Extrusion fountains are hallmarks of chromosome organization emerging upon zygotic genome activation

Aleksandra Galitsyna[*#1], Sergey V. Ulianov[*3,4], Nikolai S. Bykov[5,6], Marina Veil[7], Meijiang Gao[7,8], Kristina Perevoschikova[9], Mikhail Gelfand[5], Sergey V. Razin[3,4], Leonid Mirny[#1,2] and Daria Onichtchouk[#,7,8,10]

*- these authors contributed equally to this work
- corresponding authors: galitsyn@mit.edu, leonid@mit.edu, daria.onichtchouk@biologie.uni-freiburg.de

1 - Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
2 - Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139.
3 - Institute of Gene Biology, Russian Academy of Sciences, 119334, Russia
4 - Faculty of Biology, M.V. Lomonosov Moscow State University, Moscow, 119991, Russia
5 - Institute for Information Transmission Problems (the Kharkevich Institute), Russian Academy of Sciences, Moscow, 127051, Russia
6 - Centro Nacional de Análisis Genómico (CNAG), Baldiri Reixac 4, Barcelona, 08028 Spain
7 - Department of Developmental Biology, University of Freiburg, Freiburg, 79104, Germany.
8 - Signalling Research Centres BIOSS and CIBSS, Freiburg, 79104, Germany
9 - Faculty of Bioengineering and Bioinformatics, M.V. Lomonosov Moscow State University, Moscow, 119991, Russia
10 - Koltzov Institute of Developmental Biology RAS, Moscow, 119991, Russia.

Abstract

The first activation of gene expression during development (zygotic genome activation, ZGA) is accompanied by massive changes in chromosome organization. The connection between these two processes remains unknown. Using Hi-C for zebrafish embryos, we found that chromosome folding starts by establishing "fountains", novel elements of chromosome organization, emerging selectively at enhancers upon ZGA. Using polymer simulations, we demonstrate that fountains can emerge as sites of targeted cohesin loading and require two-sided, yet desynchronized, loop extrusion. Specific loss of fountains upon loss of pioneer transcription factors that drive ZGA reveals a causal connection between enhancer activity and fountain formation. Finally, we show that fountains emerge in early Medaka and *Xenopus* embryos; moreover, we found cohesin-dependent fountain pattern on enhancers of mouse embryonic stem cells. Taken together, fountains are the first enhancer-specific elements of chromosome organization; they constitute starting points of chromosome folding during early development, likely serving as sites of targeted cohesin loading.

Introduction

Zygotic genome activation (ZGA) is a critical point in the development of multicellular organisms. It is largely driven by the binding of pioneer transcription factors that open up chromatin and activate enhancers, leading to the awakening of zygotic transcription (reviewed in [1,2]). Dramatic changes in chromosome folding occur in conjunction with ZGA in all the animals studied so far. In *Drosophila*, *Xenopus*, zebrafish, medaka, human, and mouse, hallmarks of chromatin organization gradually emerge after ZGA, with largely featureless pre-ZGA organization[3–6].

On a sub-megabase scale, the genome is folded into topologically associated domains (TADs), and other structures that can be seen in Hi-C as stripes/lines and dots [7,8]. These hallmarks of chromosome organization are evident in all vertebrates, including mammals[9], chicken[10], *Xenopus* [6], and zebrafish [11]. These structures are highly dynamic [12,13] and are formed by the loop extrusion mechanism, where a cohesin complex (an "extruder") loads on DNA and generates a progressively larger DNA loop. Extrusion can be halted by a variety of barriers, with the DNA-bound CTCF being the most prominent [7,14,15]. Functionally, a broad range of evidence now supports the role of cohesin-mediated loop extrusion in mediating and controlling enhancer-promoter interactions [16–18].

Importantly, apart from their architectural roles in the interphase nucleus, both cohesin and CTCF have multiple functions which are essential for embryo survival. Cohesin is required for sister chromatid cohesion during mitosis and meiosis[19]. In mice and zebrafish embryos, depletion of CTCF results in very early lethality accompanied by widespread apoptosis[20]. Thus, unraveling multiple functional roles of cohesin and CTCF in embryo development through acute depletion is challenging.

While the emergence of higher-order nuclear architecture during development occurs with the major ZGA wave, it is not yet clear how these two phenomena are causally connected. TADs and compartments emerge independent of transcription in most developing animals [5,21], and it's unknown how ZGA stimulates the establishment of these structures. Chromosome organization can be established if pioneer transcription factors (TFs) can also modulate chromosome association of key chromatin-folding players such as loop-extruding factors (cohesin, and possibly condensin), and insulator proteins (e.g. CTCF). For example, in *Drosophila*, the removal of the pioneer factor Zelda, a zygotic genome activator, led to the weakening and loss of a subset of domain boundaries [3]. In zebrafish, binding regions of the zygotic genome activators Pou5f3, Sox19b, and Nanog are relatively depleted of CTCF motifs and unlikely to influence CTCF directly [22,23]. Alternatively, pioneer factors may be involved in the recruitment of loop-extruding cohesin. Indeed, cohesin is redistributed during ZGA, accumulating at sites of active transcription, and it overlaps with the binding of the genome activator Pou5f3 after ZGA [24], suggesting the role of pioneer factors in cohesin loading.

In zebrafish, ZGA critically depends on Pou5f3, Sox19b, and Nanog [25,26]. These three TFs act as pioneers to create open chromatin regions, promote H3K27 acetylation on enhancers and activate transcription [22,27,28]. The major ZGA wave occurs at 3 hours postfertilization (hpf) after a series of rapid cellular divisions. In this study, we used Hi-C on wild-type and mutant zebrafish embryos to establish the missing connection between the early activation of genes, patterns of chromatin organization, and loop extrusion. We found that while the pioneer activity is not required for TAD formation, Pou5f3 and Nanog are specifically required for the establishment of a new class of Hi-C features, "fountains". We systematically mapped and characterized fountains as they emerged in early development. Fountains are characterized by broad cohesin enrichment, are devoid of CTCF binding, emerge specifically at early enhancers, and likely reflect sites of preferential cohesin loading. Fountains appear to be starting points for the extrusion-driven folding of chromosomes at developmental enhancers during early zygotic development.

Results

Dynamics of large-scale chromatin spatial organization in zebrafish early embryogenesis

To characterize the dynamics of 3D chromatin organization, we performed Hi-C for sperm and for four stages of *Danio rerio* embryonic development: the last cell cycle before the major ZGA wave (2.75 hours post fertilization, hpf), late blastula (5.3 hpf, 2.3 hours after ZGA), 3-somite stage (11 hpf) and pharyngula (25 hpf) [29] (Fig. 1a).

While Hi-C maps were largely featureless before ZGA, the hallmarks of chromosome organization started to emerge at 5.3 hpf. At the global level, we noticed the formation of compartments (Fig. 1a) and a strong Rabl configuration (Fig. 1b) [30,31], i.e., enrichment of interactions between centromeric regions. At the local level (<1Mb), we observed the formation of features, referred to below as "fountains", that are initially distinct from TADs and morph into TADs only at later stages (Fig. 2a, see below). Compartments and TADs became most pronounced at 11 hpf, and they are already fully formed at this developmental stage (Fig. 1c, