Key role for CTCF in establishing chromatin structure in human embryos

https://doi.org/10.1038/s41586-019-1812-0

Received: 14 January 2019

Accepted: 16 October 2019

Published online: 4 December 2019

Xuepeng Chen[1,7], Yuwen Ke[1,7], Keliang Wu[2,7], Han Zhao[2,7], Yaoyu Sun[1,3], Lei Gao[1], Zhenbo Liu[1], Jingye Zhang[2], Wenrong Tao[2], Zhenzhen Hou[2], Hui Liu[2], Jiang Liu[1,3,4]* & Zi-Jiang Chen[2,5,6]*

In the interphase of the cell cycle, chromatin is arranged in a hierarchical structure within the nucleus[1,2], which has an important role in regulating gene expression[3–6]. However, the dynamics of 3D chromatin structure during human embryogenesis remains unknown. Here we report that, unlike mouse sperm, human sperm cells do not express the chromatin regulator CTCF and their chromatin does not contain topologically associating domains (TADs). Following human fertilization, TAD structure is gradually established during embryonic development. In addition, A/B compartmentalization is lost in human embryos at the 2-cell stage and is re-established during embryogenesis. Notably, blocking zygotic genome activation (ZGA) can inhibit TAD establishment in human embryos but not in mouse or *Drosophila*. Of note, CTCF is expressed at very low levels before ZGA, and is then highly expressed at the ZGA stage when TADs are observed. TAD organization is significantly reduced in CTCF knockdown embryos, suggesting that TAD establishment during ZGA in human embryos requires CTCF expression. Our results indicate that CTCF has a key role in the establishment of 3D chromatin structure during human embryogenesis.

During mouse embryogenesis, TADs are disorganized in 2-cell embryos and become increasingly established during development[7,8]. To study the process of 3D chromatin structure reprogramming in human embryos, we examined human sperm, 2-cell embryos, 8-cell embryos, morula, blastocysts and six-week-old embryos. We optimized a previous low-input Hi-C (genome-wide chromosome conformation capture) method[8] by using as few as 50–100 cells to construct Hi-C libraries in human early embryos (Methods). Our new ultra-low-input Hi-C method can detect very similar chromatin structures in a single mouse blastocyst to those obtained from pooled mouse blastocysts using the previous method[8] (Extended Data Fig. 1a–c). At least two biological replicates were sequenced for each collected developmental stage, (Supplementary Table 1) and showed high reproducibility (Extended Data Fig. 1d–f).

TAD structures in human early embryos

Next, we examined chromatin interactions in human sperm and embryos. We did not detect the characteristic 'triangle' interactions of TAD structures in human 2-cell embryos. There was a low level of these interactions in 8-cell embryos, and the level increased during human embryonic development (Fig. 1a). To rule out the possible influence of read depth on the analysis, we picked the same number of reads at random when plotting interaction heat maps for each stage, showing consistent results (Extended Data Fig. 2a).

We further investigated TAD reprogramming during human embryogenesis. We used a TAD separation-score method (Methods) to call TAD domains and TAD boundaries (Supplementary Table 2). The majority of TAD domains can be detected even with very low read depth (Extended Data Fig. 2b, c). We also calculated TAD signal and directional index[9]. Our data show that TAD signal variance and directional index are lowest at the 2-cell stage and increase gradually through development (Fig. 1b, Extended Data Fig. 2d, e). To exclude possible experimental bias in TAD analysis of human embryos, we used mouse morula embryos as spiked-in controls, which were mixed in with human 2-cell, 8-cell and morula embryos. We then constructed Hi-C libraries for these mixed samples (Methods). In parallel, we also generated Hi-C libraries for mouse morula embryos that were not previously mixed with human samples. These results show that TAD structures from the mixed samples remain obscure in human 2-cell embryos and become clearer in 8-cell embryos and morula embryos (Extended Data Fig. 2f, g), whereas all the spiked-in mouse morula embryos show clear TAD structures (Extended Data Fig. 2h). TAD signal analysis from the mixed samples supports our finding that TAD structures become established during human embryogenesis (Extended Data Fig. 2i).

In sum, these data show that TAD structures are largely absent in human 2-cell embryos, are weakly present in 8-cell embryos, and become increasingly evident during embryonic development.

[1]CAS Key Laboratory of Genome Sciences and Information, Collaborative Innovation Center of Genetics and Development, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China. [2]Center for Reproductive Medicine, Shandong University, The Key laboratory of Reproductive Endocrinology, Ministry of Education, Shandong University, Jinan, China. [3]University of Chinese Academy of Sciences, Beijing, China. [4]CAS Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China. [5]Center for Reproductive Medicine, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. [6]Shanghai Key Laboratory for Assisted Reproduction and Reproductive Genetics, Shanghai, China. [7]These authors contributed equally: Xuepeng Chen, Yuwen Ke, Keliang Wu, Han Zhao. *e-mail: liuj@big.ac.cn; chenzijiang@hotmail.com

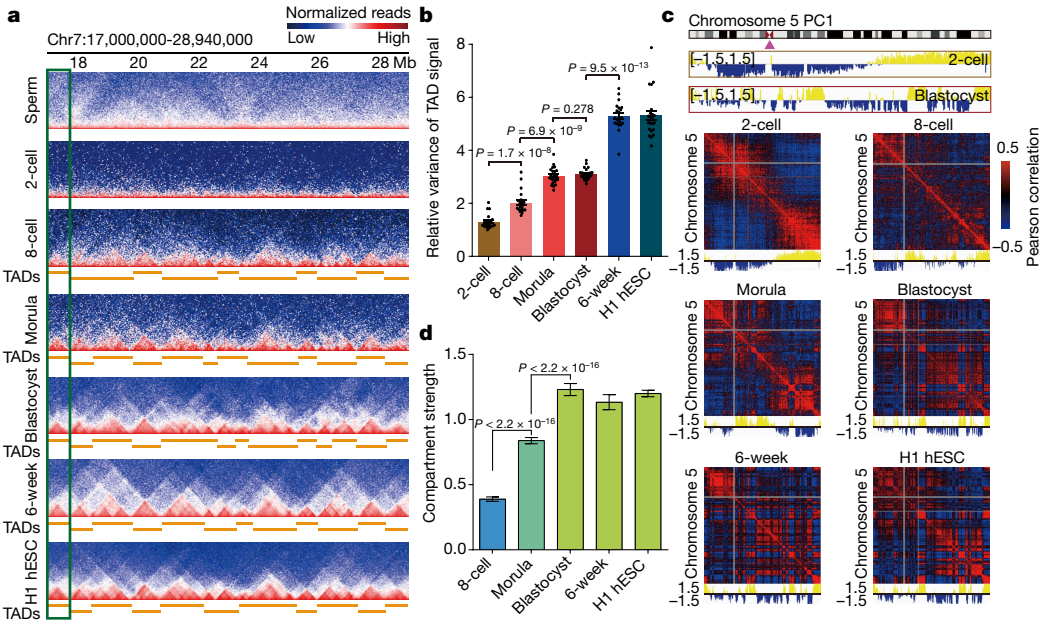


Fig. 1 | Three-dimensional chromatin structures of human sperm and embryos. a, Snapshot of the interaction heat map for high-order chromatin structures in human sperm, human embryos and H1 human ES cells (hESCs)[11] at 40-kb resolution (pooled biological replicates; $n = 2–3$). **b**, Relative variance of TAD signal in human embryos calculated with equal numbers of reads (generated from 2–3 biological replicates for each stage). Each dot represents one chromosome. *P* values are also shown; two-sided Wilcoxon rank-sum test. Data are mean ± s.e.m. **c**, PC1 value tracks and Pearson correlation heat maps for chromosome 5 at 500-kb resolution in human embryos with equal numbers of reads (generated from 2–3 biological replicates for each stage). **d**, Compartment strength in human embryos with equal numbers of reads. Data are mean ± s.d., obtained by bootstrapping ($n = 100$). *P* values were calculated by one-sided *t*-test.

The dynamics of A/B compartments

A/B compartmentalization is an important hierarchical structural feature of chromatin organization[10,11]. We found that the chromosomes of human 2-cell embryos lost A/B compartmentalization, but usually partitioned into a few large interaction domains (Fig. 1c, Extended Data Fig. 3a, b), in contrast to those of mouse zygotes and early embryos (Extended Data Fig. 3c). This pattern is similar to the mouse inactive X chromosome[12]. A/B compartmentalization is very weak at the human ZGA stage, and becomes more evident during development (Fig. 1c, Extended Data Fig. 3a). By contrast, there is clear A/B compartmentalization at the mouse ZGA stage (Extended Data Fig. 3c–e). The dynamic process of A/B compartmentalization in human embryos is further supported by compartment strength (Fig. 1d). Additionally, principal components (PC) analysis of the morula stage onwards showed bimodal distribution patterns of PC1 values, which are not observed at the 8-cell stage—suggesting poor compartment segregation at the 8-cell stage (Extended Data Fig. 3f). Together, these results suggest that A/B compartmentalization of chromosomes is not present at the 2-cell stage, is very weak at the 8-cell stage, emerges clearly at the morula stage and becomes more evident at the blastocyst stage.

Next, we investigated A/B compartment switches during human embryogenesis. We found that 13% of genomic regions show A/B compartment switches from the morula stage to the 6-week stage and 77% of them change their compartment statuses only once (Extended Data Fig. 4a). Then, we examined the potential effect of A/B compartment switches on embryonic development. Gene ontology analysis shows that genes that undergo A/B compartment switches are significantly enriched in developmental processes (Extended Data Fig. 4b, Supplementary Table 3). For example, genes undergoing a B-to-A switch between blastocyst and 6-week stage are enriched in telencephalon development and organ morphogenesis (Extended Data Fig. 4b). *NEFM* and *NEFL*, which function in maintaining the neuronal calibre, switch their B compartment status to A compartment and become highly expressed (Extended Data Fig. 4c).

Intensive genome-wide epigenetic reprogramming occurs during human embryonic development[13–15]. We analysed how A/B compartments and other epigenetic marks are orchestrated in human embryos. Consistent with the previous report in mouse embryos[8], unmethylated CpGs in human embryos are enriched in A compartments; the level of CpG demethylation in A compartments is significantly higher than that in B compartments (Extended Data Fig. 4d, e). Furthermore, our data show that DNA demethylation levels are higher in regions that maintain their A compartment status from sperm to blastocyst (Extended Data Fig. 4f). This suggests that A compartments provide a more-open local nuclear environment that may enable easier access for DNA-binding proteins.

Human sperm do not have typical TADs

Previous studies have shown that TAD structures are present in mature mouse sperm[8,16]. Surprisingly, we did not observe typical triangle TAD structures in human sperm (Fig. 1a). For example, there was no TAD boundary at the *HOXA* cluster region in human sperm, but it was present in human blastocysts (Fig. 2a) and mouse sperm (Extended Data Fig. 5a). To validate this observation, we plotted TAD signal variance at different read depths in human sperm and blastocysts. Unlike in mouse sperm[8] and human blastocysts, the *y* intercept of the human sperm line is close to 0 (Fig. 2b), suggesting an absence of TADs in human sperm. We further compared the density of interaction insert size between human sperm and mouse sperm. Our data show that human sperm presents one main peak around 4 Mb (middle-range), whereas mouse sperm has a shoulder peak around 933 kb and a long-range main peak at around 41 Mb (Extended Data Fig. 5b). Additionally, there was also a difference between human sperm and mouse sperm in the contact probability decay curve (Extended Data Fig. 5c). Together, these results suggest that human sperm do not contain TADs.

To exclude potential experimental bias, we mixed mouse sperm with human sperm and constructed a Hi-C library for the sperm mixture

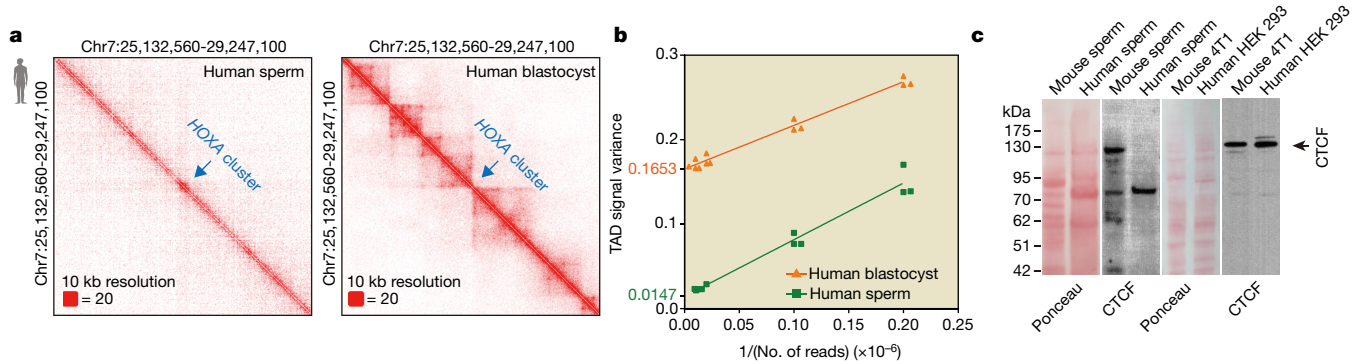


Fig. 2 | Human sperm does not have typical TAD structures. a, Interaction heat maps around the *HOXA* cluster in human sperm and blastocysts at 10-kb resolution (pooled biological replicates; $n = 3$). **b**, Linear regression curves for TAD signal variance versus read depth (1/no. of reads) in human sperm and blastocysts. The linear extrapolation of regression lines for human blastocysts and sperm are marked on the *y* axis. Downsampling analysis at each read depth was performed three times ($n = 3$). **c**, Western blot for CTCF in mouse sperm, human sperm and somatic cell lines, with Ponceau staining. Black arrow, CTCF band. Experiment repeated on two biologically independent replicates per sample. For gel source data, see Supplementary Fig. 1.

(Methods). In parallel, we mixed human HeLa cells and mouse HT22 cells. Consistent with our previous result, we observed TADs from mouse sperm—but not from human sperm—in the sperm mixture (Extended Data Fig. 5d, e). By contrast, TAD structures were observed in both human HeLa cells and mouse HT22 cells (Extended Data Fig. 5d, e). These data corroborate the absence of TADs from human sperm.

We also inspected A/B compartmentalization in human sperm and found that chromatin from human sperm, both in isolation and when mixed with mouse sperm, shows clear A/B compartmentalization (Extended Data Fig. 5f).

The CTCF–cohesin complex has important roles in high-order chromatin structures[4,17]. We investigated levels of CTCF and cohesin protein in human and mouse sperm. RAD21, a subunit of cohesin complex, is present in both human and mouse sperm (Extended Data Fig. 5g); however, CTCF was not detected in human sperm and was very weakly expressed in cell lines in which CTCF was depleted using short interfering RNA (siRNA) (Methods), but was detected in mouse sperm, human and control mouse cell lines (Fig. 2c, Extended Data Fig. 5h). Because depletion of CTCF can lead to the disruption of TAD structures[18], the lack of CTCF may underlie the loss of TAD structures in human sperm.

Establishment of TAD boundaries

TAD structure is obscure in human 2-cell embryos. However, some regions in 2-cell embryos show evidence of insulator binding, which can separate upstream and downstream interactions (Extended Data Fig. 6a). At the later embryonic stages, most of these regions become TAD boundaries (Fig. 3a); these regions can therefore be regarded as immature TAD boundaries. In this study, we defined both immature TAD boundaries and mature TAD boundaries as insulated boundaries. We then analysed the dynamics of insulated boundaries during human embryonic development. Our data show that 635, 905, 317 and 306 insulated boundaries developed primarily at the 2-cell, 8-cell, morula and blastocyst stages, respectively (Fig. 3a, Supplementary Table 4). We also found immature TAD boundaries in mouse 2-cell embryos, and identified the stage-gained insulated boundaries in mouse (Extended Data Fig. 6b, Supplementary Table 5). We compared the insulated boundaries at the 2-cell stage with the total boundaries at the blastocyst stage in human samples, showing that 2-cell boundaries overlapped with 30% of blastocyst boundaries (Fig. 3b). Our data also show that the ZGA stage contained 67% of boundaries present in the blastocyst, which is similar to the proportion in mouse models (Fig. 3b). Furthermore, we found significant overlap when we compared human boundaries at the ZGA stage with those from mouse (Fig. 3c). For example, we found insulated boundaries at the ZGA stage in both human and mouse around the *TTC1* and *CCNG1* genes (Extended Data Fig. 6c, d).

Next, we sought to identify the genomic regions that exhibit a preference for first forming insulated boundaries at an early stage. The distance of genomic regions that first gained boundaries at the 2-cell stage to housekeeping genes was smaller than that of boundaries regions that gained boundaries at a later stage (Fig. 3d). A similar result was observed in mouse embryos (Extended Data Fig. 7a). These data indicate that insulated boundaries gained at the earlier stage tend to locate around housekeeping genes in both human and mouse. We also found that the expression levels of housekeeping genes near the boundaries tended to be higher than those of other housekeeping genes (Extended Data Fig. 7b).

