

OBOX regulates mouse zygotic genome activation and early development

<https://doi.org/10.1038/s41586-023-06428-3>

Received: 23 October 2022

Accepted: 12 July 2023

Published online: 17 July 2023

 Check for updates

Shuyan Ji^{1,2,8}, Fengling Chen^{1,2,8}, Paula Stein^{3,4,8}, Jiacheng Wang^{1,2,8}, Ziming Zhou^{1,2,8}, Lijuan Wang^{1,2,8}, Qing Zhao^{1,2}, Zili Lin^{1,2,5}, Bofeng Liu^{1,2}, Kai Xu^{1,2}, Fangnong Lai^{1,2}, Zhuqing Xiong^{1,2}, Xiaoyu Hu^{1,2}, Tianxiang Kong^{1,2}, Feng Kong^{1,2}, Bo Huang⁶, Qiuju Wang^{1,2}, Qianhua Xu^{1,2}, Qiang Fan^{1,2}, Ling Liu^{1,2}, Carmen J. Williams³, Richard M. Schultz^{4,7} & Wei Xie^{1,2}

Zygotic genome activation (ZGA) activates the quiescent genome to enable the maternal-to-zygotic transition^{1,2}. However, the identity of transcription factors that underlie mammalian ZGA *in vivo* remains elusive. Here we show that OBOX, a PRD-like homeobox domain transcription factor family (OBOX1–OBOX8)^{3–5}, are key regulators of mouse ZGA. Mice deficient for maternally transcribed *Obox1/2/5/7* and zygotically expressed *Obox3/4* had a two-cell to four-cell arrest, accompanied by impaired ZGA. The *Obox* knockout defects could be rescued by restoring either maternal and zygotic OBOX, which suggests that maternal and zygotic OBOX redundantly support embryonic development. Chromatin-binding analysis showed that *Obox* knockout preferentially affected OBOX-binding targets. Mechanistically, OBOX facilitated the ‘preconfiguration’ of RNA polymerase II, as the polymerase relocated from the initial one-cell binding targets to ZGA gene promoters and distal enhancers. Impaired polymerase II preconfiguration in *Obox* mutants was accompanied by defective ZGA and chromatin accessibility transition, as well as aberrant activation of one-cell polymerase II targets. Finally, ectopic expression of OBOX activated ZGA genes and MERVL repeats in mouse embryonic stem cells. These data thus demonstrate that OBOX regulates mouse ZGA and early embryogenesis.

Zygotic genome activation (ZGA), the first transcription event after fertilization, drives the transition from maternal to embryonic control in early development^{1,2}. It often occurs in two waves—minor and major ZGA^{6–8}. In mice, minor ZGA occurs around the mid-one-cell stage when only a handful of genes are activated. Thousands of genes are then activated in late two-cell (L2C) embryos during major ZGA.

Polymerase II (Pol II) initiates widespread chromatin binding in one-cell (1C) mouse embryos, including many non-major ZGA targets⁹. It then undergoes relocation to major ZGA genes, or ‘preconfiguration’, with an intermediate state detected at the early two-cell (E2C) stage before major ZGA⁹. However, which sequence-specific factors guide Pol II preconfiguration remains unknown. In *Drosophila*, ZELDA, GAF and CLAMP were identified as master transcription factors for ZGA^{10–12}. Similar roles for NANOG, SOXB1 and POU5F1 were reported in zebrafish^{13–15}. DUX in mouse and DUX4 in human activate a subset of ZGA genes (mainly minor ZGA genes) in embryonic stem (ES) cells^{16–18}. However, *Dux* knockout barely affected ZGA in mouse embryos and roughly half of *Dux* knockout mice survived to term^{19,20}. NRSA2 was suggested to regulate mouse ZGA and development beyond the 2C stage, although its precise role remains under discussion^{21–23}. We recently showed that PRD-like homeobox transcription factors TPRXs regulate human ZGA

and early development²⁴. However, this finding remains to be tested in a genetic knockout model, and whether equivalent transcription factors in mice have similar roles in ZGA remains unknown.

Obox is highly expressed around ZGA

We first searched for transcription factors highly translated before and during major ZGA based on our transcriptome data²⁵ in mouse oocytes and early embryos (Supplementary Table 1). The most highly translated transcription factors in 1C, E2C (pre-major ZGA) and L2C (major ZGA) embryos were overwhelmingly dominated by the OBOX family (Fig. 1a, RPF). Moreover, five out of the top six transcription factor motifs enriched in accessible chromatin²⁶ (OBOX, OTX2, GSC, CRX and PITX1) at the L2C stage were all PRD-like homeobox transcription factors sharing the TAATCC-binding motif²⁷ (Fig. 1a, ATAC-seq motif). Because OTX2, GSC, CRX and PITX1 were either not or lowly expressed in mouse oocytes and early embryos, we focused on the OBOX family as potential ZGA regulators.

Previous phylogenetic analyses showed 66 *Obox* loci in mouse, all located in a single cluster on chromosome 7 (refs. 3,5) (Extended Data Fig. 1a). Based on transcriptome²⁸ and translational data²⁵,

¹Center for Stem Cell Biology and Regenerative Medicine, MOE Key Laboratory of Bioinformatics, New Cornerstone Science Laboratory, School of Life Sciences, Tsinghua University, Beijing, China. ²Tsinghua-Peking Center for Life Sciences, Beijing, China. ³Reproductive and Developmental Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA. ⁴Department of Biology, University of Pennsylvania, Philadelphia, PA, USA. ⁵College of Animal Science and Technology College, Beijing University of Agriculture, Beijing, China. ⁶Zhejiang Provincial Key Laboratory of Pancreatic Disease, the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. ⁷Department of Anatomy, Physiology and Cell Biology School of Veterinary Medicine University of California, Davis, Davis, CA, USA. ⁸These authors contributed equally: Shuyan Ji, Fengling Chen, Paula Stein, Jiacheng Wang, Ziming Zhou, Lijuan Wang. [✉]e-mail: rschultz@sas.upenn.edu; xiewei121@tsinghua.edu.cn

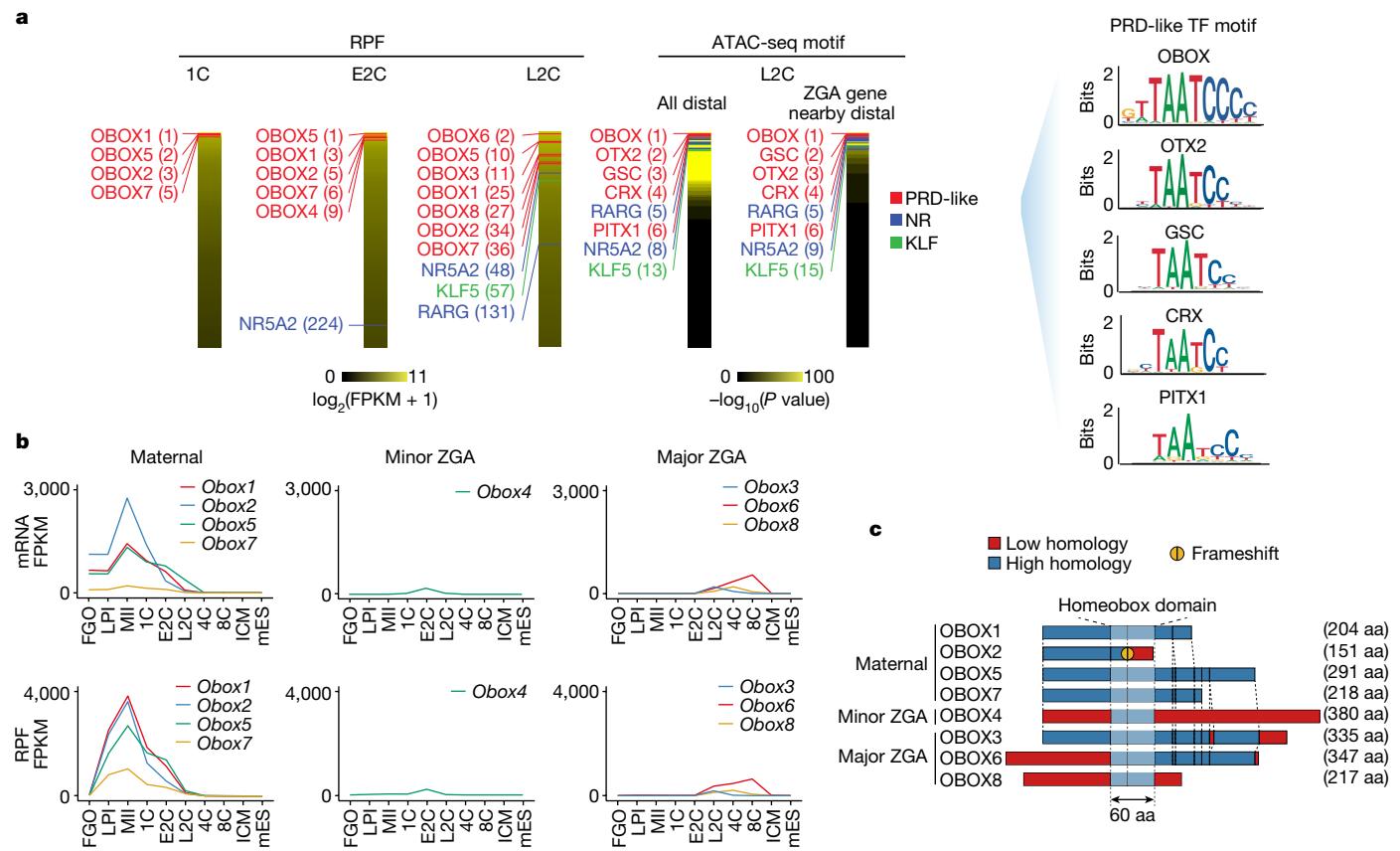


Fig. 1 | OBOX expression in mouse oocytes and preimplantation embryos.

a, Left, the top 250 transcription factors (TFs) based on translation levels (ribosome-protected fragments, RPF)²⁵ or motif enrichment in all distal-accessible regions or those near ZGA genes based on assay for transposase-accessible chromatin using sequencing (ATAC-seq)²⁶ in embryos. PRD-like homeobox family (red), nuclear receptor (NR, blue) and Kruppel-like factor

(KLF, green) transcription factors and their ranks are indicated. Right, OBOX-, OTX2-, GSC-, CRX- and PITX1-binding motifs are shown. **b**, Line plots showing mRNA²⁸ and translation²⁵ levels of maternal, minor and major ZGA *Obox* genes in oocytes and early embryos (two biological replicates). ICM, inner cell mass. **c**, Sequence alignment of OBOX proteins based on Clustal Omega. aa, amino acids.

we classified them into four groups (Supplementary Tables 2 and 3). (1) Maternal *Obox* genes include *Obox1/2/5/7*, which showed high RNA levels in oocytes and early embryos before their expression declined after ZGA (Fig. 1b and Extended Data Fig. 1b). They were not translated in full-grown oocytes but became highly translated from late prometaphase I stage until 2C stage, consistent with their transcripts containing cytoplasmic polyadenylation elements in the proximity of polyadenylation signal sites and undergoing poly(A) tail lengthening during oocyte maturation^{25,29} (Extended Data Fig. 1c,d). Maternally expressed OBOX are highly similar in protein sequences but with different lengths, because they arise through different premature stop codons (Fig. 1c). OBOX2 has a truncated homeobox domain due to a frameshift mutation, raising the possibility of impaired DNA-binding ability (Fig. 1c, frameshift). (2) Minor ZGA *Obox* include *Obox4* and its pseudogenes ($n = 51$), whose transcripts and translation were low in oocytes but increased markedly in E2C embryos before rapidly declining in L2C embryos (Extended Data Fig. 1b). (3) Major ZGA *Obox* include *Obox3* and its pseudogenes ($n = 7$), *Obox8* and *Obox6*, which were primarily activated during major ZGA and peaked at L2C, 4C and 8C, respectively (Fig. 1b). The expression of *Obox3* was detected as early as mid-2C and can be considered as minor-major ZGA genes (Extended Data Fig. 1e,f). (4) A proportion of *Obox4* pseudogenes showed little or no expression in oocytes/early embryos and were hence excluded from further analyses (Extended Data Fig. 1b). Examination of previously published transcriptome and translome datasets^{30,31} showed largely similar results (Extended Data Fig. 2a,b). OBOX proteins detected by specific antibodies showed consistent expression patterns with the corresponding translation levels

(Extended Data Fig. 2c–h). Therefore, *Obox* genes are dynamically regulated at both transcriptional and posttranscriptional levels in mouse oocytes and early embryos.

Obox knockout caused 2C–4C arrest

We asked whether OBOX regulates mouse early development and ZGA. Knocking down individual *Obox* genes did not affect embryo development (Extended Data Fig. 3a–d and Supplementary Table 4). Considering their possible redundancy, we sought to knock out multiple *Obox* genes simultaneously. Maternally expressed *Obox1/2/5/7*, minor ZGA *Obox4* and minor-major ZGA *Obox3* showed the highest expression levels before or around major ZGA (Fig. 1b). We therefore removed a region that encompasses *Obox1/2/3/4/5/7*, including all expressed *Obox3/4* pseudogenes, in mice (referred to as *Obox*^{-/-} hereafter). We first confirmed knockout of *Obox* genes (Extended Data Fig. 4a–d). Heterozygous females and males were fertile, with offspring numbers comparable to those of wild type (Extended Data Fig. 4e). However, no *Obox* maternal-zygotic knockout (mzKO) pups were born when crossing *Obox*^{-/-} females with *Obox*^{-/-} males (Extended Data Fig. 4f). Morphology of the *Obox*^{-/-} ovary was comparable to that of wild-type ovary (Extended Data Fig. 4g). *Obox*^{-/-} female mice could ovulate normally, and *Obox*^{-/-} oocytes underwent meiosis with correct spindle configuration and no apparent transcriptome alterations (Extended Data Fig. 4h–j). However, when we isolated embryos *in vivo* at a time when the control embryos developed to blastocysts, the *Obox* mzKO embryos were still arrested at the 2C–4C stage (with a small percentage

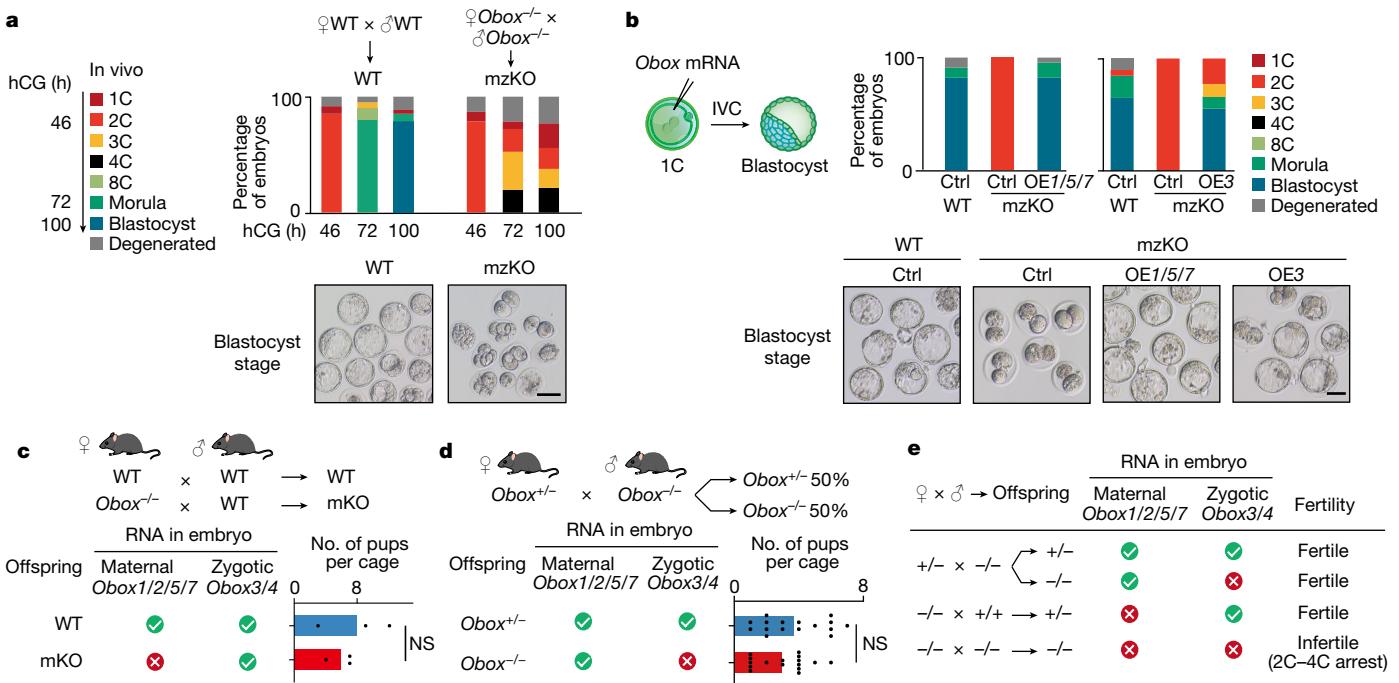


Fig. 2 | Maternal and zygotic OBOX redundantly supported embryo development. **a**, Embryo morphology and developmental rates of wild-type (WT) and *Obox* mzKO embryos dissected in vivo (four biological replicates). hCG, human chorionic gonadotropin. h, hour. Scale bar, 75 μm . **b**, OBOX rescue by overexpression of *Obox1/5/7* (OE1/5/7) or *Obox3* (OE3) mRNA, and the resulting embryo morphology and developmental rates (three biological replicates). IVC, in vitro culture. Ctrl, control. Scale bar, 75 μm . **c**, Offspring

numbers for either wild-type or *Obox-/-* female mice crossed with wild-type male mice (three litters for each group). The presence or absence of *Obox* mRNAs in embryos is indicated. NS, not significant ($P = 0.52$, two-sided *t*-test). **d**, Offspring types and numbers for *Obox-/-* female mice crossed with *Obox-/-* male mice (total of 108 pups from 16 litters). $P = 0.07$, two-sided paired *t*-test. **e**, Summary of genotypes and phenotypes from different *Obox* mutant mouse crossings.

arrested at 1C; Fig. 2a). A similar result was obtained for in vitro cultured embryos from 1C (Extended Data Fig. 4k), suggesting that OBOX proteins are required for development beyond 4C.

