# Supplementary information

## Material and Methods

### Animals

Specific-pathogen-free (SPF) mice were housed in the animal facility at Tongji University, Shanghai, China. All animal maintenance and experimental procedures were performed according to Tongji University Guide for the use of laboratory animals.

### Creation of *Dux* KO mice

C57BL/6 female mice (6-8 weeks old) were super-ovulated by injection with 6 IU of pregnant mare serum gonadotropin (PMSG), followed by injection of 8 IU of human chorionic gonadotropin (hCG) (San-Sheng Pharmaceutical Co., Ltd.) 48 h later. The super-ovulated female mice were mated with C57BL/6 male mice and zygotes were collected 20 h later. Cas9 mRNA (100 ng/μL) and sgRNAs (50 ng/μL each) were injected into zygotes by a Piezo-driven micromanipulator and then cultured in G1 medium with amino acids at 37 °C in humidified atmosphere of 5% $CO_2$. At ~28 h later, 2-cell embryos were transferred into oviducts of pseudo-pregnant ICR mothers. mMESSAGE mMACHINE T7 Ultra Kit (Ambion, Thermo Fisher Scientific, AM1345) and MEGA shortscrip T7 Transcription Kit (Ambion, Thermo Fisher Scientific, AM1354) were used for the synthesis of Cas9 mRNA and sgRNAs, respectively. All Cas9 mRNA and sgRNAs were purified according to the standard protocol by phenol:chloroform extraction and ethanol precipitation, and then dissolved in

DNase/RNase-free water. For genotyping, tail tips from mice were lysed in SENT lysis buffer (1% SDS, 5 mM EDTA, 0.4 M NaCl, 20 mM Tris-HCl, 400 μg/mL Proteinase K) and genomic DNA was extracted as template for PCR. The purified PCR products with large fragment deletion from Dux-KO mice were cloned into pLB vectors using Lethal Based Fast Cloning Kit (Tiangen) to sequence. SgRNA and primer sequences are listed in Supplementary information, Table S9.

**Third-generation long-read whole-genome sequencing and analyses**

High quality genome DNA was extracted from liver and muscles of Dux-KO mouse according to the standard protocol by phenol:chloroform extraction and ethanol precipitation. Genome DNA library was prepared for sequencing using standard PacBio Sequel protocol. Long-read (30 kb) sequencing was performed on PacBio Sequel II system following the manufacturer's instructions at Zhejiang Annoroad Biotechnology Co., Ltd. PacBio reads were aligned to the mouse genome (mm9) using pbmm2 (version 1.0.0), a wrapper of minimap2 with default settings. We used variantCaller (version 2.3.3) to get the consensus and used the algorithm "arrow" to get variant calling. We calculated the read counts of the repeat class using the analyzeRepeats.pl script of HOMER with the option of 'repeat' and '-noadj'. Total reads count of each sample was normalized to 1 million and replicates were averaged for comparison.

**Plasmid construction and mRNA microinjection in embryos**

Dux altered codon sequence was amplified from the plasmid pCW57.1-mDux-CA

(Addgene #99284), and EGFP and RFP sequences were amplified from pSico (Addgene #11578) and pSicoR (Addgene #11579), respectively. EGFP sequence flanked by restriction enzyme cutting sites of *Eco*RI and *Sal*I was inserted into plasmid pCW57.1 (Addgene #41393), resulting in complete restriction enzyme cutting sites of *Nhe*I and *Eco*RI between tight TRE promoter and EGFP. Stop codon removed Dux altered codon sequence then was cloned into pCW57.1-EGFP by *Nhe*I and *EcoR*I sites, resulting in a Dux-EGFP fusion recombinant vector pCW57.1-Dux-EGFP. mRNAs of EGFP, RFP and Dux-EGFP were expressed using mMESSAGE mMACHINE T7 Ultra Kit (Ambion, Thermo Fisher Scientific, AM1345). To overexpress Dux in embryos, 260 ng/μL mRNA of Dux-EGFP mixed with 160 ng/μL RFR mRNA and 260 ng/μL mRNA of EGFP mixed with 160 ng/μL RFP mRNA as control, were microinjected into zygotes (hCG 23 h) , late 2-cell（hCG 48 h）or 4-cell (hCG 56 h) embryos from B6D2F1 mice.