Repeat elements are reported to associate with TAD boundaries in cell lines[9]. We thus analysed the enrichment of repeat elements around stage-specifically gained boundaries in embryos. Our data illustrates that Alu repeats, but not LINE or MIR repeats (Fig. 3e, Extended Data Fig. 7c–e) repeats, are enriched around the insulated boundaries at the earlier stage in human. Similar results are observed in mouse embryos (Extended Data Fig. 7f–h). For example, insulated boundaries gained at the 2-cell stage around the human or mouse *RAB5A* gene are established at an Alu-dense region (Extended Data Fig. 7i). Moreover, our data show that AluS elements are highly enriched around insulated boundaries gained at the 2-cell stage (Fig. 3e). Moreover, AluS repeats around insulated boundaries gained at the 2-cell stage were highly expressed at the cleavage stage compared with other stages (Extended Data Fig. 7j, k). Collectively, these results show a preference for insulated boundaries to locate around Alu-dense regions in human embryos.

TAD establishment depends on ZGA

Previous reports show that TAD establishment is independent of ZGA in mouse and *Drosophila* embryos[7,8,19]. We investigated whether these characteristics are conserved in human. We treated human zygotes with α-amanitin to inhibit ZGA (Extended Data Fig. 8a, b) and collected embryos at the 8-cell stage (Methods, Supplementary Table 6). Surprisingly, TAD structure was obscure in α-amanitin-treated 8-cell embryos (Fig. 4a, Extended Data Fig. 8c, d). Relative variance of the TAD signal in α-amanitin-treated embryos was also significantly lower than that in untreated 8-cell embryos (Fig. 4b). Therefore, TAD establishment in human embryos requires ZGA.

CTCF regulates the chromatin landscape

Next, we aimed to identify the protein involved in TAD establishment during ZGA. The cohesin complex and CTCF have important roles

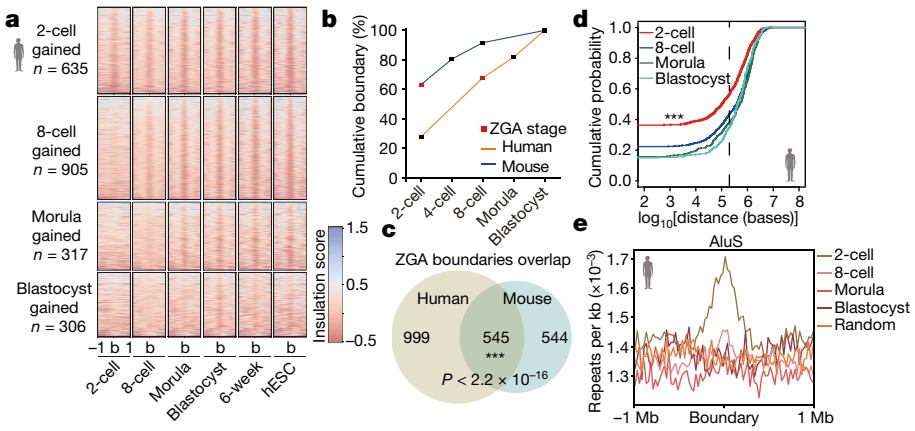


Fig. 3 | Establishment of insulated boundaries during human embryogenesis. a, Heat map for insulation score, centred at insulated boundaries gained at specific stages (±1 Mb range). *n*, number of stage-specific gained insulated boundaries; b, insulated boundary centre. **b**, The cumulative percentage of insulated boundary number in each stage relative to the blastocyst boundary number. **c**, Venn diagram showing overlap between human ZGA boundaries and mouse ZGA boundaries (χ^2 test). **d**, Cumulative distribution function plot for the distance of stage-specific gained boundaries to the closest housekeeping genes in human embryos (pooled data from 2–3 biological replicates). The dash line shows a distance of 200 kb. $P = 3.24 \times 10^{-8}$ (2-cell versus 8-cell stage); $P = 9.95 \times 10^{-8}$ (2-cell stage versus morula); $P = 3.83 \times 10^{-11}$ (2-cell stage versus blastocyst); two-sided Kolmogorov–Smirnov test. **e**, Enrichment of AluS elements at insulated boundaries gained at specific stages in human embryos.

in higher-order chromosome structures[4,17]. Thus, we investigated differences in expression of these proteins[20–22]. Subunits of cohesin complex, such as RAD21, are already highly expressed before ZGA in human embryos (Extended Data Fig. 9a). By contrast, CTCF expression is very limited before the ZGA stage, and is sharply increased at the 8-cell stage when TAD structure is first observed in human embryos (Fig. 4c, Extended Data Fig. 9b). In the α-amanitin-treated 8-cell embryos, CTCF expression is inhibited (Extended Data Fig. 9c). Consistently, immunostaining images show that CTCF protein is barely observed in the 2-cell nucleus (Fig. 4d). CTCF is clearly present in untreated 8-cell nucleus, but it is absent in the α-amanitin-treated 8-cell nucleus (Fig. 4d). These results suggest that CTCF expression requires human ZGA.

Next, we investigated the requirement of TAD establishment in human embryos for CTCF. We inhibited CTCF expression by injecting *CTCF* siRNA (si*CTCF*) into human zygotes, and collected embryos at the morula stage (Fig. 4d). Notably, triangle TAD structures were scarcely observed in the si*CTCF* morula (Fig. 4e, Extended Data Fig. 9d). Relative TAD signal variance supports inhibition of TAD establishment in the si*CTCF* morula (Extended Data Fig. 9e). Consistently, most of the TAD boundaries vanish in the control morula and become weaker in the si*CTCF* morula (Extended Data Fig. 9f). Thus, our data suggest

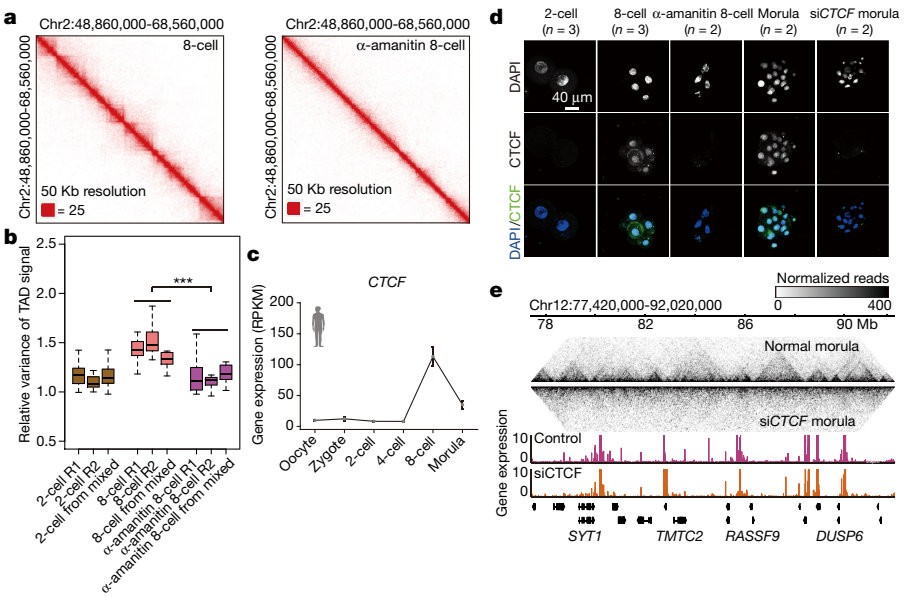


Fig. 4 | CTCF regulates the establishment of chromatin landscape in human embryos. a, Interaction heat maps in human 8-cell and α-amanitin-treated 8-cell embryos. **b**, Box plot for relative variance of TAD signal in human 2-cell ($n = 3$), 8-cell ($n = 3$) and α-amanitin-treated 8-cell embryos ($n = 3$). Boxes show 25th, 50th and 75th percentiles and whiskers show 1.5× the interquartile range. ***Adjusted $P < 0.001$ for all pairwise comparisons between 8-cell and α-amanitin-treated 8-cell (two-sided Wilcoxon rank-sum test with Benjamini–Hochberg multiple testing correction). **c**, Dynamics of CTCF expression during human embryonic development (expression data from ref. [22]; 3–20 cells for each stage). Data are mean ± s.e.m. RPKM, reads per kilobase of transcript per million mapped reads. **d**, Immunofluorescence of CTCF in human embryos ($n = 2$–3). Scale bar, 40 μm. **e**, Track snapshot for TAD structures in untreated control morula and si*CTCF* morula, overlaid with gene expression.

that CTCF expression during ZGA is required for TAD establishment in human embryos.

Frequently interacting region (FIRE) is a recently defined hierarchy of chromatin high-order structure[23]. We identified FIREs at 40-kb resolution in human embryos. Human embryo FIREs are enriched in A compartments (Extended Data Fig. 10a–c). Our data show that both assay for transposase-accessible chromatin (ATAC) peaks[24] and DNase I hypersensitive sites (DHSs)[25] are enriched in FIREs (Extended Data Fig. 10d, e). To check whether *CTCF* knockdown has an effect on FIRE scores in human embryos, we analysed the FIRE score in the si*CTCF* morula and found that FIRE scores were significantly decreased (Extended Data Fig. 10f), consistent with previous reports in cell lines[25].

Additionally, we analysed the gene expression profile of the si*CTCF* morula and identified 565 downregulated genes in the si*CTCF* morula (Extended Data Fig. 10g, h). Our data show that highly downregulated genes in the si*CTCF* morula usually have CTCF-binding peaks around transcription start sites (TSSs) in human embyonic stem (ES) cells (Extended Data Fig. 10i).

Finally, we wanted to know whether CTCF expression is the sole determinant of TAD establishment. To address this question, we recovered CTCF expression by injecting *CTCF* mRNA into α-amanitin-treated human zygotes and collected the treated embryos at the 8-cell stage (Extended Data Fig. 11a, Methods). We did not observe clear triangle TAD patterns in 8-cell embryos that had been treated with α-amanitin and *CTCF* mRNA, in contrast to the patterns observed in untreated human 8-cell embryos and spiked-in mouse morula embryos (Extended Data Fig. 11b–e). Further insulation boundary analysis and TAD strength analysis[26] validated that TAD was not established in embryos treated with α-amanitin and *CTCF* mRNA (Extended Data Fig. 11f, g). In summary, our data suggest that CTCF expression is required, but is not the only factor needed, for the establishment of chromatin structure during human ZGA.

Although both human and mouse embryos show genome-wide reprogramming of high-order chromatin structure, there are substantial differences in chromatin structure between human and mouse embryos. Our data provide a valuable resource and mechanistic insights into the establishment of chromatin structure during mammalian embryonic development.

Online content

Any methods, additional references, Nature Research 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/s41586-019-1812-0.

1. Fullwood, M. J. et al. An oestrogen-receptor-α-bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
2. Atlasi, Y. & Stunnenberg, H. G. The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* **18**, 643–658 (2017).
3. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
4. Tang, Z. et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015).
5. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572 (2017).
6. Hsieh, T. H. et al. mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell* **162**, 108–119 (2015).
7. Du, Z. et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232–235 (2017).
8. Ke, Y. et al. 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. *Cell* **170**, 367–381 (2017).
9. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
10. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
11. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
12. Giorgetti, L. et al. Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575–579 (2016).
13. Li, C. et al. DNA methylation reprogramming of functional elements during mammalian embryonic development. *Cell Discov.* **4**, 41 (2018).
14. Smith, Z. D. et al. DNA methylation dynamics of the human preimplantation embryo. *Nature* **511**, 611–615 (2014).
15. Guo, H. et al. The DNA methylation landscape of human early embryos. *Nature* **511**, 606–610 (2014).
16. Jung, Y. H. et al. Chromatin states in mouse sperm correlate with embryonic and adult regulatory landscapes. *Cell Rep.* **18**, 1366–1382 (2017).
17. Baranello, L., Kouzine, F. & Levens, D. CTCF and cohesin cooperate to organize the 3D structure of the mammalian genome. *Proc. Natl Acad. Sci. USA* **111**, 889–890 (2014).
18. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 (2017).
19. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* **169**, 216–228 (2017).
20. Xue, Z. et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597 (2013).
21. 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).
22. Yan, L. et al. Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131–1139 (2013).
23. Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
24. Wu, J. et al. Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256–260 (2018).
25. Gao, L. et al. Chromatin accessibility landscape in human early embryos and its association with evolution. *Cell* **173**, 248–259 (2018).
26. Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).

Methods

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

Human subjects

The regulatory framework about the use of human gametes and embryos for this research was based on the policies of the Human Biomedical Research Ethics Guidelines (set by National Health Commission of the People's Republic of China on 1 December 2016),the 2016 Guidelines for Stem Cell Research and Clinical Translation (issued by the International Society for Stem Cell Research, ISSCR) and the Human Embryonic Stem Cell Research Ethics Guidelines (set by China National Center for Biotechnology Development on 24 December 2003). These policies and guidelines permit human gametes, and/or human embryos created or genetic manipulated in vitro no more than 14 days, can be used specifically for scientific researches. All the experiments in this study are in compliance with these relevant ethical regulations.

The aim and protocols of this study has been reviewed and approved by the Institutional Review Board of Reproductive Medicine, Shandong University and Beijing Institute of Genomics. The protocols include the injection of siRNA into human embryos. Oocytes donated from patients taking in-vitro fertilization treatments, were fertilized using donated sperm by intracytoplasmic sperm injection (ICSI). Written informed consent was obtained from all oocyte and sperm donors. The donor women are 25–38 years old with tubal-factor infertility and their partners have healthy semen. The embryos with high quality were randomly assigned to experimental groups without examination about the gender of embryos owing to the limited cell number. All embryos are collected under standard clinical protocols. The number of embryos used at each developmental stage used for Hi-C in this study is shown in Supplementary Table 1. All control and siRNA-injected embryos were cultured no more than 7 days and only used for molecular research analyses. No statistical methods were used to predetermine sample size.

Human embryos collection

We collected human 2-cell, 8-cell, morula and morphological AA grade blastocysts at appropriate time after in vitro fertilization (IVF). The six-week embryos (6-week) were obtained from IVF couples by transvaginal multifetal pregnancy reduction surgery at the six-week stage of pregnancy.

For 2-cell embryos, they were from immature metaphase I (MI) oocytes that were not used clinically. MI oocytes matured in vitro and were fertilized by donated sperm. These fertilized embryos were then cultured in G1.5 medium (Vitrolife) in a humidified atmosphere at 37 °C with 6% CO_2 in air. The 2-cell embryos were collected and vitrified around 27 h after routine fertilization. The embryo vitrification was carried out as previously described[27]. In brief, the embryos were incubated in vitrification solution 1 consisting of 8% ethylene glycol and 8% dimethyl sulfoxide in Cryobase (10 mM HEPES-buffered medium containing 20 mg ml^{-1} human serum albumin and 0.01 mg ml^{-1} gentamicin) at room temperature for 11 min. After initial shrinkage, embryos with original volume were transferred into vitrification solution 2 (16% ethylene glycol, 16% dimethyl sulfoxide and 0.68 M trehalose in Cryobase) for 1–1.5 min. Embryos were finally transferred onto Cryotop strip in an extremely small volume of solution (<0.1 ml) and plunged into liquid nitrogen. After addition of the protective cover, the Cryotop was transferred into liquid nitrogen for storage.

The embryos of 8-cell, morula and blastocyst stages were from donated frozen embryos with signed informed consent. These embryos were vitrified at 3 day, 4 day and 5 day after routine in vitro fertilization, respectively. The vitrified embryos on Cryotop strip were thawed rapidly by taking them from the liquid nitrogen after removal of the protective cover, then immersed in 2.5 ml of 37 °C warming solution 1 (1 M trehalose in Cryobase) for 1 min on a heated stage. Embryos were then transferred to 0.5 ml of warming solution 2 (0.5 M trehalose in Cryobase) for 3 min, and then placed into 0.5 ml Cryobase for 5 min followed by fresh 0.5 ml Cryobase for 1 min. Embryos were finally transferred to G1.5 or G2 medium (Vitrolife) to evaluate embryo quality.

The thawed embryos with high quality were picked randomly for experimental groups. The zona pellucida was removed by mechanical dissection with a glass needle. Embryos were washed several times by gentle pipetting with a narrow-bore glass pipette to remove the attached cumulus or polar bodies. After final washing with 0.1% BSA/PBS three times, the embryos were transferred into microcentrifuge tubes (Sorenson BioScience, 39640T). Experiments were performed on at least two biological replicates for each embryonic stage.

