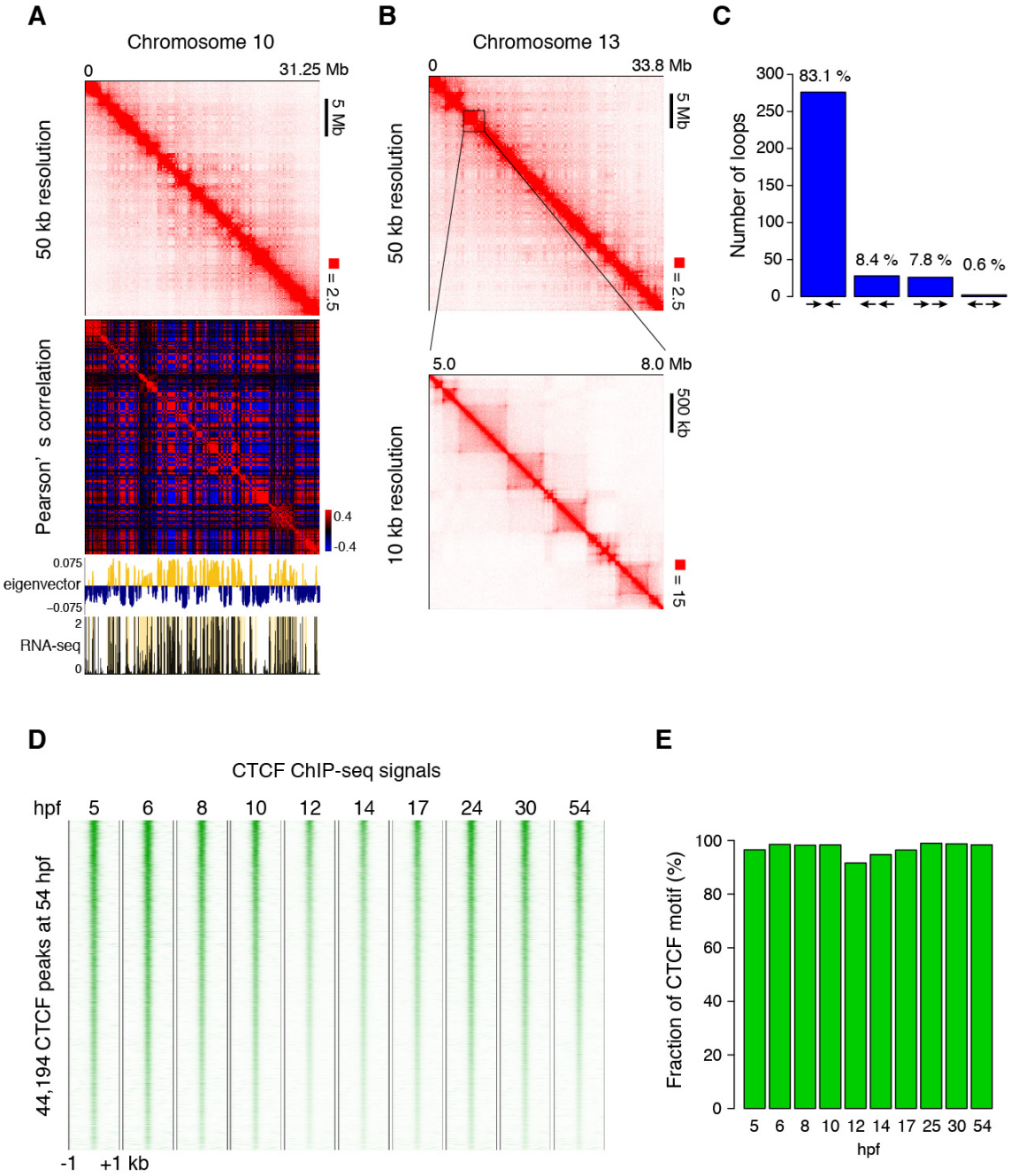
Supplemental Figures

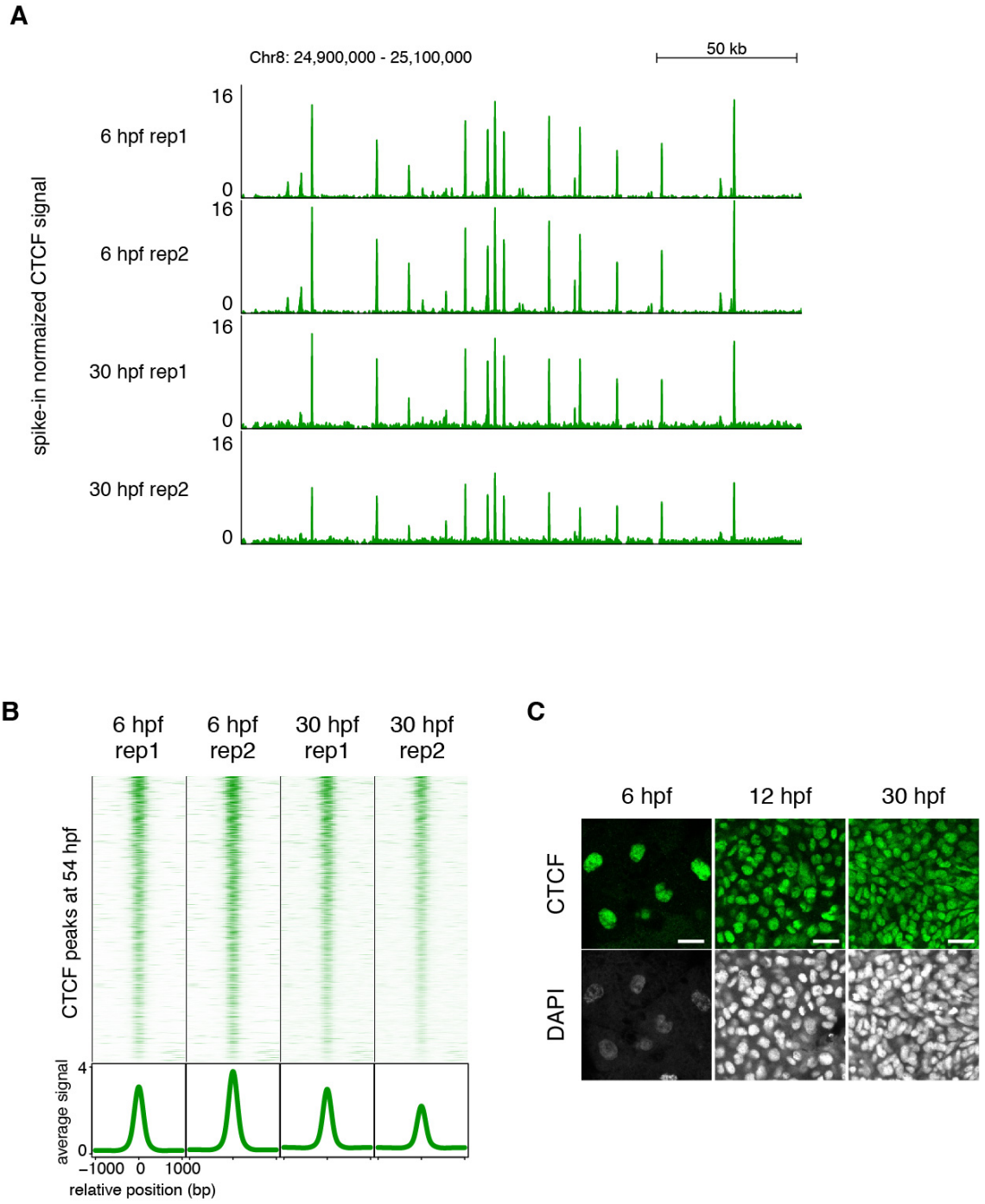
CTCF looping is established during gastrulation in medaka embryos