Maternal and zygotic *Obox* are redundant

We asked whether restoration of OBOX could rescue the developmental defects of *Obox* mzKO embryos. When maternal *Obox1/5/7* mRNAs (*Obox2* omitted due to a truncated homeobox domain) were introduced back into *Obox* mzKO zygotes, these embryos successfully developed to blastocysts (Fig. 2b) with the transcriptome properly restored (Extended Data Fig. 5a). A similar rescue was achieved by the introduction of zygotic *Obox3* mRNA at 1C or mid-2C (Fig. 2b and Extended Data Fig. 5a,b), suggesting that ZGA regulators are not limited to maternally deposited and 1C-expressed genes. By contrast, *Obox4* mRNA partially rescued development with a small portion of embryos developed to blastocysts (16.7%; Extended Data Fig. 5c,d).

We then asked whether maternal or zygotic *Obox* could further support development to term. Because zygotic *Obox* were expressed in embryos, maternal *Obox* knockout mice derived from *Obox-/-* female \times wild-type male mice would still express *Obox3/4* but not maternal *Obox1/2/5/7* (Extended Data Fig. 5e,f). Indeed, maternal *Obox* knockout embryos developed to blastocysts with normal gene expression (Extended Data Fig. 5g,h) and further survived to term (Fig. 2c), suggesting that *Obox* mzKO embryos could be fully rescued by zygotic *Obox3/4*. We then asked whether supplementation of maternal OBOX could also support mzKO embryo development to term. By crossing *Obox-/-* female with *Obox-/-* male mice, all embryos (including half of *Obox-/-* and half of *Obox-/-*) carried maternal *Obox1/2/5/7* mRNAs supplied from oocytes (Extended Data Fig. 5i). However, *Obox-/-* embryos, unlike *Obox-/-* embryos, did not express *Obox3* and *Obox4*. We found that *Obox-/-* embryos could also survive to term (Fig. 2d), suggesting

that the defects of mzKO embryos were fully rescued by maternal OBOX. Therefore, maternal and zygotic OBOX redundantly regulate mouse early development (Fig. 2e).

Obox knockout impaired ZGA

We next asked whether ZGA in *Obox* mzKO embryos was affected. At E2C, 32% (21 put of 65) of minor ZGA genes (Methods and Supplementary Tables 5 and 6) were downregulated in *Obox* mzKO embryos (Fig. 3a and Extended Data Fig. 6a,b), including MERVL repetitive elements, a marker of E2C³²⁻³⁴ (Fig. 3b and Extended Data Fig. 6c). *Dux* was not downregulated (Fig. 3c), suggesting that ZGA defects resulting from OBOX depletion were not through DUX. At L2C, *Obox* mzKO embryos exhibited a widespread decrease in major ZGA genes (530 out of 1,107, 48%; Fig. 3a and Supplementary Table 6). Downregulated genes preferentially function in essential pathways such as ribosomal RNA, mRNA processing and translation (Extended Data Fig. 6b), and also include transcription factors *Dppa2*, *Gata1/4* and *Nr5a2* (Fig. 3c). Such ZGA defects were not due to developmental delay, because maternal transcript clearance was not globally altered (Extended Data Fig. 6d,e). Minor ZGA genes were upregulated in *Obox* mutants at L2C, probably reflecting delayed downregulation (Extended Data Fig. 6c). In sum, the OBOX family regulates both minor and major ZGAs in mouse embryos.

OBOX bound ZGA genes

We then asked how OBOX regulates ZGA, by probing the binding targets of OBOX1/5 and OBOX3 (representing maternal and zygotic ZGA OBOX, respectively). Overexpression of Flag-tagged *Obox* in wild-type embryos had no or only moderate effects on the transcriptome (Extended Data Fig. 7a). As negative controls, no significant binding was detected for either OBOX2 (with a truncated homeobox domain) or OBOX5 with a

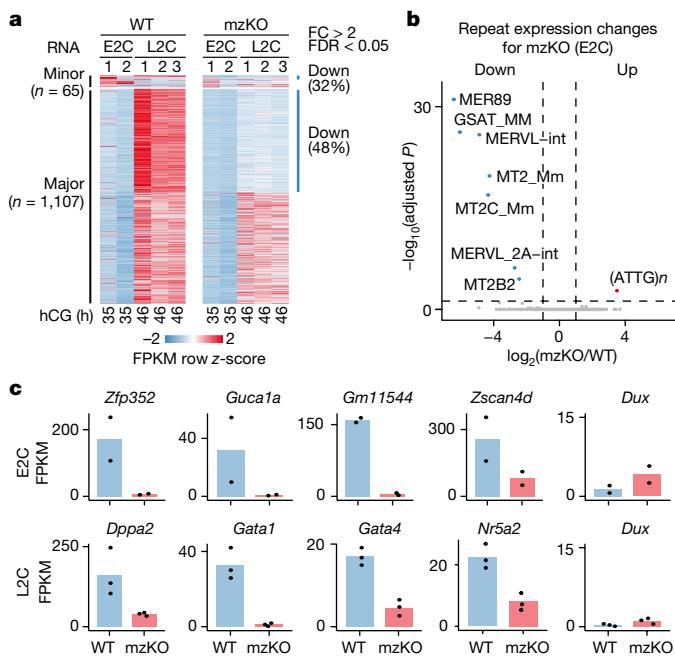


Fig. 3 | Loss of OBOX caused defective minor and major ZGAs. a, Heatmap showing minor and major ZGA gene expression in wild-type and *Obox* mzKO embryos (two biological replicates for E2C and three for L2C). n, ZGA gene number. FC, fold change; FDR, false discovery rate. b, Volcano plot showing repeat expression changes comparing *Obox* mzKO and wild-type E2C embryos (two biological replicates). Dashed line, adjusted P threshold of 0.05. c, Bar charts showing minor and major ZGA gene expression in wild-type and *Obox* mzKO embryos (two biological replicates for E2C and three for L2C; ten embryos per group).

single amino acid mutation (OBOX5^{R98E}, with a mutation in a key amino acid that contacts the minor groove of DNA)^{35,36} that abolished gene activation ability in a luciferase reporter assay (Extended Data Fig. 7b), or for embryos without *Obox* injection (Fig. 4a and Extended Data Fig. 7c). By contrast, OBOX1, OBOX5 and OBOX3 showed strong (with 48,592, 33,422 and 32,125 binding peaks, respectively; Supplementary Table 7) and similar binding in the genome in L2C embryos (Extended Data Fig. 7d), as exemplified at *Dppa2*, *Nr5a2* and MERVL (Fig. 4a). OBOX preferentially enriched at enhancers, promoters, MERVL and B1/B2/B4 repetitive elements (Fig. 4b and Extended Data Fig. 7c,e,f). The top two de novo motifs for OBOX-binding peaks matched well with the reported OBOX-binding motif (TAATCCC)²⁷ (Fig. 4c). These two motifs were actually adjacently present in OBOX-binding peaks (51.1% of OBOX1 peaks, 71.4% of OBOX5 peaks and 68.1% of OBOX3 peaks), leading to the identification of an extended OBOX-binding motif (ACNCCTTTAAC-TCCAG), with OBOX1 showing the longest consecutive version (CCTT-TAATCCCAG), which was chosen for subsequent analysis (Extended Data Fig. 7g). About 95.9% of this extended OBOX motif located in a B1 element, and it rendered stronger gene activation than the reported seven-base-pair (bp) motif in reporter assays in embryos and HEK293 cells (Extended Data Fig. 7h,i). Reporter activity was abolished in *Obox* mzKO L2C embryos but was rescued by the reintroduction of *Obox* mRNAs (Extended Data Fig. 7j). Of note, 2C-specific genes contained more OBOX-binding motifs at promoters compared with those specifically activated at other stages (Extended Data Fig. 7k). ZGA genes containing more OBOX motifs showed higher OBOX binding, both for promoters and distal regions (putative enhancers), and stronger downregulation in *Obox* mzKO embryos (Fig. 4d,e and Extended Data Fig. 7l–n). Genes containing both promoter and distal OBOX motifs were most affected (Fig. 4f), arguing against the likelihood that developmental delay was the basis for the differences noted. In sum, OBOX

preferentially binds and regulates ZGA genes with the OBOX motif in the mouse embryo.

OBOX guided Pol II preconfiguration

Pol II undergoes ‘loading, preconfiguration, and production’ during mouse ZGA⁹. Pol II binding correlates with cytosine-guanine (CG) density for 1C-specific and 1C–L2C-shared peaks before deviating from such correlation at L2C-specific peaks (Fig. 4g, Pol II), raising the possibility that CG-rich promoters are naturally accessible³⁷ for initial Pol II loading in 1C embryos whereas L2C-specific Pol II peaks, which are CG-poor, require additional transcription factors. In L2C-specific Pol II targets, the OBOX motif was the top motif enriched and *Obox* showed the highest expression among the inferred TFs (Fig. 4g, right). Importantly, OBOX1 and OBOX5 binding preceded Pol II recruitment in these regions in E2C embryos (Fig. 4g, arrows), raising the possibility that OBOX guides Pol II to these targets.

Next, we performed small-scale Tn5-assisted chromatin cleavage with sequencing (stacc-seq) for Pol II in wild-type and *Obox* mzKO 1C, E2C and L2C embryos. Among all Pol II peaks in wild-type embryos, 21% were present at promoters and 79% were away from promoters (distal) (30% intergenic and 49% intragenic; Fig. 5a and Supplementary Table 8). For distal Pol II peaks, we focused on intergenic peaks to avoid the confounding elongating Pol II in gene bodies. Pol II binding in wild-type and *Obox* mzKO 1C embryos was similar (Extended Data Fig. 8a), suggesting that the initial Pol II binding is independent of OBOX and may be recruited to CG-rich regions ‘by default’. However, at E2C, whereas Pol II already initiated recruitment to L2C-specific sites in wild-type embryos, this process was impaired in *Obox* mzKO mutants (Extended Data Fig. 8a–c). Such defects were exacerbated in L2C embryos (Fig. 5a, red arrows). Decreased Pol II peaks in *Obox* mutants were more likely to contain the OBOX motif compared with unaffected and increased peaks (Extended Data Fig. 8d). Moreover, genes showing decreased Pol II at promoters or distal regions (potential enhancers), but not those with unaffected Pol II, preferentially exhibited downregulation (Fig. 5b,c). Notably, failure of Pol II recruitment to 2C-specific targets was accompanied by its aberrant retention at 1C targets in both E2C and L2C *Obox* mzKO embryos (Fig. 5d and Extended Data Fig. 8a,b); 435 genes were ectopically activated in L2C embryos (Fig. 5e, Extended Data Fig. 8e and Supplementary Table 9). Approximately 50% of ectopically activated genes (compared with 27% of all genes) showed strong Pol II binding in 1C embryos (Extended Data Fig. 8f). These ectopically activated genes were normally inactive in early development and enriched for developmental genes, transcription factors and Polycomb targets³⁸ (Extended Data Fig. 8e–g), consistent with their CG-rich promoters. Of note, by identification of trophectoderm- and inner cell mass-enriched genes from a published dataset²⁶ we found 27 (out of 340) trophectoderm-enriched genes and 22 (out of 360) inner cell mass-enriched genes aberrantly activated in *Obox* mzKO embryos (Extended Data Fig. 8h,i).

We then asked whether OBOX might drive chromatin opening in early embryos. Using ATAC-seq, we found that OBOX depletion decreased chromatin accessibility at L2C-specific Pol II binding sites (Fig. 5a, green arrows). About 21% of active enhancers (9,191 out of 43,995, defined by distal H3K27ac) showed substantial decreases in chromatin accessibility in *Obox* mzKO L2C embryos. Failure to open promoters and enhancers also correlated with the downregulation of nearby ZGA genes (Fig. 5b,c). Together, these data indicate that OBOX guides timely preconfiguration of Pol II and chromatin accessibility at regulatory elements, and that loss of OBOX results in defective ZGA and aberrant activation of Pol II 1C targets.

OBOX activated ZGA genes in mES cells

To ask whether OBOX can activate ZGA genes beyond early embryos, we transiently overexpressed *Obox5* or *Obox3* (representing maternal or

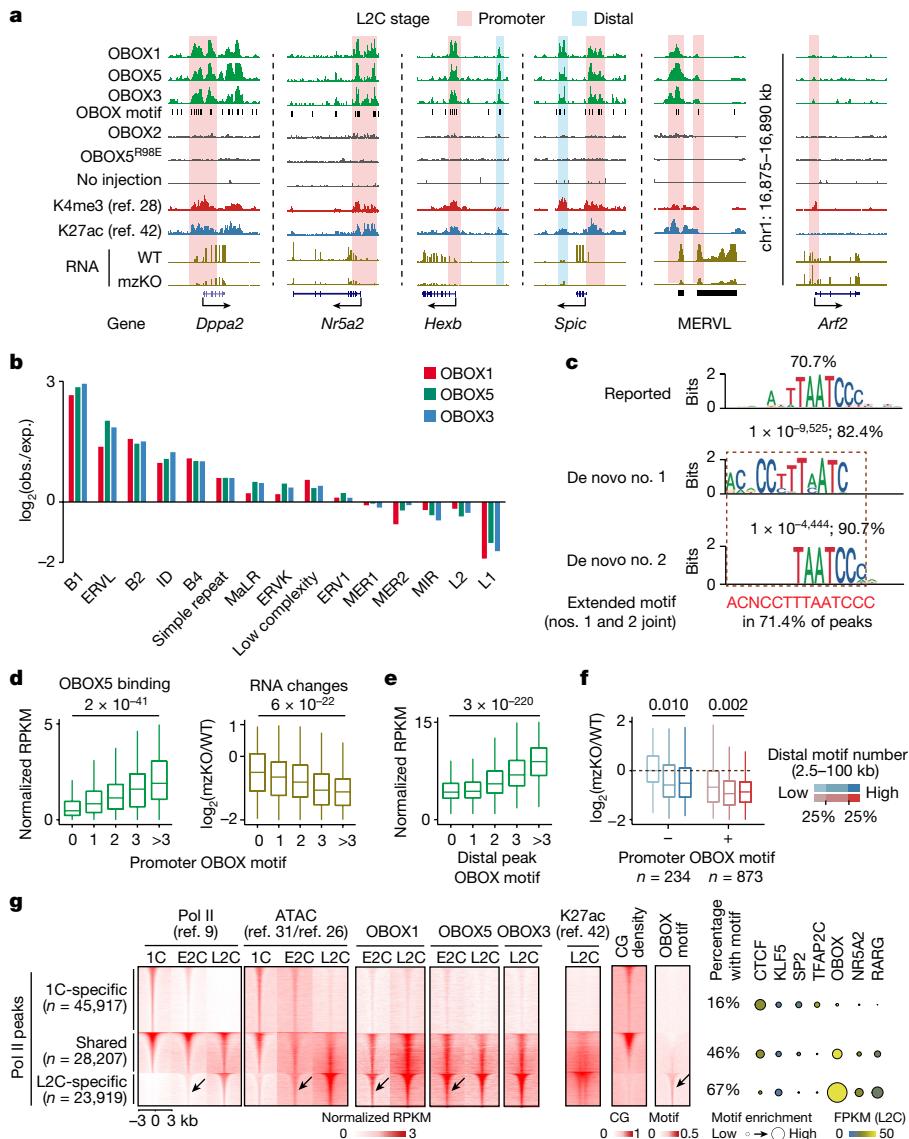


Fig. 4 | OBOX binding in 2C embryos. **a**, UCSC browser snapshots showing OBOX1/5/3 binding at example genes and repeats in L2C embryos. Stacc-seq of OBOX2, OBOX5^{R98E} and non-injected embryos are all negative controls. H3K4me3 (ref. 28), H3K27ac (ref. 42), OBOX motif and RNA levels in wild-type and *Obox* mzKO embryos are shown. **b**, Bar chart showing repeat enrichment at OBOX-binding peaks at L2C. obs., observed; exp., experimental. **c**, Reported motif²⁷, de novo top two and combined extended motif identified by OBOX5-binding peaks in embryos are shown. Percentages of peaks containing these motifs and *P* values are shown. **d**, Box plots showing OBOX5 binding at major ZGA gene promoters in wild-type L2C embryos (left) and major ZGA gene expression FC following OBOX depletion (right, three biological replicates). *P* values, two-sided Wilcoxon rank-sum test; 234, 272, 232, 169 and 201 genes have zero, one, two, three and more than three OBOX motifs on promoters, respectively. Centre line, median; box, 25th and 75th percentiles; whiskers,

1.5 × interquartile range (same for **e**, **f**). **e**, Box plot showing OBOX5-binding enrichment at distal binding peaks in L2C embryos. *P* value, two-sided Wilcoxon rank-sum test; 5,885, 16,226, 5,314, 1,370 and 561 distal OBOX5-binding peaks have zero, one, two, three and more than three OBOX motifs, respectively. **f**, Box plots showing ZGA gene expression changes following *Obox* knockout (three biological replicates). *P* values, two-sided Wilcoxon rank-sum test. *n*, major ZGA gene number. **g**, Left, heatmaps showing Pol II binding⁹, chromatin accessibility (ATAC)^{26,31}, OBOX binding, H3K27ac (ref. 42), CG density and OBOX motif enrichment at 1C-specific, shared and L2C-specific Pol II peaks in wild-type embryos. Percentages of peaks with at least one OBOX motif are shown. Arrows indicate Pol II, accessible chromatin, OBOX1/5 binding and OBOX motif at L2C-specific Pol II peaks in E2C embryos. Right, enrichment of known transcription factor motifs. Motif *P* values indicated by circle size.

zygotic *Obox*, respectively) in naive (2i) mouse embryonic stem (mES) cells (Extended Data Fig. 9a). Genes activated by OBOX in 2i mES cells included substantial numbers of ZGA genes (132 out of 449 for OBOX5 and 188 out of 728 for OBOX3; Fig. 5f, Extended Data Fig. 9b–d and Supplementary Table 10), and showed strong OBOX5/3 binding in 2C embryos (Extended Data Fig. 9e,f). These genes were preferentially activated in wild-type 2C–8C embryos (Extended Data Fig. 9g) and downregulated in *Obox* knockout embryos (73.5%, *n* = 97 for OBOX5 and 71.3%, *n* = 134 for OBOX3; Extended Data Fig. 9h). ZGA genes activated

by ectopic OBOX5 and OBOX3 also exhibited a strong overlap (*n* = 118, *P* = 7 × 10⁻⁵), again supporting the functional redundancy of OBOX proteins. MERVL elements (including MT2C_Mm, MT2B2, MT2_Mm and MERVL-int) were also activated by OBOX5/3 (Fig. 5g and Extended Data Fig. 9i). About 70% of MT2_Mm repeats and 41% of MT2C_Mm repeats contain the extended OBOX-binding motif. Intriguingly, several pluripotency genes were downregulated by *Obox*5/3 expression (for example, *Sox2*, *Klf3*4/5; Fig. 5g), raising the possibility that OBOX proteins promote totipotency and suppress pluripotency programmes.