Ultra-low-input Hi-C

The Hi-C library generation with ultra-low number of cells was optimized according to previous protocols[8,26]. Samples were fixed with 20 µl of freshly made 1% formaldehyde solution and incubated at room temperature for 10 min. To quench the reaction, 2.5 M glycine solution was added to a final concentration of 0.2 M. Samples were incubated at room temperature for 5 min and then were lysed directly in 100 µl of ice-cold Hi-C lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA630) with a protease inhibitor cocktail on ice for 15 min. Samples were then centrifuged at 3,000*g* for 5 min and the supernatants carefully discarded. Pelleted nuclei were washed once with 100 µl of 1× NEBuffer 2; discard the supernatant and remain 4.5 µl, and add 0.5 µl of 5% sodium dodecyl sulfate (SDS). Tubes were gently tapped to mix the pellet and were incubated at 62 °C for 5 min. After incubating, 14.5 µl of water and 2.5 µl of 10% Triton X-100 were added to quench the SDS. Tubes were gently tapped to mix well, avoiding excessive foaming and then incubated at 37 °C for 15 min. Two and a half microlitres of 10× NEBuffer 2 and 10 U of MboI restriction enzyme (NEB, R0147) were added and chromatin was digested at 37 °C for 5 h. Samples were incubated at 62 °C for 20 min to inactivate MboI and then cooled to room temperature. To fill in the restriction fragment overhangs and mark the DNA ends with biotin, 5 µl of fill-in master mix (3.75 µl of 0.4 mM biotin–14-dATP, 0.45 µl of 10 mM dCTP/dGTP/dTTP mix, 0.8 µl of 5 U µl^{-1} DNA polymerase I, large) was added. Samples were mixed by pipetting and incubated at 37 °C for 30 min. Ligation master mix (66.3 µl of water, 12 µl of 10× NEB T4 DNA ligase buffer, 10 µl of 10% Triton X-100, 1.2 µl of 10 mg ml^{-1} bovine serum albumin, 1 µl of 400 U/µl T4 DNA ligase) was added and samples were incubated at 16 °C for 18 h. Nuclei were pelleted by centrifugation for 5 min at 3,000*g* and were washed with 100 µl of 10 mM Tris buffer, pH 8.0. Pellets were then resuspended in 50 µl of 10 mM Tris buffer with 2 µl of 20 mg ml^{-1} proteinase K and incubated at 65 °C for 18 h. Proteinase K was inactivated by incubation at 75 °C for 30 min. To fragment the DNA ligation products, 50 ng of λ DNA was added, and the DNA samples were digested with 1 U AluI at 37 °C for 4 h and inactivated at 65 °C for 30 min. Then the fragmented DNA was treated with the End Repair/dA-Tailing Module (NEB, E7442L) and Ligation Module (NEB, E7445L) following the operation manual. Samples were prepared for biotin pull-down by washing with 50 µl of 10 mg ml^{-1} Dynabeads MyOne Streptavidin T1 beads (Life Technologies, 65602) in 100 µl of 1× Tween washing buffer (1× TWB: 5 mM Tris-HCl, pH 7.5; 0.5 mM EDTA; 1 M NaCl; 0.05% Tween 20). Samples were separated on a magnet and the solution was discarded. Ligation products were mixed with 83.5 µl of 2× binding buffer (2× BB: 10 mM Tris-HCl, pH 7.5; 1 mM EDTA; 2 M NaCl) and centrifuged at 10,000*g*, 1 min. Beads were resuspended with supernatant and incubated at room temperature for 30 min with rotation to bind the biotinylated DNA to the streptavidin beads. Samples were again separated on a magnet and the solution was discarded. Beads were washed by adding 200 µl of 1× TWB and tubes were heated on a Thermomixer at 55 °C for 2 min with mixing. Beads were pelleted again using a magnet and the supernatant was discarded. This wash step was repeated twice. Beads were then resuspended in

100 µl of 10 mM Tris buffer and transferred to a new tube. Beads were repelleted and the buffer discarded. Beads were resuspended in 20 µl of 10 mM Tris buffer. The Hi-C library was amplified for 12 cycles of PCR with Q5 master mix (NEB, M0492L) following the operation manual. PCR products were confirmed by analysing 1 µl of product using the FlashGel System (Lonza, 57063). PCR was continued with additional cycles until bright DNA bands were seen. A bottle of Agencourt AMPure XP beads (Beckman Coulter, A63881) was warmed to room temperature and gently shaken to resuspend the magnetic beads. One hundred microlitres of beads was added to 200 µl of diluted PCR product (0.5 volumes). Samples were mixed by pipetting and incubated at room temperature for 10 min. Beads were pelleted on a magnet and the clear solution was transferred to a new tube. Another 30 µl of beads was added to the clear solution (0.65× volume), mixed by pipetting and incubated at room temperature for 10 min. Keeping the beads on the magnet, samples were washed twice with 200 µl of 70% ethanol (freshly made) without mixing. Ethanol was then completely removed. Beads were left on the magnet for 5 min to allow the remaining ethanol to evaporate. DNA was eluted by adding 20 µl of ddH_2O, mixing by pipetting and incubating at room temperature for 5 min. After separating on a magnet, the solution was transferred to a new tube. DNA was then quantified and sequenced using an Illumina sequencing platform.

For human 2-cell, 8-cell, morula, blastocyst and 6-week embryos, we performed ultra-low-input Hi-C experiments. At least two replicates are prepared for each stage. For spike-in experiments, we performed ultra-low-input Hi-C experiments in human 2-cell mixed sample, human 8-cell mixed sample, human morula mixed sample, human α-amanitin-treated 8-cell mixed sample and human 8-cell (α-amanitin plus *CTCF* mRNA) mixed samples. In these spike-in experiments, the indicated stages of human embryos were mixed with six mouse morula embryos respectively.

Mouse sperm and morula embryos collection

Mouse sperm cells were collected from the cauda epididymis in 8-week old C57BL/6J male mice (Beijing Vital River Laboratory Animal Technology, stock number 213). The cauda epididymis was isolated and washed with 1× PBS three times, and cut into pieces and placed in a 5 ml 1× EmbryoMax Human Tubal Fluid (HTF) (EMDmillipore, MR-070-D) with 10% FBS (Gibco, 16140071) at a 6-well plate (Corning). Two cauda epididymis in one well. Sperms were incubated in HTF for 30 min in a CO_2 incubator at 37 °C to reduce somatic cells contaminants. Then, the supernatant with swimming active sperm was transferred to a new well without disturbing the settled somatic cells and low-activity sperm on bottom. The incubation–transfer cycle was repeated 4–5 times. The quality of samples was assessed under microscopy before snap freezing. The purity of sperm was over 99%. The sperms were then collected at minimal volume and frozen.

The morula embryos were from the cross of C57 BL/6J female and PWK/PhJ male. C57BL/6J female mice aged 4 to 6 weeks were intraperitoneally injected with pregnant mare serum gonadotropin (PMSG, 5UI) followed by human chorionic gonadotropin (hCG, 5UI) 48 h later. Morula embryos were flushed with M2 medium (Sigma) from the reproductive tract and the uteri at about 70–82 h after hCG administration. The morula embryos with high quality were confirmed with microscope. The embryos were serially washed to remove contaminants and were treated with acid Tyrode's solution (Sigma) for a few minutes to remove zona pellucida by pipetting through a fine pulled-glass needle. The embryos were then manually picked into PCR thin-wall tubes and stored at −80 °C.

All animal maintenance and experimental procedures were carried out according to guidelines of Institutional Animal Care and Use Committee (IACUC) of Beijing Institute of Genomics, CAS.

Sperm in situ Hi-C library generation

The generation of Hi-C libraries with sperm was optimized according to previous protocols[3]. Three million human sperm cells were used for each sample. For spike-in experiment, 3 million human sperm and 1–3 million mouse sperm were mixed together and use for one sample. As a control experiment, 3 million human cells HeLa (ATCC CCL-2; tested for mycoplasma contamination but not authenticated) and 3 million mouse cells HT22 (EMD Millipore, SCC129; tested for mycoplasma contamination but not authenticated) were mixed and used for one sample, and the Hi-C libraries generation of mixed somatic cell sample were followed the methods as previously published[3].

Sperm cells were fixed with 1 ml of freshly made 1% formaldehyde solution and incubated at room temperature for 10 min. To quench the reaction, 2.5 M glycine solution was added to a final concentration of 0.2 M. Samples were incubated at room temperature for 5 min and then centrifuge for 5 min at 3,000*g* at 4 °C. Supernatant was discarded. The pellet was washed with ice-cold 1× PBS and spun for 5 min at 3,000*g* at 4 °C. Supernatant was discarded and cell pellets were flash-frozen in liquid nitrogenand stored at −80 °C until used, or used immediately. The sperm pellet was resuspend with 1 ml of ice-cold Hi-C lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA630, 0.05% L-α-lysophosphatidylcholine) with a protease inhibitor cocktail. The cells were homogenized 5 times with a Dounce on ice and incubate for 30 min. Samples were then centrifuged at 3,000*g* for 5 min and the supernatants carefully discarded. Pelleted nuclei were washed once with 1 ml of ice-cold Hi-C lysis buffer. Supernatant was discarded and pellet was resuspended with 50 µl of 0.5% SDS and incubated at 62 °C for 10 min. After incubating, 145 µl of water and 25 µl of 10% Triton X-100 were added to quench the SDS. Tubes were gently tapped to mix well, avoiding excessive foaming and then incubated at 37 °C for 15 min. Twenty-five microlitres of 10× NEBuffer 2 and 100 U of MboI restriction enzyme (NEB, R0147) were added and chromatin was digested at 37 °C for overnight with rotation. Samples were incubated at 62 °C for 20 min to inactivate MboI and then cooled to room temperature. To fill in the restriction fragment overhangs and mark the DNA ends with biotin, 50 µl of fill-in master mix (37.5 µl of 0.4 mM biotin–14-dATP, 4.5 µl of 10 mM dCTP/dGTP/dTTP mix, 8 µl of 5 U μl^{-1} DNA polymerase I) was added. Samples were mixed by pipetting and incubated at 37 °C for 1.5 h. Ligation master mix (663 µl of water, 120 µl of 10× NEB T4 DNA ligase buffer, 100 µl of 10% Triton X-100, 12 µl of 10 mg ml^{-1} bovine serum albumin, 10 µl of 400 U μl^{-1} T4 DNA ligase) was added and samples were incubated at 16 °C for over 10 h with rotation. Nuclei were pelleted by centrifugation for 5 min at 3000*g* and were washed with 1× PBS. Pellets were then resuspended in 400 µl 1× PBS, and 16.7 µl of 20 mg ml^{-1} proteinase K and 40 µl 10% SDS were added, incubated at 55 °C for 30 min. The sample was added to 43.3 µl 5 M Nacl, 12 µl 0.5M EDTA, 24 µl 1M DTT, 50 µl 10%SDS and incubated at 68 °C overnight. The sample was cooled to room temperature, and 1,250 µl pure ethanol and 50 µl 3 M sodium acetate, pH 5.2 were added, mixed by inverting and incubated at −80 °C for over 1 h. Sample was centrifuged at 13,000*g*, 4 °C for 20 min. Tubes were kept on ice after spinning and the supernatant was carefully removed by pipetting. The pellet was washed with 800 µl of 70% ethanol twice and centrifuged at 13,000*g* for 5 min, dissolved in 100 µl of 1× Tris buffer (10 mM Tris-HCl, pH 8) and incubated at 37 °C for 15 min to fully dissolve the DNA. DNA was quantified by Qubit dsDNA High Sensitivity Assay (Life Technologies, Q32854). Biotin-14-dCTP at non-ligated DNA ends was removed with the exonuclease activity of T4 DNA polymerase (For 5 µg DNA: 1 µl 10 mg ml^{-1} BSA, 10 µl 10× NEBuffer 2, 1 µl 10 mM dNTP, and 5 U T4 DNA polymerase (NEB) in a total volume of 100 µl). The reaction was incubated at 12 °C for 2 h and stopped by adding 2 µl 0.5 M EDTA pH 8.0. All reaction products were combined. To make the biotinylated DNA suitable for high-throughput sequencing using Illumina sequencers, it was sheared to a size of 300–500 bp. Sheared DNA was transferred to a fresh 1.5-ml tube, and an equal volume of 2× binding buffer (2X BB: 10mM Tris-HCl (pH 7.5); 1mM EDTA; 2M NaCl) was added, then centrifuged at 13,000*g* for 5 min. Sample was prepared for biotin pull-down by washing 50 µl of 10 mg ml^{-1} Dynabeads MyOne Streptavidin T1 beads (Life Technologies, 65602) with 400 µl of 1× Tween washing buffer (1× TWB: 5mM Tris-HCl (pH 7.5);

0.5mM EDTA; 1M NaCl; 0.05% Tween 20), separated on a magnet and solution was discarded. Beads were resuspended with sheared DNA supernatant, incubated at room temperature for 30 min with rotation to bind biotinylated DNA to the streptavidin beads, separated on a magnet and solution was discarded. Beads were washed by adding 500 µl of 1× TWB. The tubes were heated on a Thermomixer at 55 °C for 2 min with mixing. The beads were reclaimed the beads using a magnet and the supernatant discarded. The wash was repeated twice. Beads were resuspended in 100 µl 1× Tris buffer (10 mM Tris-HCl, pH 8) and transferred to a new PCR tube. Beads were reclaimed and buffer discarded. Beads were resuspended in 50 µl 1× Tris buffer (10 mM Tris-HCl, pH 8). Then, the the beads binding with fragmented DNA were treated with the End Repair/dA-Tailing Module (NEB, E7442L) and Ligation Module (NEB, E7445L) following the operation manual. After ligation, the beads were washed with 1× TWB twice and 1× Tris buffer (10 mM Tris-HCl, pH 8). Beads were resuspended in 20 µl of 1× Tris buffer (10 mM Tris-HCl, pH 8). The sperm Hi-C library was amplified for 10 cycles of PCR with Q5 master mix (NEB, M0492L) following the operation manual. PCR products were confirmed by analysing 1 µl of product using the FlashGel System (Lonza, 57063). PCR was continued with additional cycles until bright DNA bands were seen. A bottle of Agencourt AMPure XP beads (Beckman Coulter, A63881) was warmed to room temperature and gently shaken to resuspend the magnetic beads. One hundred microlitres of beads was added to 200 µl of diluted PCR product (0.5× volumes). Samples were mixed by pipetting and incubated at room temperature for 10 min. Beads were pelleted on a magnet and the clear solution was transferred to a new tube. Another 30 µl of beads was added to the clear solution (0.65× volume), mixed by pipetting, and incubated at room temperature for 10 min. Keeping the beads on the magnet, samples were washed twice with 200 µl of 70% ethanol (freshly made) without mixing. Ethanol was then completely removed. Beads were left on the magnet for 5 min to allow the remaining ethanol to evaporate. DNA was eluted by adding 20 µl of ddH2O, mixing by pipetting and incubating at room temperature for 5 min. After separating on a magnet, the solution was transferred to a new tube. DNA was then quantified and sequenced using an Illumina sequencing platform.

CTCF western blot

To confirm the CTCF or RAD21 antibody, 0.5 million HEK 293 cells (obtained from ATCC: ATCC CRL-1573; tested for mycoplasma contamination but not authenticated) and 0.5 million 4T1 cells (obtained from ATCC: ATCC CRL-2539; tested for mycoplasma contamination but not authenticated) were resuspended separately in 50 µl 4% SDS and boiled for 5 min. To exclude the possibility of the CTCF antibody false positive detection on the alternate paralogue protein, we transfected HEK 293 cells and HeLa cells with human *CTCF* siRNA no. 1 (GGAG CCUGCCGUAGAAAUUTT)[28] and human *CTCF* siRNA no. 2 (CAGAG AAAGUGGUUGGUAA)[29] or control oligonucleotides for two-rounds (denoted as si*CTCF* cell lines and control cell lines). The HEK 293 cells and HeLa cells were also resuspended separately in 50 µl 4%SDS, boiled for 5 min. Protein concentration was assessed by BCA Protein Assay Kit (Cat no: CW0014, CWBIO). SDS–PAGE loading buffer (5×) (cat no: CW0027, CWBIO) was added into sample before loading to SDS–PAGE gel. The same total protein amount, 20 µg, was loaded.

For mouse and human sperm samples, 10 million sperm cells were resuspended separately in 50 µl 4%SDS, and sonicated using Diagenode Bioruptor sonication device for 10 cycles by 30 s on and 30 s off, then boiled for 30 min. Protein concentration was assessed by BCA Protein Assay Kit (cat no: CW0014, CWBIO). SDS–PAGE loading buffer (5×) (cat no: CW0027, CWBIO) was added into sample before loading to SDSPAGE gel. The same total protein amount, 30 µg, was loaded.

Samples were run on 4–12% SDSPAGE and transferred to 0.2-µm nitrocellulose membrane (Bio-Rad). Transfer was carried out in 25mM Tris base, 192 mM glycine, 20% methanol solution. The membranes were were stained with Ponceau red and photographed using a camera, then were blocked with 5% (w/v) milk buffer PBS/0.1% Tween 20) for 1 h at room temperature. The primary rabbit monoclonal antibody used was: CTCF (cat no: ab188408, Abcam) or RAD21 (cat no: ab992, Abcam) at a dilution of 1:1,000, 4 °C for overnight. Membranes were washed three times with TBST. The secondary antibody used was: goat anti-rabbit IgG, HRP conjugated (cat no: CW0103, CWBIO) at a dilution of 1:1,000, 1 h at room temperature. Membranes were washed three times with TBST. Signal development, followed the kit manufacturer's recommendations (cat no: CW0049, CWBIO). Excess reagent was removed and membrane was covered in transparent plastic wrap. Image was developed in a dark room for chemiluminescence or with digital imaging equipment.