Ryohei Nakamura[1]†, Yuichi Motai[2]†, Masahiko Kumagai[3], Candice L. Wike[5], Haruyo Nishiyama[1], Yoichiro Nakatani[6], Neva C. Durand[4], Kaori Kondo[7], Takashi Kondo[7], Tatsuya Tsukahara[8], Atsuko Shimada[1], Bradley R. Cairns[5], Erez Lieberman Aiden[4]*, Shinichi Morishita[2]*, and Hiroyuki Takeda[1]*

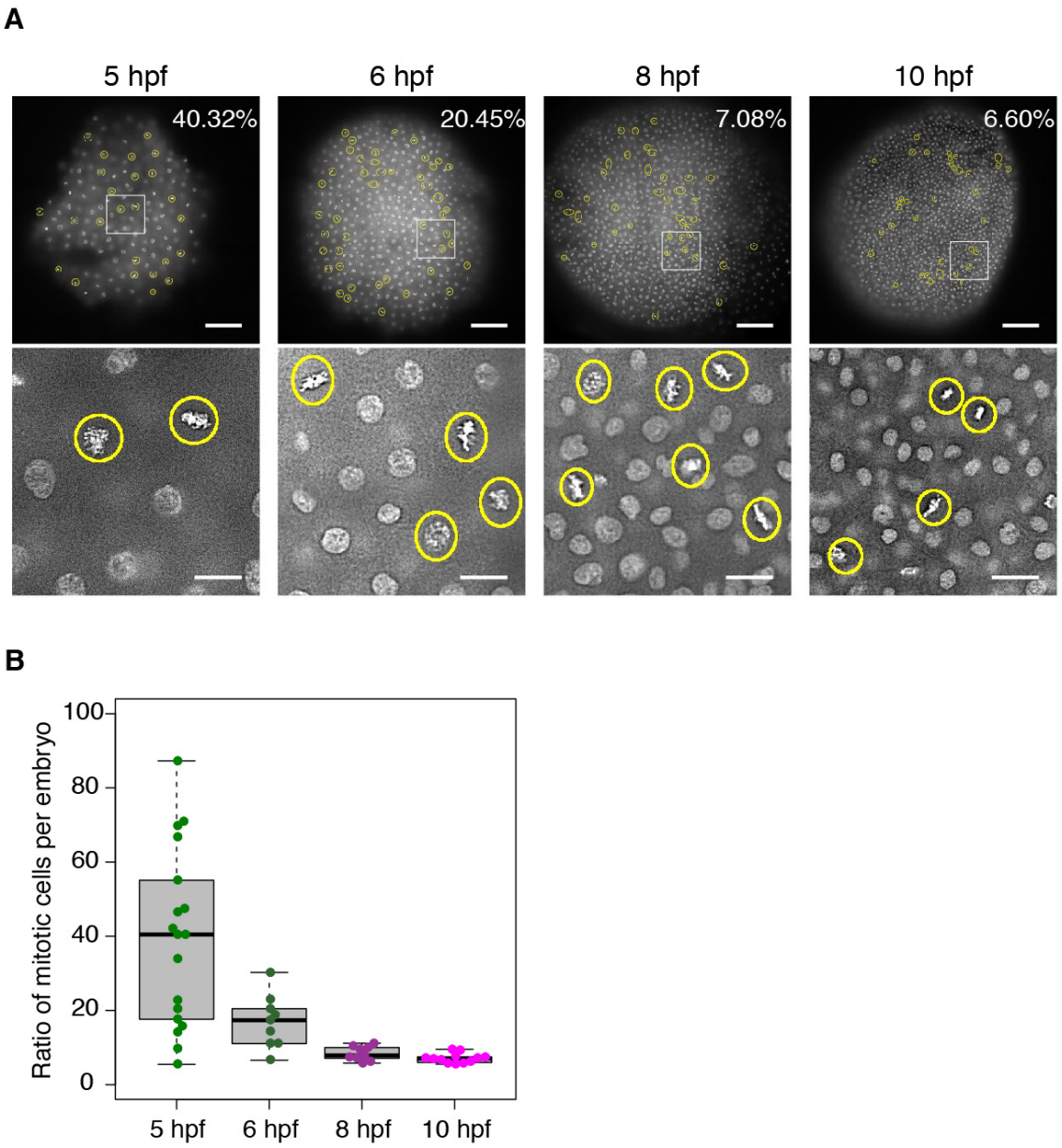
[1]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo 113-0033 Japan.
[2]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa 277-8562, Japan.
[3]Advanced Analysis Center, National Agriculture and Food Research Organization, Tsukuba, Ibaraki 305-8602, Japan
[4]The Center for Genome Architecture, Baylor College of Medicine, Houston, TX 77030, USA
Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. Department of Computer Science, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005, USA. Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. Center for Theoretical Biological Physics, Rice University, Houston, TX 77030, USA.
[5]Howard Hughes Medical Institute, Department of Oncological Sciences and Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT 85112, USA.
[6]Department of Cancer Genome Informatics, Graduate School of Medicine, Osaka University, Osaka 565-0871, Japan
[7]RIKEN-IMS, Laboratory for Developmental Geneticsm, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan.
[8]Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA



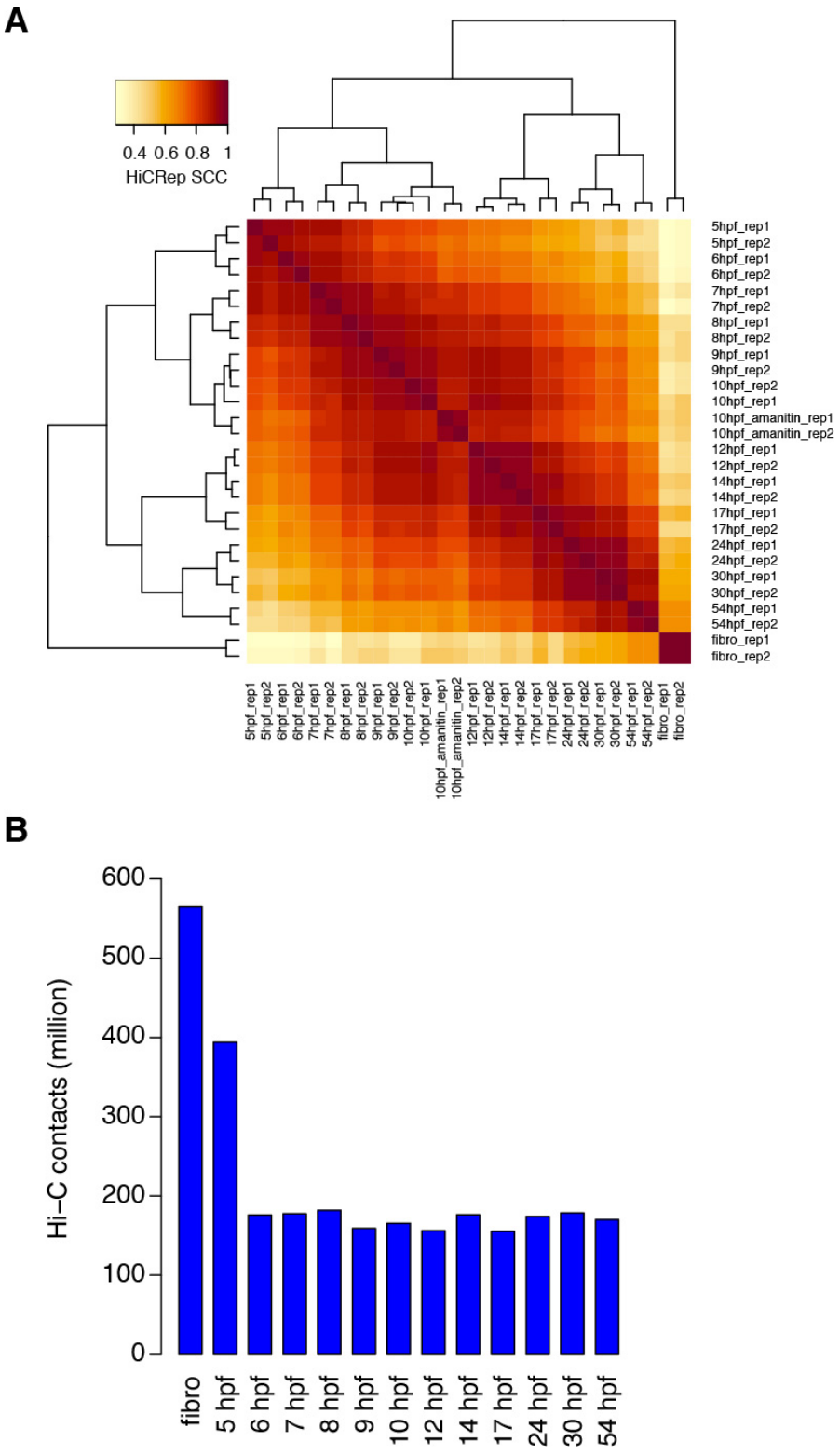
Supplemental Fig. S1.
Compartments and CTCF loops organize 3D genome structure of medaka fibroblast cells, and CTCF binds stably across all developmental stages
(A) From top to bottom, observed contact matrix of whole Chromosome 10 at 50 kb resolution, Pearson' s correlation matrix, eigenvector, and RNA-seq tracks are shown. The A compartments are highlighted with yellow in eigenvector and RNA-seq (rpkm) tracks.
(B) An example of compartments and loops. Observed Hi-C contact matrix of medaka fibroblast cells. Whole Chromosome 13 at 50 kb resolution showing plaid pattern (top) and zoomed view at 10 kb resolution showing loop domains.
(C) Number of loops with indicated combination of unique CTCF motif orientation.
(D) CTCF ChIP-seq signals aligned to the positions of the peaks at 54 hpf. Rows correspond the 44,194 CTCF signal peaks called by MACS2 at 54 hpf, ordered by fold-enrichment.
(E) Percentages of CTCF motifs in the convergent orientation observed at fibroblast loop anchor overlapped with CTCF ChIP-seq peak.



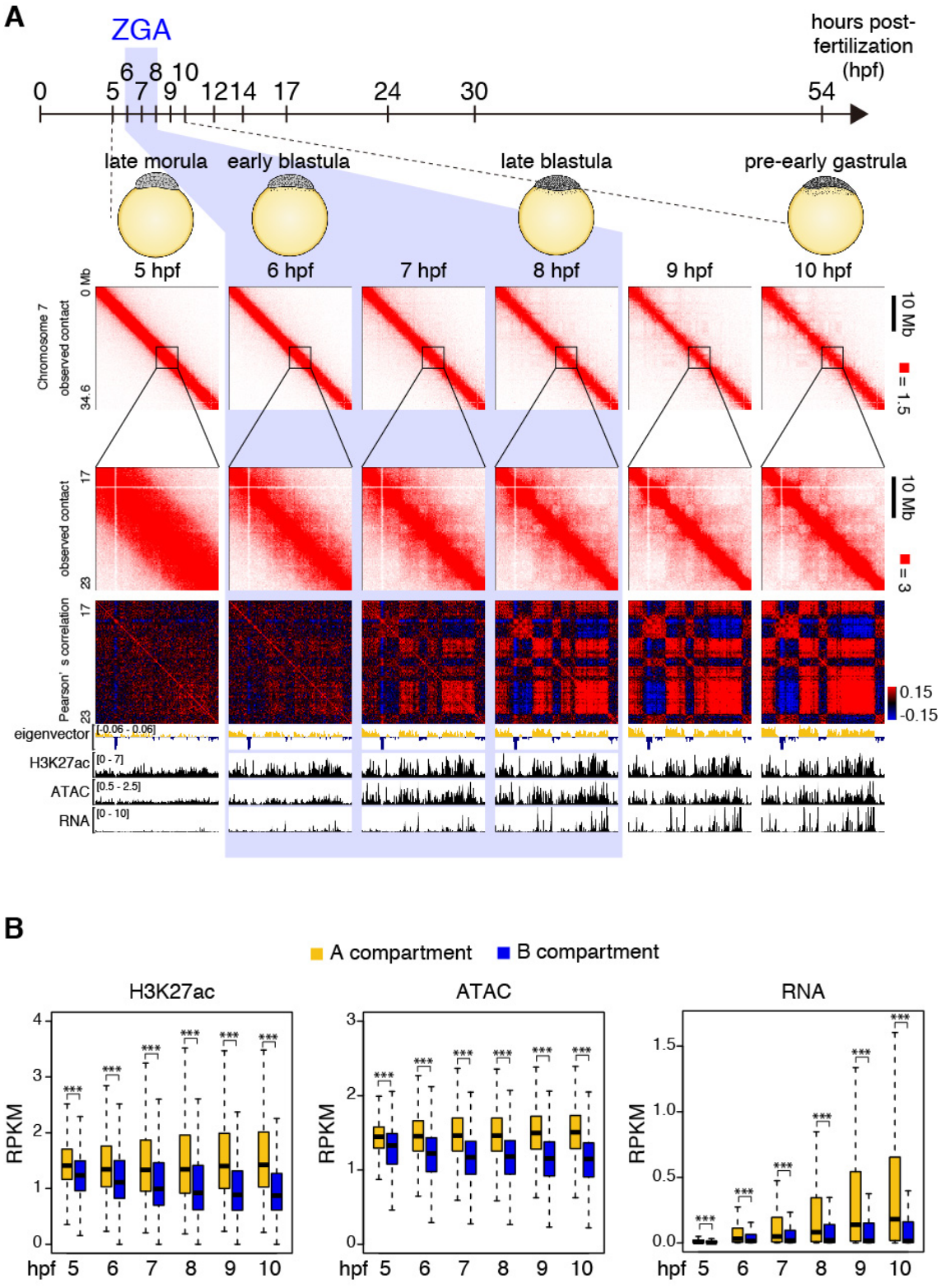
Supplemental Fig. S2.
Quantification of CTCF binding by spike-in ChIP-seq and immunofluorescence
(A) Representative view of spike-in normalized CTCF ChIP-seq tracks at 6 and 30 hpf. Peak height was normalized using reads of spike-in chromatin.
(B) Heatmap showing spike-in normalized CTCF ChIP-seq signals aligned to the positions of the peaks at 54 hpf. Rows correspond the 44,194 CTCF signal peaks called by MACS2 at 54 hpf, ordered by fold-enrichment. In the lower portion, averages of all peaks are shown.
(C) Immunofluorescence staining of CTCF antibody and DAPI at 6, 12, and 30 hpf. Scale bar: 20μm.



