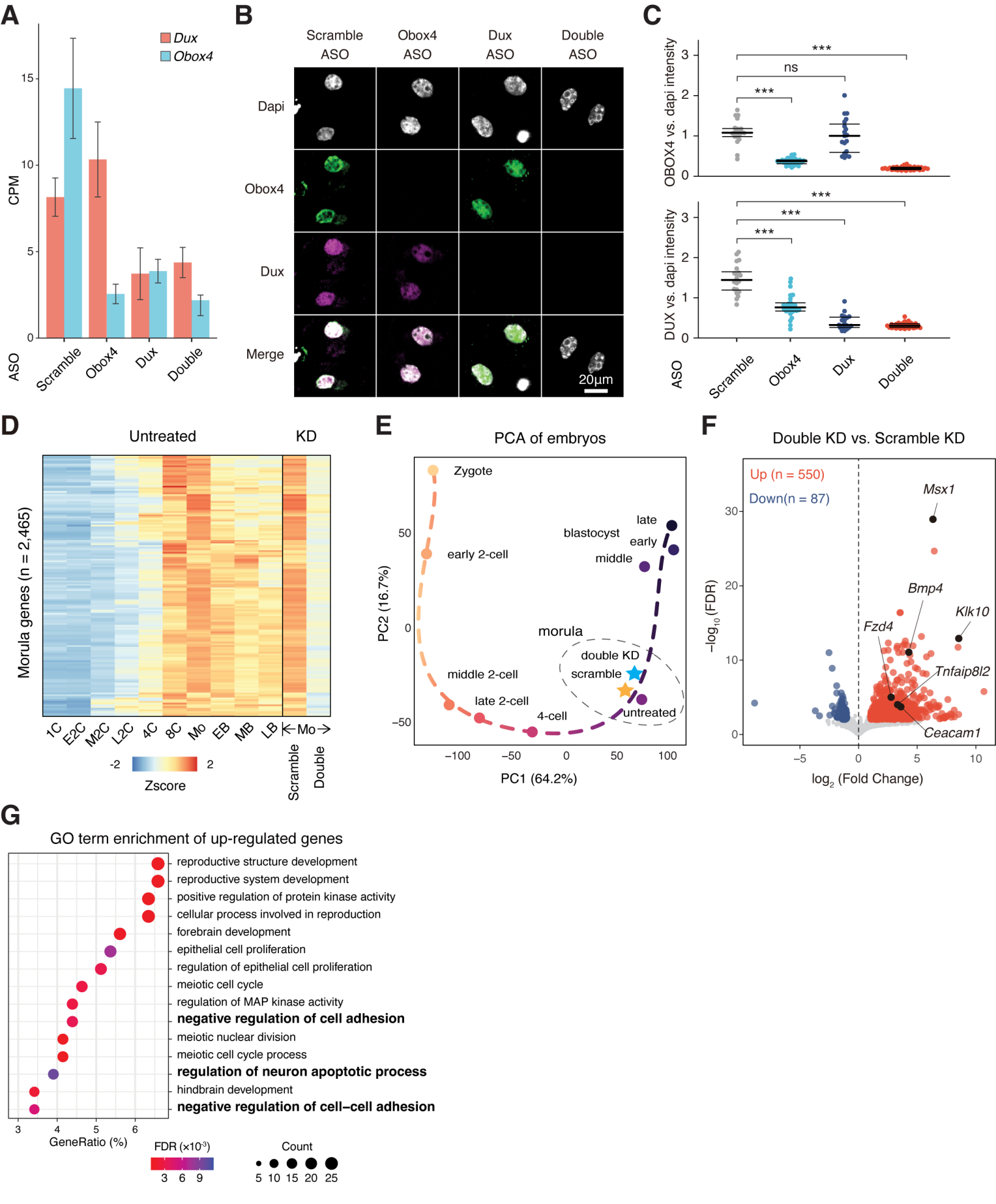


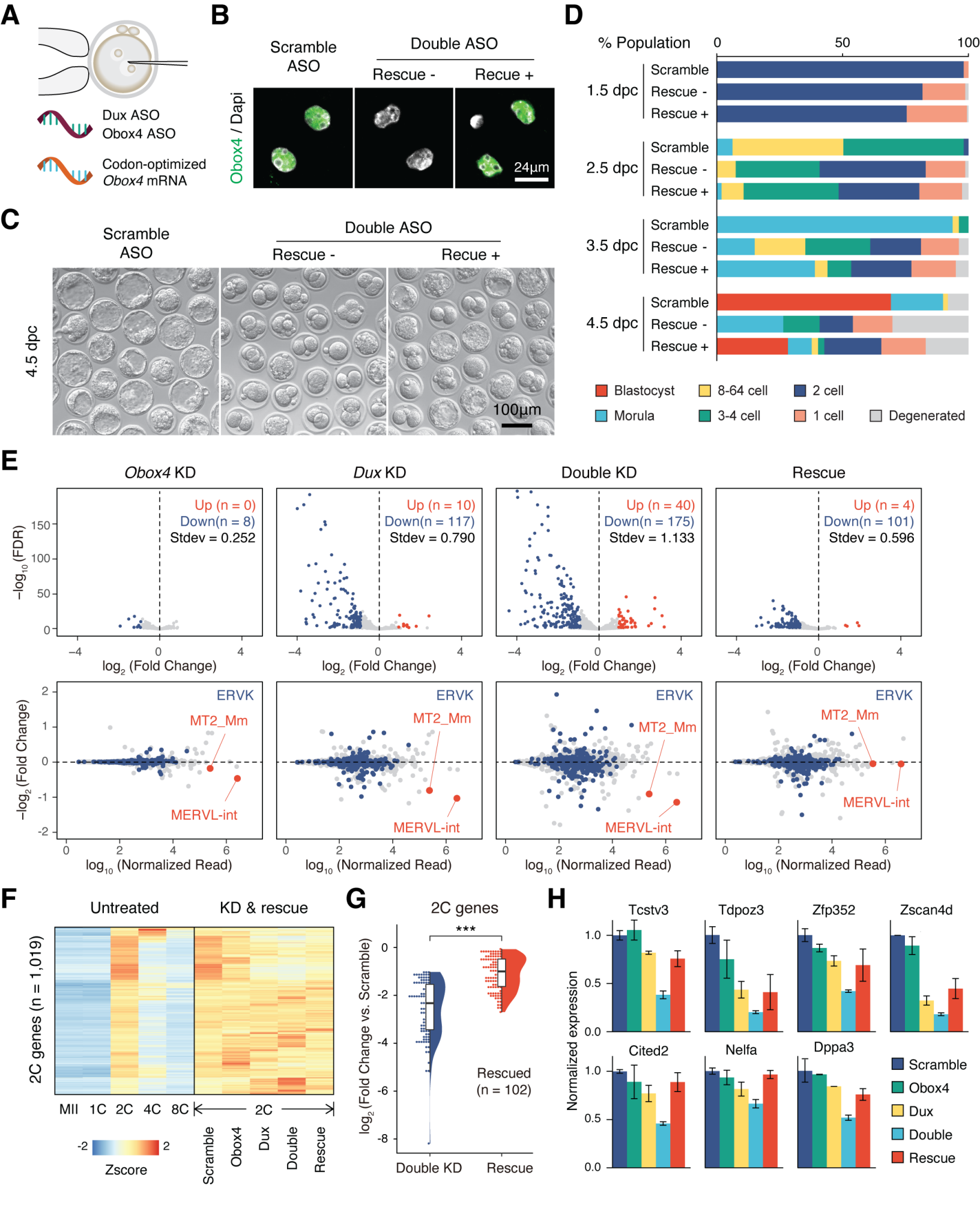

A
M 109 111 113 115 117 119 121 123 125 127 129 F0
108 110 112 114 116 118 120 122 124 126 128 WT
A' + D'
M 132 134 136 138 140 142 144 146 148 150 152 F0
131 133 135 137 139 141 143 145 147 149 151 WT
A' + D'
M 200 202 204 F0
M 201 203 WT
KO allele
Obox4 cluster
A' → ~6,300bp ← D'
Protein-coding
Non-protein-coding
B
M 109 111 113 115 117 119 121 123 125 127 129 F0
108 110 112 114 116 118 120 122 124 126 128 WT
A + D
M 132 134 136 138 140 142 144 146 148 150 152 F0
131 133 135 137 139 141 143 145 147 149 151 WT
A + D
M 200 202 204 F0
M 201 203 WT
KO allele
Dux cluster
A → ~4,200bp ← D
C
M 113 116 129 F0
108 115 121 WT
A' + B'
131 135 141 143 147 151 DKO
M 132 138 142 146 150 WT
200 202 204 DKO
M 201 203 WT
WT allele
Obox4 cluster
A' → ← B'
1,257bp
D
M 111 122 124 128 DKO
109 113 123 127 WT
A + B
132 135 139 142 144 146 DKO
M 134 138 140 143 145 WT
200 202 204 DKO
M 201 203 WT
WT allele
Dux cluster
A → ← B
4,971bp
E
M 113 127 WT
109 123 128 DKO
E + F
M 134 140 145 WT
132 139 143 146 DKO
M 201 203 WT
200 202 204 DKO
Duxf3
5' UTR
3' UTR
E → 679bp ← F
F
Obox4
5' UTR
3' UTR
E' →
970bp
← F'
E' + F' amplicon
Protein coding
KO targets
36
21
6
4
0
Copy per diploid
WT DKO 109 123 127 145 132 143 146 135 138 142 116 121 129 131 141
Sex ♂ ♂ ♂ ♂ ♀ ♂ ♂ ♀ ♂ ♀ ♀ ♂ ♂ ♀ ♀ ♀ ♀
Mating
Genotype
Dux^KO
Dux^KO Obox4^Het
Dux^Het Obox4^Het
Obox4^Het