ZGA block with α-amanitin treatment in human embryos

Zygotes were collected after IVF or ICSI. Zygotes were transferred to G1 media in the presence or absence of α-amanitin (25 ng µl, Sigma-Aldrich). After the control and α-amanitin-treated embryos reached the 8-cell stage, embryos of high quality were collected. Others with fragments or those that were arrested at other stages were discarded. The zona pellucida was gently removed by mechanical dissection with a glass needle. The cells were washed three times with 1× PBS to avoid potential contamination. They were then prepared for Hi-C, RNA-seq and immunostainning.

CTCF knockdown by siRNA injection and CTCF overexpression by mRNA injection in human embryos

To investigate the function of CTCF in human embryonic TAD establishment, two independent siRNAs targeting human *CTCF* were used for microinjection: human *CTCF* siRNA no. 1 (GGAGCCUGCCGUAGAAAUU TT)[28]; human *CTCF* siRNA no. 2 (CAGAGAAAGUGGUUGGUAA)[29] and negative control siRNA (UUCUCCGAACGUGUCACGUdTdT). For this assay, the donated IVM oocytes were used. They were in vitro-fertilized and cultured in G1.5 medium as described above. siRNA solution (10 mM) was loaded into injection pipette and injected into the zygote before pronuclear fading using Eppendorf PiezoXpert and Eppendorf CellTram vario microinjector. The injected embryos were cultured in G1.5 medium (Vitrolife) in a humidified atmosphere at 37 °C with 6% CO_2 in air. The injected embryos with normal morphology were collected at morula stage after fertilization. For RNA-seq library preparation, only one morula embryo was used for each assay. Morula embryos were collected for RNA-seq, Hi-C library generation and immunostainning, respectively. We collected three replicates in total for si*CTCF* morula: one replicate with human *CTCF* siRNA no. 1 (si*CTCF* morula no. 1) and two replicates with human *CTCF* siRNA no. 2 (si*CTCF* morula no. 2 and si*CTCF* morula no. 3).

To investigate whether human embryonic TAD establishment during ZGA is soley due to CTCF expression, we overexpressed CTCF in α-amanitin-treated embryos. The cDNA of complete open reading frame of human *CTCF* gene was cloned into the pCDNA3.1 vector. These plasmids were first linearized with NotI and mRNA was synthesized with SP6 polymerase using the mMessage mMachine kit (Ambion). The 1.5 µg μl^{-1} mRNA was injected into the zygotes. Zygotes were transferred to G1 medium in the presence of α-amanitin (25 ng μl^{-1}, Sigma-Aldrich). After these embryos reached the 8-cell stage, embryos with high quality were collected. Others with fragments or arrested at other stages were discarded. The zona pellucida was gently removed by mechanical dissection with a glass needle. The cells were washed three times with 1× PBS to avoid potential contamination. They were then prepared for Hi-C library generation and immunostainning.

SMART-seq2 library preparation

The human embryos were lysed directly and used for cDNA synthesis using SMART-Seq v.4 Ultra Low Input RNA Kit for Sequencing (Takara, 634888). In brief, the sample volume was adjusted to 9.5 µl with nuclease-free water. After adding 1 µl 10× reaction buffer (0.95 µl 10× lysis buffer, 0.05 µl RNase inhibitor), samples were incubated at room temperature for 5 min, then placed on ice. Next, 2 µl of 3′ SMART-Seq CDS

Primer II A (12 mM) was added. Following incubation at 72 °C for 3 min, samples were placed on ice for 2 min. cDNA synthesis reaction was set up by adding 4 µl 5× ultra low first-strand buffer, 1 µl SMART-Seq v.4 Oligonucleotide (48 mM), 0.5 µl RNase Inhibitor (40 U ml^{-1}) and 2 µl SMARTScribed reverse transcriptase. The reaction was performed in a thermal cycler with following program: 42 °C for 90 min, 70 °C for 10 min, 4 °C hold. The first-strand cDNA product was amplified by adding 25 µl 2× SeqAmp PCR Buffer, 1 µl PCR Primer II A (12 mM), 1 µl SeqAmp DNA Polymerase and 3 µl nuclease-free water. Sixteen rounds of PCR amplification were used with the following program: 95 °C for 1 min; 98 °C for 10 s, 65 °C for 30 s and 68 °C for 3 min, repeat these 3 steps 15 times; 72 °C for 10 min; 4 °C hold. The amplified cDNA was purified using 1 volume SPRIselect beads, then fragmented to 200–400 bp using a Covaris sonicator (Covaris). Sequencing libraries were prepared with NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645S) as described above. To obtain an adequate amount of DNA for sequencing, the cycle of PCR amplification was determined according to the amount of 1 µl amplified DNA, which was evaluated using FlashGel System (Lonza, 57063). The libraries were sequenced on Hiseq X10 with paired-end 150 bp (Illumina).

Immunostaining

After removal of zona pellucida, human embryos were washed three times in 1×PBS, and fixed in PFA solution (4% paraformaldehyde, 0.04% Triton, 0.3% Tween in 1×PBS) for 20 min at RT. After washing three times in 1×PBS for 5 min, the embryos were permeabilized in 1×PBS containing 0.5% Triton for 20 min. Embryos were incubated with blocking buffer (3% BSA, 0.01% Tween-20 in 1× PBS) for 2 h at room temperture, then incubated with CTCF antibodies (cat. no.: ab188408, Abcam, 1:500 dilution) for about 12 h at 4 °C. After washing 6 times in 0.1%Triton in 1× PBS (PBST), embryos were blocked in blocking buffer again for 30 min and then incubated for 1 h at room temperature with Alexa Fluor 488-labelled goat anti-rabbit IgG (Beyotime, P0176). Embryos were washed in PBST 3 times, then stained with DAPI (Beyotime, C1002) for 10 min. After washing in PBST 3 times, embryos were mounted in Prolong Gold antifade reagent (Thermo Fisher, P10144). Confocal images were obtained by Zeiss LSM 710 confocal microscope using a 63× oil objective. All the staining assays were repeated independently at least twice. Images were processed and quantified by ZEN software (2010) and ImageJ software (v.1.52a).

Hi-C data processing

The HiCExplorer suite[30,31] (v.2.1) was used for the processing of valid mapped Hi-C reads. Ultra-low-input Hi-C data were first mapped to human genome hg19 using bwa mem with parameters '-E50 -L0'. Then, we filtered out read pairs which were not uniquely mapped (mapping score <15). Dangling end reads, same fragment reads, self-circled reads, self-ligation reads and other invalid Hi-C reads were also discarded. Details about data quality are summarized in Supplementary Table 1. After removing duplication, reads were used to generate raw Hi-C matrix at 10 kb, 40 kb, 100 kb and 200 kb resolution using hicBuildMatrix. In the raw contact matrix, rows and columns with zero or small total counts were removed because these bins were mostly from repetitive regions. After filtering low-count bins, the matrices were corrected by the ICE method[32] using hicCorrectMatrix. We also converted valid read pairs into .hic format files with the juicer tool[33] pre command and use these .hic files for heat map visualization.

In addition, to avoid read depth influence on data interpretation, we also randomly picked up equal reads for each stage in the analysis when mentioned.

Reproducibility score

Reproducibility score is calculated from GenomeDISCO software[34]. It uses random walks on the contact map graph for smoothing before comparing the interaction maps, resulting in a reproducibility score that can be used for quality control of biological replicates. We calculated GenomeDISCO reproducibility score at 200-kb resolution for biological replicates.

TAD and TAD boundary identification

TADs and TAD boundaries (immature and mature TAD boundaries) were identified based on public TAD separation score method[30,31] with hicFindTADs command in HiCExplorer suite (v.2.1)[30,31]. The TAD separation scores in human embryos were calculated at 40 kb resolution using parameters '–minDepth 300000 –maxDepth 3000000 –step 300000–minBoundaryDistance 400000 –thresholdComparisons 0.01 –correctForMultipleTesting fdr –delta 0.01'.

The principle for TAD separation score and boundary detection is based on a graph clustering called the conductance of a cut. The conductance of a cut measure calculates the interactions between two clusters and the interactions within the clusters. It allows the identification of good clustering characterized by few interactions between clusters. It is similar to the TAD feature. Thus, TADs can be regarded as graph clusters of consecutive nodes (bins), and domain boundaries can be regarded as genomic positions that best separate two clusters (TADs).

The hicFindTADs command first transformed the Hi-C contact matrix into a z-score matrix $A = a_{ij}$. In the z-score matrix, each contact frequency is transformed into a z-score from the input contact matrix based on the distribution of all contacts at the same genomic distance. For an arbitrary bin l, the contacts between an upstream and downstream region of length w are all included in the z-score $\mathbf{A}[\alpha_l, \beta_l]$with row index $\alpha_l \in \{l - w, \ldots, l\}$ and column index $\beta_l \in \{l, \ldots, l + w\}$. For each matrix bin, we compute the TAD separation score (w) as the mean value of $\mathbf{A}[\alpha_l, \beta_l]$. To reduce noise and improve robustness, a multi-scale version of the TAD-separation score is computed for different values of w that are averaged per bin from 300,000 to 30,000,000 with step 300,000 in an exponential size-growth way (minDepth + (step × int(x)$^{1.5}$) for x in [0,1,2,...]). Genomic bins with a low TAD-separation score (local minima) are indicative of boundaries and stronger boundaries will have lower scores. For the local minima bin i, $i - v$ and $i + v$ are the bins at distance v upstream and downstream of i, respectively. We use the Wilcoxon rank-sum test to compare the values of $\mathbf{A}[\alpha_i, \beta_i]$ with the values of the other two submatrices $A[\alpha_{i-v}, \beta_{i-v}]$ and $A[\alpha_{i+v}, \beta_{i+v}]$, respectively. The higher of the two P values is used. We corrected the P values using the false discovery rate (FDR) method, and regions with $q < 0.01$ were reported as boundaries. TADs were identified as regions between two boundaries.

Directional index and insulation score

Directional index[9] can quantify the degree of upstream or downstream bias of a given bin. We calculated it for each bin based on the ICE-normalized and depth-normalized matrix as previously described[8].

Insulation score was calculated as previously described[12] with the public code on Github (matrix2insulation.pl; https://github.com/dekkerlab/giorgetti-nature-2016). A sliding 480 kb × 480-kb square along the matrix diagonal was used. The IQR mean signal within the square was then assigned to the each 40-kb diagonal bin. This procedure was then repeated for all 480-kb diagonal bins. The insulation score was normalized relative to all of the insulation scores across each chromosome by calculating the $\log_2$ ratio of each bin's insulation score versus the mean of all insulation scores. Valleys or minima along the normalized insulation score vector represent the loci of reduced Hi-C interactions that occur across the bin. These valleys or minima are interpreted as areas of high local insulation.

TAD structure strength

TAD structure strength was quantified as the ratio of the interactions within-TAD over the between-TAD intensity as previously described[26,35]. For 90 × 90 contact enrichment map, the interactions within-TAD was the sum of the central square of the enrichment map, rows 30 to 59 and

columns 30 to 59 ([30:60, 30:60]) and the interactions in upstream and downstream neighbouring blocks was the between-TAD intensity, half of the sums of the regions [0:30, 30:60] and [30:60, 60:90]. TAD structure strength was both calculated for replicates and the merged.

TAD signal variance

TAD signal variance can indicate the strength of the TAD structure and calculated as previously described[36]. We used intra-chromosomal maps at 40-kb resolution for each embryo stage with equal reads. In brief, TAD signal was calculated as the $\log_2$ ratio of the number of corrected upstream-to-downstream interactions within a 2-Mb window in each chromosome. Regions which contained less than 10 counts or gaps within a 2-Mb distance were filtered out. The variance was calculated for individual chromosomes. Comparisons between different stages were performed only for the bins shared by all of the different stages. Relative TAD signal variance was calculated relative to the minimum TAD signal variance. We also performed a downsampling analysis for human sperm and human blastocysts. Variance due to sampling noise was confirmed to be proportional to 1/reads. The Wilcoxon rank-sum test was applied to test statistical significance.

Hi-C data mapping for mixed sample

After sequencing for the mixed samples, the human reads and mouse reads were separated from the mixture by using human genome reference and mouse genome reference. We mapped the mixed sample data to human hg19 genome and mouse mm10 genome with bwa mem with parameters '-E50 -L0', respectively. Separated data for human and mouse were processed as mentioned in the section 'Hi-C data processing'. In addition, we also combined the hg19 and mm10 genome together as reference genome. We mapped the reads to the combined reference genome. Those reads, with one end mapped to human and the other mapped to mouse genome, were all discarded.

A/B compartmentalization, compartment strength and A/B compartment dynamics

PC1 values from principal component analysis on correlation heat maps have been used to indicate the A/B compartment status[10]. We used HOMER[37] software with parameters '–res 100,000, -superRes 400,000' to obtain the PC1, PC2 and PC3 value and parameters '–res 500,000, -superRes 500,000' for equal read depth samples. Because sometimes PC2 or PC3 values reflect A/B compartments, we manually checked of PC1, PC2 and PC3 track with gene density and the plaid pattern in the correlation heat maps along each chromosome and got final 'PC1' list. The direction of the Eigen (PC) values is arbitrary, and therefore positive values were set to 'A' and negative values were set to 'B' based on their association with gene density.

Compartment strength was defined as the natural logarithm of the $AA \times BB/AB^2$ according to the public description[26]. Enrichment contact maps of A/B interactions in Hi-C data were calculated at 500-kb resolution. A GC content profile was separated into 5 bins: minimum to 20th percentile, 20th percentile to 40th percentile and so on. For each pixel of the 5 × 5 matrix, O/E values were then calculated for loci belonging to each pair of bins as contact enrichment. GC content is known to highly correlate with the compartment profile. Thus GC content was used for compartment strength instead of PC1 values. We used the top 20% of GC content for A, and the bottom 20% GC content for B. For the error bar in evaluating the compartment strength in Fig. 1d, we obtained 100 5 × 5 compartment enrichment matrices by bootstrapping. For each pixel of the 5 × 5 compartment enrichment map, we took all the observed-over-expected values that contributed to this pixel and took a random sample with replacement of the same size that the contributing values. We then proceeded with downstream for each of the 100 reshuffled maps.

We identified A/B compartment dynamics regions as statistically significant variability in PC1 values across developmental stages using analysis of variance with adjusted P value < 0.05. Genes with A/B compartment switches were analysed for Gene Ontology (GO) enrichment with DAVID 6.8 (with parameters count ≥3; ease ≤0.05).

Density of paired loci at different interaction insert sizes

We examined the relative abundance of paired loci at different insert sizes in human sperm and mouse sperm as previously described[8]. We assigned the unique paired reads to restriction enzyme fragments and plotted the distribution of the $\log_{10}$ of insert size between fragments.

Contact probability

Contact probability was calculated as previously described[36] and implemented with command line tool 'pairsqc.py' (https://github.com/4dn-dcic/pairsqc). It first divides distances into logarithmically spaced bins at $\log_{10}$ scale at interval of 0.1 (growing by about 1.25-fold). For each bin, contact probability is computed as number_of_reads / number_of_possible_reads/bin_size. Number_of_reads is the number of interactions at corresponding distances. Number_of_possible_reads is computed as the sum of L_chr − s_mid − 1 over all chromosomes, where L_chr is the length of a chromosome. S_mid is the mid point of the bin at $\log_{10}$ scale (bin $10^{5.0} - 10^{5.1}$ has mid point $10^{5.05}$). Bin_size is computed as max distance − min distance (for example, for bin 10**5.0 ~ 10**5.1, the binsize is 10**5.1 − 10**5.0).

Region comparison between human genome and mouse genome

For the region comparison between human and mouse, regions were shifted using the UCSC liftOver tool online for hg19 and mm10 genome with default parameters. For ZGA boundary overlap calculation, we lifted the mouse ZGA boundaries to human hg19 genome.

Identification of stage-specific gained boundaries

To analyse insulated boundaries between embryo stages, we first merged boundaries within 200 kb of each other across all examined embryo stages as common boundary reference (200-kb collapsed boundaries). It increased the robustness of boundary comparison. For stage-specific gained boundaries, we selected those boundaries that did not exist in previous stages, first emerged in the indicated stage and persisted until in blastocysts. For example, human 8-cell gained boundaries refer to this kind of boundary: they do not function as insulated boundaries in the 2-cell stage, and at the 8-cell stage they can be observed with evident insulation

Enrichment analysis of repeats on stage-specific gained boundaries

The repeat frequency in the genome was calculated for every 1-kb bin with bedtools coverage command. For generating the enrichment plots for repeats, the mid-point of each stage-specific gained boundary region was identified, and repeat frequency was calculated in 25-kb bins for ±1 Mb from the boundary mid-point centre.