Supplemental Fig. S3.
Cell cycle synchronization measured by counting mitotic cells per embryo and number of Hi-C contacts generated in this study
Anamniote embryos generally exhibit short and synchronous cell cycles until the blastula stage, followed by the transition into longer and asynchronous ones. In medaka, synchronous cell divisions take place with a cycle length of 30 to 35 minutes before the blastula stage. As mitotic chromosomes have unique 3D organization characterized by mega-base contact patterns, which may affect the global view of Hi-C contacts from whole embryos especially at early stages, we evaluated the proportion of mitotic cells in morula (5 hpf) to pre-early gastrula (10 hpf) embryos by microscope observation. Collected embryos contained approximately 39%, 17%, 8%, and 7% of mitotic cells on average at 5 hpf, 6 hpf, 8 hpf and 10 hpf, respectively. Although the mixture of mitotic and interphase cells in an embryo prohibited us to remove mitotic chromatin from our analyses, we reasonably assumed that the majority of cells were at interphase cells, especially after 5 hpf.
(A) Typical representations of counting mitotic cells by DAPI staining before, during and after ZGA. The square windows at the top pictures were enlarged in the bottoms. The scale bars are 100 μm (top) and 20 μm (bottom). The yellow circles indicate mitotic cells.
(B) Plots of the ratio of mitotic cells per embryo, showing that cell cycle synchronization already becomes less effective at 5 hpf.

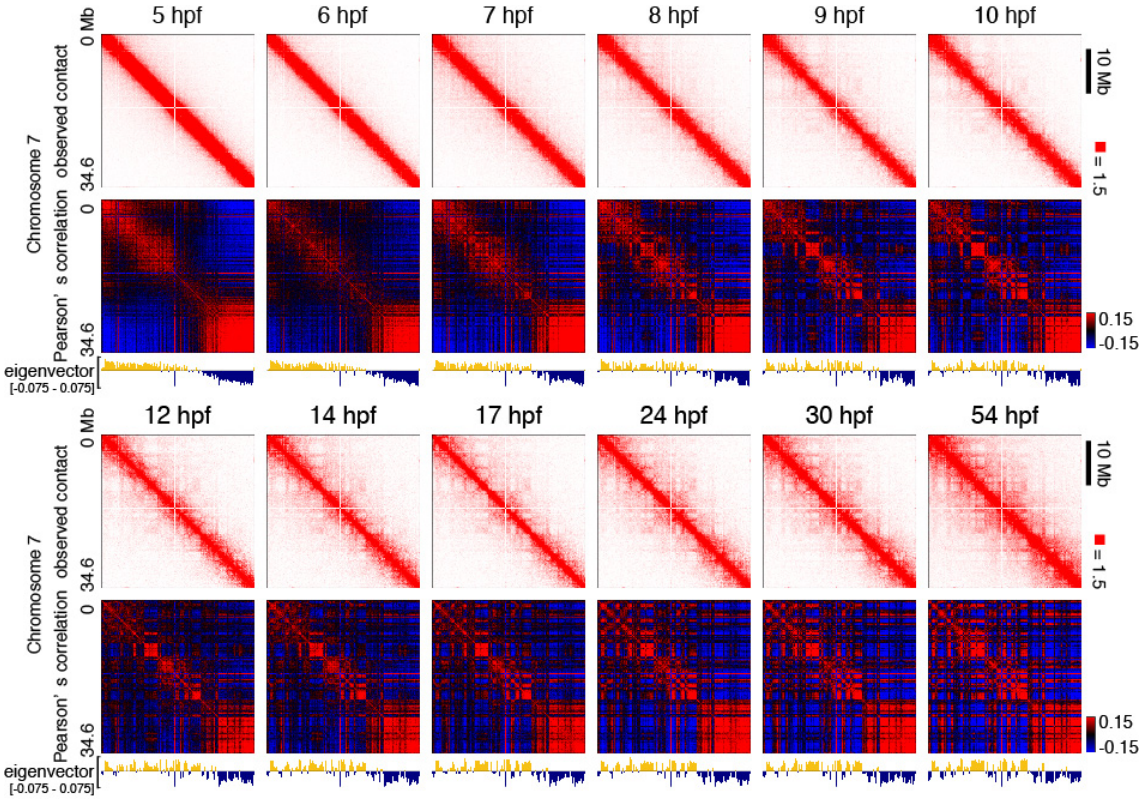


Supplemental Fig. S4.
Reproducibility of the Hi-C data and number of Hi-C contacts generated in this study
(A) Heatmap showing clustering of Hi-C matrix by HiCRep. Normalized contacts at 100 kb resolution were used to calculate similarity.
(B) The bargraph shows the number of Hi-C contacts at each stage after filtering using Juicer (replicates combined).

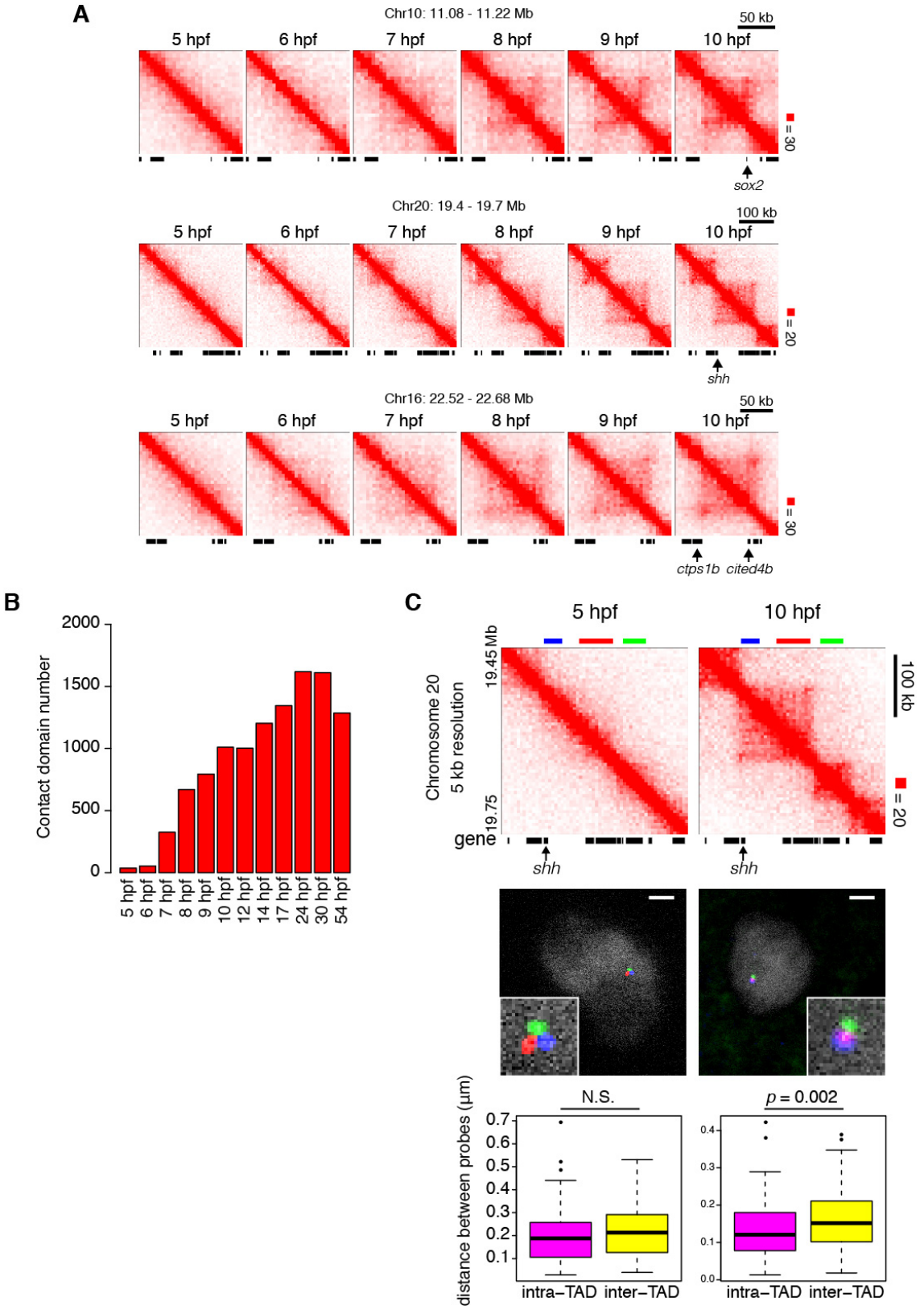


Supplemental Fig. S5.
Compartmentalization during medaka ZGA
(A) Compartmentalization at early developmental stages. Observed contact maps at 50 kb resolution are shown for whole Chromosome 7 and enlarged views with corresponding Pearson's correlation matrices, eigenvectors, H3K27ac ChIP-seq, ATAC-seq, and RNA-seq tracks for all indicated stages. For RNA-seq track, we assayed zygotic transcription. To facilitate the discrimination of maternal transcripts, which are ubiquitous in the egg throughout early