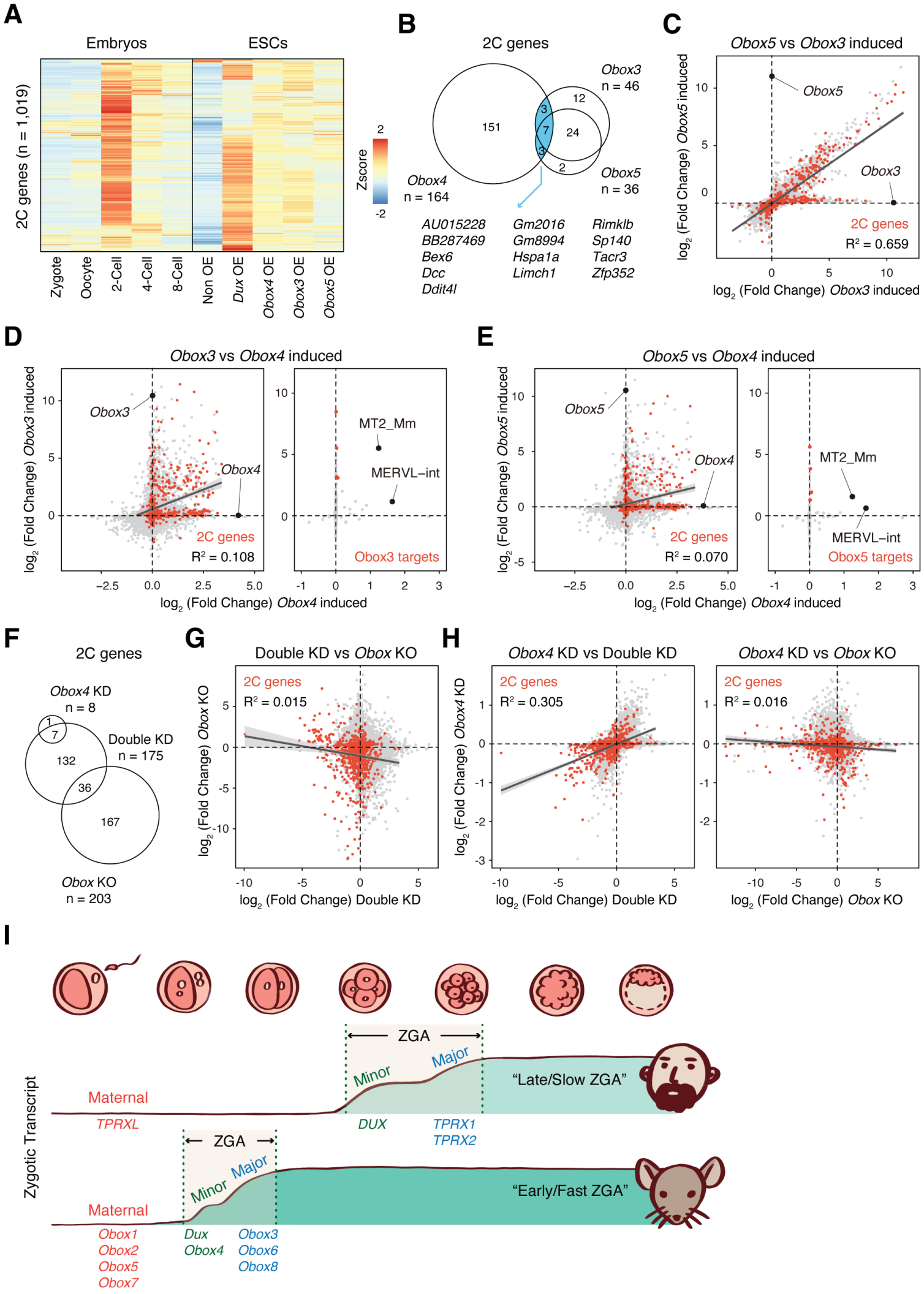

A
M 181 183 185 WT
180 182 184 186 DKO
M 188 190 192 WT
187 189 191 193 DKO
M 195 197 199 DKO
194 196 198 WT
205 207 209 211 213 DKO
M 206 208 210 212 WT
A' + D'
A' + B'
B
E' + F' amplicon
Protein coding
KO targets
36
21
6
4
0
Copy per diploid
WT DKO 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 205 206 207 208 209 210 211 212 213
Sex
Mating
Parents
116 × 121
116 × 131
116 × 141
116 × 121


A
B
C
Litter size
*
Dux KO Obox4 Het × Dux Het Obox4 Het
Dux KO Obox4 Het × Dux KO Obox4 Het
M 188s WT 190s 192s 194s WT 196s 198s 200s WT 202s 204s 206s
187s 189s M 191s 193s 195s M 197s 199s 201s M 203s 205s WT
A' + D'
A' + B'
E + F
M 208s 210s 212s 214s WT M 217s 219s 221s WT 223s 225s 227s WT 229s 231s 233s 235s WT 237s 239s WT
207s 209s 211s 213s 215s 216s 218s 220s 222s M 224s 226s 228s M 230s 232s 234s 236s M 238s 240s
E' + F' amplicon
Protein coding
KO targets
Copy per diploid
36
21
6
4
0