FIREs calling

FIREs calling was based on the previous method[23]. For each 40-kb bin, the raw count was calculated as the total number of chromatin interactions within 200 kb. A normalization pipeline, 'HiCNormCis'[23], was applied to normalize raw count. A Poisson regression model was fitted for each 40-kb bin taking the raw count as the outcome variable, and three local genomic features, such as effective fragment length, GC content and mappability score, as the covariates. The residuals from the Poisson regression model were used as the HiCNormCis output. Next, for each 40-kb bin, z-score was calculated based on the HiCNormCis output. FIREs are defined as bins with a one-sided P value less than 0.05, corresponding to $-\ln(P)$ greater than 3. $-\ln(P)$ was chosen as the FIRE score.

DNA methylation data analysis

The human early embryos and sperm DNA-methylation data were downloaded from CRA000114[38], which had been deposited in the Genome

Sequence Archive (GSA). Low-quality reads were removed by Trimmomatic v.0.32. The filtered reads were aligned to hg19 genome by Bismark with default parameters. PCR duplicates were removed by Picard and overlapped parts of pair-end reads were trimmed from one end by bamUtil. The methylation level of each CpG site was calculated by a custom script.

DNase-seq data analysis and ATAC-seq data analysis

DNase-seq data was analysed as previously described[25]. Low-quality reads were removed and the remaining reads were cropped to 100 bp by Trimmomatic v.0.32. Paired reads1 and unpaired reads were used for mapping to human genome hg19 with Bowtie v.1.2.0[39] with parameter '-m 1'. Low mapping quality (MAPQ < 10) and PCR-duplicated reads were removed by SAMtools and Picard[40,41]. DHSs were called by hotspot algorithm with FDR <0.01 (ref. [42]). DHSs covered by less than 8 reads were filtered out. DNase-seq signal tracks for visualization were generated by bamCoverage in Deeptools2 suite[43]. ATAC-seq data were downloaded and analysed from GEO: GSE101571[24].

RNA-seq data processing

Human early embryo RNA-seq raw data were downloaded from GSE36552[22]. We used Refseq genes from the UCSC genome browser and Ensemble genes as gene annotation. Housekeeping gene annotation are downloaded from a previous publication[44]. Low quality RNA-seq reads and adaptor sequences were removed by Trimmomatic v.0.32. Then reads were mapped to the hg19 genome by STAR (v.2.5.2b)[45] with default parameters. Gene expression FPKM was calculated by HOMER software. The ZGA genes were defined as those obviously upregulated in human 8-cell embryos (FPKM >1 in 8-cell embryos and fold change >3 comparing with 2-cell embryo)[25]. To analyse the expression of Alu repeats in human early embryos, the total RNA-seq data were downloaded from GSE85632[21]. The reference repeat annotation was downloaded from rmsk in UCSC Table Browser. The FPKM value of each repeat or transposon was calculated by HOMER analyseRepeats.pl with parameters "-fpkm" and plotted as tracks in Extended Data Fig. 7j.

To find differential expression genes between α-amanitin 8-cell and control 8-cell or between si*CTCF* treatment and siControl morula, HOMER software command getDiffExpression.pl with –DESeq2 was employed. FDR <0.05, |fold change| > 2 were used as cut-off for significant differently expressed genes.

RNA-seq tracks for visualization were generated by bamCoverage program in DEEPtools2 with parameter '–normalizeUsingRPKM'.

Statistics and reproducibility

R and Prism were used for statistics analysis. Wilcoxon rank-sum test was used for statistical significance (Figs. 1b, 4b, Extended Data Figs. 2i, 4b, left, 4d–f, 7b, 8a, 9e, 10d, e) with the wilcox.test function in R. χ^2 test was applied for the statistical significance of the overlap between human ZGA boundaries and mouse ZGA boundaries relative to random (Fig. 3c) and the enrichment of FIREs in A compartments (Extended Data Fig. 10c). Two-sided Kolmogorov–Smirnov test were used for comparison the distance distribution of stage-gained insulated boundaries to the closet housekeeping genes in Fig. 3d and Extended Data Fig. 7a. One-sided *t*-test is used to calculate statistical significance for A/B compartment strength (Fig. 1d) and differential FIRE score (si*CTCF* – control) in FIREs and non-FIREs (Extended Data Fig. 10f). Fisher's Exact test was used for GO enrichment analysis by DAVID tool (v.6.8) with $P < 0.05$ cut-off in Extended Data Fig. 4b. For box plots, boxes represent the 25th, 50th and 75th percentiles and whiskers show 1.5× the interquartile range. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$; NS, $P > 0.05$.

Reporting summary

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

Data availability

Data generated for this study have been deposited to the Genome Sequence Archive with the accession number CRA000852 and CRA000108. Hi-C data of unmixed human sperm sample have been deposited in CRA000108, and the other data have been deposited in CRA000852. Raw image files used in the figures that support the findings of this study are available from the corresponding authors upon reasonable request.

Code availability

Codes used for the analysis reported in this study are available at https://github.com/ChenXP0310/2019-humanembryo3D.

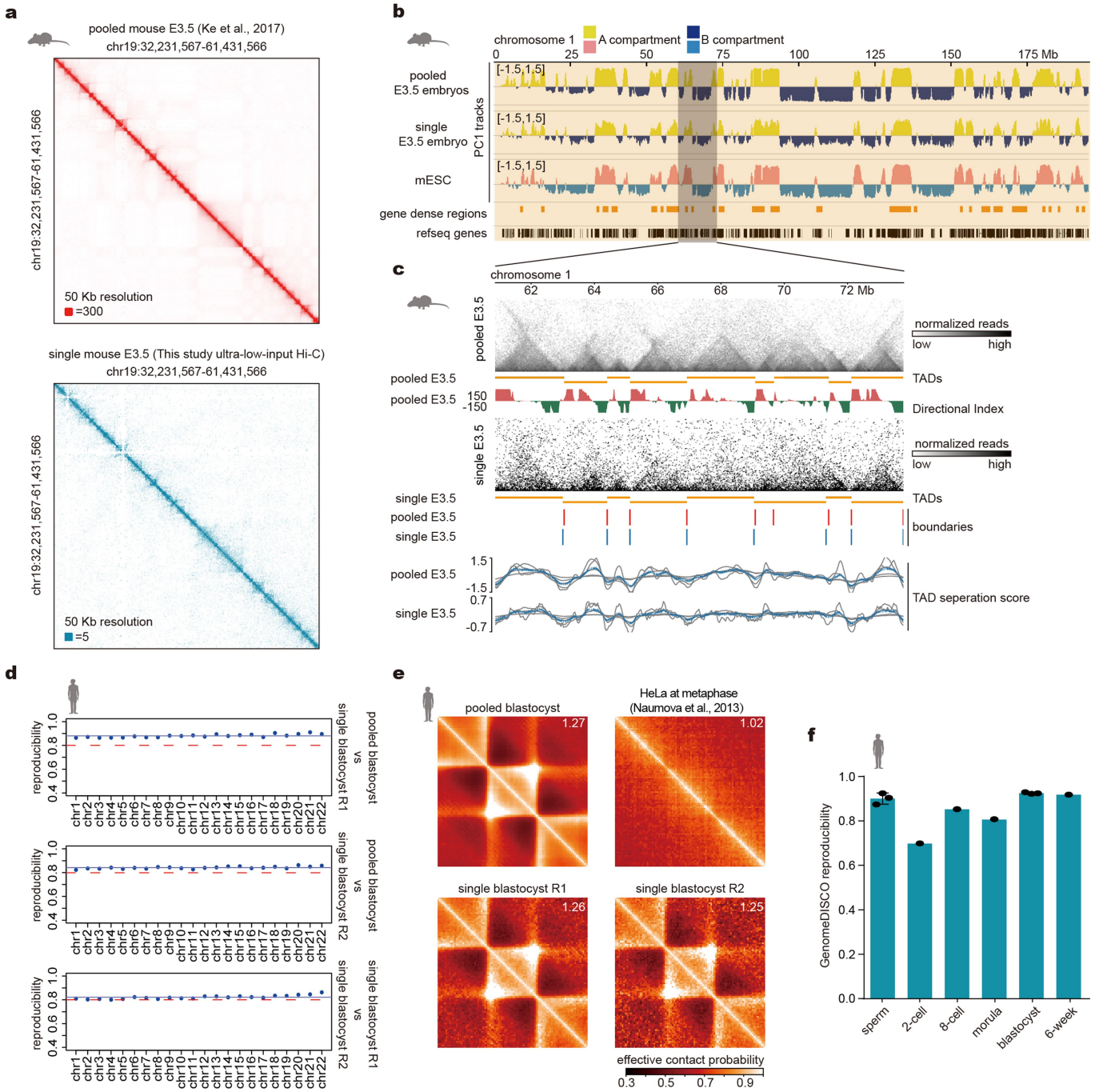
27. Roy, T. K., Bradley, C. K., Bowman, M. C. & McArthur, S. J. Single-embryo transfer of vitrified-warmed blastocysts yields equivalent live-birth rates and improved neonatal outcomes compared with fresh transfers. *Fertil. Steril.* **101**, 1294–1301 (2014).
28. Zuin, J. et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl Acad. Sci. USA* **111**, 996–1001 (2014).
29. Guillou, E. et al. Cohesin organizes chromatin loops at DNA replication factories. *Genes Dev.* **24**, 2812–2822 (2010).
30. Ramírez, F. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
31. Wolff, J. et al. Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **46** (W1), W11–W16 (2018).
32. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
33. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
34. Ursu, O. et al. GenomeDISCO: a concordance score for chromosome conformation capture experiments using random walks on contact map graphs. *Bioinformatics* **34**, 2701–2707 (2018).
35. Gassler, J. et al. A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J.* **36**, 3600–3618 (2017).
36. Naumova, N. et al. Organization of the mitotic chromosome. *Science* **342**, 948–953 (2013).
37. 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).
38. Li, G. et al. Genome wide abnormal DNA methylome of human blastocyst in assisted reproductive technology. *J. Genet. Genom.* **44**, 475–481 (2017).
39. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
40. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
41. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
42. John, S. et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* **43**, 264–268 (2011).
43. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44** (W1), W160–W165 (2016).
44. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends Genet.* **29**, 569–574 (2013).
45. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
46. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
47. Consortium, E. P.; ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

Acknowledgements This work was supported by the grants from The Ministry of Science and Technology of China (2018YFC1004000 and 2018YFC1003300), National Natural Science Foundation of China (81430029, 91731312, 81871171, 31425015, 31630040, 31871454, 81622021 and 81871168), CAS funding (QYZDY-SSW-SMC016) and Strategic Priority Research Program of the Chinese Academy of Sciences (XDB13040200).

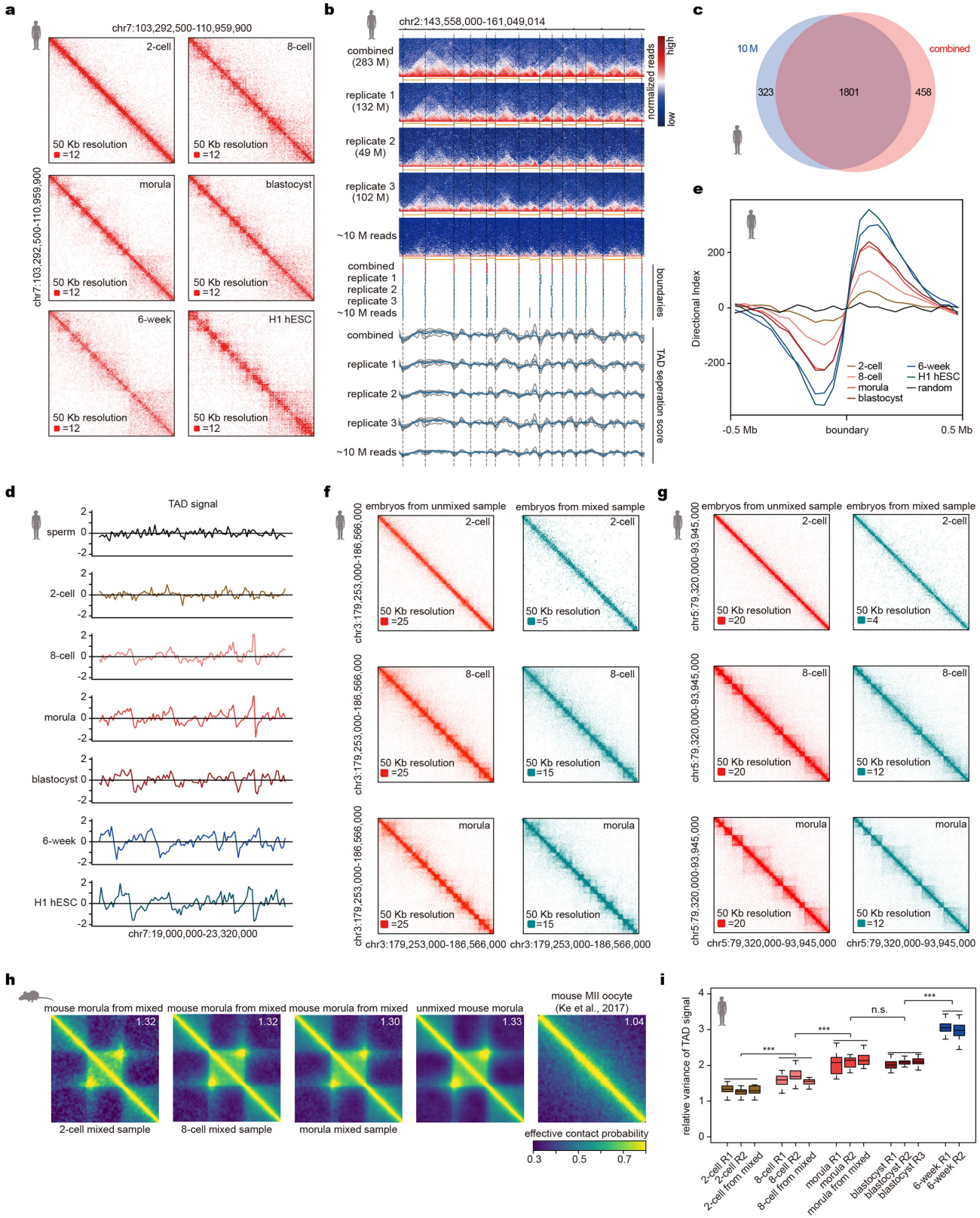
Author contributions J.L. and Z.-J.C. conceived the study. X.C., Y.K. and K.W. facilitated its designs. K.W., H.Z., J.Z., W.T. and H.L. collected human embryos. K.W. and Z.H. performed siRNA-microinjection in human embryos. Y.K., X.C. and Y.S. performed Hi-C library construction. Z.L. performed RNA-seq library construction. X.C., Y.S., and L.G. performed the bioinformatics analyses. X.C., Y.K., K.W., Z.-J.C. and J.L. interpreted the data. X.C., Y.K., Z.-J.C., and J.L. wrote the paper with the assistance of the other authors.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1812-0.
Correspondence and requests for materials should be addressed to J.L. or Z.-J.C.
Reprints and permissions information is available at http://www.nature.com/reprints.