**scRNA-seq**

RNA-seq libraries were generated from individual WT and Dux-KO (maternal and zygotic KO) embryos including zygotes (hCG 28 h), early 2-cell (hCG 31-32 h), middle 2-cell (hCG 41-42 h), late 2-cell (hCG 47-48 h) and mRNA-injected embryos including zygote (5 h after injection, hCG 28 h), early 4-cell (6 h after injection, hCG 54 h) and late 4-cell (17 h after injection, hCG 65 h) using the Covaris DNA shearing protocol for Smart-seq sequence library generation as previously described[1]. Briefly, RNAs with a polyadenylated tail were captured, reverse transcribed and pre-amplified. After

fragmentation, the sequence libraries were generated using the KAPA Hyper Prep Kit for the Illumina platform, following the manufacturer's instructions. Paired-end 150-bp sequencing was performed on NovaSeq (Illumina) platform at the Berry Genomics Co., Ltd.

**RNA-seq analyses**

RNA-seq raw reads were processed by BBDuk (version 38.34) to remove adapters and low-quality reads. For gene analysis, reads (> 35 bp) were then aligned to the mouse genome (mm9) using STAR (version 0.6.0)[2] with the parameters "--outFilterScoreMinOverLread 0.4 --outFilterMatchNminOverLread 0.4". Expression levels for all Refseq genes were quantified to FPKM using Stringtie (version 1.3.3b) [3]. To perform differential gene expression analysis, we first calculated the read counts of each RNA-seq sample using the script prepDE.py of Stringtie and Deseq2[4] was used to perform differential analysis. Genes with a Benjamini-Hochberg adjusted $P \leq 0.05$ and a fold change > 2 were considered as differentially expressed. For the repeat analyses, RNA-seq reads were re-mapped to the mm9 genome using STAR, allowing at most three mismatches and discarding reads that map to more than 500 positions in the genome. The makeTagDirectory algorithm with the -keepOne option of HOMER[5] were applied to get the tag directories of the mapped sam files. We calculated the read counts and expression level of the repeat class using the analyzeRepeats.pl script of HOMER with the options of '-noadj' and 'FPKM', respectively. The getDiffExpression.pl script of HOMER was applied to identify differentially expressed repeat class. Repeat class

with Benjamini-Hochberg adjusted $P \leq 0.05$ and a fold change > 2 were defined as differentially expressed repeat class. Batch effects of samples from removed using 'removeBatchEffect' function of the R library 'edgeR'. The negative values in the normalized data are set to zero.

**Quantitative PCR**

1 μL scRNA-seq preamplification libraries of each embryo was taken and diluted in 19 μL Nuclease-free water as template to determine target gene transcript abundance. Quantitative PCR was performed using SYBR qPCR Master Mix (Vazyme Biotech Co., Ltd, Q711-02) and signals were detected with 7500 Fast Real-Time PCR System (Applied BioSystems). H2A.Z was used as endogenous control. Primers were shown in Supplementary information, Table S9.

**Immunofluorescence assay**

Embryos injected with Dux-EGFP mRNA mixed with RFP mRNA or control EGFP mRNA mixed with RFP mRNA were fixed with 4% paraformaldehyde, permeabilized for 15 min with 0.5% Triton X-100, and blocked for 30 min with 3% BSA at room temperature. Embryos were then incubated overnight at 4 °C with anti-OCT4 (Santa Cruz Biotechnology, sc-5279), anti-NANOG (BETHYL, A300-397A), anti-CDX2 (Biogenex, MU392A-UC), anti-ZSCAN4 (EMD Millipore, AB4340) or anti-MERVL-gag (EpiGentek, A-2801) antibodies diluted in PBS with 1% BSA. After being washed three times, the embryos were incubated with anti-mouse (for OCT4 and CDX2) or

anti-rabbit (for NANOG, ZSCAN4 and MERVL-gag) Alexa Fluor 647-conjugated secondary antibodies for 1 h and washed again three times. Embryos then were incubated for 15 min with DAPI and transferred to PBS with 1% BSA on the glass dish. The embryos were viewed with a Zeiss LSM880 confocal microscope.