development, and paternal transcripts, which were only derived from the paternal loci of the zygotic genome, we performed RNA-Seq on a cross between two medaka polymorphic inbred strains, d-rR and HNI. The rate of SNPs between the two strain is 2.5% (1 in 40 bp), making it possible to assign most reads to a paternal or maternal genome of origin. Using the level of paternal transcripts seen in RNA-seq as a proxy for overall zygotic transcription, we found that ZGA takes place between 6hpf and 8hpf.
(B) Boxplots showing the enrichment of H3K27ac ChIP-seq, ATAC-seq, or RNA-seq reads in A and B compartments. *** indicates P-value < 2.2e-16. P-values were calculated by Wilcoxon rank sum test.

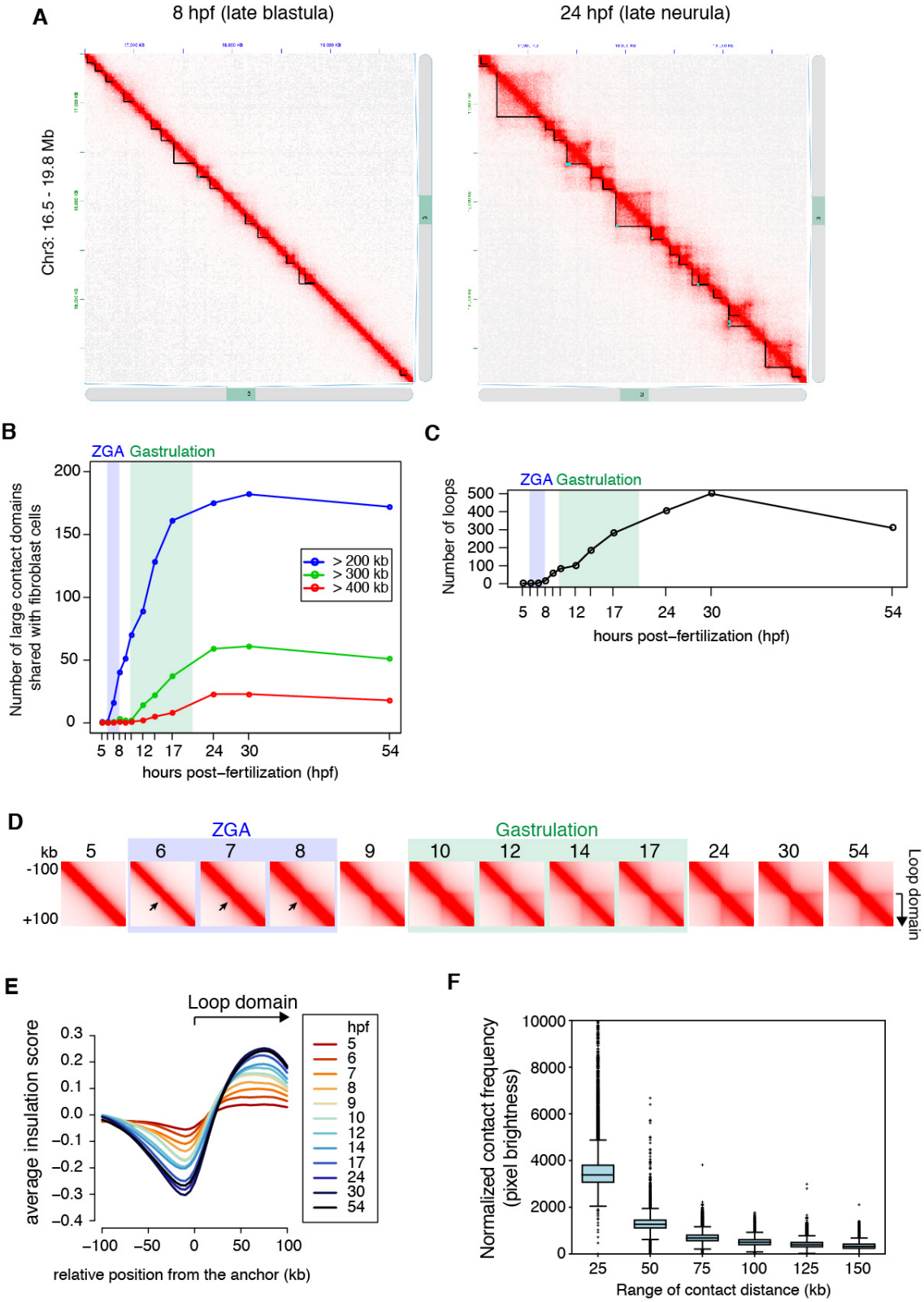


Supplemental Fig. S6.
Compartmentalization through medaka development
Observed contact matrices, Pearson's correlation matrices, eigenvector tracks are shown for whole Chromosome 7 for all developmental stages used in this study.