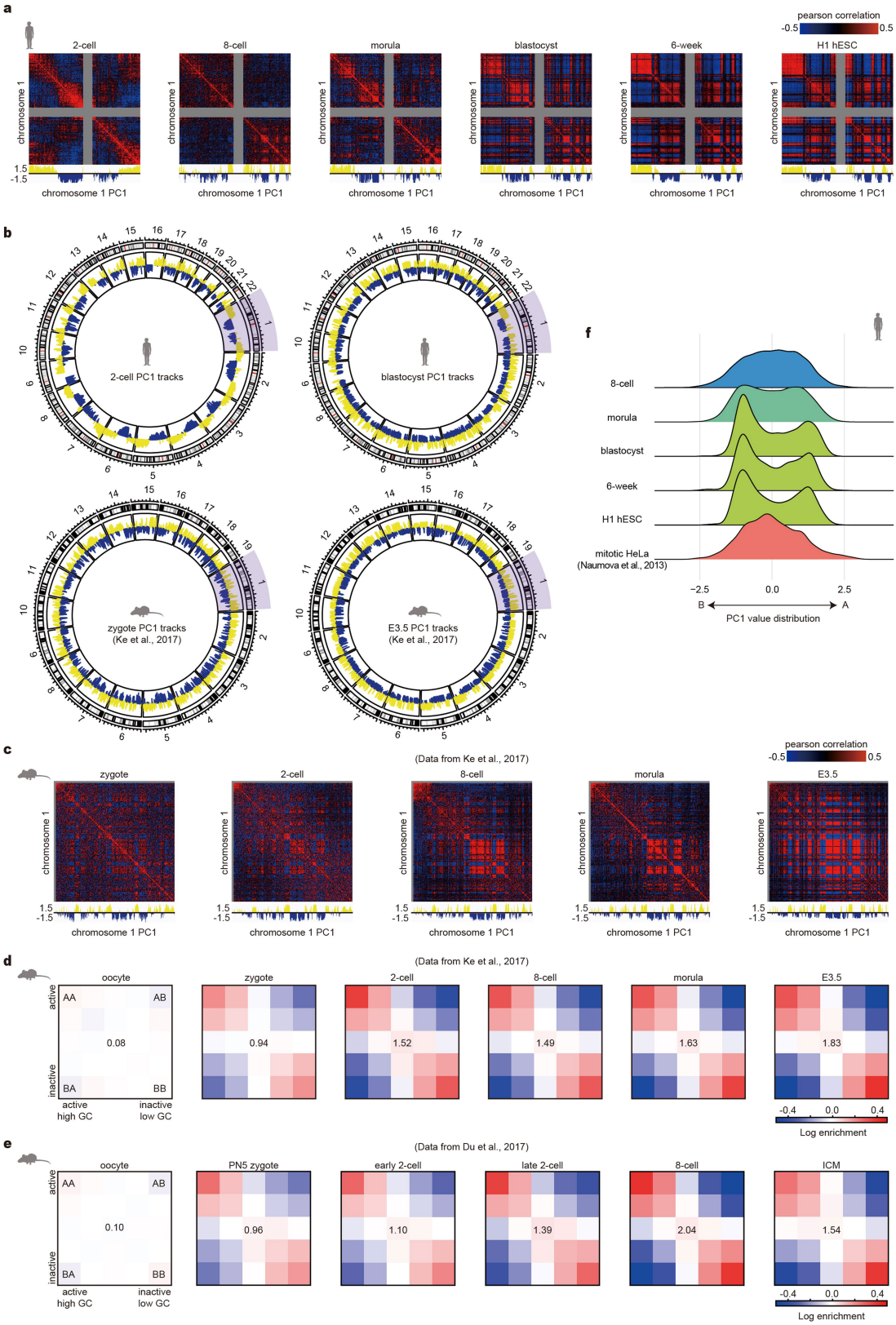


Extended Data Fig. 1 | Validation of optimized ultra-low-input Hi-C. **a**, Interaction heat maps for mouse embryonic day (E)3.5 embryos. Top, pooled mouse E3.5 embryos Hi-C[8]. Bottom, a single mouse E3.5 embryo with ultra-low-input Hi-C from this study. **b**, A track snapshot of PC1 values for chromosome 1 in pooled mouse E3.5 embryos and single mouse E3.5 embryo along with public mouse ES cell (mESC) data[9]. **c**, Interaction heat maps for pooled mouse E3.5 embryos and single mouse E3.5 embryo for a zoomed-in region from chromosome 1 overlaid with TADs, directional index, boundaries and TAD separation scores. **d**, GenomeDISCO reproducibility (Methods) among the Hi-C data for each chromosome in pooled human blastocysts and two human single-blastocyst replicates. **e**, The strength of average TADs in pooled human blastocyst data, single human blastocyst data and metaphase HeLa data[36]. TAD positions are annotated in combined human blastocyst data. TAD structure strength (Methods) is shown on each corresponding panel. **f**, GenomeDISCO reproducibility for biological replicates of human sperm and embryos. $n = 2–3$ independent biological replicates for each stage. Each dot represents one comparison between two biological replicates. Data are mean ± s.d.

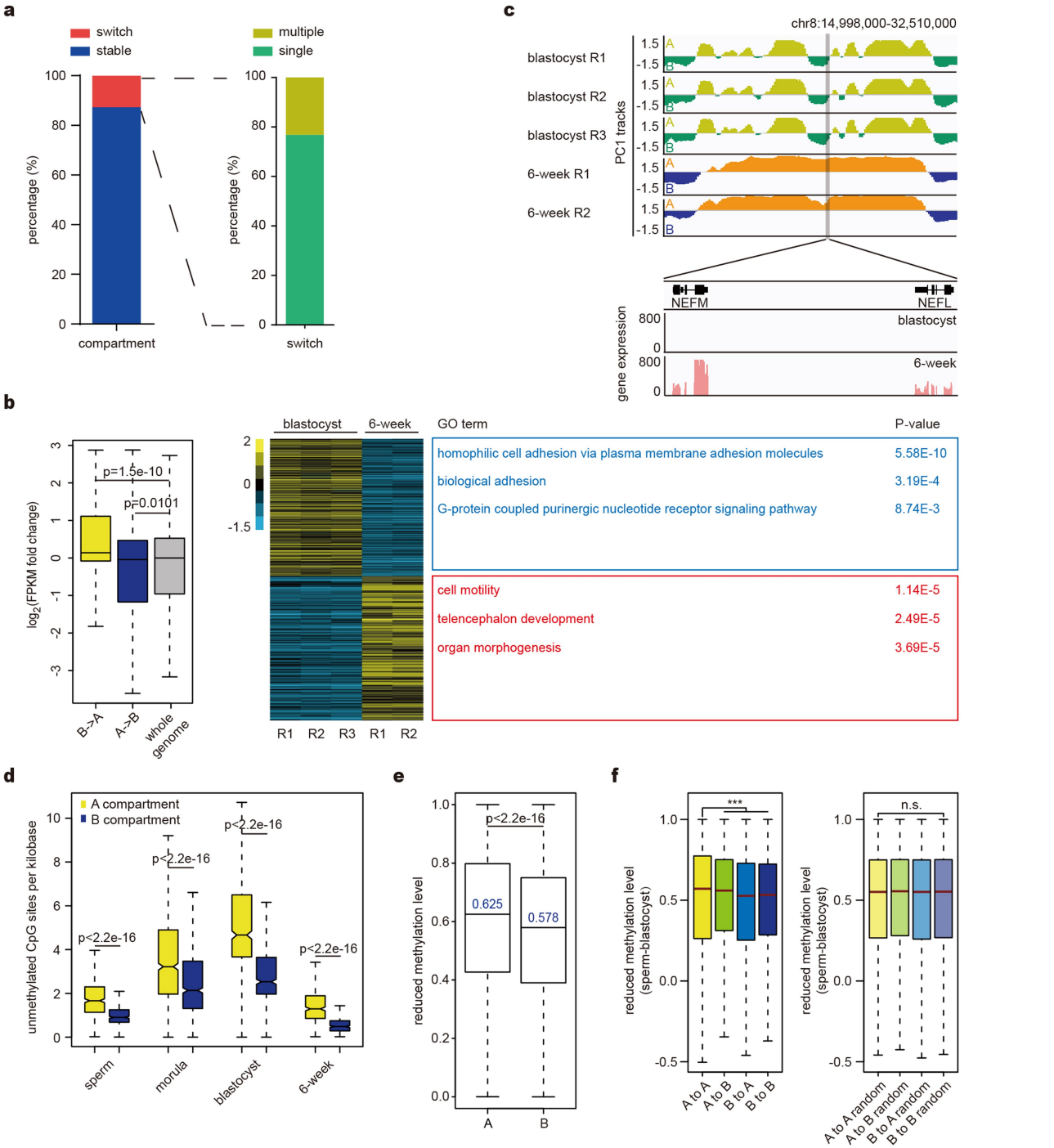


Extended Data Fig. 2 | See next page for caption.

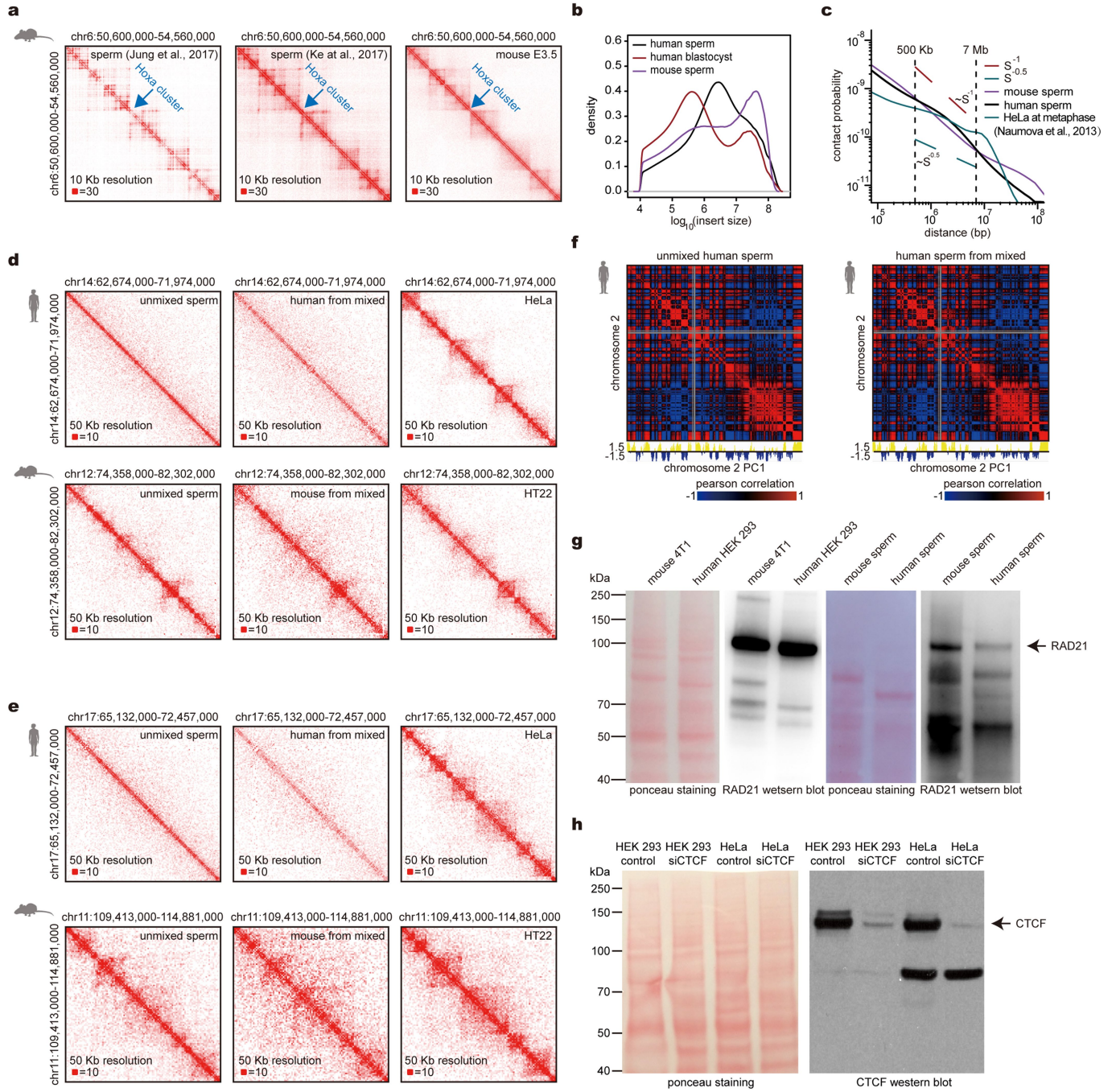
Extended Data Fig. 2 | TAD establishment during human embryonic development. a, Interaction heat maps for human embryos with equal reads along with human H1 ES cells. **b**, An example for TAD calling by using TAD separation score method in human blastocysts with different read depths. The orange track under each heat map represents TAD domains. The bottom panels are TAD boundaries and TAD separation score tracks. M, million reads. **c**, Venn diagram for the overlap of TAD boundaries of human blastocyst between 10 million read depth data (10 M, blue) and total combined data (total, red). **d**, An equal-read-depth TAD signal track snapshot for human sperm and embryos along with human H1 ES cells. **e**, Average directional index around human TAD boundaries(0.5 Mb) at different embryonic stages. Directional index generated by a random blastocyst valid read pair dataset is also shown as a control. **f**, **g**, Interaction heat map examples at 50-kb resolution for unmixed human embryos and human embryos from mixed samples. **h**, The strength for average TADs in mouse morula embryos from mixed samples (six mouse morulae for each mixed sample), unmixed mouse morula and mouse MII oocyte. TAD positions are annotated in combined mouse E3.5 embryos. TAD structure strength is shown on each corresponding panel. **i**, Relative variance of TAD signal for replicates of human embryos including human embryos from mixed samples. We used the chromosome as the unit for this analysis. Two-cell ($n = 3$), 8-cell ($n = 3$), morula ($n = 3$), blastocyst ($n = 3$) and 6-week embryos ($n = 2$). Boxes represent the 25th, 50th and 75th percentiles and whiskers show 1.5× the interquartile range. ***adjusted $P < 0.001$ for all pairwise comparison between two stage samples (two-sided Wilcoxon rank-sum test with Benjamini–Hochberg multiple testing correction). NS, not significant.



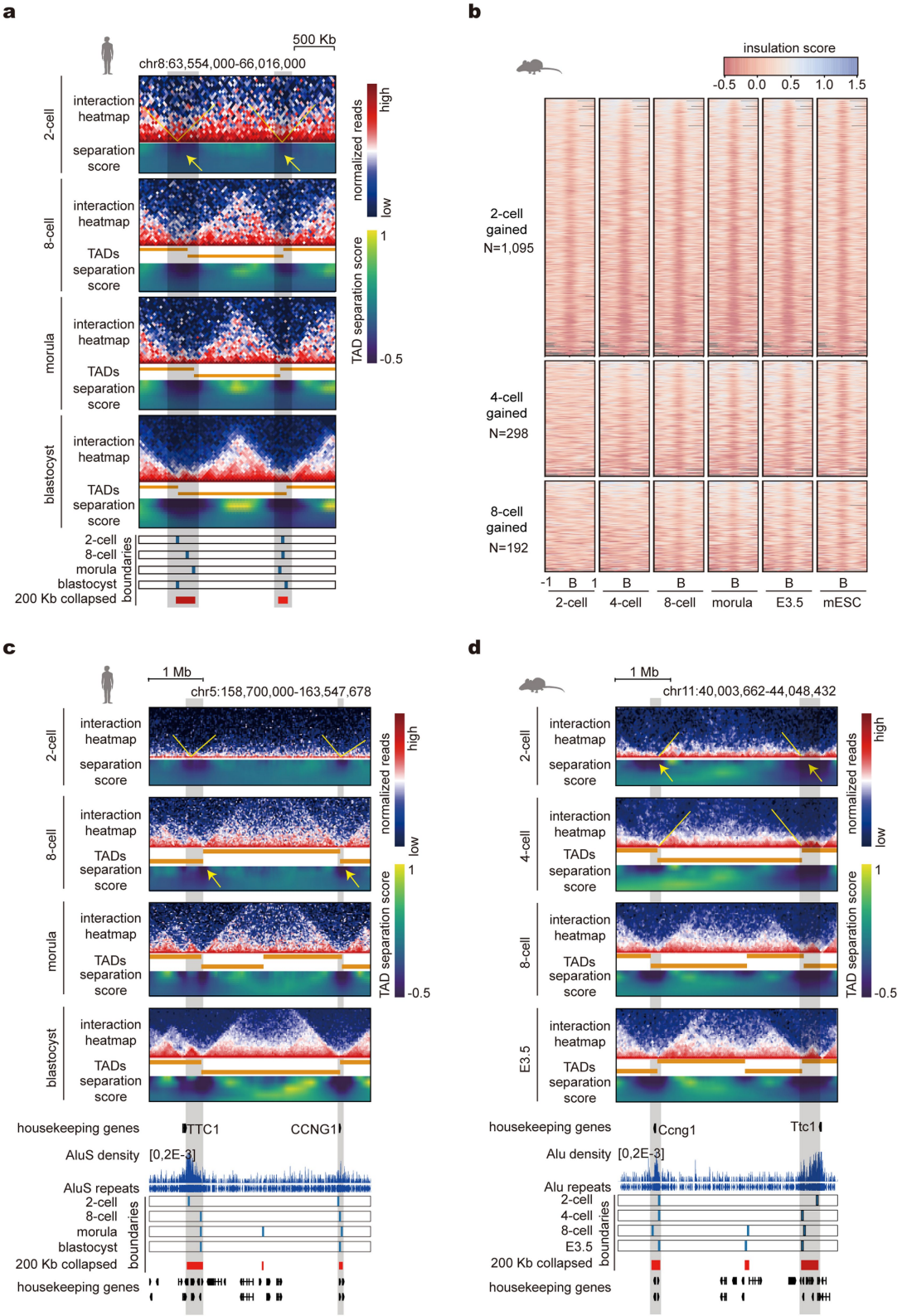
Extended Data Fig. 3 | A/B compartmentalization dynamics during human embryonic development. a, Pearson correlation heat maps of chromosome 1 at 500-kb resolution in human 2-cell, 8-cell, morula and blastocysts, 6-week embryos and H1 ES cells (equal no. of reads generated from 2–3 biological independent replicates for each stage). **b**, Top, whole-genome PC1 value circle plots for human 2-cell embryos and human blastocysts (with R package circlize[46]). Bottom, PC1 tracks in mouse zygotes and mouse E3.5 embryos. **c**, Pearson correlation heat maps at 500-kb resolution of chromosome 1 in mouse zygotes and early embryos[8] (equal number of reads generated from 2–3 biological independent replicates for each stage). **d**, **e**, Five-by-five contact enrichment maps of A–B compartments averaged over genomic positions using GC content in mouse early embryo Hi-C data[8] (**d**) and ref. [7] (**e**). **f**, PC1 value distribution in human embryos and mitotic HeLa cells[36].