**Immunoprecipitation and mass-spectrometry**

NIH3T3 cells were transfected with pCW57.1-Dux-EGFP plasmid and pCW57.1-EGFP empty vector using Vigofect (Vigorous Biotechnology) according to the manufacturer's protocol, respectively. Briefly, for a 10 cm plate containing ~$2.5 \times 10^6$ NIH3T3 cells, 8 μL reagent and 10 μg DNA were diluted in 800 μL of 0.9% NaCl, mixed and incubated at room temperature for 15 min. The mixture was then applied to the cells cultured in 8 mL fresh medium. The cells were harvested after 24 h with Dox treatment, and protein interactions were examined by immunoprecipitation. Immunoprecipitation experiments were performed using DPA bead-conjugated EGFP antibody (Invitrogen A11122). Fresh protein extracts from $5 \times 10^7$ cells were incubated with the beads overnight at 4 °C, in rotation. After being washed from the unbound proteins, the beads were incubated for 10 min at 98 °C. The elution was analyzed by western blot and then mass-spectrometry.

**References**

1. Picelli S, et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**: 171-81 (2014).

2. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21 (2013).

3. Pertea M, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**: 290-5 (2015).

4. 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).

5. 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**: 576-89 (2010).

# Supplementary information, Figure S1

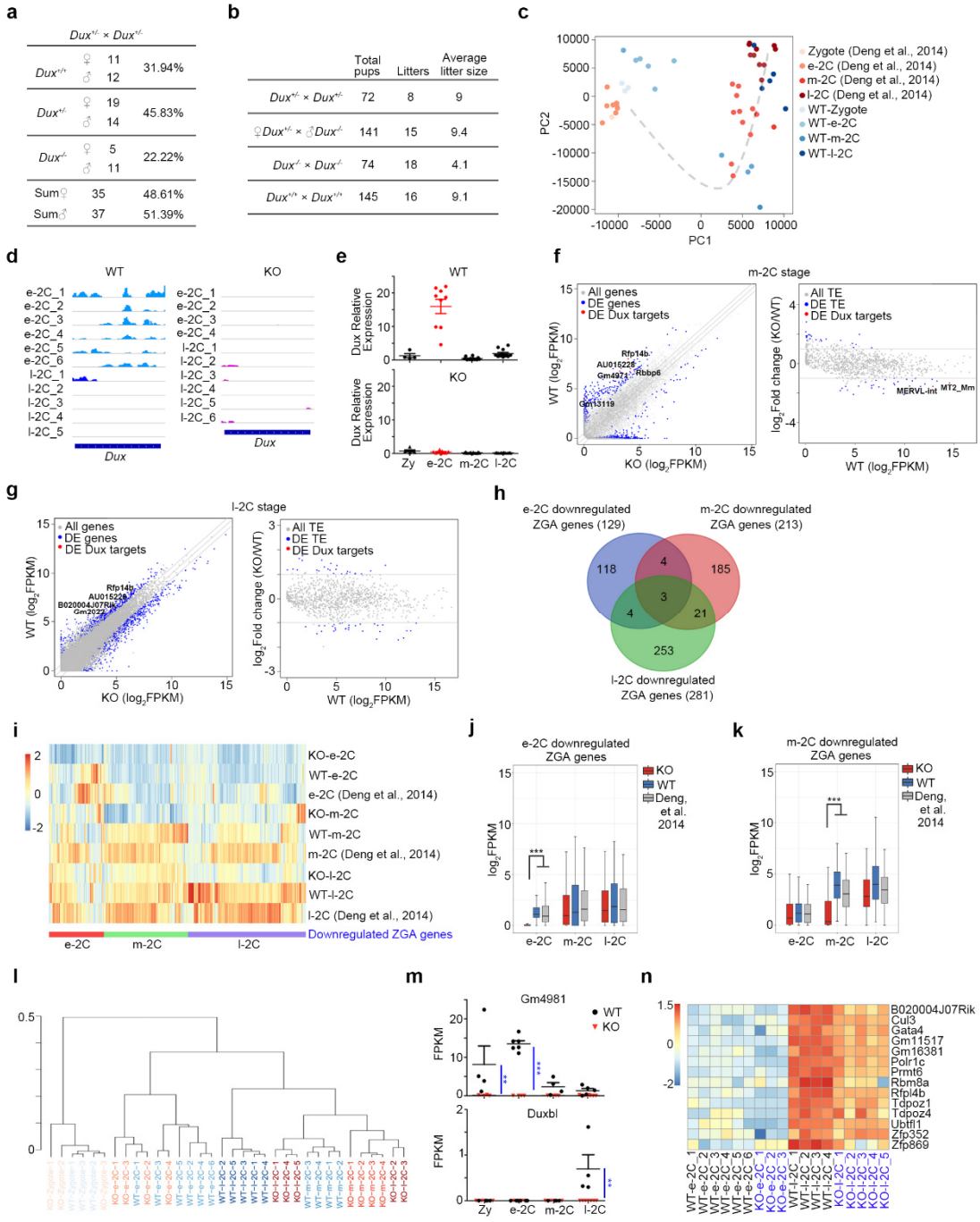



**Fig. S1 Dux is important but not essential for both ZGA and embryo development.