WT DKO 187s 188s 189s 190s 191s 192s 193s 194s 195s 196s 197s 198s 199s 200s 201s 202s 203s 204s 205s 206s 207s 208s 209s 210s 211s 212s 213s 214s 215s 216s 217s 218s 219s 220s 221s 222s
Parents
143 × 146
138 × 132
135 × 132
143 × 146
138 × 132
135 × 132
WT DKO 223s 224s 225s 226s 227s 228s 229s 230s 231s 232s 233s 234s 235s 236s 237s 238s 239s 240s
138 × 132
135 × 132
143 × 146


A
♂ Dux^KO/Obox4^Het
♀ Dux^Het/Obox4^Het
Embryo collection
4.5 dpc
Assess development & genotyping
B
M 2 4 6 8 10 12 14 16 18 20 22 24
1 3 5 7 9 11 13 15 17 19 21 23
M 26 28 30 32 34 36 38 40 42 44 46 48
25 27 29 31 33 35 37 39 41 43 45 47
M 50 52 54 56 58 60 62 64 66 68 70 72
49 51 53 55 57 59 61 63 65 67 69 71
M 74 76 78 80 82 84 86 88 90 92 94 WT
73 75 77 79 81 83 85 87 89 91 93 DKO
A' + B'
E + F
Tardbp
Artifact
Blastocyst
Morula
8-cell
4-cell
2-cell
Degenerated
C
77
80
82
83
88
89
90
100µm
D
♂ Dux^KO/Obox4^Het
♀ Dux^KO/Obox4^Het
Embryo collection
1.5 dpc
Seperate blastomeres
Genotyping
RNA-seq
E
M 2 4 6 8 10 12 14
1 3 5 7 9 11 13 15
A' + D'
A' + B'
M 17 19 21 23 25 27 29 31
16 18 20 22 24 26 28 30