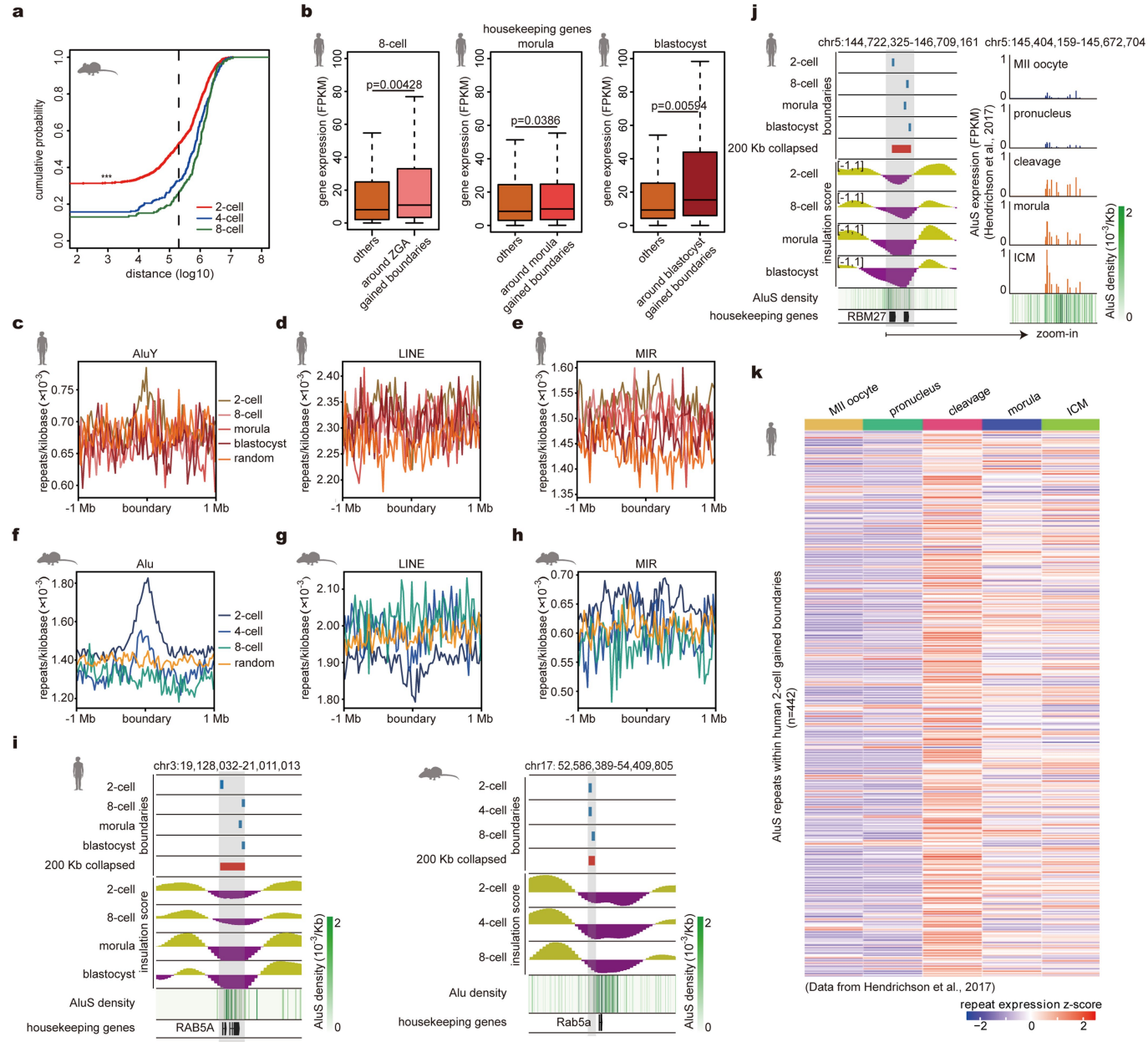
Extended Data Fig. 4 | A/B compartment switches during human embryonic development. a, The percentage of genomic regions with A/B compartment switches from morula to 6-week stage, and the percentage of single A/B switched regions relative to total switched regions. **b**, Left, the box plot illustrates the gene expression dynamics with A/B compartment switches between human blastocysts and 6-week embryos (biological replicates pooled; blastocyst, $n = 3$; 6-week embryo, $n = 2$). *P* values are shown (one-sided Wilcoxon rank-sum test). Boxes represent the 25th, 50th and 75th percentiles and whiskers show 1.5× the interquartile range (**b**, **d**–**f**). Middle, PC1 heat map for A/B compartment switched regions. Heat map is sorted according to A/B compartment switched PC1 values. Right, GO enrichment for genes with A/B compartment switches between blastocysts and 6-week embryos using DAVID v.6.8. **c**, A/B compartment status around the *NEFM*–*NEFL* locus in human blastocysts and 6-week embryos overlaid with gene expression. **d**, Unmethylated CpG (methylation level <0.25) density within A or B compartments in human sperm and embryos (biological replicates pooled; sperm, $n = 3$; morula, $n = 2$; blastocyst, $n = 3$; 6-week embryos, $n = 2$). *P* values are shown (one-sided Wilcoxon rank-sum test). **e**, Reduced CpG methylation level (ML) ($ML_{sperm} - ML_{blastocyst}$) in A compartments and B compartments of human blastocyst (biological replicates pooled; $n = 3$). *P* value is calculated by one-sided Wilcoxon rank-sum test. **f**, Reduced methylation level ($ML_{sperm} - ML_{blastocyst}$) for regions with A-to-A, A-to-B, B-to-A, and B-to-B compartment switches from human sperm to blastocysts (replicates pooled; $n = 3$). ***$P < 0.001$ for comparisons between 'A-to-A' and other groups (two-sided Wilcoxon rank-sum test). Right, randomly shuffled A/B compartment switched regions as control.



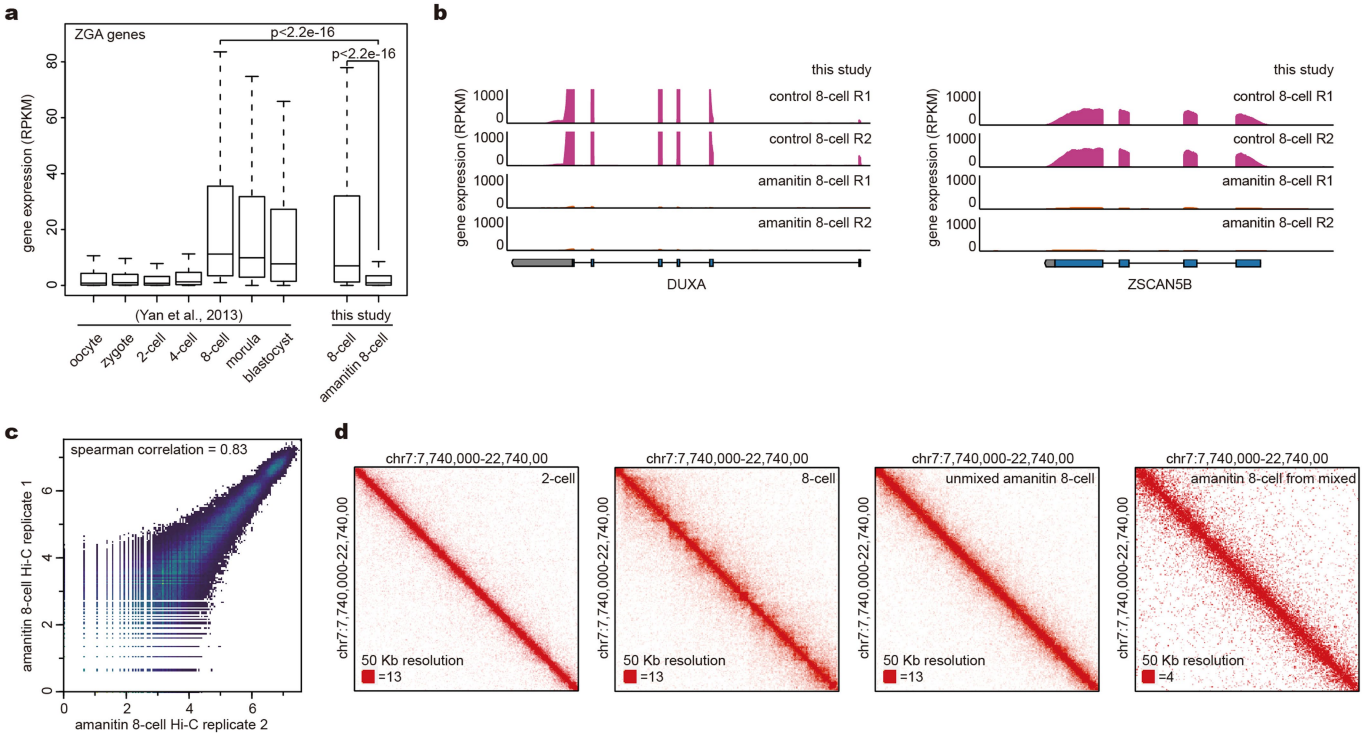
Extended Data Fig. 5 | Human sperm have no typical TADs. a, Interaction heat maps around *Hoxa* cluster in mouse sperm[8,16] and mouse E3.5 embryos[8]. **b**, Density plot of interaction insert size for human sperm, human blastocysts and mouse sperm (Methods). **c**, Contact probability decay curve for human sperm, mouse sperm and metaphase HeLa cells. **d**, **e**, Examples of interaction heat maps at human and mouse conserved syntenic regions from unmixed sperm and mixed human and mouse sperm sample. Interaction heat maps in **d** (right) and **e** (right) are for HeLa and HT22 cells somatic cell mixed samples, respectively. **f**, Pearson correlation heat maps of chromosome 2 at 500-kb resolution in human sperm. Left, unmixed human sperm Hi-C. Right, human sperm from the mixed sperm sample (replicates pooled). **g**, RAD21 western blots and Ponceau staining of samples from somatic cell lines and human and mouse sperm. Black arrow indicates RAD21 band. Experiment repeated on two biologically independent replicates per sample. For gel source data, see Supplementary Fig. 1. **h**, Right, CTCF western blots in the control and knockdown cell lines (control and si*CTCF* HEK 293; control and si*CTCF* HeLa). Left, Ponceau staining. Black arrow indicates CTCF band. Experiment repeated on two biologically independent replicates per sample. For gel source data, see Supplementary Fig. 1.



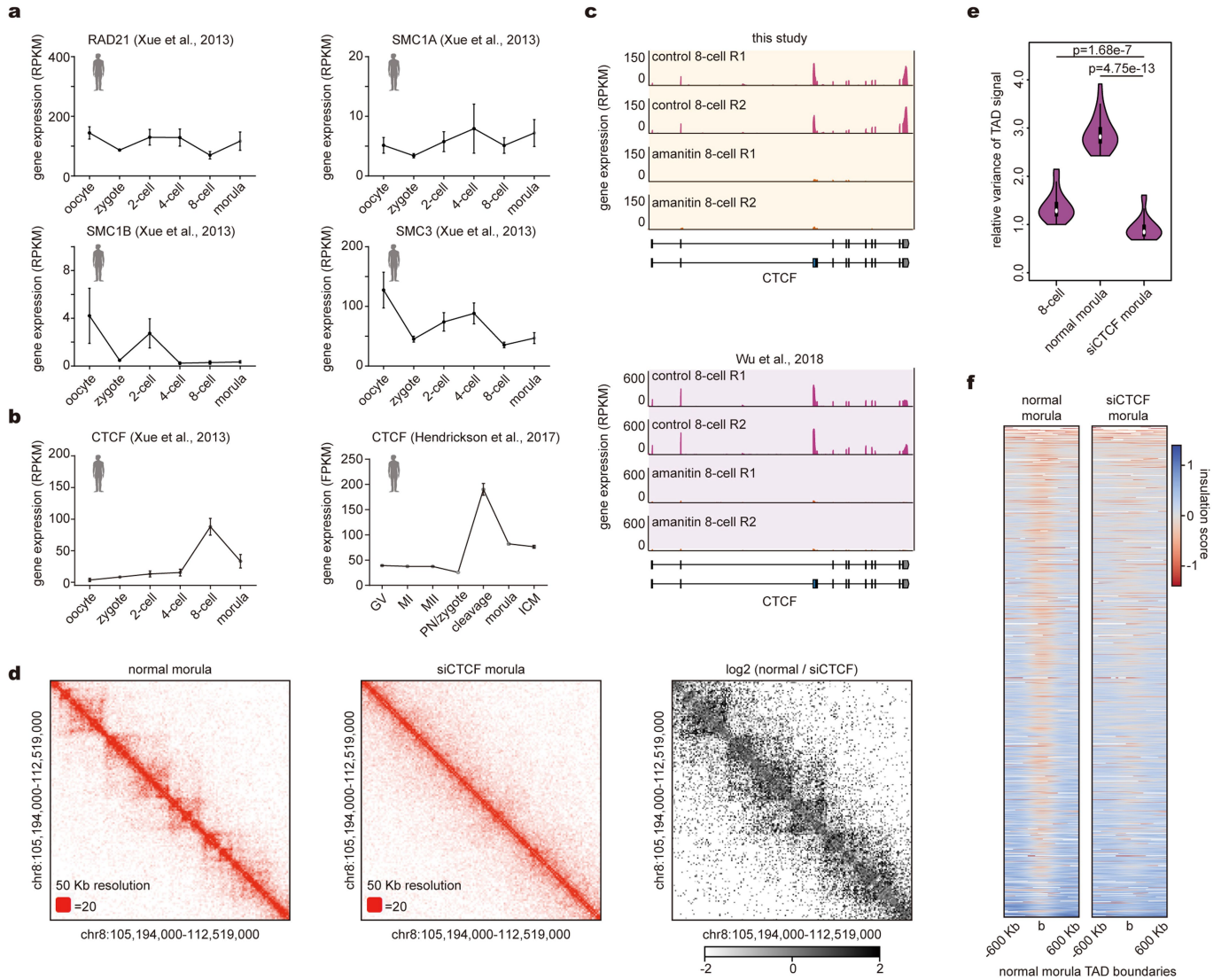
Extended Data Fig. 6 | Insulated boundary dynamics in human embryos and mouse embryos. a, An example for insulated boundaries in human embryos as yellow arrows point in the 2-cell stage. Dark coloured regions in TAD separation score heat maps mean strong insulation. The bottom blue bars are insulated boundaries of each stage. The red bars are 200-kb collapsed boundaries. Grey boxes highlight the 200-kb collapsed boundaries. **b**, Heat map illustrating the insulation score at stage-specific gained insulated boundaries in mouse embryos. **c**, **d**, An example of shared ZGA boundaries in human embryos (**c**) and mouse embryos (**d**). Grey boxes highlight the shared 200-kb collapsed ZGA boundaries.