**

**a** Statistical analysis of mouse numbers generated by *Dux*$^{+/-}$ × *Dux*$^{+/-}$ crosses. **b** Statistical analysis of litter size generated by WT and *Dux*-KO pairs. **c** PCA analysis of

WT embryos from zygotic stage to 2C stage in our dataset and published dataset. e-2C, early 2C stage; m-2C, middle 2C stage; l-2C, late 2C stage. **d** RNA-seq signals at *Dux* locus in WT and *Dux*-KO embryos at early and late 2C stage. **e** Relative expression of Dux in WT and Dux-KO embryos from zygotic stage to late 2C stage by qPCR. Each dot represents an embryo. Center lines and error bars indicate means and SD, respectively. **f** Scatter plots of gene expression (left panel) and transposable element (TE) expression (right panel) in WT and *Dux*-KO embryos at middle 2C stage. FC > 2, FDR < 0.05. **g** Scatter plots of gene expression (left panel) and TE expression (right panel) in WT and *Dux*-KO embryos at late 2C stage. FC > 2, FDR < 0.05. **h** Venn diagram illustrating the overlapping of downregulated ZGA genes in early, middle and late 2C Dux-KO embryos. e-2C, early 2C stage; m-2C, middle 2C stage; l-2C, late 2C stage. **i** Heatmap illustrating the expression of the 588 downregulated ZGA genes from early to late 2C stages of WT and Dux-KO embryos. e-2C, early 2C stage; m-2C, middle 2C stage; l-2C, late 2C stage. **j** Box plots of expression of 129 ZGA genes downregulated at early 2C stage caused by *Dux*-KO in WT and *Dux*-KO embryos from early to late 2C stages. e-2C, early 2C stage; m-2C, middle 2C stage; l-2C, late 2C stage. ***$P < 0.001$; Mann-Whitney-Wilcoxon two-sided test. The middle lines in the boxes represent medians. Box hinges indicate the twenty-fifth and seventy-fifth percentiles, and the whiskers indicate the hinge ± 1.5× interquartile range. **k** Box plots of expression of 213 ZGA genes downregulated at middle 2C stage caused by *Dux*-KO in WT and *Dux*-KO embryos from early to late 2C stages. e-2C, early 2C stage; m-2C, middle 2C stage; l-2C, late 2C stage. ***$P < 0.001$; Mann-Whitney-Wilcoxon two-sided test. The