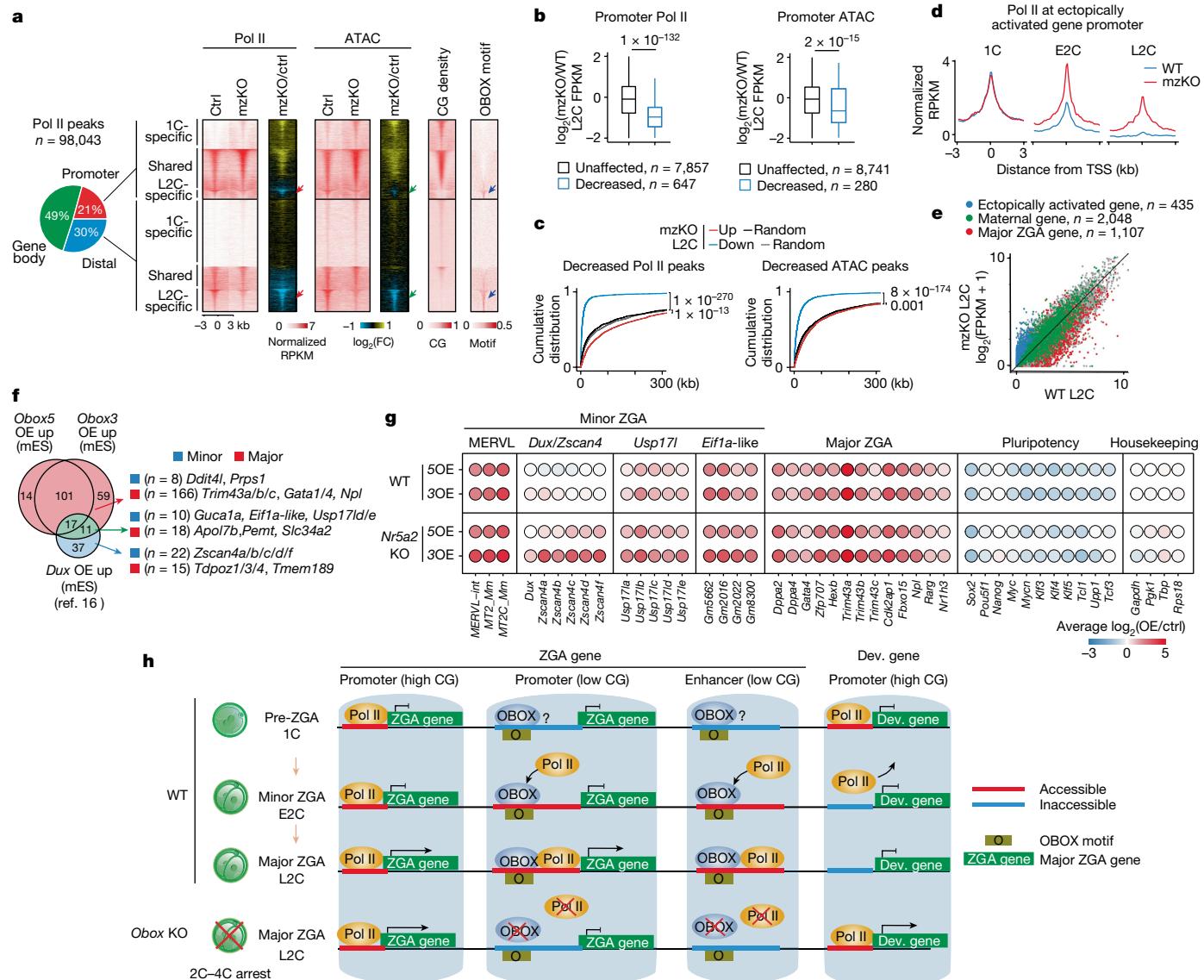


Fig. 5 | OBOX regulated Pol II preconfiguration in embryos and its overexpression activated ZGA genes and MERVL in mES cells. **a.** Left, pie chart showing Pol II peak distribution in the genome. Right, heatmaps showing Pol II binding and ATAC signals at 1C-specific, shared and L2C-Pol II peaks at L2C (two biological replicates). CG density and OBOX motif enrichment are shown. **b.** Box plots showing expression changes for genes with promoter Pol II binding or ATAC enrichment decreased or unaffected in *Obox* mzKO embryos (two biological replicates). *P* values, two-sided Wilcoxon rank-sum test; *n*, gene number. **c.** Empirical cumulative density function of distance from downregulated or upregulated gene transcription start sites (TSS) to the nearest decreased distal Pol II peaks or ATAC peaks (two biological replicates). *P* values, two-sided Wilcoxon rank-sum test. Downregulated genes, *n* = 2,026; upregulated genes, *n* = 1,486. Equal numbers of random control genes are included. Decreased distal Pol II and ATAC peak numbers are 23,039 and 14,364, respectively. **d.** Promoter Pol II enrichment (*z*-score normalized,

In line with *Dux* expression being unaffected in *Obox* mzKO embryos, neither OBOX binding nor OBOX motif was present at the *Dux* promoter in 2C embryos, and *Dux* was barely activated following *Obox5/3* over-expression (fragments per kilobase exon per million mapped reads (FPKM) < 2, $P = 0.14$) in 2i mES cells (Extended Data Fig. 10a,b). Conversely, Re-analyses of published data showed that *Obox* genes, apart from *Obox4*, were not or only moderately affected by *Dux* overexpression in mES cells¹⁶ and *Dux* knockout in embryos²⁰ (Extended Data

two biological replicates) for ectopically activated genes. **e**, Scatter plots comparing gene expression between wild-type and *Obox* mzKO embryos (three biological replicates). **f**, Venn diagram showing the overlap of *Obox* (four biological replicates) or *Dux* (two biological replicates) overexpression (OE) of upregulated ZGA genes in 2i mES cells. *n*, ZGA gene number. **g**, Balloon plot showing average gene expression changes following overexpression of *Obox5* or *Obox3* in wild-type (top, four biological replicates) or *Nr5a2* knockout (bottom, two biological replicates) 2i mES cells. Housekeeping genes were used as a control. **h**, Model illustrating the role of OBOX in ZGA. Before ZGA (IC), promoters with high CG density are initially accessible and bound by Pol II. Later (E2C and L2C), Pol II leaves IC-specific targets (with mechanisms unclear) and OBOX guides Pol II to CG-poor ZGA gene promoters and enhancers. The loss of OBOX leads to impaired Pol II binding at ZGA gene promoters and enhancers, defective ZGA and aberrant Pol II retention in IC targets, accompanied by ectopic gene activation and 2C–4C arrest.

Fig. 10c,d). In total, 28 ZGA genes were commonly activated (10 minor and 18 major ZGA genes) in both *Obox*- and *Dux*-overexpressed mES cells (Fig. 5f). The majority of OBOX-activated ZGA genes were not activated by DUX in mES cells. OBOX-specifically-activated genes in mES cells were enriched for major ZGA genes (166 out of 174, 95.4%) whereas DUX preferentially activated minor ZGA genes (22 out of 37, 59.5%). Only 39 ZGA genes were commonly downregulated in *Obox* and *Dux* knockout^{19,20} embryos (Extended Data Fig. 10e), suggesting that OBOX

and DUX largely function in parallel. Whereas partial overlap was found between ‘2C genes’ in 2CLCs (2-cell-like cells)³⁹ and *Obox5*/3-activated genes (14 minor and 42 major ZGA genes), OBOX also activated a set of ZGA genes that were not enriched in 2CLCs (4 minor and 142 major ZGA genes) (Extended Data Fig. 10f). Of note, OBOX bound regions near *Zscan4a/d* and the expression of *Zscan4a/b/c/d/f* was downregulated in *Obox* mzKO E2C embryos (Extended Data Fig. 10a,b). *Dppa2*, *Dppa3* and *Dppa4* were bound by OBOX5 and OBOX3 in embryos and activated by *Obox5*/3 overexpression in mES cells (Extended Data Fig. 10a,b), suggesting that these may be downstream targets of OBOX. Finally, *Nr5a2* was also bound by OBOX1/5/3 in embryos and was downregulated in *Obox* mzKO embryos (Figs. 3c and 4a). On the other hand, *Nr5a2* knockdown²¹ did not affect *Obox* expression and OBOX could still activate ZGA genes in mES cells in the absence of *Nr5a2* (Fig. 5g and Extended Data Fig. 10g,h), raising the possibility that NR5A2 may function downstream of OBOX. Overall, ectopic expression of OBOX can directly activate ZGA genes and MERVL in mES cells.

Discussion

How mammalian ZGA is regulated remains poorly understood. In this study, we identified the OBOX family as critical regulators of mouse ZGA, in part by facilitating Pol II preconfiguration and chromatin opening preferentially at CG-poor promoters and enhancers. Depletion of OBOX compromised both mouse preimplantation development and ZGA, accompanied by ectopic gene activation of Pol II 1C targets (Fig. 5h). Intriguingly, such defects can be rescued by restoration of either combined maternal OBOX1/5/7 or zygotic OBOX3, suggesting redundancy among OBOX members. It is puzzling why *Obox* undergoes rapid evolution and frequent duplications in the genome. Given that gene families with multiple copies, such as *Dux* and *Zscan4*, are also linked to ZGA, we speculate that such redundancy may have evolved as a fail-safe mechanism to ensure the successful launch of ZGA. It remains to be further explored whether individual OBOX members execute specific functions.

Of note, a handful of PRD-like transcription factors in human independently arose from the same ancestor gene, *Crx*, that gave rise to *Obox* genes in rodents, although they share limited protein similarities with OBOX (13.4–28.7%)^{3,40,41}. We recently found that PRD-like members TPRXs regulate human ZGA and early development²⁴. However, how they function is unknown due to the inaccessibility of human embryos for molecular characterization. Our study now convincingly demonstrates the essential role of PRD-like transcription factors in mouse ZGA and early development with a knockout genetic model, thus illuminating the molecular circuitry underlying the fundamental question of how life begins. We envision that this work will also pave the way for understanding mammalian ZGA and PRD-like transcription factors in other species.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-023-06428-3>.

- Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic genome activation in vertebrates. *Dev. Cell* **42**, 316–332 (2017).
- Lee, M. T., Bonneau, A. R. & Giraldez, A. J. Zygotic genome activation during the maternal-to-zygotic transition. *Annu. Rev. Cell Dev. Biol.* **30**, 581–613 (2014).
- Wilming, L. G., Boychenko, V. & Harrow, J. L. Comprehensive comparative homeobox gene annotation in human and mouse. *Database (Oxford)* **2015**, bav091 (2015).
- 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).
- Royall, A. H., Maeso, I., Dunwell, T. L. & Holland, P. W. H. Mouse Obox and Crxos modulate preimplantation transcription profiles revealing similarity between paralogous mouse and human homeobox genes. *EvoDevo* **9**, 2 (2018).
- Aoki, F., Worrad, D. M. & Schultz, R. M. Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Dev. Biol.* **181**, 296–307 (1997).
- Bouniol, C., Nguyen, E. & Debey, P. Endogenous transcription occurs at the 1-cell stage in the mouse embryo. *Exp. Cell. Res.* **218**, 57–62 (1995).
- Schulz, K. N. & Harrison, M. M. Mechanisms regulating zygotic genome activation. *Nat. Rev. Genet.* **20**, 221–234 (2019).
- Liu, B. et al. The landscape of RNA Pol II binding reveals a stepwise transition during ZGA. *Nature* **587**, 139–144 (2020).
- Liang, H. L. et al. The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature* **456**, 400–403 (2008).
- Gaskell, M. M., Gibson, T. J., Larson, E. D. & Harrison, M. M. GAF is essential for zygotic genome activation and chromatin accessibility in the early *Drosophila* embryo. *eLife* **10**, e66668 (2021).
- Duan, J. et al. CLAMP and Zelda function together to promote *Drosophila* zygotic genome activation. *eLife* **10**, e69937 (2021).
- Lee, M. T. et al. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**, 360–364 (2013).
- Leichsenring, M., Maes, J., Mössner, R., Driever, W. & Onichtchouk, D. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* **341**, 1005–1009 (2013).
- Miao, L. et al. The landscape of pioneer factor activity reveals the mechanisms of chromatin reprogramming and genome activation. *Mol. Cell* **82**, 986–1002 (2022).
- Hendrickson, P. G. et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
- Whiddon, J. L., Langford, A. T., Wong, C. J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* **49**, 935–940 (2017).
- De Iaco, A. et al. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **49**, 941–945 (2017).
- Guo, M. et al. Precise temporal regulation of Dux is important for embryo development. *Cell Res.* **29**, 956–959 (2019).
- Chen, Z. & Zhang, Y. Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nat. Genet.* **51**, 947–951 (2019).
- Gassler, J. et al. Zygotic genome activation by the totipotency pioneer factor Nr5a2. *Science* **378**, 1305–1315 (2022).
- Lai, F. et al. NR5A2 connects genome activation to the first lineage segregation in early mouse development. Preprint at bioRxiv <https://doi.org/10.1101/2022.11.25.518012> (2022).
- Festuccia, N. et al. Nr5a2 is essential for morula development. Preprint at bioRxiv <https://doi.org/10.1101/2023.01.16.524255> (2022).
- Zou, Z. et al. Transcriptome and transcriptome co-profiling reveals a role of TPRXs in human zygotic genome activation. *Science* **378**, abo7923 (2022).
- Xiong, Z. et al. Ultrasensitive Ribo-seq reveals translational landscapes during mammalian oocyte-to-embryo transition and pre-implantation development. *Nat. Cell Biol.* **24**, 968–980 (2022).
- Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
- Berger, M. F. et al. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266–1276 (2008).
- Zhang, B. et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553–557 (2016).
- Dai, X. X. et al. A combinatorial code for mRNA 3'-UTR-mediated translational control in the mouse oocyte. *Nucleic Acids Res.* **47**, 328–340 (2019).
- Luong, X. G., Daldello, E. M., Rajkovic, G., Yang, C. R. & Conti, M. Genome-wide analysis reveals a switch in the translational program upon oocyte meiotic resumption. *Nucleic Acids Res.* **48**, 3257–3276 (2020).
- Zhang, C., Wang, M., Li, Y. & Zhang, Y. Profiling and functional characterization of maternal mRNA translation during mouse maternal-to-zygotic transition. *Sci. Adv.* **8**, eabj3967 (2022).
- Svoboda, P. et al. RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev. Biol.* **269**, 276–285 (2004).
- Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
- Sakashita, A. et al. Transcription of MERVL retrotransposons is required for preimplantation embryo development. *Nat. Genet.* **55**, 484–495 (2023).
- Chi, Y. I. Homeodomain revisited: a lesson from disease-causing mutations. *Hum. Genet.* **116**, 433–444 (2005).
- Katayama, S. et al. Phylogenetic and mutational analyses of human LEUTX, a homeobox gene implicated in embryogenesis. *Sci. Rep.* **8**, 17421 (2018).
- Fenouil, R. et al. CpG Islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res.* **22**, 2399–2408 (2012).
- Zheng, H. et al. Resetting epigenetic memory by reprogramming of histone modifications in mammals. *Mol. Cell* **63**, 1066–1079 (2016).
- Eckersley-Maslin, M. A. et al. MERVL/Zscan4 network activation results in transient genome-wide DNA demethylation of mESCs. *Cell Rep.* **17**, 179–192 (2016).
- Maeso, I. et al. Evolutionary origin and functional divergence of totipotent cell homeobox genes in eutherian mammals. *BMC Biol.* **14**, 45 (2016).
- Zhong, Y. F. & Holland, P. W. The dynamics of vertebrate homeobox gene evolution: gain and loss of genes in mouse and human lineages. *BMC Evol. Biol.* **11**, 169 (2011).
- Dahl, J. A. et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548–552 (2016).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023

Article

Methods

No statistical methods were used to predetermine sample size. Experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Animal maintenance

Wild-type C57BL/6 and ICR strain mice were purchased from Vital River and Tsinghua Animal Center. PWK/PhJ mice were originally purchased from Jackson Laboratory. Both wild-type and knockout mice were raised at Tsinghua Animal Center. Mice were maintained under specific-pathogen-free conditions with a 12/12 h light/dark cycle in an environment of 20–22 °C and 55 ± 10% humidity. All animals were maintained according to the guidelines of the Institutional Animal Care and Use Committee of Tsinghua University, Beijing, China.