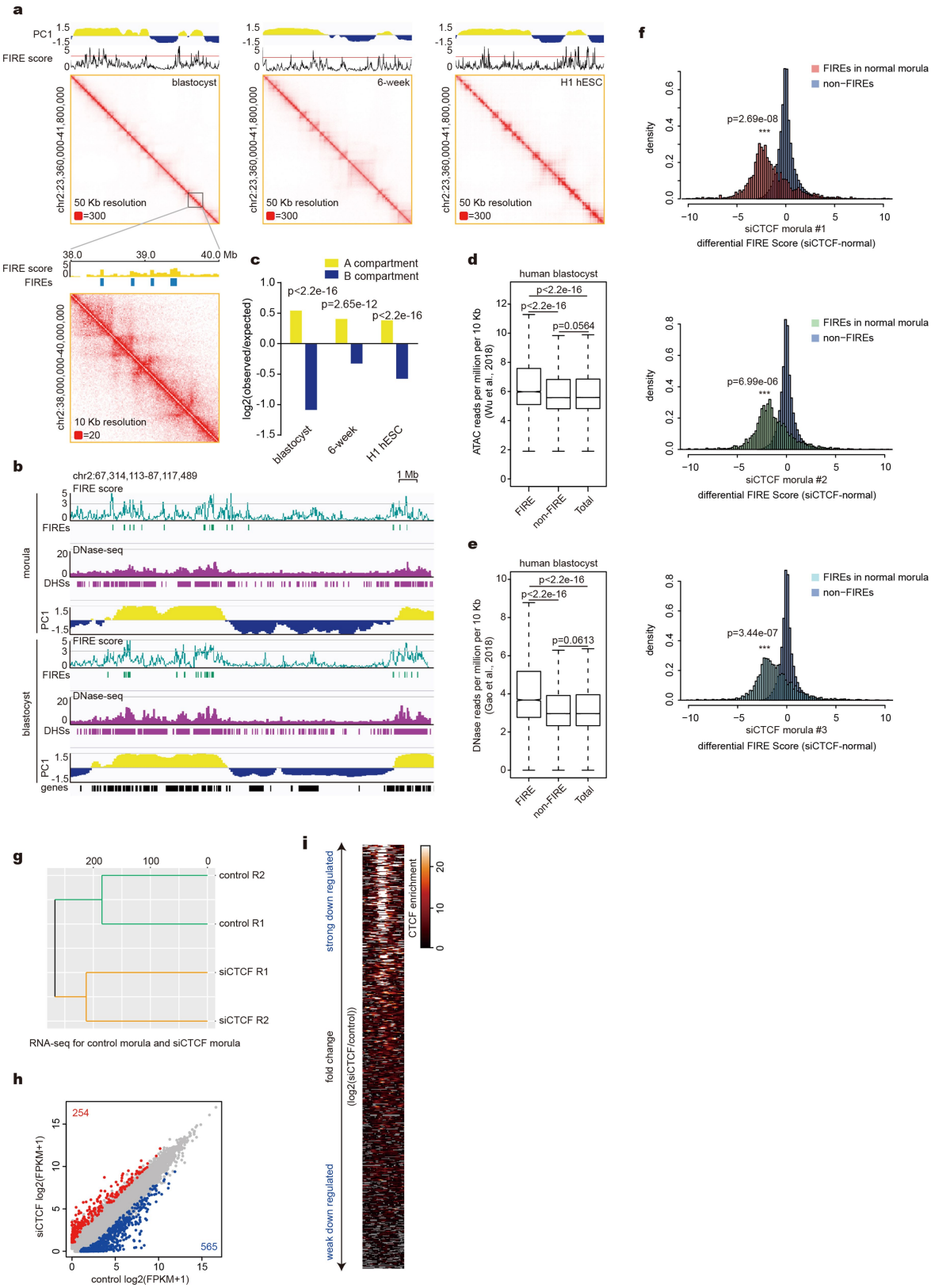
Extended Data Fig. 7 | Analysis of housekeeping genes and repeat elements at insulated boundaries. a, Cumulative distribution function plot for the distance of stage-specific gained boundaries to the closest housekeeping genes in mouse embryos[8] (pooled data from 2–3 biological replicates). The dash line marks a distance of 200 kb; 2-cell versus 4-cell, $P = 1.25 \times 10^{-9}$; 2-cell versus 8-cell, $P = 1.59 \times 10^{-11}$ (two-sided Kolmogorov–Smirnov test). **b**, Gene expression (expression data from ref. [22]; 8-cell, $n = 20$; morula, $n = 16$; blastocyst, $n = 30$) for housekeeping genes within 200 kb of the stage-specific gained boundaries and the other housekeeping genes at the stages when insulated boundaries are being formed. Boxes represent the 25th, 50th and 75th percentiles and whiskers show 1.5× the interquartile range. *P* values are shown (two-sided Wilcoxon rank-sum test). **c**–**e**, Enrichment analysis of repeat elements at stage-specific gained insulated boundaries in human embryos. Enrichment analysis of AluY repeats (**c**), LINE repeats (**d**) and MIR repeats (**e**). **f**–**h**, Enrichment analysis of repeat elements at stage-specific gained insulated boundaries in mouse embryos. Enrichment analysis of Alu repeats (**f**), LINE repeats (**g**) and MIR repeats (**h**). **i**, Examples of the earlier stage gained insulated boundary locating around Alu dense regions in human and mouse. **j**, Left, snapshot of insulated boundaries overlaid with insulation score, AluS density and housekeeping genes. Grey box highlights the 200-kb collapsed boundary. Right, expression of AluS repeats at this boundary. **k**, Expression *z*-score of AluS repeats locating at human 2-cell gained boundaries during human embryonic development (repeat expression data from ref. [21]; $n = 2$ for each stage; for repeat expression calculation, see Methods). Average gene expression of AluS repeats within a boundary was used. The number of 2-cell gained boundaries with AluS repeats is shown ($n = 442$).



Extended Data Fig. 8 | TAD establishment depends on ZGA in human embryos. a, ZGA gene expression in human embryos and α-amanitin treated 8-cell embryos (pooled replicates; for data from ref. [22]: oocyte, $n = 3$; zygote, $n = 3$; 2-cell, $n = 6$; 4-cell, $n = 12$; 8-cell, $n = 20$; morula, $n = 16$; blastocyst, $n = 30$; for data from this study: 8-cell, $n = 2$; α-amanitin treated 8-cell, $n = 2$). ZGA genes were identified as in ref. [22] (Methods). *P* values are calculated by the one-sided Wilcoxon rank-sum test. Boxes represent the 25th, 50th and 75th percentiles and whiskers show 1.5× the interquartile range. **b**, Gene expression for human ZGA genes *DUXA* and *ZSCAN5B* in control 8-cell and α-amanitin treated 8-cell embryos. **c**, The Spearman correlation for two α-amanitin treated 8-cell Hi-C replicates. **d**, Interaction heat maps of human 2-cell, 8-cell, unmixed α-amanitin treated 8-cell embryos and α-amanitin 8-cell from the mixed sample (Methods).

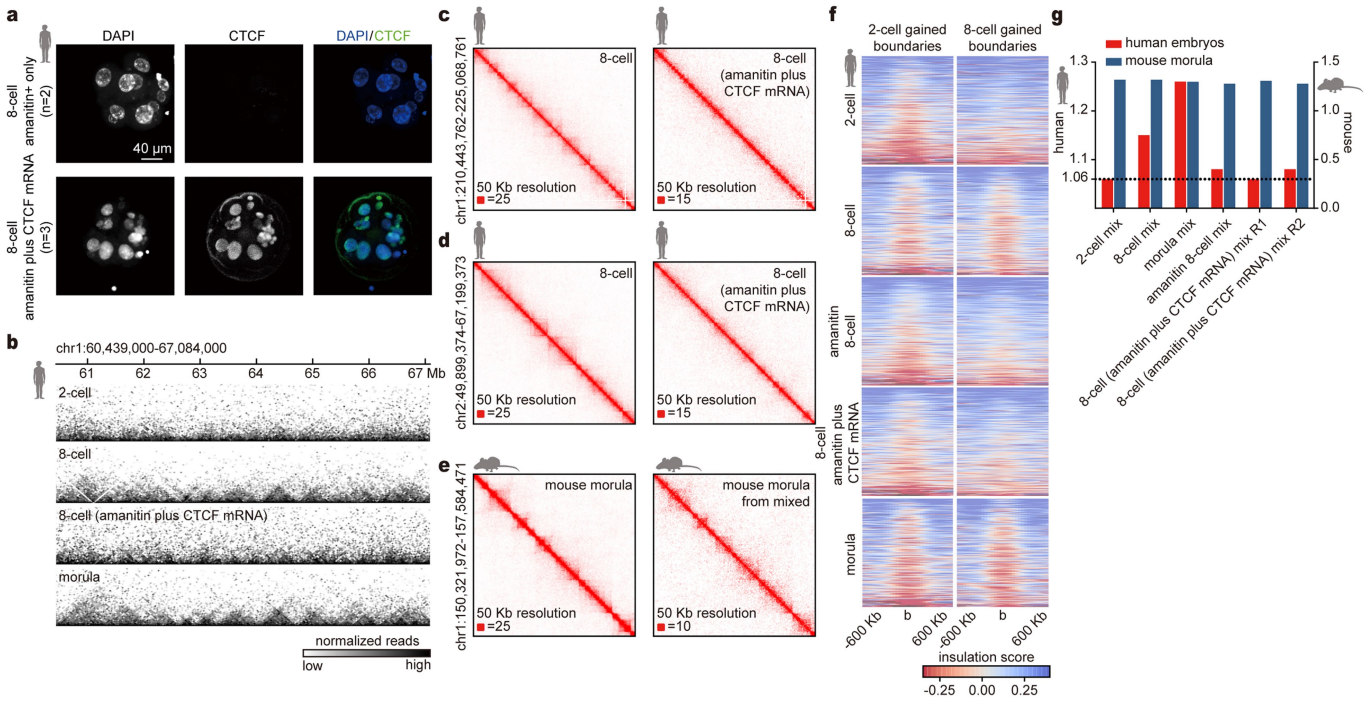


Extended Data Fig. 9 | CTCF regulates TAD establishment in human embryos. a, Gene expression for different cohesin complex subunits during human embryonic development. Expression data from ref. [20]: oocyte, $n = 3$; zygote, $n = 2$; 2-cell, $n = 3$; 4-cell, $n = 4$; 8-cell, $n = 11$; morula, $n = 3$. Data are mean ± s.e.m. **b**, Dynamics of *CTCF* expression during human embryonic development. Expression data from refs. [20,21]. Data are mean ± s.e.m. **c**, *CTCF* gene expression in control 8-cell and α-amanitin treated 8-cell embryos. **d**, Interaction heat maps of untreated control morula and si*CTCF* morula with equal reads at 50-kb resolution. Right, contrast heat map between control and si*CTCF* morula. Dark colour represents increased interactions in untreated morula compared with si*CTCF* morula. **e**, Violin plot for relative variance of TAD signal in 8-cell, control morula and si*CTCF* morula (equal number of reads generated from 2–3 biological replicates for each stage). *P* values are shown (two-sided Wilcoxon rank-sum test). The white boxes in the violin plot represent median values. **f**, Enrichment heat maps for the insulation score of control morula and si*CTCF* morula at control morula TAD boundaries (±600 kb). b, TAD boundary centre.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | The change of FIRE score and gene expression in si*CTCF* morula. a, Snapshot of FIREs in human blastocysts, 6-week embryos and H1 human ES cells overlaid with PC1 tracks and FIRE score tracks. One zoomed-in region in human blastocysts is also shown. **b**, Snapshot of FIREs in human morula and blastocysts overlaid with DHSs[25] and A/B compartments. **c**, Bar plot showing FIREs enrichment and depletion in A/B compartments (replicates pooled; blastocyst, $n = 3$; 6-week, $n = 2$; H1 human ES cell, $n = 2$). *P* values are also shown (χ^2 test). **d**, Box plot showing ATAC-seq reads[24] signal at FIREs, non-FIREs and whole genome in human blastocysts. Boxes represent the 25th, 50th and 75th percentiles and whiskers show 1.5× the interquartile range. *P* values are also shown (two-sided Wilcoxon rank-sum test). **e**, Box plot showing DNase-seq reads[25] signal at FIREs, non-FIREs and whole genome in human blastocysts. Boxes represent the 25th, 50th and 75th percentiles and whiskers show 1.5× the interquartile range. *P* values are also shown (two-sided Wilcoxon rank-sum test). **f**, Histograms of FIRE score difference between si*CTCF* and untreated control morula (si*CTCF* − untreated) in FIREs and non-FIREs for si*CTCF* morula no. 1, si*CTCF* morula no. 2 and si*CTCF* morula no. 3. *P* values for differential FIREscore (si*CTCF* − untreated) in FIREs and non-FIREs are also shown (one-sided *t*-test). **g**, The hierarchical cluster of gene expression in control morula and si*CTCF* morula analysed using the package ggdendro. **h**, Scatter plot of gene expression between control morula and si*CTCF* morula. Red dots refer to upregulated genes (254 genes) in si*CTCF* morula. Blue dots refers to downregulated genes (565 genes) in si*CTCF* morula. **i**, Heat map for human ES cell *CTCF* ChIP–seq signal[47] around gene TSSs downregulated in si*CTCF* morula. Strongly downregulated genes ($\log_2$(fold change(siCTCF/control)) < −5).



Extended Data Fig. 11 | TADs cannot re-establish in 8-cell (α-amanitin plus *CTCF* mRNA) embryos. a, Immunofluorescence confocal images of CTCF in α-amanitin treated 8-cell embryos ($n = 2$) and 8-cell (α-amanitin plus *CTCF* mRNA) embryos ($n = 3$). Scale bar, 40 µm. **b**, Track snapshot for TAD structures in 8-cell (α-amanitin plus *CTCF* mRNA) embryos along with untreated 2-cell, 8-cell and morula embryos. **c**, **d**, Interaction heat map examples at 50-kb resolution for human 8-cell embryos and 8-cell (α-amanitin plus *CTCF* mRNA) embryos. **e**, Interaction heat map examples for mouse morula embryos without mix and mouse morula from the α-amanitin plus *CTCF* mRNA mixed sample (Methods). **f**, Enrichment heat maps for the insulation score of human 2-cell embryos, 8-cell embryos, α-amanitin 8-cell embryos, 8-cell (α-amanitin plus *CTCF* mRNA) embryos and morula embryos around boundaries (± 600 kb). b, boundary centre. Left, boundaries gained at 2-cell stage; right, boundaries gained at 8-cell stage. **g**, Bar plots for TAD structure strength of human embryos and mouse morula from mixed samples. Left *y* axis, human embryos; right *y* axis, mouse morula.

nature research

Corresponding author(s): Jiang Liu and Zi-jiang Chen

Last updated by author(s): Aug 27, 2019

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	ZEN software (2010) was used to collect confocal microscopy data
Data analysis	HiCExplorer suite (version 2.1);juicer tool;GenomeDISCO software;'matrix2insulation.pl' https://github.com/dekkerlab/giorgetti-nature-2016;HOMER;'pairsqc.py' (https://github.com/4dn-dcic/pairsqc);UCSC liftOver tool online;Deeptools2 suite;R;Prism;ZEN software (2010);ImageJ 1.52a; See details in Methods.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

CRA000852 and CRA000108

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	The number of embryos used in this study is shown in Supplementary Table 1 and Supplementary Table 6. About as few as 50-100 cells can be used for a Hi-C assay, and 1 embryos can be used to prepare a RNA-seq library. In our analysis, our ultra-low-input Hi-C using 50-100 cells and SMART-seq2 RNA-seq can provide robust results (Extended Data Figure 1). No statistical method were used to predetermine the sample size.
Data exclusions	No data were excluded from the analyses.
Replication	Hi-C were performed at least 3 biological replicates for each stage of preimplantation embryos. And RNA-seq were performed at least 2 biological replicates for each sample. Detailed information is shown in Supplementary Table 1 and Supplementary Table 6. All replications were consistent for data results.
Randomization	All the embryos were collected and randomly allocated for experiements.
Blinding	During the siRNA injection , investigators injected the siRNA to the embryos without knowing the types of the siRNA. For other experiments, variations were controlled through replicates.

Behavioural & social sciences study design

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

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

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

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*

Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access and import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

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
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Antibodies used in western blotting: Anti-CTCF antibody, abcam, cat#: ab188408, lot#GR3253930-3, 1:1000 dilution Anti-RAD21 antibody, abcam, cat#: ab992, lot#GR239192-1, 1:1000 dilution Goat Anti-Rabbit IgG, CWBIO, Cat no: CW0103, lot#013341/50402, 1:3000 dilution Immunostaining Antibodies: Anti-CTCF antibody, abcam, cat#: ab188408, lot#GR3253930-3, 1:500 dilution Alexa Fluor 488-labeled goat anti-rabbit IgG, Beyotime, cat#: P0176, 1:1000 dilution
Validation	All the antibodies used in this study were commercial antibodies and were only used for applications, with validation procedures

described on the following sites of the manufacturers: https://www.abcam; https://www.beyotime.com; https://www.cwbiotech.com.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK 293 cells (ATCC: ATCC® CRL-1573™) ;HeLa (ATCC: ATCC® CCL-2™);4T1 (ATCC: ATCC® CRL-2539™);HT22(EMD Millipore, SCC129)
Authentication	All these cell lines were used in several papers published by our lab, no further authentication were performed for thess cell lines.
Mycoplasma contamination	All cell lines tested negetive for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used

Palaeontology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mouse sperm cells were collected from the cauda epididymis in 8 weeks old C57BL/6J male mice (Vital River). The mouse morula embryos were from the cross of C57 BL/6J female (Vital River; 4-6 weeks) and PWK/PhJ male (Jackson Laboratory;10 weeks in average).
Wild animals	No wild animals were used.
Field-collected samples	No field collected samples
Ethics oversight	All animal maintenance and experimental procedures were carried out according to guidelines of Institutional Animal Care and Use Committee (IACUC) of Beijing Institute of Genomics, CAS, Beijing, China.

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	The oocyte donors are 25-38 year-old women with normal BMI. The healthy sperm donors are 22-50 year-old men with normal semen. All volunteers have no family heredity case history, contagion case and smoke history. The gametes and embryos are checked under microscopy and are with high quality. The donated frozen embryos as well as the donated oocytes and sperm in this study are from Center for Reproductive Medicine, Shandong University.
Recruitment	Research donors were recruited from patients taking in vitro fertilization treatments. Before giving consent, donors have a suitable opportunity to receive proper counselling about the implications of the donation and potential risks. Written informed consent was obtained from all oocyte and sperm donors, respectively. They were informed that the donation would not affect their IVF process.
Ethics oversight	Institutional Review Board of Reproductive Medicine, Shandong University and Beijing Institute of Genomics

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

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 *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*

Cell population abundance — *Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*

Gating strategy — *Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type — *Indicate task or resting state; event-related or block design.*

Design specifications — *Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*

Behavioral performance measures — *State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s) — *Specify: functional, structural, diffusion, perfusion.*

Field strength — *Specify in Tesla*

Sequence & imaging parameters — *Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*

Area of acquisition — *State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software — *Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*

Normalization — *If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template — *Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal — *Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring — *Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings — *Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*

Effect(s) tested — *Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016) — *Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

Correction — *Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.