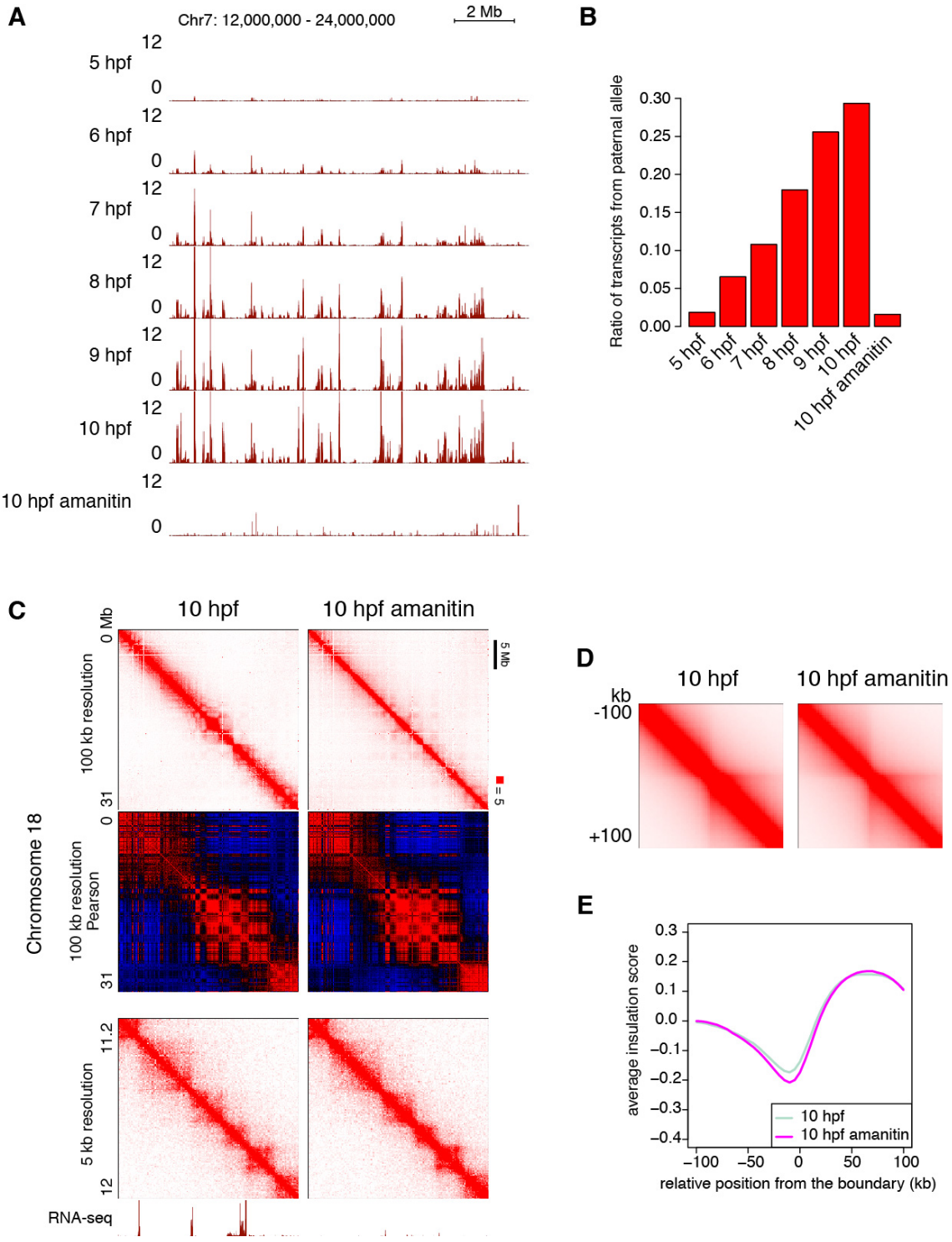
Supplemental Fig. S7.
Small contact domains emerge during ZGA

(A) By manual inspection, we noticed some small contact domains which were overlooked by Arrowhead program because these contact domains were hard to detect within tight proximal contacts along the diagonals of contact matrices. Shown are several examples of relatively clear ones of those domains. These domains are absent at 5 hpf, but emerge during ZGA, and become clear at early gastrula stage (10 hpf).
(B) Number of contact domains
(C) Validation of small contact domain by microscopy. (top contact maps) A relatively small contact domain including *shh* gene was obviously established as development proceeded from 5 to 10 hpf (top). To validate the establishment of a small contact domain during ZGA *in vivo*, we performed 3D-FISH. Three probes were designed, two in a same contact domain to quantify an intra-TAD distance, and one in a neighboring contact domain to measure an inter-TAD distance. (middle microscope photos) We estimated the distance between a pair of two probes from a distance distribution among multiple microscope photos in interphase nuclei of 5 hpf and 10 hpf embryos. The representative views of their fluorescent signals were shown at the middle. The scale bar indicates 2 μm. (bottom box plots) Consistent with the Hi-C data, we observed no significant difference between the distance distributions between intra-TAD probes and inter-TAD probes at 5 hpf, but a significance increase in inter-TAD distance distribution at 10 hpf.

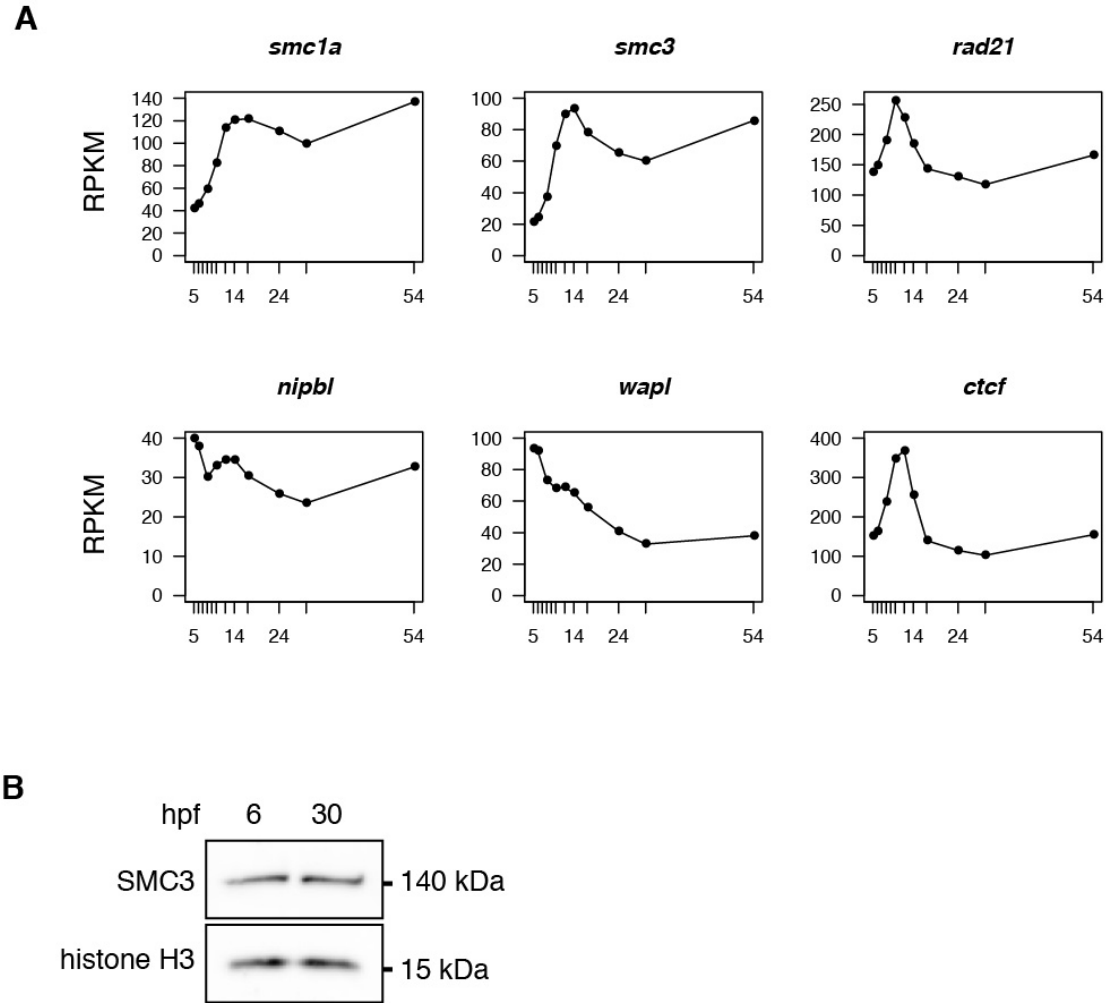


Supplemental Fig. S8.
Emergence of large contact domains, loops, and loop domain boundaries
(A) Screen shots of Juicebox showing the representative loops and contact domains before gastrulation (8 hpf; late blastula) and after gastrulation (24 hpf; late neurula).

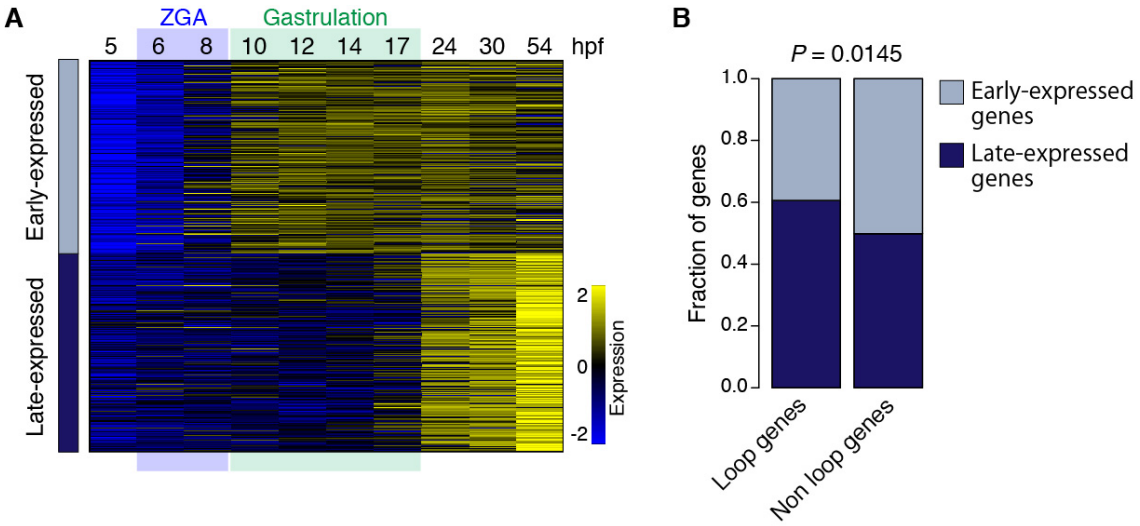
(B) The majority of large contact domains associated with mature fibroblast cells emerge during gastrulation. The number of contact domains present in mature fibroblast cells is shown for different size ranges; $\geq$ 200 kb, $\geq$ 300 kb, and $\geq$ 400 kb.
(C) Number of loops associated with contact domains. Loops smaller than 100 kb were difficult to distinguish from false positives due to the resolution limitation. Thus, we counted the number of loops (associated with contact domains) that are 100 kb or larger.
(D) Aggregation of Hi-C contacts at all loop domain anchors identified in fibroblast cells. Arrows indicate the emergence of boundaries.
(E) Average insulation scores for each of twelve stages centered at loop anchors called in fibroblast cells. The insulation effect is illustrated by the steepness of a valley in an insulation score curve. Even at 5 hpf, a small valley is seen, and valleys gradually get deeper through embryogenesis, indicating that the insulation effect of loop domains already appears at the blastula stage and become stronger continuously while development proceed.
(F) Frequency distribution of pixel brightness in the fibroblast contact frequency map. Each loop in a contact frequency map is identified as a local maximum peak in the region of size 35kb for 5kb resolution (50kb for 10kb resolution) from the center according to the peak calling method (9). It becomes harder to detect smaller loops that are closer to the diagonal of contact frequency map. To confirm this tendency, the box plots show the frequency distribution of pixel brightness in the fibroblast contact frequency map, where the brightness of a pixel represents a normalized contact frequency between pairs of positions at a distance. Distances are partitioned into ranges of length 25kb, which almost accords with the radius used for calling local maximum peaks. “25kb” in the x-axis, for example, shows the range smaller than 25kb, and “50kb” ranges from 25kb to 50kb. Of note, most of values in the 50kb box plot are smaller than the median in the 25kb box plot, and hence, many local maximum peaks in proximity to the diagonal are difficult to treat as loops. By contrast, highly bright pixels with extremely large contact frequencies in the 125kb box plot are also outliers in the 100kb box plot. Put another way, most of > 100kb loops are detectable as local maximum peaks, and hence, we focus on >100kb loops.



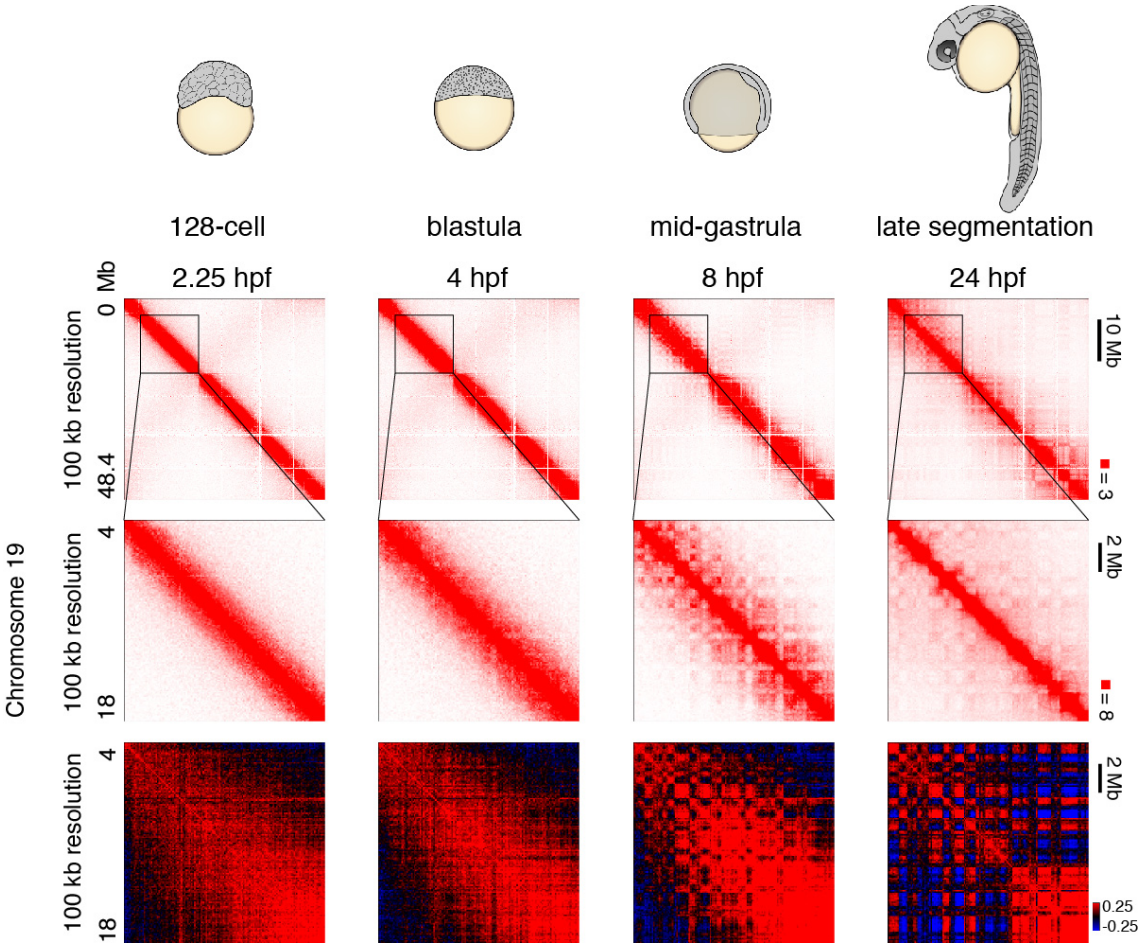
Supplemental Fig. S9.
Emergence of compartments, small contact domains, and loop domain boundaries do not require ZGA
(A) Representative view of RNA-seq track of zygotic transcripts (from paternal allele).
(B) Ratio of RNA-seq reads mapped to the paternal allele.
(C) Representative Hi-C map of control and ZGA-blocked (alpha-amanitin-injected) 10 hpf embryos. Upper portion: normalized observed contact matrices and Pearson's correlation matrices at 100 kb resolution. Lower portion: normalized observed contact matrices at 5 kb resolution showing small contact domains.
(D) Aggregation of Hi-C contacts at all loop domain anchors identified in fibroblast cells.
(E) Average insulation scores centered at loop anchors called in fibroblast cells.