middle lines in the boxes represent medians. Box hinges indicate the twenty-fifth and seventy-fifth percentiles, and the whiskers indicate the hinge ± 1.5× interquartile range. **l** Hierarchical clustering analysis of WT and *Dux*-KO embryos from zygotic stage to 2C stage. e-2C, early 2C stage; m-2C, middle 2C stage; l-2C, late 2C stage. **m** Dot plots of the expression of Dux functional homologue genes in WT and *Dux*-KO embryos from zygotic stage to late 2C stage. Each dot represents an embryo. $^{**}P < 0.01$, $^{***}P < 0.001$; two-tailed Student's *t*-test. Center lines and error bars indicate means and SD, respectively. **n** Heatmap illustrating the expression of the representative downregulated ZGA genes caused by Dux-KO at early and late 2C stages of WT and Dux-KO embryos.

# Supplementary information, Figure S2

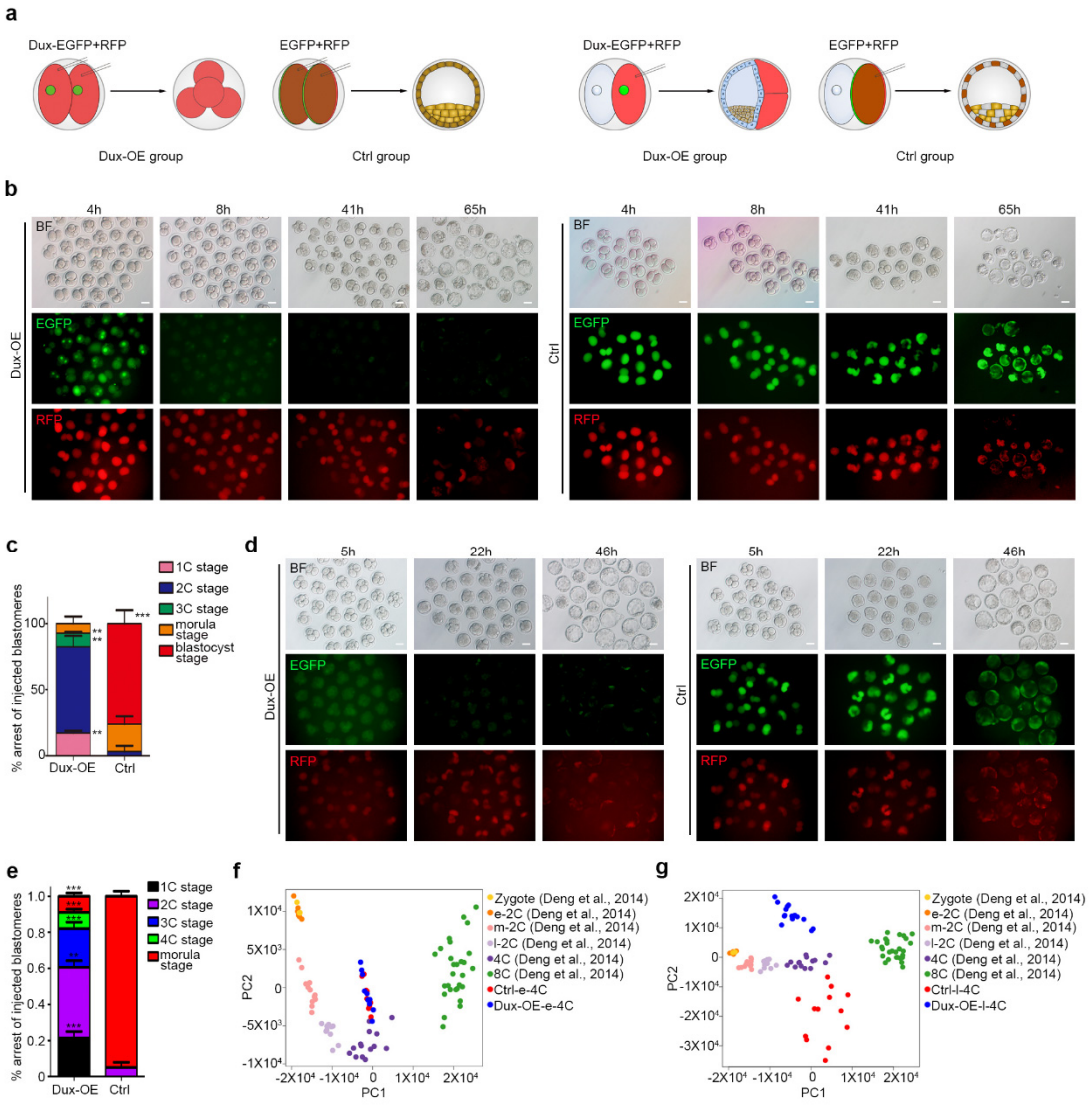
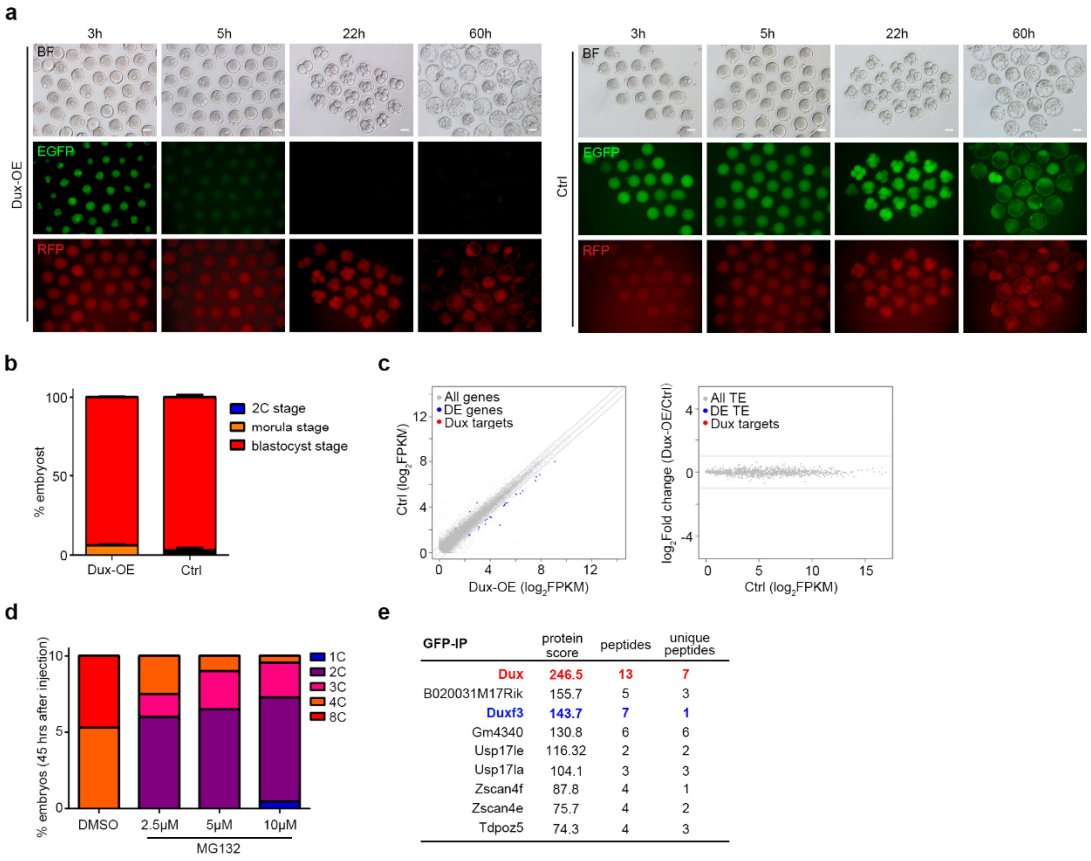