A
Dux
Obox4
CPM
ASO
Scramble
Obox4
Dux
Double
B
Scramble ASO
Obox4 ASO
Dux ASO
Double ASO
Dapi
Obox4
Dux
Merge
20μm
C
OBOX4 vs. dapi intensity
DUX vs. dapi intensity
***
ns
D
Untreated
KD
Morula genes (n = 2,465)
1C
E2C
M2C
L2C
4C
8C
Mo
EB
MB
LB
Mo
Zscore
Scramble
Double
E
PCA of embryos
Zygote
early 2-cell
middle 2-cell
late 2-cell
4-cell
morula
double KD
scramble
untreated
blastocyst
late
early
middle
PC1 (64.2%)
PC2 (16.7%)
F
Double KD vs. Scramble KD
Up (n = 550)
Down(n = 87)
Msx1
Bmp4
Klk10
Fzd4
Tnfaip8l2
Ceacam1
log2 (Fold Change)
−log10 (FDR)
G
GO term enrichment of up-regulated genes
reproductive structure development
reproductive system development
positive regulation of protein kinase activity
cellular process involved in reproduction
forebrain development
epithelial cell proliferation
regulation of epithelial cell proliferation
meiotic cell cycle
regulation of MAP kinase activity
negative regulation of cell adhesion
meiotic nuclear division
meiotic cell cycle process
regulation of neuron apoptotic process
hindbrain development
negative regulation of cell–cell adhesion
GeneRatio (%)
FDR (×10-3)
Count


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
−log2 (Fold Change)
log10 (Normalized Read)
F
Untreated
KD & rescue
2C genes (n = 1,019)
MII
1C
2C
4C
8C
2C
Scramble
Obox4
Dux
Double
Rescue
-2
2
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


A
Embryos
ESCs
2C genes (n = 1,019)
Zscore
2
-2
Zygote
Oocyte
2-Cell
4-Cell
8-Cell
Non OE
Dux OE
Obox4 OE
Obox3 OE
Obox5 OE
B
2C genes
Obox3
n = 46
Obox5
n = 36
Obox4
n = 164
151
3
7
3
12
24
2
AU015228
BB287469
Bex6
Dcc
Ddit4l
Gm2016
Gm8994
Hspa1a
Limch1
Rimklb
Sp140
Tacr3
Zfp352
C
Obox5 vs Obox3 induced
log2 (Fold Change) Obox5 induced
log2 (Fold Change) Obox3 induced
Obox5
Obox3
2C genes
R² = 0.659
D
Obox3 vs Obox4 induced
log2 (Fold Change) Obox3 induced
log2 (Fold Change) Obox4 induced
Obox3
Obox4
2C genes
R² = 0.108
MT2_Mm
MERVL–int
Obox3 targets
E
Obox5 vs Obox4 induced
log2 (Fold Change) Obox5 induced
log2 (Fold Change) Obox4 induced
Obox5
Obox4
2C genes
R² = 0.070
MT2_Mm
MERVL–int
Obox5 targets
F
2C genes
Obox4 KD
n = 8
Double KD
n = 175
Obox KO
n = 203
1
7
132
36
167
G
Double KD vs Obox KO
log2 (Fold Change) Obox KO
log2 (Fold Change) Double KD
2C genes
R² = 0.015
H
Obox4 KD vs Double KD
log2 (Fold Change) Obox4 KD
log2 (Fold Change) Double KD
2C genes
R² = 0.305
Obox4 KD vs Obox KO
log2 (Fold Change) Obox KO
2C genes
R² = 0.016
I
Zygotic Transcript
ZGA
Minor
Major
Maternal
TPRXL
DUX
TPRX1
TPRX2
"Late/Slow ZGA"
ZGA
Minor
Major
Maternal
Obox1
Obox2
Obox5
Obox7
Dux
Obox4
Obox3
Obox6
Obox8
"Early/Fast ZGA"