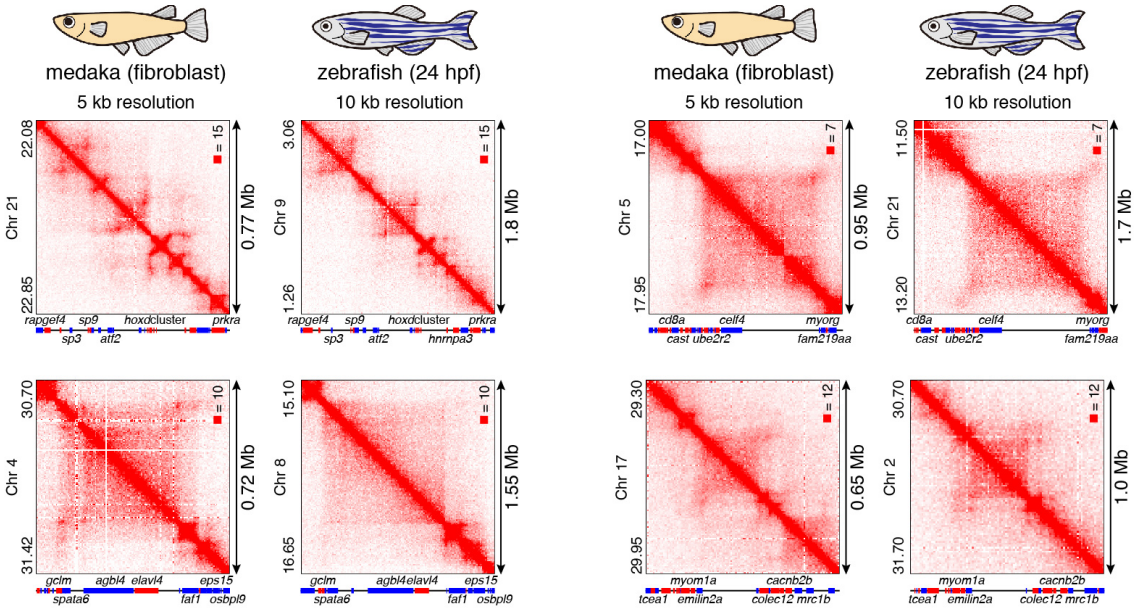
Supplemental Fig. S10.
Expression of cohesin-related genes
(A) Expression levels of cohesin-related genes quantified by RNA-seq.
(B) Western blot showing expression of SMC3 protein at 6 and 30 hpf.



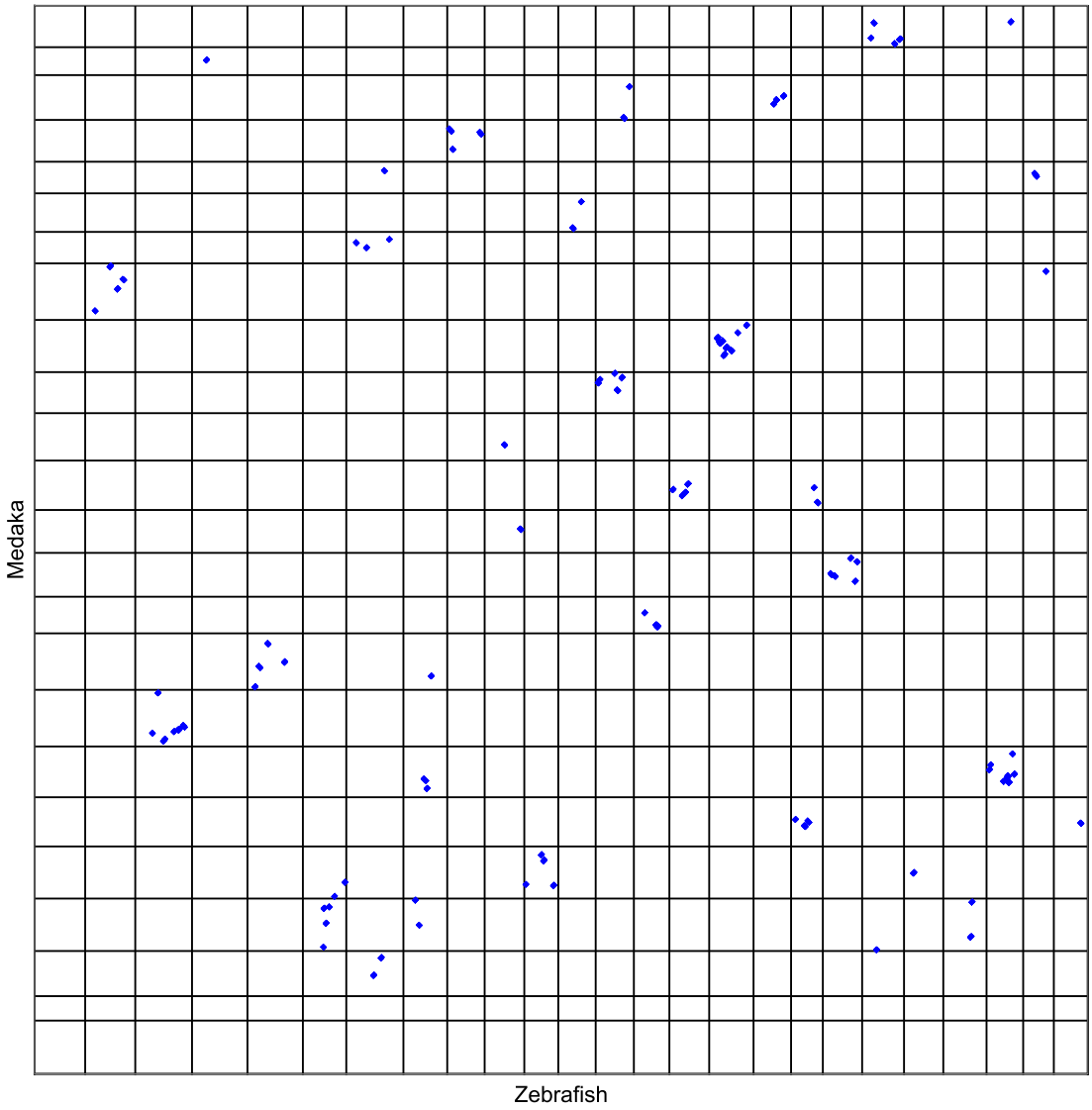
Supplemental Fig. S11.
CTCF-mediated loops are associated with late activating genes
(A) Classification of zygotically expressed genes into two groups by expression dynamics.
(B) Fraction of early-expressed genes and late-expressed genes for loop genes and non loop genes. P-value was calculated by Fisher's exact test.



Supplemental Fig. S12.
Emergence of compartments in zebrafish embryos
Representative example of Hi-C contact matrices of zebrafish embryos. Pearson's correlation matrices are shown at the bottom.



Supplemental Fig. S13.
Conservation of 3D structure in synteny blocks between medaka and zebrafish
Additional examples of conserved 3D structure in synteny blocks. Hi-C maps of medaka fibroblast (5 kb resolution), and 24 hpf zebrafish embryos (10 kb resolution) are shown. The orientations of genes are indicated by different colors (forward: red, reverse: blue).



Supplemental Fig. S14.
Synteny blocks between medaka and zebrafish.
Protein-coding genes were sorted with respect to the chromosomal position, and arranged along the x-axis (zebrafish) and y-axis (medaka). The zebrafish chromosomes are shown from 1 to 25 (from left to right), and the medaka chromosomes are shown from 1 to 24 (from bottom to top). Lines indicate chromosome boundaries, and dots indicate orthologs in the high-scoring chains (score>9).