**Fig. S2 Prolonged expression of Dux significantly induces developmental arrest of blastomeres with 2C signature. a** Schematic summary of early embryonic development in Dux-OE group with Dux-EGFP (green, nuclei) and RFP (red, cytoplasm) mRNA injection, and in Ctrl group with EGFP (green, cytoplasm) and RFP (red, cytoplasm) mRNA injection in both blastomeres (left panel) or single blastomere (right panel) of late 2C embryos, respectively. Brown color indicates mixture of EGFP

and RFP. **b** Development progression of embryos 4 h, 8 h, 41 h and 65 h after mRNA injection in 1-blastomere of 2C embryos. h, hours; Dux-OE, group of embryos injected with mRNA mixture of Dux-EGFP and RFP (left panels); Ctrl, control group injected with mRNA mixture of EGFP and RFP (right panels). Scale bars, 50 μm. **c** Bar graph comparing the percentage of blastomeres arrested at different developmental stages between Dux-OE and Ctrl group 65 h after mRNA injection into 1-blastomere of 2C embryos. ***$P < 0.001$, **$P < 0.01$; two-tailed Student's *t*-test. Error bars indicate SD. **d** Development progression of embryos 5 h, 22 h, and 46 h after mRNA injection in 1-blastomere of 4C embryos. h, hours; Dux-OE, group of embryos injected with mRNA mixture of Dux-EGFP and RFP (left panels); Ctrl, control group injected with mRNA mixture of EGFP and RFP (right panels). Scale bars, 50 μm. **e** Bar graph comparing the percentage of blastomeres arrested at different developmental stages between Dux-OE and Ctrl group 22 h after mRNA injection into 1-blastomere of 4C embryos. ***$P < 0.001$, **$P < 0.01$; two-tailed Student's *t*-test. Error bars indicate SD. **f** PCA analysis of WT, Ctrl group and Dux-OE group embryos at different developmental stages in our RNA-seq dataset and published RNA-seq dataset. e, early; m, middle; l, late; Dux-OE-e-4C, embryos of Dux-OE group collected at early 4C stage (blue dots); Ctrl-e-4C, embryos of Ctrl group collected at early 4C stage (red dots). **g** PCA analysis of WT, Ctrl group and Dux-OE group embryos at different developmental stages in our RNA-seq dataset and published RNA-seq dataset. e, early; m, middle; l, late; Dux-OE-l-4C, embryos of Dux-OE group collected at late 4C stage (blue dots); Ctrl-l-4C, embryos of Ctrl group collected at late 4C stage (red dots).

## Supplementary information, Figure S3




| GFP-IP | protein score | peptides | unique peptides |
|---|---|---|---|
| Dux | 246.5 | 13 | 7 |
| B020031M17Rik | 155.7 | 5 | 3 |
| Duxf3 | 143.7 | 7 | 1 |
| Gm4340 | 130.8 | 6 | 6 |
| Usp17le | 116.32 | 2 | 2 |
| Usp17la | 104.1 | 3 | 3 |
| Zscan4f | 87.8 | 4 | 1 |
| Zscan4e | 75.7 | 4 | 2 |
| Tdpoz5 | 74.3 | 4 | 3 |

**Fig. S3 Exogenous Dux-EGFP was dramatically degraded in zygote. a** Development progression of embryos 3 h, 5 h, 22 h, and 60 h after mRNA injection in zygotes. h, hours; Dux-OE, group of embryos injected with mRNA mixture of Dux-EGFP and RFP (left panels); Ctrl, control group injected with mRNA mixture of EGFP and RFP (right panels). Scale bars, 50 μm. **b** Percentage of developmental stages 60 h after Dux-EGFP mRNA injection in zygote. **c** Scatter plots of gene expression (left panel) and transposable elements (TE) expression (right panel) in Dux-OE and Ctrl embryos at zygotic stage. FC > 2, FDR < 0.05. **d** Percentage of embryo developmental stages 45 h after Dux-EGFP mRNA injection in zygote treated with a concentration gradient of MG132. **e** List of representative Dux-interacting proteins by IP-mass spectrometry analysis.