Oocyte and early embryo collection

Full-grown oocytes (diameter over 70 µm) were collected from the ovaries of 4-week-old female C57BL/6 mice or 8-week-old ICR mice 46–48 h following injection of 5 IU pregnant mare serum gonadotropin (Ningbo Hormone Product Co., 110254564). For MII and embryo collection, C57BL/6 female mice were injected with pregnant mare serum gonadotropin followed 48 h later by 5 IU hCG (Ningbo Hormone Product Co., 110251283). For embryo collection, females were mated with PWK/PhJ or C57BL/6 males following hCG administration. Zygotes, E2C, mid-2C, mid-2C–L2C, L2C, 4C, 8C and blastocyst-stage embryos were collected at 20, 35, 40, 44, 46, 56, 68 and 100 h post hCG, respectively. Oocytes and embryos were collected in M2 medium (Sigma-Aldrich, M7167).

Parthenogenetic activation and embryo culture

Full-grown oocytes were collected from ICR strain mice and cultured to MII stage in an atmosphere of 5% CO₂ in air at 37.0 °C in Medium 199 (Gibco, 11150-059) supplemented with 10% (v/v) KSR (Gibco, A3181501), 0.1% bovine serum albumin (BSA, Sigma-Aldrich, A1933), 3.05 mM d-glucose (Sigma-Aldrich, G7012), 0.91 mM sodium pyruvate (Sigma-Aldrich, P4562), 0.05 IU ml⁻¹ FSH (Millipore, 869001), 0.05 IU ml⁻¹ LH (Millipore 869003), 20 ng ml⁻¹ EGF (Gibco, PHG0311), 100 µM cysteamine (Sigma-Aldrich, M9768 (freshly added)) and 200 µM cystine (Sigma-Aldrich, C7602 (freshly added)). MII eggs were activated for 6 h in calcium-ion-free Chatot-Ziomek-Bavister (Ca²⁺-free CZB) medium with 2.5 mM SrCl₂ and 2.5 µg ml⁻¹ cytochalasin B (Sigma-Aldrich, C6762). Ca²⁺-free CZB medium comprises the following: 85.35 mM NaCl, 4.83 mM KCl, 1.18 mM KH₂PO₄, 0.11 mM EDTA-2Na, 25.12 mM NaHCO₃, 0.27 mM sodium pyruvate (Gibco, 11360070), 1× GlutaMAX (Gibco, 35050061) and 5 mg ml⁻¹ BSA (Sigma-Aldrich, A1933). Parthenogenetic 1C embryos were then cultured in KSOM medium (Millipore, MR-121-D) to blastocyst stage.

Immunostaining

All steps were performed at room temperature. Mouse oocytes and embryos were fixed with 4% paraformaldehyde (Sigma-Aldrich, P6148) for 30 min and then permeabilized with 0.5% Triton X-100 in PBS for 30 min. Samples were blocked with 1% BSA for 1 h and incubated with primary antibodies (1:100 dilution for Flag antibody and 1:500 for all OBOX antibodies) for 1 h. OBOX antibodies were generated in house with peptides CERNLLKQESQGPSSR for OBOX1/2/3, NLQNIEQVLPES for OBOX1/5, EVLDQSKPYSHEEVC for OBOX3 and ASTQGPEYAQDS for OBOX6 (due to their similarities in protein sequences, some of the epitopes were present in more than one OBOX protein). The primary antibody was washed out with PBST (0.1% Triton X-100 in PBS) and samples were then incubated with secondary antibody and Hoechst 33342 for 30 min. Samples were then washed with PBST three times. All immunofluorescence images were recorded using a Zeiss LSM880 confocal microscope.

In vitro transcription and microinjection

For mRNA samples, pRK5 vectors containing a T3 promoter were linearized and transcribed with the T3 mMESSAGE Kit (Invitrogen, AM1348) following the manufacturer's instructions. mRNAs were recovered by RNA Clean XP beads (Beckman, A63987). For *Obox* mzKO rescue experiments, *Obox1*, *Obox5* and *Obox7* mRNA (100 ng µl⁻¹ each) were used for the combined *Obox1/5/7* rescue experiments, 100 ng µl⁻¹ *Obox3* mRNA for *Obox3* rescue and 5, 20 and 100 ng µl⁻¹ *Obox4* mRNA for *Obox4* rescue. For *Obox* stacc-seq, wild-type zygotes were injected with *Flag-Obox1*, *Flag-Obox5*, *Flag-Obox3*, *Flag-Obox2* or *Flag-Obox3^{R98E}* mRNA (500 ng µl⁻¹). For knockdown of individual maternal *Obox* genes, small interfering RNAs were injected into full-grown oocytes followed by in vitro maturation and parthenogenetic activation. Minor and major *Obox* genes were knocked down from the zygote stage. Note that *Obox1/2* were knocked down together because their mRNA sequences are highly similar (despite divergent protein sequences due to a frameshift mutation of *Obox2*). siRNA-targeting *Obox3* also partially reduced *Obox1/2/6/7* transcripts, again due to their sequence similarities. All injections were performed with an Eppendorf Transferman NK2 micromanipulator. Samples were injected at 5–10 pl per zygote or 2C embryo. For the knockdown experiment, 25 µM siRNA was used for each siRNA and non-targeting siRNA as a control. siRNA sequences are included in Supplementary Table 4.

Generation of *Obox* knockout mice

Obox knockout mice were generated by GemPharmatech. Cas9 mRNA (100 ng µl⁻¹) and single-guide RNA (50 ng µl⁻¹ each) were injected into the cytoplasm of zygotes. Following injection, zygotes were cultured in KSOM to the 2C stage at 37 °C under 5% CO₂ in air. Two-cell embryos were transferred into the oviducts of surrogate ICR strain mothers. *Obox* mutant mice were crossed with wild-type C57BL/6J mice for two generations before conducting the related experiments, to reduce the risk of off-target. To genotype colonies, a mouse tail tip was lysed in 70 µl of solution A (25 mmol l⁻¹ NaOH, 0.2 mmol l⁻¹ EDTA) at 95 °C for 50 min before cooling, followed by the addition of solution B (40 mmol l⁻¹ Tris-HCl). Supernatants were used as templates for PCR (wild-type and knockout alleles are 558 and 200 bp, respectively). The sgRNA sequences used in generation of *Obox* knockout mice and genotyping of primer sequence are provided in Supplementary Table 11.

Cell culture

Naive (2i) mES cells were cultured on feeder-free dishes coated with 0.1% gelatin in N2B27 medium supplemented with 1 µM PD0325901, 3 µM Chir99021 and 1 × 10³ units ml⁻¹ leukemia inhibitory factor (LIF). Cells were passaged at 1:10–1:20.

Plasmid construction and transfection

In regard to plasmids used for *Obox* overexpression in embryos or cell lines (2i mES cells or HEK293), *Obox* complementary DNA was cloned into piggyBac vector between 3× FLAG and P2A (self-cleaving peptide). Luciferase reporters were constructed with the pGL4.23 (Promega) plasmid as previously described³⁶ with minor modifications. The 4 × extended motif (113 bp sequence containing four of the extended de novo OBOX motif with 12 bp spacer between them), 4 × 7 bp motif (94 bp sequence containing four of the 7 bp motif with 12 bp spacer between them) or 4 × no-motif (61 bp sequence without motif sequence) was inserted between KpnI and XbaI. Sequences were generated through T4 polynucleotide kinase phosphorylation followed by primer annealing and ligation. The primers used for reporter plasmid construction are listed in Supplementary Table 12. Green fluorescent rotein (GFP) reporter plasmids were constructed by replacement of luciferase with GFP in luciferase reporter plasmids. For OBOX overexpression assays in mES cells, *Obox3* and *Obox5* plasmids were transiently overexpressed in 2i mES cells using Lipofectamine 3000

(Invitrogen, L3000015). GFP⁺ cells were selected following 24 h of transfection by flow cytometry (BD FACSAria II or Beckman MoFlo Astrios EQ). RNA sequencing (RNA-seq (Smart-seq)) was conducted for sorted cells with GFP-OBOX expression to measure gene expression. For the *Nr5a2* knockout mES cell line, four sgRNAs (Supplementary Table 13) were cloned into a pX330 plasmid (Addgene, 42230) and then cotransfected into mES cells with Lipofectamine 3000 (Thermo Fisher Scientific). Two to three days after transfection, cells were manually sorted into a gelatinized 96-well plate for single-clone selection. The obtained clones were genotyped by PCR and validated by Sanger sequencing.

Reporter assay in HEK293 and embryos

For the reporter assay in HEK293, firefly luciferase, renilla luciferase and *Obox* plasmids were transfected to HEK293 with Lipofectamine 3000 (Invitrogen, L3000015). Luciferase activity was measured at 16 h after transfection using the Dual-Luciferase Reporter Assay System (Promega) following the manufacturer's protocol with the following modification: 30 µl of lysis buffer was added to HEK293; samples were centrifuged and 8 µl of cell lysis supernatant was collected; 40 µl of firefly substrate and then 40 µl of stop buffer were added to cell lysis at the measuring step.

For the reporter assay in wild-type and *Obox* mzKO embryos, two rounds of microinjection were performed. *Obox1/5* mRNAs (100 ng µl⁻¹ each) were injected into zygotes and then 50 ng µl⁻¹ reporter plasmids (pGL4.23-GFP plasmid with or without 4 × extended OBOX motif) were injected into the nuclei of E2C embryos (because the OBOX motif is mainly associated with genes activated in 2C embryos and plasmids would diffuse away following mitosis and nuclear membrane breakdown if injected in 1C embryos). Embryos were cultured to the L2C stage for imaging analyses. For the reported 7 bp and extended OBOX motif comparison, reporter plasmids (pGL4.23-GFP plasmid with 4 × extended OBOX motif, 4 × 7 bp OBOX motif or without the OBOX motif) were injected into one of the nuclei of E2C embryos at a concentration of 200 ng µl⁻¹ and then imaged at the L2C stage.

RNA-seq library preparation and sequencing

All RNA-seq libraries were generated following the Smart-seq2 protocol as described previously⁴³. The zona pellucida was gently removed by treatment with Tyrode's solution (Sigma, T1788). Oocytes and embryos were washed three times in M2 medium and then lysed in 2 µl of lysis buffer containing RNase inhibitor.

Whole-genome sequencing

Tail-tip DNA was extracted using the isopropanol precipitation method. DNA libraries were generated via the Tn5-based method⁴³.

Stacc-seq library generation and sequencing

Stacc-seq libraries were constructed as previously described with minor modifications⁹. Embryo samples (total volume with buffer less than 1 µl) were prepared freshly into a 1.5 ml low-binding tube. The zona pellucida was removed with Tyrode's solution and the polar body removed with a sharp glass pipette.

For Pol II stacc-seq, DB1 buffer was prepared freshly (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 2% glycerol, 1× EDTA-free with Roche complete protease inhibitor, 0.01% digitonin and 2 mM DTT). For each sample, 2.5 µl (0.2 µg µl⁻¹) of anti-Pol II antibody (active motif 102660), 0.5 µl (1 µg µl⁻¹) of pG-Tn5 (Vazyme Biotech, TD901) and 9.5 µl of DB1 buffer were added to a 200 µl low-binding tube and the mixture was incubated at 4 °C for 30 min. DB1 buffer (37.5 µl) was added to embryos. The mixture was incubated for 10 min at 4 °C and vortexed gently every 2.5 min. Embryo samples, 12.5 µl of preincubated antibody-pG-Tn5 mixture and 12.5 µl of prewarmed (37 °C) 5× TTBL (Vazyme Biotech, TD502) were mixed and incubated in an Eppendorf Thermomixer at 37 °C for 30 min. Then, 2 µl of 10% SDS, 2 µl of carrier

RNA and 2 µl of spike-in DNA were added to the tube after complete mixing and incubation at room temperature for 5 min, followed by further incubation at 55 °C for 10 min. DNA was purified by 3x Ampure XP beads. PCR was performed to amplify libraries (Vazyme Biotech, TD601) using the following PCR conditions: 72 °C for 3 min; 98 °C for 30 s; thermocycling for 16 cycles at 98 °C for 15 s, 60 °C for 30 s and 72 °C for 3 min; and 72 °C for 5 min. Following the PCR reaction, libraries were purified by 0.4–1.7× AMPure beads size selection and subjected to next-generation sequencing.

For OBOX1, OBOX2 and OBOX3, DB1 buffer was freshly prepared (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 2% glycerol, 1× EDTA-free with Roche complete protease inhibitor and 0.02% digitonin). For each sample, 0.5 µl (1 µg µl⁻¹) of anti-Flag antibody (Sigma-Aldrich, F1804), 0.5 µl (1 µg µl⁻¹) of pG-Tn5 (Vazyme Biotech, TD901) and 11.5 µl of DB1 buffer were added to a 200 µl low-binding tube and the mixture was incubated at 4 °C for 30 min. Tgmentation, DNA purification and PCR steps were performed as for those used in Pol II profiling.

For OBOX5 and OBOX5^{R98F}, stacc-seq with washing was performed as previously described⁹. DB1 buffer was prepared freshly (as for OBOX1, OBOX2 and OBOX3 stacc-seq). For each sample, 0.5 µl (1 µg µl⁻¹) of anti-Flag antibody (Sigma-Aldrich, F1804), 0.5 µl (1 µg µl⁻¹) of pG-Tn5 (Vazyme Biotech, TD901) and 11.5 µl of DB1 buffer were added to a 200 µl low-binding tube and the mixture incubated at 4 °C for 30 min. For each sample, 10 µl of concanavalin A beads was washed twice in binding buffer (20 mM HEPES-KOH pH 7.5, 10 mM KCl, 1 mM CaCl₂, 1 mM MnCl₂) and resuspended in 10 µl of binding buffer. After collection of embryos in a 1.5 ml low-binding tube, 50 µl of buffer 1 (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 2% glycerol, 1× EDTA-free with Roche complete protease inhibitor) and 10 µl of washed concanavalin A beads were added and gently mixed. Following incubation at room temperature for 10 min, bead-bound embryos were washed once with 100 µl of DB1 buffer. Next, 37.5 µl of DB1 buffer and 12.5 µl of preincubated antibody-pG-Tn5 mixture were added to the sample. Following incubation at 4 °C for 2 h, the embryo sample with 12.5 µl of preincubated antibody-pG-Tn5 mixture was washed twice with 200 µl DB1 buffer. 12.5 µl prewarmed (37 °C) 5× TTBL (Vazyme Biotech, TD502) was then added and the sample was incubated in an Eppendorf Thermomixer at 37 °C for 30 min. DNA purification and PCR were then performed as for Pol II binding profiling.

ATAC-seq library preparation and sequencing

ATAC-seq libraries of wild-type and *Obox* mzKO embryos were prepared as previously described with minor modifications^{26,44}. Briefly, samples were lysed in 11 µl of lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl₂ and 0.05% digitonin) for 10 min at 4 °C. Samples were then incubated with 5 µl of Tn5 transposase and 4 µl of TTBL tgmentation buffer at 37 °C for 30 min (Vazyme Biotech). Following tgmentation, 2 µl of 10% SDS was added directly to the reaction to end tgmentation. Next, 2 µl each of carrier RNA and spike-in DNA were added and PCR was performed to amplify the library for 17 cycles using the following PCR conditions: 72 °C for 3 min; 98 °C for 30 s; thermocycling at 98 °C for 15 s, 60 °C for 30 s and 72 °C for 3 min; and 72 °C for 5 min. Following the PCR reaction, libraries were purified by 0.4–1.7× AMPure bead size selection and subjected to next-generation sequencing.

Data analyses

RNA-seq data processing. Paired-end RNA-seq reads were trimmed and then mapped to the mm9 genome by HISAT2 v.2.2.1 (ref. 45). StringTie v.2.1.2 (ref. 46) was used to calculate FPKM per gene based on mm9 refFlat from the UCSC genome annotation database⁴⁷. HTSeq v.0.6.0 (ref. 48) was applied to calculate counts per gene with default parameters. Trimmed RNA-seq data were also mapped to the reference *Obox* mRNA sequence by Magic-BLAST⁴⁹. Reads mapped to each *Obox* gene were counted and normalized by total reads and gene length for

Article

estimation of FPKM. *Obox* translation levels were calculated using StringTie v.2.1.2 based on Ribo-lite data²⁵.

DEG analysis. Differentially expressed genes (DEGs) were identified with adjusted $P < 0.05$ and $FC > 2$ by DESeq2 v.1.24.0 (ref. 50). Gene Ontology terms of DEGs were analysed by DAVID v.6.8 (ref. 51). FeatureCounts v.2.0.1 (ref. 52) was used to count reads mapped to annotated repeats (RepeatMasker). Differentially expressed repetitive elements were identified with adjusted $P < 0.05$ and $FC > 2$ by DESeq2 v.1.24.0, with total reads as size factors.

Identification of stage-specific genes, ZGA genes and maternal genes. Minor/major ZGA genes, maternal genes and stage-specific genes were defined based on the reference RNA-seq data using staged mouse embryos dissected *in vivo*²⁶. ZGA genes were defined as those not expressed, or lowly expressed, in full-grown oocytes and MII oocytes ($FPKM < 5$) but that had become upregulated ($FPKM > 5$, at least threefold upregulation) in either E1C or E2C embryos (minor ZGA genes) or L2C embryos (major ZGA genes). Genes expressed in oocytes ($FPKM > 5$) but highly upregulated at L2C (over fivefold upregulation) were also included in major ZGA genes ($n = 99$). No such genes exist for minor ZGA genes. Note that a small number of major ZGA genes was already moderately activated in our wild-type E2C samples, which were collected at a slightly later time point (35 h post hCG) compared with that described previously²⁸ (30 h post hCG) used to define the ZGA gene list ($n = 159$, E2C (hCG 35 h)/E2C (hCG 30 h) > 2). Among these genes, 87 were downregulated in E2C *Obox* mutants. Maternal genes were defined as those expressed in MII or full-grown oocytes ($FPKM > 5$) but that had become downregulated (at least threefold) at L2C.

Genes specifically activated at each stage during early development were defined using stricter criteria to ensure their stage specificity. These genes were activated at a defined stage ($FPKM > 5$) but remained silenced at all preceding stages from full-grown oocytes ($FPKM < 1$).

Stacc-seq and ATAC-seq data processing. Paired-end stacc-seq and ATAC-seq reads were aligned to the mm9 genome with the following parameters: -t -q -N1 -L 25 -X 2000 --no-mixed --no-discordant by Bowtie2 v.2.3.5 (ref. 53). Aligned reads were filtered with a minimum mapping quality of 20 and PCR duplicates were removed. Read coverages over the mm9 genome were estimated by bamCoverage from deepTools v.3.3.1 (ref. 54) with parameters --binSize 100 --normalizeUsing RPKM and visualized by UCSC browser⁵⁵. To minimize batch and cell type variation in comparisons, RPKM values of stacc-seq and ATAC-seq data were further normalized through z-score transformation.

Peak analyses. Peaks were called using MACS v.1.4.2 (ref. 56) with the parameters --nolambda --nomodel. Peaks in all heatmaps were sorted according to peak enrichment in each group. Promoters were defined as ± 2.5 kb around the TSS. Pol II peaks at least 2.5 kb away from TSS and excluded from the gene body were defined as distal peaks by BEDTools v.2.29.0 (ref. 57). Differential peaks were identified by those showing fold change (normalized RPKM + 0.5) > 2 .

OBOX-binding site feature annotation. Genomic distributions of OBOX stacc-seq peaks and randomly shuffled peaks were calculated by ChIPseeker v.1.20.0 (ref. 58). Stacc-seq peaks and randomly shuffled peaks were compared with annotated repeats (RepeatMasker) to estimate the enrichment of repetitive elements. The numbers of observed peaks overlapping with a certain type of repeat were compared with the average numbers of a set of random shuffled peaks (100 rounds) that overlapped with those repeats, and a log ratio (\log_2) was generated as 'observed/expected' enrichment.

Motif analyses. Motif analyses were carried out with HOMER v.4.11.1 (ref. 59). De novo motifs of OBOX stacc-seq peaks were identified by

findMotifsGenome.pl. The percentages of peaks or promoters containing motifs were estimated by overlapping them with genome-wide motif locations determined by scanMotifGenomeWide.pl based on position weight matrices of de novo or reported OBOX motifs. For distal ATAC-seq peaks and stage-specific Pol II peaks, findMotifsGenome.pl was applied to enrich known motifs. OBOX motif density heatmaps were created by annotatePeaks.pl.

OBOX protein sequence alignment. Sequence alignment of OBOX proteins was based on Clustal Omega⁶⁰. Pairwise correlation between sequences of aligned regions was calculated by RcpI v.1.30.0 (ref. 61), with BLOSUM62 as scoring function. Those regions with correlation above 0.8 were considered high homology regions. Homeobox locations were identified and confirmed with SMART⁶² and UniProt knowledgebase⁶³.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data are available within the article and its Supplementary Tables. All data have been deposited at GEO with accession no. GSE215813. Accession codes of the published data in GEO used in this study are as follows: RNA-seq of oocytes and early embryos and L2C H3K4me3, GSE71434; Ribo-lite data of oocytes and early embryos, GSE165782; RiboTag data of oocytes, GSE135525; total RNA-seq of oocytes and early embryos, GSE169632; 1C ATAC-seq, GSE169632; E2C and L2C ATAC-seq, GSE92605; Pol II stacc-seq of early embryos, GSE135457; L2C H3K27ac, GSE72784; RNA-seq of *Dux*-overexpressed and control mES cells, GSE85632; RNA-seq of *Dux* KO embryos, GSE121746 and GSE134832; RNA-seq of 2C-like cells and control mES cells, GSE75751; RNA-seq of *Nr5a2* knockdown and control 2C embryos, GSE178661. Source data are provided with this paper.

Code availability

Software used to analyse these data are listed in Methods and are all publicly available.

43. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
44. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
45. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
46. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
47. Karolchik, D. et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**, D493–D496 (2004).
48. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
49. Boratyn, G. M., Thierry-Mieg, J., Thierry-Mieg, D., Busby, B. & Madden, T. L. Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics* **20**, 405 (2019).
50. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
51. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
52. 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).
53. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
54. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
55. Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. *Brief. Bioinform.* **14**, 144–161 (2013).
56. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

57. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
58. 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).
59. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
60. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
61. Cao, D.-S., Xiao, N., Xu, Q.-S. & Chen, A. F. Rcp: R/Bioconductor package to generate various descriptors of proteins, compounds and their interactions. *Bioinformatics* **31**, 279–281 (2015).
62. Letunic, I. & Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **46**, D493–D496 (2018).
63. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **46**, 2699 (2018).

Acknowledgements We thank members of the Schultz and Xie laboratories for discussion and comments during the OBOX study and preparation of the manuscript, and the Animal Center and Biocomputing Facility at Tsinghua University for their support. We thank H. Wang for advice on embryo culture. This work was funded by the National Natural Science Foundation of China (grant 31988101 to W.X.), the National Key R&D Program of China (nos. 2021YFA1100102 and 2019YFA0508900 to W.X.), the National Natural Science Foundation of China (nos. 31830047 and 31725018 to W.X.) and the Tsinghua-Peking Center for Life Sciences (W.X.). This work was also supported in part by NIH grant HD022681 (R.M.S.) and the Intramural Research Program of the NIH, National Institutes of Environmental Health Sciences grant 1ZIAES102985 (C.J.W.). S.J. and F.C. are supported by postdoctoral fellowships from the Tsinghua-Peking

Center for Life Sciences. W.X. is a recipient of an HHMI International Research Scholar award and is a New Cornerstone Investigator.

Author contributions R.M.S., P.S. and W.X. conceived the project. C.J.W., R.M.S. and W.X. supervised the project. S.J., F.C., P.S., R.M.S. and W.X. designed the experiments. S.J. performed embryo experiments with help from L.W., Z.L. and B.H. F.C. performed bioinformatics analysis. S.J. and P.S. generated OBOX antibodies. P.S. conducted pioneering experiments for OBOX expression and function that laid the foundation for this study. L.W. and Z.L. performed microinjection. J.W., Z.Z., S.J. and B.L. performed OBOX and PolII stacc-seq and ATAC-seq, with help from L.W., F.L., F.K., Q.W. and Q.X. Q.Z., J.W. and Z.Z. constructed plasmids and prepared mRNAs. Z.L. and K.X. established the IVM-PA embryo culture condition. L.W. and Q.Z. performed mouse genotyping with help from Q.F. and L.L.. S.J., J.W., Z.Z., Q.Z. and Z.L. performed RNA-seq. L.W. and S.J. performed immunofluorescence with help from Q.Z. and T.K. Z.X. provided Ribo-seq data. X.H. generated the Nr5a2 knockout cell line. S.J., F.C. and W.X. prepared most of the figures and wrote the manuscript with help from all authors. Q.Z. and Z.L. are co-second authors who contributed equally.

Competing interests The authors declare no competing interests.

Additional information

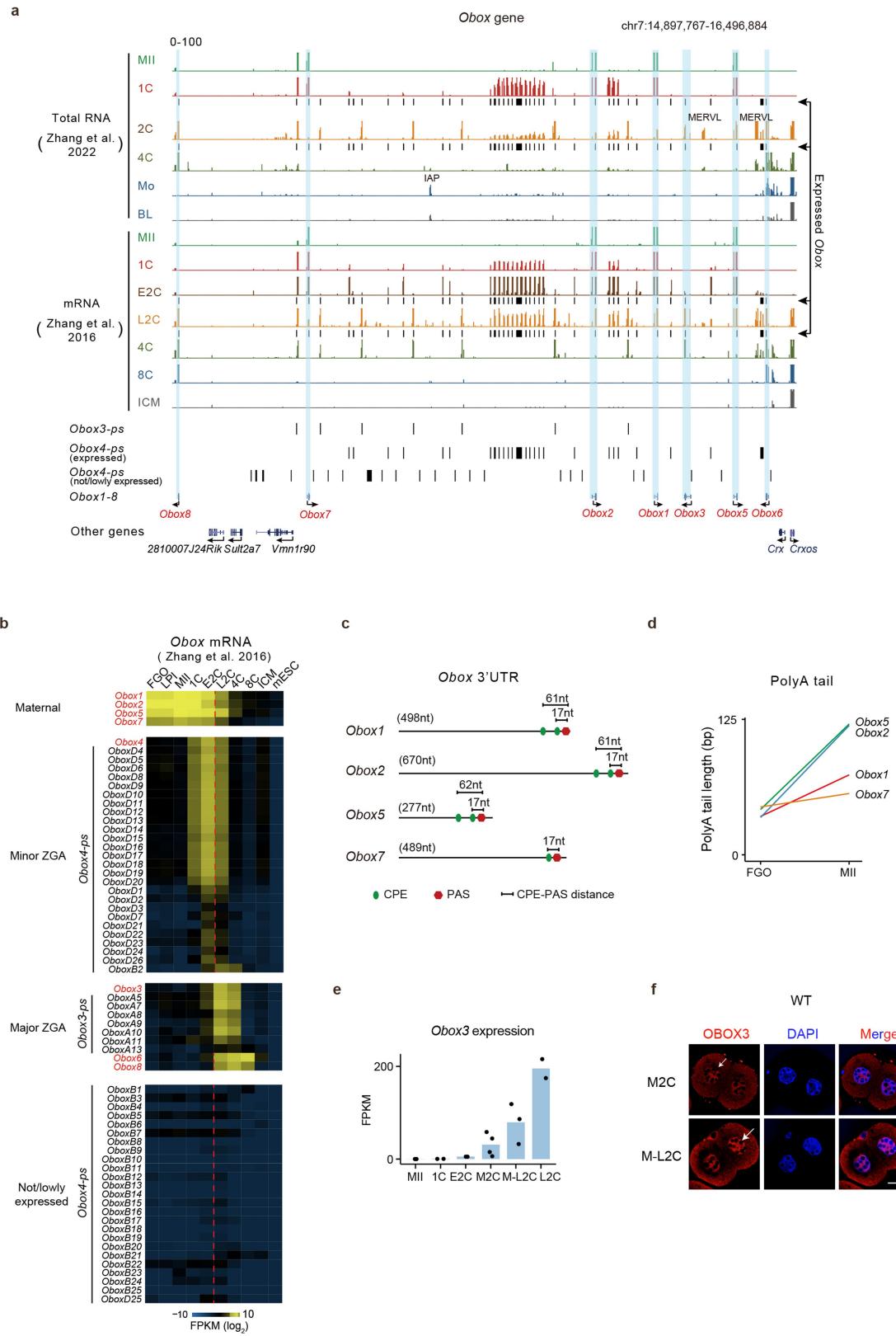
Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-023-06428-3>.

Correspondence and requests for materials should be addressed to Richard M. Schultz or Wei Xie.

Peer review information *Nature* thanks Joshua Brickman, Abdenour Soufi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

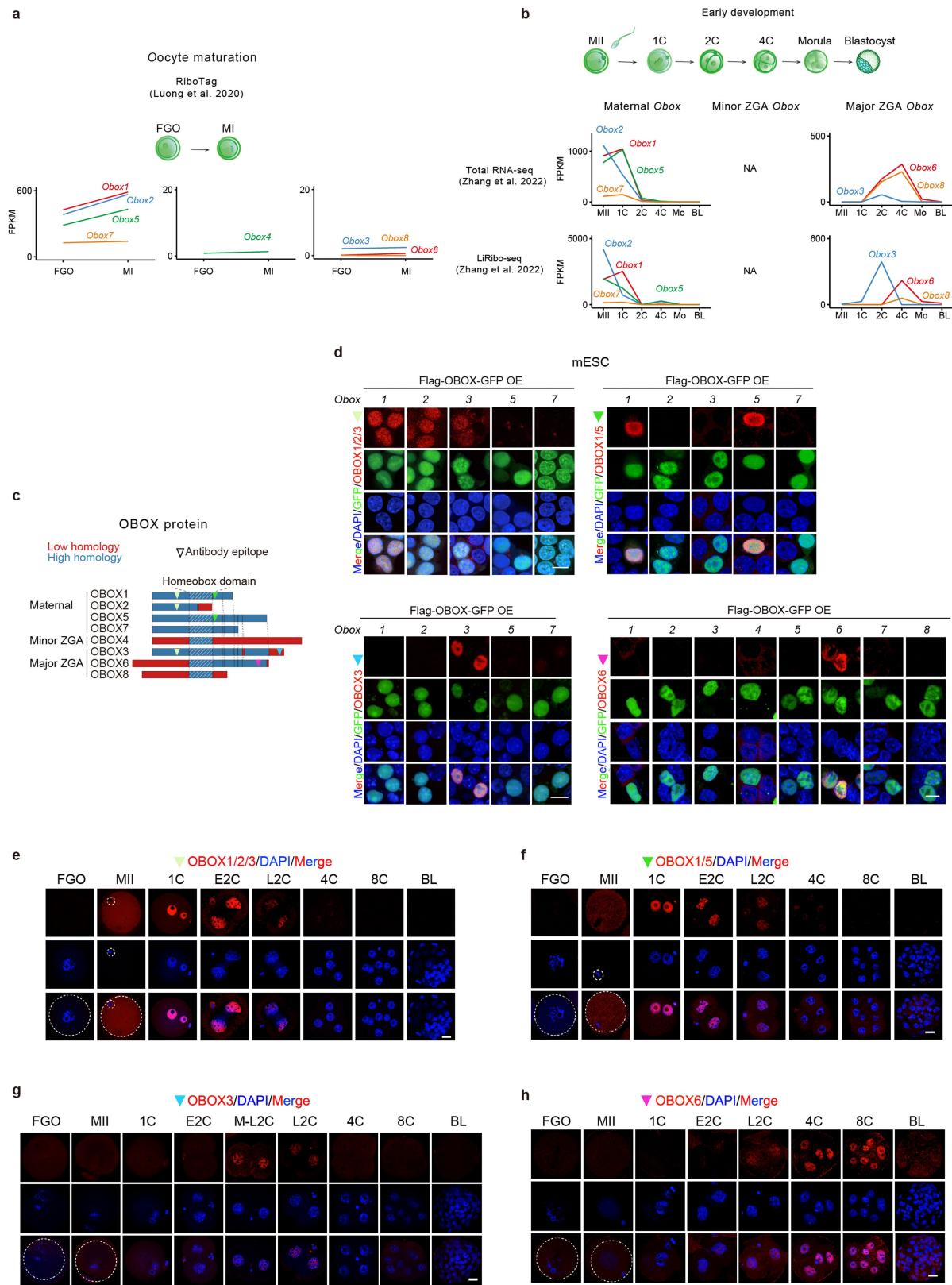
Reprints and permissions information is available at <http://www.nature.com/reprints>.

Article



Extended Data Fig. 1 | The location and expression of Obox genes. **a**, The UCSC genome browser snapshots showing *Obox* location and expression. **b**, Heatmap showing *Obox* mRNA levels in oocytes and embryos. **c**, CPE and PAS locations in maternal *Obox* 3'UTRs. **d**, Line plots showing poly(A) tail lengths²⁵ of maternal *Obox* during oocyte maturation. **e**, Bar chart showing *Obox*3 mRNA

levels in WT (2–4 biological replicates; 10 oocytes or embryos for each group). **f**, OBOX3 immunofluorescence in 2C embryos. M2C, mid-2-cell; M-L2C, mid-to-late 2-cell (3 biological replicates). Scale bar, 20 μ m. Arrow, nuclear OBOX3.

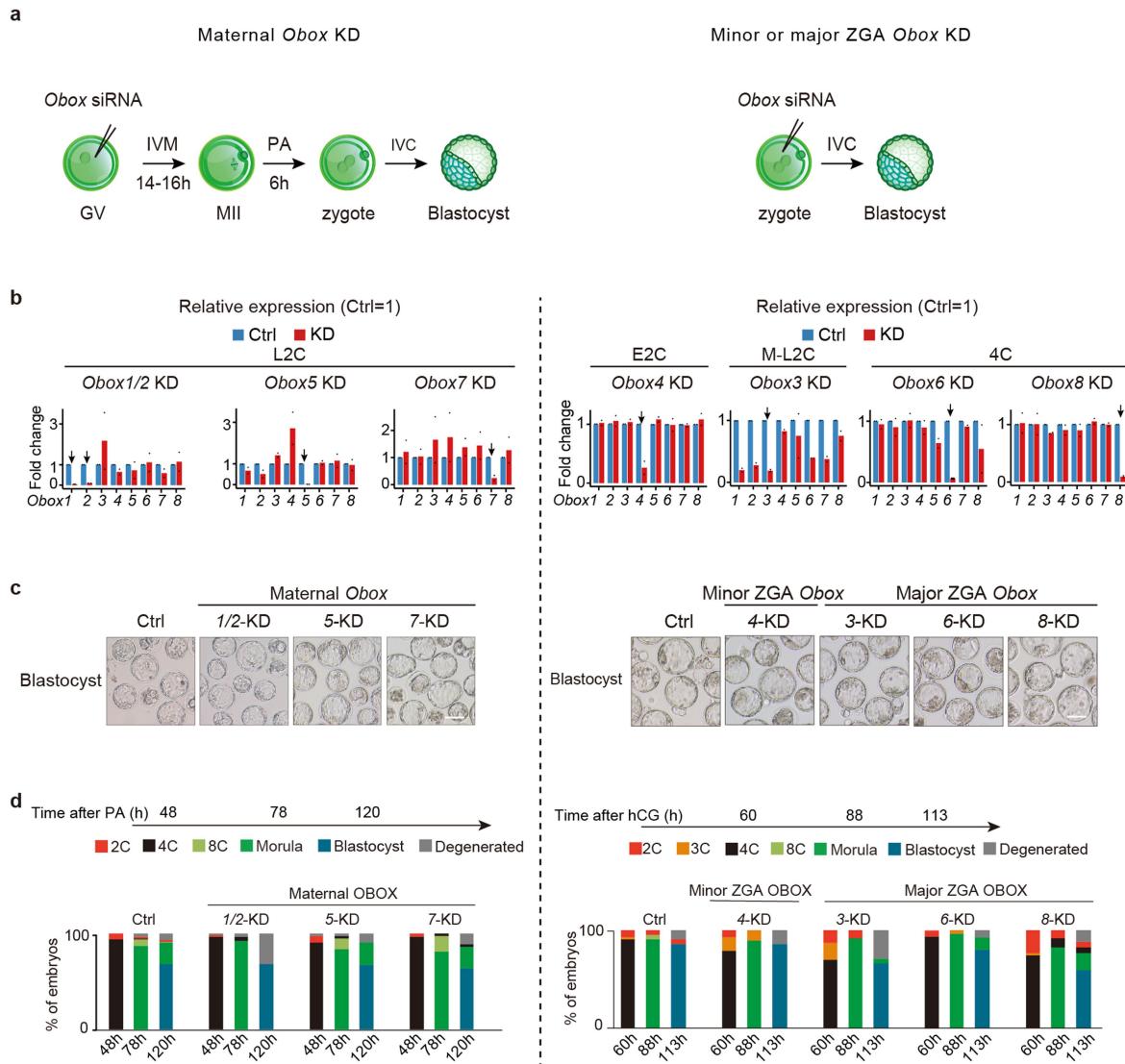


Extended Data Fig. 2 | OBOX protein levels in oocytes and early embryos.

a-b. Line plots showing *Obox* mRNA and translation levels during oocyte maturation (2 biological replicates) and early embryo development (2 biological replicates) based on datasets from the previous publications^{30,31}. NA, data not available. **c.** OBOX antibody epitope locations. **d.** Immunofluorescence showing

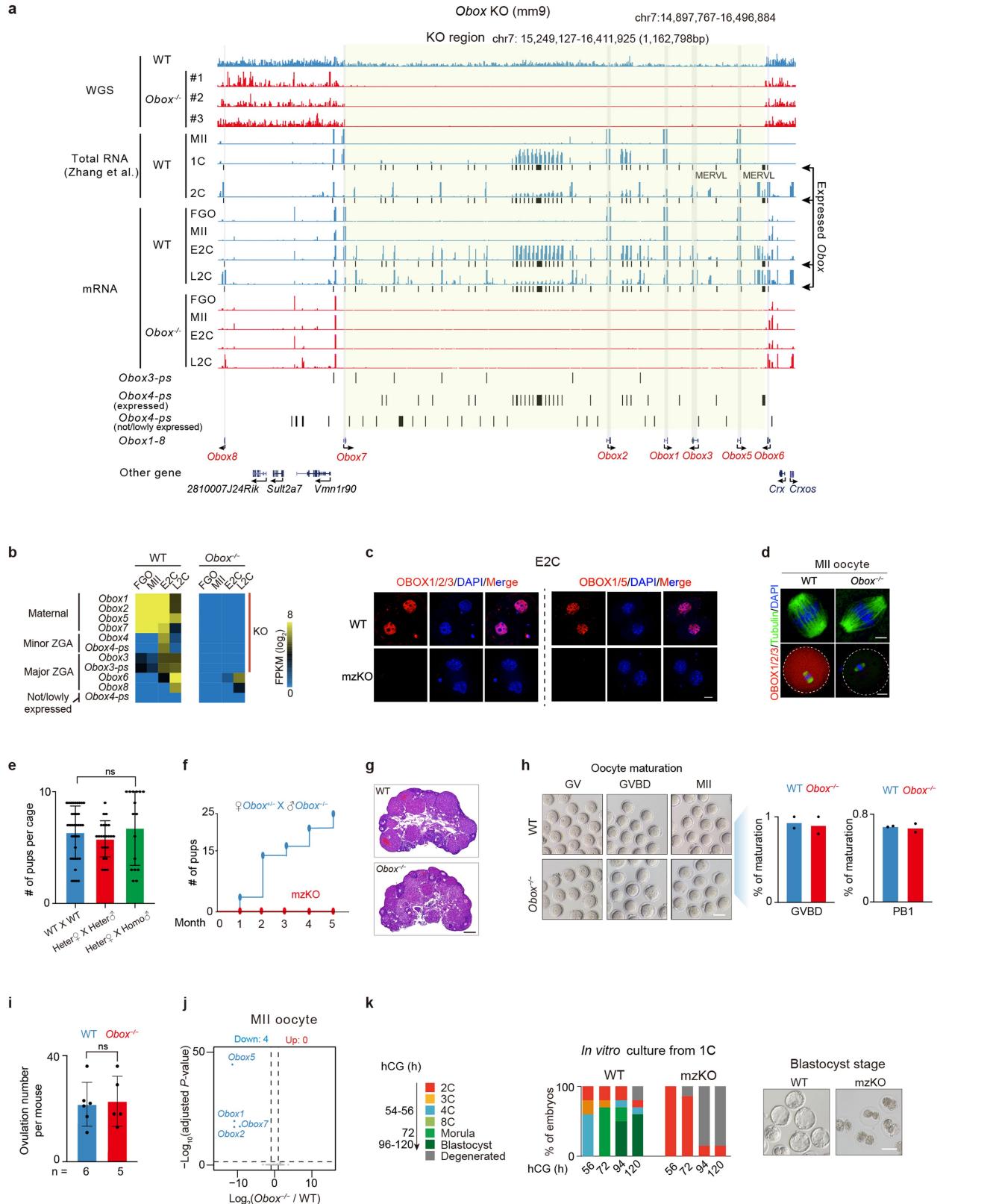
OBOX signals detected by OBOX antibodies upon Flag-OBOX-GFP overexpression in mESCs (2 biological replicates). Scale bar, 10 μm. e–h, OBOX immunofluorescence in mouse oocytes and embryos (3 biological replicates). BL, blastocyst. Scale bar, 20 μm.

Article



Extended Data Fig. 3 | Individual *Obox* knockdown had limited effects on preimplantation development. **a**, Schematic of individual *Obox* knockdown. IVM, in vitro maturation. PA, parthenogenetic activation. IVC, in vitro culture. h, hour. **b**, Bar chart showing the *Obox* knockdown efficiency in embryos (2 biological replicates; 10 embryos for each group). The control RNA levels

were normalized to 1. Arrow, targeted *Obox*. **c**, Embryo morphology upon individual *Obox* knockdown at the blastocyst stage (2 biological replicates). Scale bars, 100 μ m. **d**, Developmental rate upon individual *Obox* knockdown (2 biological replicates).



Extended Data Fig. 4 | See next page for caption.

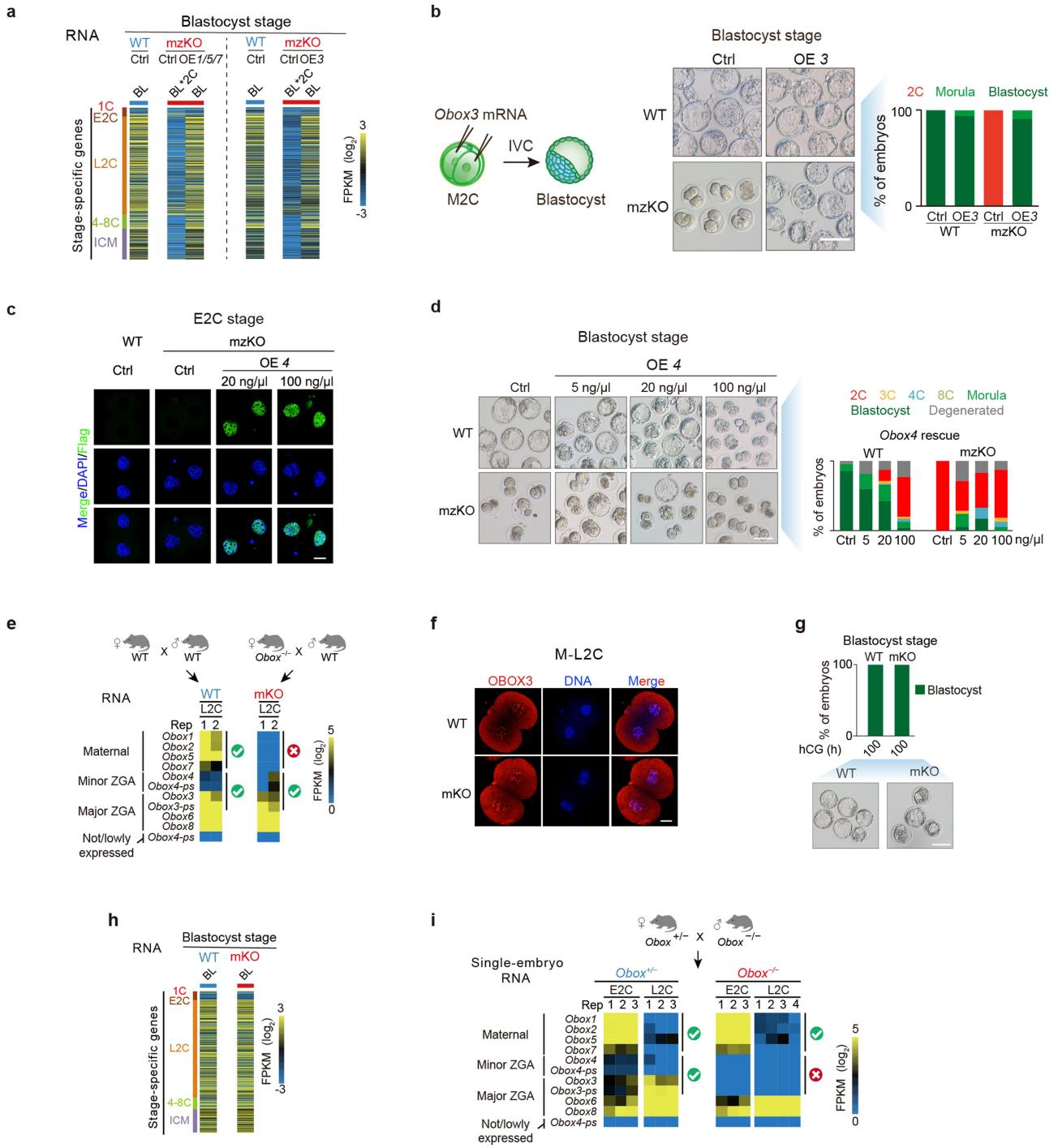
Article

Extended Data Fig. 4 | *Obox* depletion did not affect oocyte maturation.

a, Whole-genome sequencing (WGS) and RNA-seq showing *Obox* genes and expression. Yellow shade, the deleted *Obox*. #1/#2/#3, three *Obox* mzKO mice. **b**, RNA-seq showing *Obox* levels (2 or 3 biological replicates). KO, the knocked out *Obox* genes. **c**, OBOX staining in WT and *Obox* mzKO embryos (2 biological replicates). Scale bar, 20 μ m. **d**, Tubulin and OBOX staining in WT and *Obox*^{-/-} oocytes (3 biological replicates). Scale bars, 5 μ m (top) and 20 μ m (bottom). **e**, Bar chart showing offspring numbers with different crossing strategies. 37, 23, and 16 cages for WT \times WT, heterozygote \times heterozygote, and homozygote \times homozygote, respectively. ns, not significant (P -value = 0.69, two-sided t-test). Data are presented as mean values \pm SD. **f**, Fertility test of mzKO (four female

mice per group). **g**, HE staining (3 biological replicates). Scale bar, 0.25 mm.

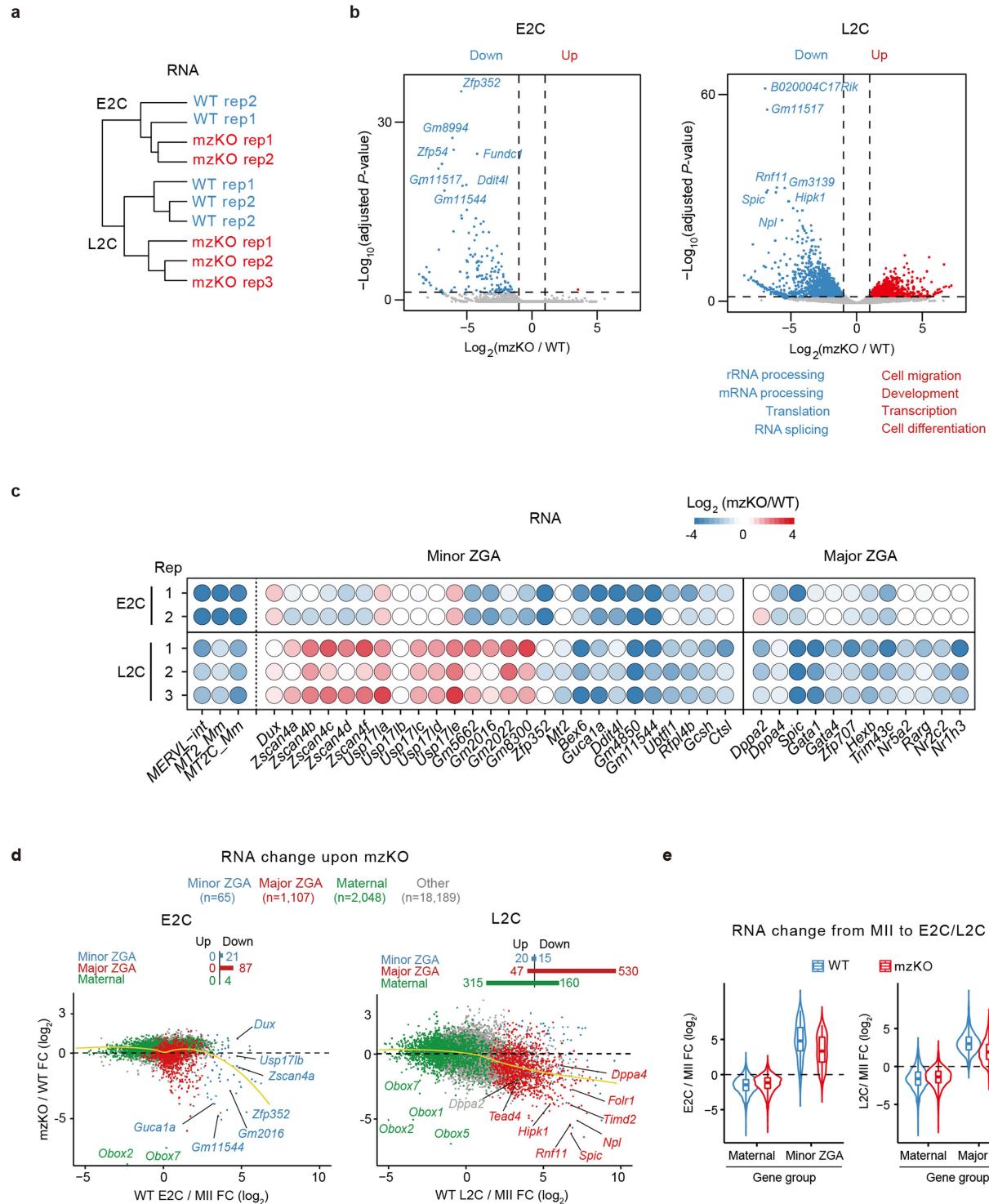
h, Bright-field images and bar charts showing oocyte morphology and maturation percentages upon OBOX depletion (2 biological replications). GVBD, germinal vesicle breakdown; PBL, the first polar body. Scale bar, 75 μ m. **i**, Bar chart showing the numbers of ovulated oocytes per mouse. n, number of mice used. P -value = 0.84, two-sided t-test. Data are presented as mean values \pm SD. **j**, Volcano plot showing gene expression changes between *Obox*^{-/-} and WT oocytes (2 biological replicates). Dashed line, adjusted P -value threshold 0.05. **k**, Embryo morphology and developmental rate *in vitro* (5 biological replicates). Scale bar, 75 μ m.



Extended Data Fig. 5 | Maternal and zygotic OBOX redundantly support early development. **a**, Expression of stage-specific genes in WT, *Obox* mutant, and rescued embryos. BL⁺2C, *Obox* mKO embryos arrested at 2C when WT developed to blastocyst. **b**, Schematic of OBOX3 rescue in *Obox* mKO embryos with embryo morphology and developmental rates shown (3 biological replicates). Scale bar, 100 μm. **c-d**, OBOX4 expression (c), embryo morphology, and developmental rate (d) with or without *Obox4* rescue (3 biological

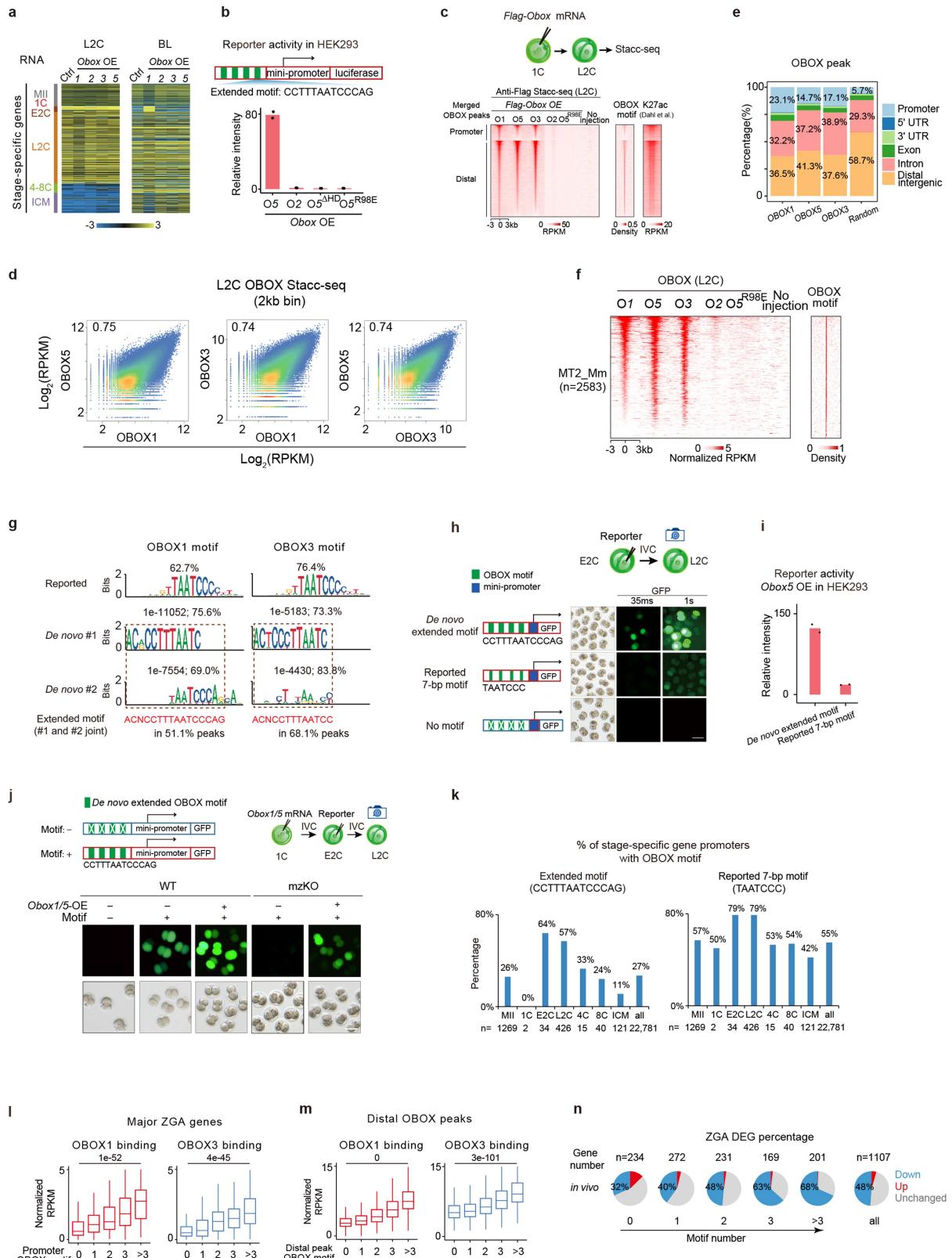
replicates). Scale bar, 75 μm. **e**, RNA-seq showing *Obox* levels in WT and maternal *Obox* knockout embryos. Check and cross, the presence or absence of *Obox* mRNAs. **f**, OBOX3 immunofluorescence in WT and *Obox* mKO embryos (3 biological replicates). Scale bar, 20 μm. **g-h**, Embryo morphology, developmental rate (g), and expression of stage-specific genes (h) for WT and *Obox* mKO embryos *in vivo* at the blastocyst stage (2 biological replicates). Scale bar, 100 μm. **i**, *Obox* expression levels in *Obox* mutant embryos.

Article



Extended Data Fig. 6 | *Obox* depletion impaired ZGA and MERVIL activation. **a**, Hierarchical clustering based on RNA-seq (2 biological replicates for E2C and 3 for L2C). **b**, Volcano plot showing gene expression changes upon *Obox* depletion (2 biological replicates for E2C and 3 for L2C). Dashed line, adjusted *P*-value threshold 0.05. GO terms are shown. **c**, Balloon plot showing gene expression changes (mzKO/WT) for MERVIL and ZGA genes at 2C (2 biological

replicates for E2C and 3 for L2C). **d**, Scatter plot showing gene expression fold-changes upon *Obox* depletion (2 biological replicates for E2C and 3 for L2C). FC, fold-change. Yellow lines, local regression fitting. **e**, Violin plot showing maternal and ZGA gene expression changes from oocytes to E2C or L2C in WT and *Obox* mzKO embryos (2 biological replicates for MII, E2C and 3 for L2C). Centre line, median; box, 25th and 75th percentiles; whiskers, 1.5 × IQR.

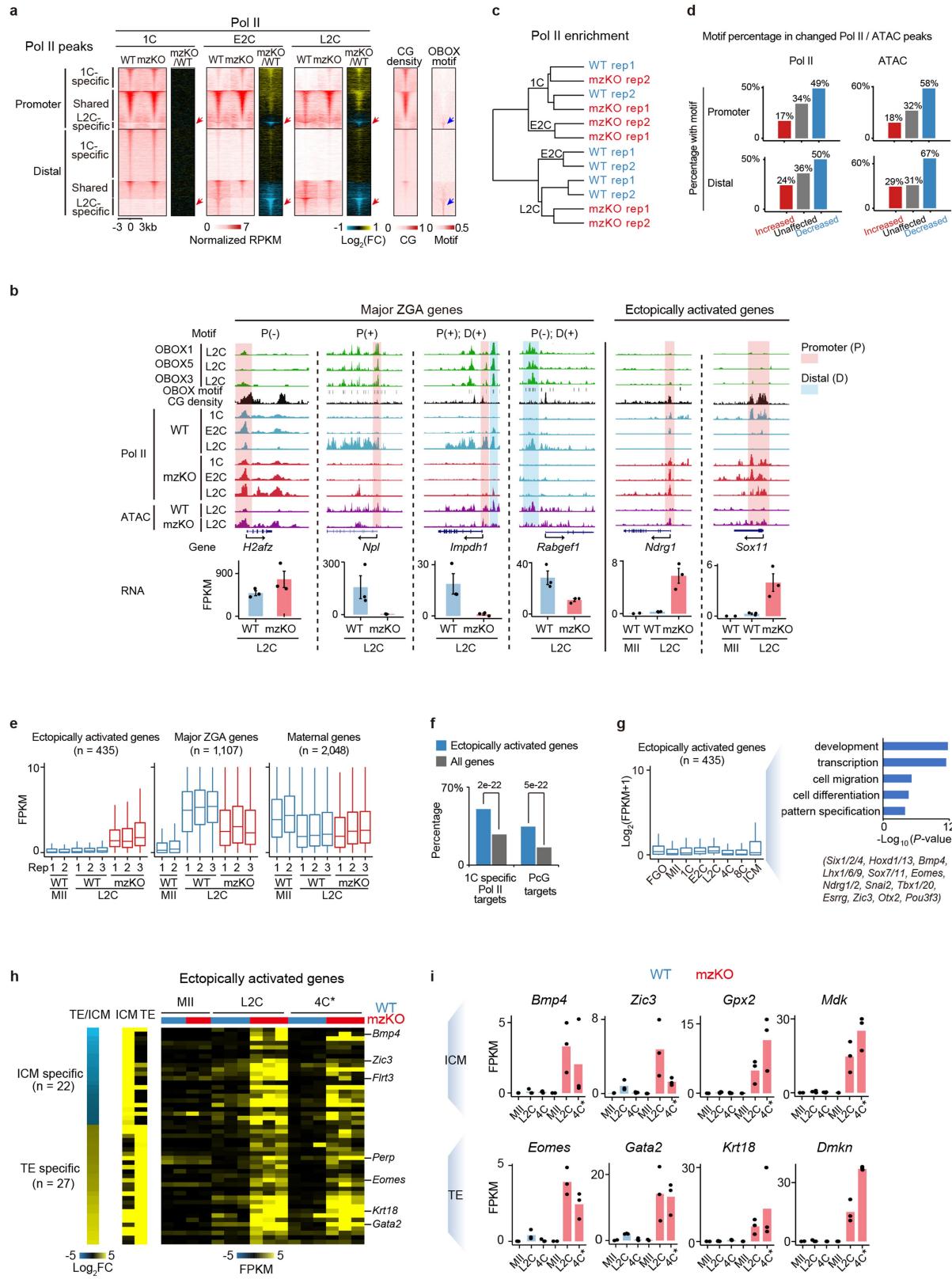


Extended Data Fig. 7 | See next page for caption.

Article

Extended Data Fig. 7 | OBOX binding in 2C embryos. **a**, Stage-specific gene expression upon *Obox* overexpression in WT embryos. **b**, Luciferase reporter assay showing OBOX gene activation abilities in HEK293 cells (2 biological replicates). ΔHD, homeobox domain deletion. **c**, Heatmap showing OBOX binding at L2C. OBOX motif densities and H3K27ac⁴² are shown. **d**, Scatter plot comparing OBOX binding at L2C. **e**, Bar chart showing the genomic distribution of OBOX binding at L2C. **f**, Heatmap showing OBOX binding on MERVL at L2C. OBOX motif is shown. **g**, Motif identified in OBOX binding sites in embryos. Percentages and *P*-values are shown. **h**, OBOX motif reporter assay in WT mouse embryos (2 biological replicates). Exposure time is shown. **i**, Luciferase reporter intensities in HEK293 cells (2 biological replicates). **j**, OBOX motif reporter assay in WT and *Obox* mzKO embryos (3 biological

replicates). + and –, presence and absence of *Obox1/5* mRNAs or extended motif, respectively. Scale bar, 75 μm. **k**, Bar chart showing OBOX motif occurrence at the stage-specific gene promoters. **l–m**, Box plots showing OBOX binding enrichment at major ZGA gene promoters (l) and distal regions (m) in WT L2C. 234, 272, 232, 169, and 201 genes have 0, 1, 2, 3, and >3 OBOX motifs on promoters, respectively. 9,855, 18,416, 7,142, 2,135, and 1,257 distal OBOX1 binding peaks have 0, 1, 2, 3, and >3 OBOX motifs, respectively. 5,918, 15,350, 4,795, 1,000, and 261 distal OBOX3 binding peaks have 0, 1, 2, 3, and >3 OBOX motifs, respectively. *P*-values, two-sided Wilcoxon rank-sum test. Centre line, median; box, 25th and 75th percentiles; whiskers, 1.5 × IQR. **n**, Percentages of ZGA genes that showed gene expression changes upon *Obox* depletion at L2C (3 biological replicates).

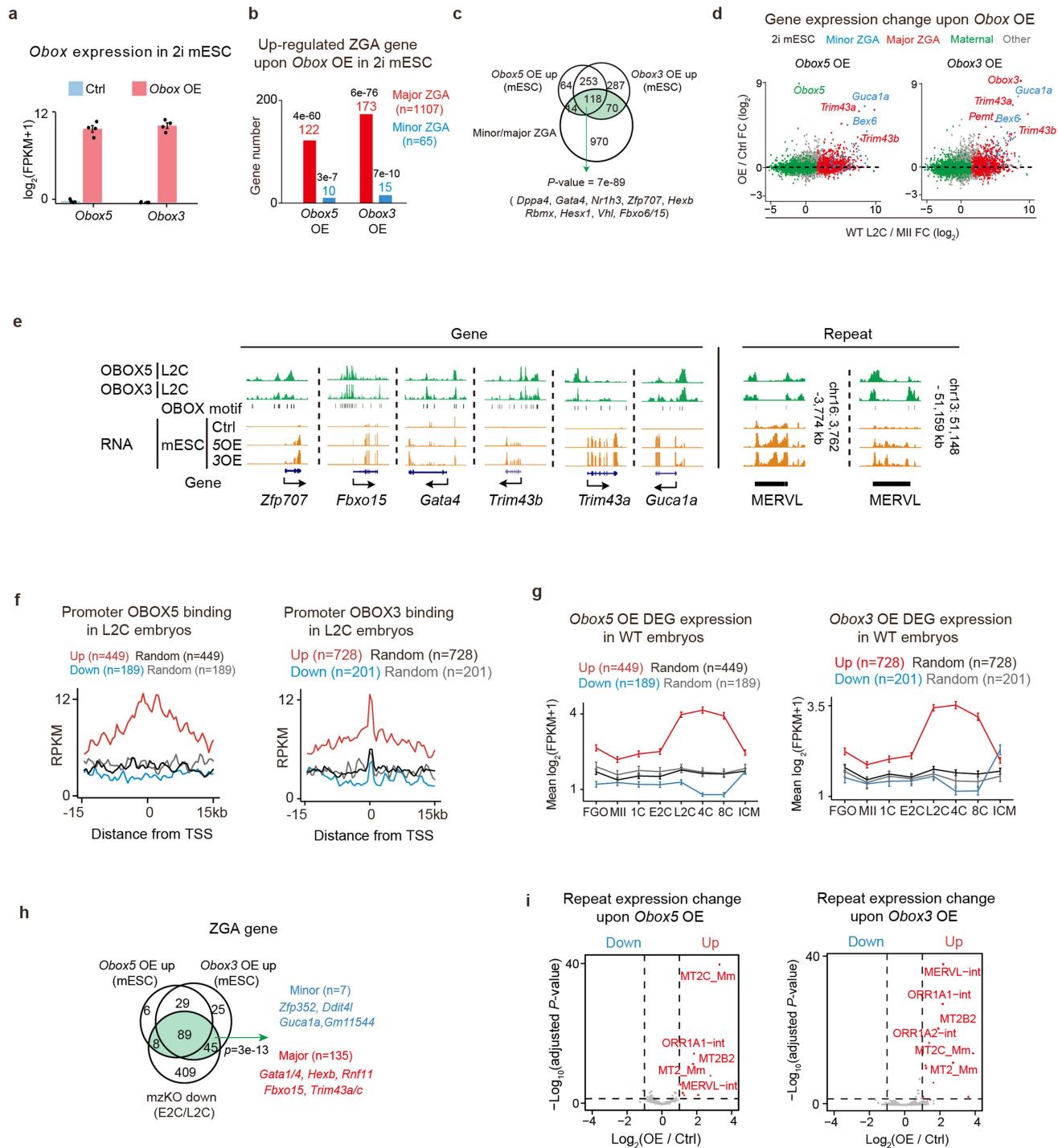


Extended Data Fig. 8 | See next page for caption.

Article

Extended Data Fig. 8 | Depletion of OBOX led to Pol II preconfiguration defects and ectopic activation of 1C Pol II targets. **a**, Pol II binding, CG density, and OBOX motif enrichment at 1C-specific, shared, and L2C-specific Pol II peaks in WT and *Obox* mzKO embryos. Red and blue arrows indicate L2C-specific Pol II binding and enrichment of the OBOX motif, respectively. **b**, Top, OBOX binding at example genes in WT embryos. OBOX motif and CG density are shown. Middle, Pol II binding and ATAC enrichment in WT and *Obox* mzKO embryos (2 biological replicates). P (+/-), promoter with or without the OBOX motif; D (+), distal enhancer with the OBOX motif. Bottom, bar charts showing gene expression (2 biological replicates for MII and 3 for L2C). Error bars, mean \pm SE. **c**, Hierarchical clustering based on Pol II Stacc-seq (2 biological replicates). **d**, Percentages of Pol II or ATAC peaks with OBOX motif at the promoters or distal regions at L2C. **e**, Box plot showing RNA levels of ectopically activated genes, major ZGA genes, and maternal genes. n, gene number. Centre

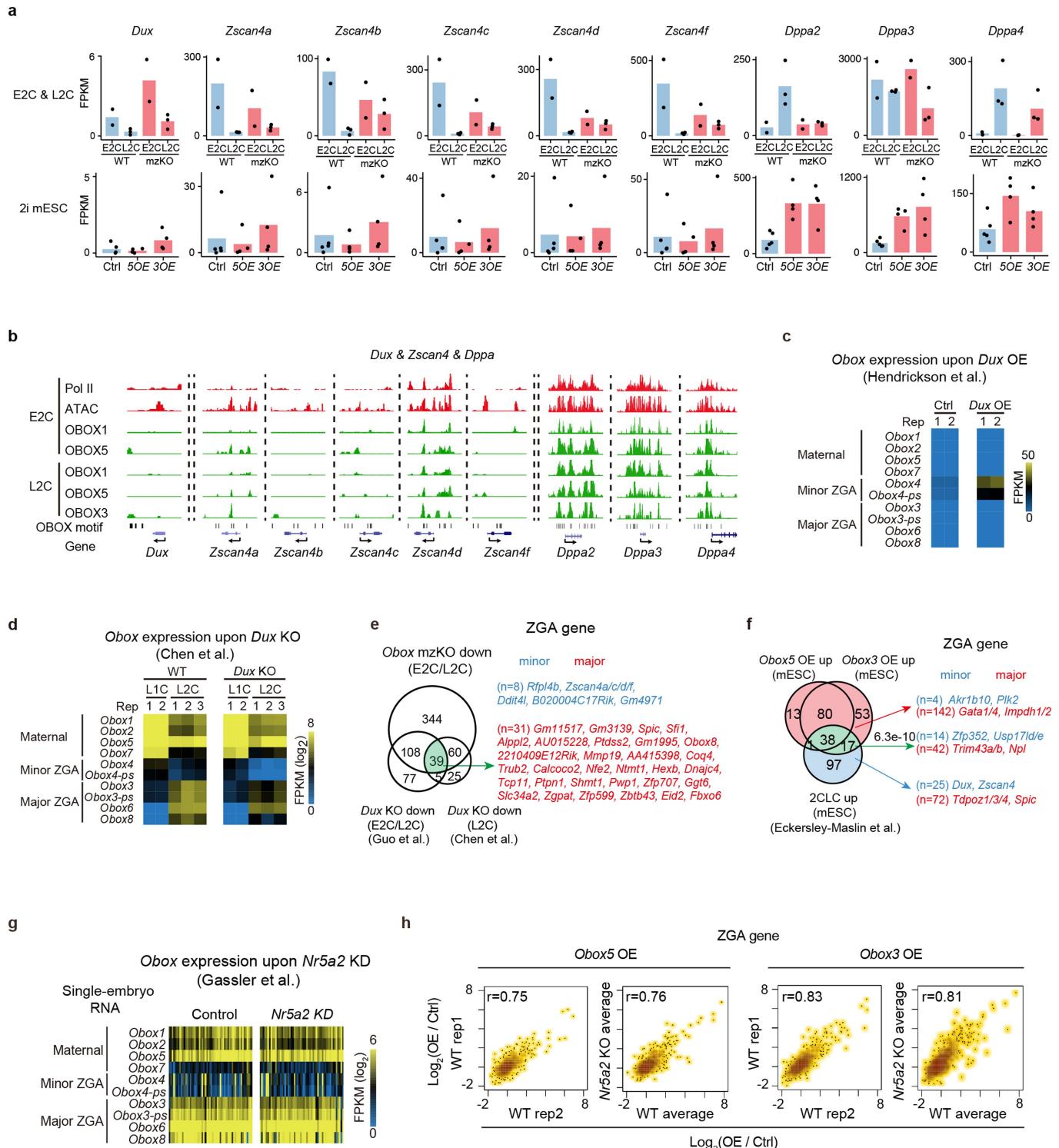
line, median; box, 25th and 75th percentiles; whiskers, $1.5 \times$ IQR. **f**, Percentages of ectopically activated genes or all genes (control) that are 1C-specific Pol II targets or Polycomb targets (PcG). P-values, two-sided Fisher's exact test. **g**, RNA levels in WT oocytes and embryos for ectopically activated genes (left). GO terms and example genes are shown (right). Centre line, median; box, 25th and 75th percentiles; whiskers, $1.5 \times$ IQR. **h**, Heatmap showing gene expression in ICM (inner cell mass), TE (trophectoderm), and the ratio of TE/ICM in WT embryos for ectopically activated ICM and TE genes in *Obox* knockout embryos. Gene expression for W and *Obox* mzKO MII oocytes (2 biological replicates) and embryos (3 biological replicates) is mapped. n indicates gene number. 4C*, the stage when WT developed to 4C and *Obox* mzKO embryos arrested at 2-4C. **i**, Bar chart showing gene expression of example ICM and TE genes from h.



Extended Data Fig. 9 | Obox overexpression activated ZGA genes and MERVL in 2i mESCs. **a**, Obox expression levels upon overexpression in 2i mESCs (4 biological replicates). Error bars, mean \pm SE. **b**, Bar chart showing the activated ZGA gene numbers upon Obox overexpression in 2i mESCs (4 biological replicates). P -values, two-sided Fisher's exact test. Green indicates the combined ZGA gene list activated by OBOX3/5. **c**, Scatter plot showing gene expression fold-changes upon Obox overexpression in 2i mESCs (4 biological replicates). **e**, OBOX binding at example OBOX-activated ZGA genes and MERVL in embryos. OBOX motif and RNA levels are shown. **f**, OBOX binding enrichment in embryos at the promoters of differentially expressed genes (DEGs) upon Obox overexpression in 2i mESCs. **g**, Line charts showing DEG upon Obox overexpression in 2i mESCs (4 biological replicates) for their expression in oocytes and embryos. Error bars, mean \pm SE. n, gene number. **h**, Venn diagram showing the overlap between Obox activated ZGA genes in 2i mESCs (4 biological replicates) and downregulated ZGA genes in Obox mZKO embryos (2 biological replicates for E2C and 3 for L2C). P -value, two-sided Fisher's exact test. Green indicates the combined ZGA gene list activated by OBOX3/5 and downregulated in Obox mZKO embryos. **i**, Volcano plot showing the repeat expression changes upon Obox overexpression in 2i mESCs (4 biological replicates). Dashed line, adjusted P -value threshold 0.05.

promoters of differentially expressed genes (DEGs) upon Obox overexpression in 2i mESCs. **g**, Line charts showing DEG upon Obox overexpression in 2i mESCs (4 biological replicates) for their expression in oocytes and embryos. Error bars, mean \pm SE. n, gene number. **h**, Venn diagram showing the overlap between Obox activated ZGA genes in 2i mESCs (4 biological replicates) and downregulated ZGA genes in Obox mZKO embryos (2 biological replicates for E2C and 3 for L2C). P -value, two-sided Fisher's exact test. Green indicates the combined ZGA gene list activated by OBOX3/5 and downregulated in Obox mZKO embryos. **i**, Volcano plot showing the repeat expression changes upon Obox overexpression in 2i mESCs (4 biological replicates). Dashed line, adjusted P -value threshold 0.05.

Article



Extended Data Fig. 10 | OBOX activated ZGA genes in mESCs independent of DUX and NR5A2. **a**, Bar charts showing *Dux*, *Zscan4*, and *Dppa* expression in 2C embryos (top, 2 biological replicates for E2C and 3 for L2C) and mESCs (bottom, 4–5 biological replicates). **b**, The UCSC browser snapshots showing OBOX binding at 2C. Pol II, ATAC, and OBOX motif are shown. **c**, Heatmap showing *Obox* expression upon *Dux* overexpression¹⁶ in 2i mESCs (2 biological replicates). **d**, Heatmap showing *Obox* expression upon *Dux* knockout²⁰ (2 biological replicates for late 1-cell (L1C) and 3 for L2C). **e**, Venn diagram

showing the overlap of downregulated ZGA genes between *Obox* knockout and *Dux* knockout embryos^{19,20}. **n**, ZGA gene number. **f**, Venn diagram showing the overlap of OBOX-activated ZGA genes and upregulated ZGA genes in 2CLCs compared to mESCs. 2CLC, 2-cell-like cell. *P*-value, two-sided Fisher's exact test. Green indicates the combined ZGA gene list activated by OBOX3/5 and in 2CLCs. **g**, Heatmap showing *Obox* expression upon *Nr5a2* knockout²¹ in embryos. **h**, Scatter plot comparing the ZGA gene expression changes upon *Obox* overexpression between WT and *Nr5a2* knockout mESCs (2 replicates).

Corresponding author(s): Wei Xie

Last updated by author(s): Apr 19, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No special software was used.
Data analysis	HISAT2 v2.2.1, StringTie v2.1.2, HTSeq v0.6.0, Magic-BLAST v1.5.0, DESeq2 v1.24.0, DAVID v6.8, FeatureCounts v2.0.1, Bowtie2 v2.3.5, samtools v1.3.1, deepTools v3.3.1, MACS v1.4.2, BEDTools v2.29.0, ChIPseeker v1.20.0, HOMER v4.11.1 (findMotifsGenome.pl, scanMotifGenomeWide.pl, annotatePeaks.pl) , Clustal Omega, Rcp1 v1.30.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The raw and processed data from this study are available in GEO with accession number GSE215813.

Published data used in this study are as follows: RNA-seq of oocytes and early embryos and late 2-cell H3K4me3, GSE71434; Ribo-lite data of oocytes and early

embryos, GSE165782; total RNA-seq of oocytes and early embryos, GSE169632; 1-cell ATAC-seq, GSE169632; early and late 2-cell ATAC-seq, GSE92605; Pol II Stacc-seq of early embryos, GSE135457; late 2-cell H3K27ac, GSE72784; RNA-seq of Dux overexpressed and control mESCs, GSE85632; RNA-seq of Dux KO embryos, GSE121746 and GSE134832; RNA-seq of 2C-like cells and control mESCs, GSE75751; RNA-seq of Nr5a2 knockdown and control 2-cell embryos, GSE178661.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	<input checked="" type="checkbox"/> none
Population characteristics	<input checked="" type="checkbox"/> none
Recruitment	<input checked="" type="checkbox"/> none
Ethics oversight	<input checked="" type="checkbox"/> none

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. Sample sizes for 2 to 3 biological replicates including all RNA-seq, Pol II Stacc-seq, and ATAC in this study were used according to common practice in the field. For OBOX1, OBOX5, and OBOX3 binding, the sample size is 1 as they share similar protein sequences and binding regions.
Data exclusions	There is no data that were excluded from the analyses.
Replication	Samples were collected in at least two replicates to confirm the consistency except for the OBOX1, OBOX5, and OBOX3 binding sharing similar targets. Immunofluorescence was performed 2 to 3 times, and similar observation was made for each replicate. One representative result was shown. Embryo development was performed 2 to 5 times. The average development rate was calculated with all replicates and one representative picture is shown. Replication of sequencing data was confirmed by calculating correlation and reproducibility between replicates, as shown in figure and figure legend.
Randomization	Animals were randomly divided into experiment group for OBOX binding Stacc-seq in wild-type embryos. Animals were not randomly divided into experiment group for OBOX mKO function experiments. Mice were genotyped two times and those with OBOX cluster deletion was divided into experimental groups.
Blinding	The investigators were not blinded to allocation during experiments and outcome assessment.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	OBOX antibodies were generated in cooperation with Abclone company (use a concentration of 1:500 for immunofluorescence). Anti-Pol II (active motif 102660, use 0.5ug for Stacc-seq) or anti-Flag antibody (Sigma-Aldrich, F1804, use 0.5ug for Stacc-seq)
Validation	OBOX antibodies (Abclone) are suitable for IF in mouse cells. Anti-Pol II (active motif, 102660) or anti-Flag antibody (Sigma-Aldrich, F1804) are suitable for IF and ChIP-seq in mouse cells;

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	mouse mES cells and HEK293 cells were from ATCC.
Authentication	Pluripotency and naïve markers were expressed in our mouse mES cells validated by smart-seq. Similarity RNA expression of our HEK293 comparing with reported HEK293 cell lines validated by smart-seq.
Mycoplasma contamination	All cell lines tested negative for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	None

Animals and other research organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	mus musculus strain: C57/BL6, male and female, 4 weeks, from Vital River and Tsinghua Animal Center; ICR, female, 8 weeks, from Vital River and Tsinghua Animal Center; PWK/PhJ, from Jackson Laboratory.
Wild animals	None
Reporting on sex	Oocytes and embryos were collected from female mice as only female mice have ovaries and could pregnant. We assign the sex based on whether the mice contained nipples.
Field-collected samples	None
Ethics oversight	All animal maintenance and experimental procedures used in this study were carried out according to guidelines and Institutional Animal Care and Use Committee (IACUC) of Tsinghua University, Beijing, China

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links	GEO accession GSE215813 at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE215813 <small>May remain private before publication.</small>
-------------------	---

Files in database submission	Fastq and bigwig files of ChIP-seq data. ATAC_L2C_Ctrl_rep1_r1.fq.gz ATAC_L2C_Ctrl_rep1_r2.fq.gz ATAC_L2C_mzKO_rep1_r1.fq.gz ATAC_L2C_mzKO_rep1_r2.fq.gz ATAC_L2C_Ctrl_rep2_r1.fq.gz ATAC_L2C_Ctrl_rep2_r2.fq.gz ATAC_L2C_mzKO_rep2_r1.fq.gz ATAC_L2C_mzKO_rep2_r2.fq.gz Pol2_1C_Ctrl_rep1_r1.fq.gz Pol2_1C_Ctrl_rep1_r2.fq.gz Pol2_1C_Ctrl_rep2_r1.fq.gz Pol2_1C_Ctrl_rep2_r2.fq.gz Pol2_1C_mzKO_rep1_r1.fq.gz Pol2_1C_mzKO_rep1_r2.fq.gz Pol2_1C_mzKO_rep2_r1.fq.gz
------------------------------	--

```

Pol2_1C_mzKO_rep2_r2.fq.gz
Pol2_E2C_Ctrl_rep1_r1.fq.gz
Pol2_E2C_Ctrl_rep1_r2.fq.gz
Pol2_E2C_Ctrl_rep2_r1.fq.gz
Pol2_E2C_Ctrl_rep2_r2.fq.gz
Pol2_E2C_mzKO_rep1_r1.fq.gz
Pol2_E2C_mzKO_rep1_r2.fq.gz
Pol2_E2C_mzKO_rep2_r1.fq.gz
Pol2_E2C_mzKO_rep2_r2.fq.gz
Pol2_L2C_Ctrl_rep1_r1.fq.gz
Pol2_L2C_Ctrl_rep1_r2.fq.gz
Pol2_L2C_Ctrl_rep2_r1.fq.gz
Pol2_L2C_Ctrl_rep2_r2.fq.gz
Pol2_L2C_mzKO_rep1_r1.fq.gz
Pol2_L2C_mzKO_rep1_r2.fq.gz
Pol2_L2C_mzKO_rep2_r1.fq.gz
Pol2_L2C_mzKO_rep2_r2.fq.gz
Stacc_Flag_E2C_Obox1OE_r1.fq.gz
Stacc_Flag_E2C_Obox1OE_r2.fq.gz
Stacc_Flag_E2C_Obox5OE_r1.fq.gz
Stacc_Flag_E2C_Obox5OE_r2.fq.gz
Stacc_Flag_L2C_NT_r1.fq.gz
Stacc_Flag_L2C_NT_r2.fq.gz
Stacc_Flag_L2C_Obox1OE_r1.fq.gz
Stacc_Flag_L2C_Obox1OE_r2.fq.gz
Stacc_Flag_L2C_Obox2OE_r1.fq.gz
Stacc_Flag_L2C_Obox2OE_r2.fq.gz
Stacc_Flag_L2C_Obox3OE_r1.fq.gz
Stacc_Flag_L2C_Obox3OE_r2.fq.gz
Stacc_Flag_L2C_Obox5OE_r1.fq.gz
Stacc_Flag_L2C_Obox5OE_r2.fq.gz
Stacc_Flag_L2C_Obox5R98EOE_r1.fq.gz
Stacc_Flag_L2C_Obox5R98EOE_r2.fq.gz
ATAC_L2C_Ctrl_rep1.bw
ATAC_L2C_mzKO_rep1.bw
ATAC_L2C_Ctrl_rep2.bw
ATAC_L2C_mzKO_rep2.bw
Pol2_1C_Ctrl_rep1.bw
Pol2_1C_Ctrl_rep2.bw
Pol2_1C_mzKO_rep1.bw
Pol2_1C_mzKO_rep2.bw
Pol2_E2C_Ctrl_rep1.bw
Pol2_E2C_Ctrl_rep2.bw
Pol2_E2C_mzKO_rep1.bw
Pol2_E2C_mzKO_rep2.bw
Pol2_L2C_Ctrl_rep1.bw
Pol2_L2C_Ctrl_rep2.bw
Pol2_L2C_mzKO_rep1.bw
Pol2_L2C_mzKO_rep2.bw
Stacc_Flag_E2C_Obox1OE.bw
Stacc_Flag_E2C_Obox5OE.bw
Stacc_Flag_L2C_NT.bw
Stacc_Flag_L2C_Obox1OE.bw
Stacc_Flag_L2C_Obox2OE.bw
Stacc_Flag_L2C_Obox3OE.bw
Stacc_Flag_L2C_Obox5OE.bw
Stacc_Flag_L2C_Obox5R98EOE.bw

```

Genome browser session
(e.g. [UCSC](#))

https://genome.ucsc.edu/s/ChenFengling/OBOX_Project

Methodology

Replicates

2 replicates for Pol II Stacc-seq in WT and mzKO 1C, E2C and L2C embryos.

Sequencing depth

Varies in different Stacc-seq samples and can be checked at GEO accession GSE215813

Antibodies

Pol II (Active Motif, 102660)
Flag (Sigma-Aldrich, F1804)

Peak calling parameters

MACS v1.4.2 with the parameters nolambda –nomodel

Data quality

Reads with a Phred quality score of <20 were removed. Non-unique reads were removed by SAMtools. Quality were accessed by deeptools and UCSC Genome Browser.

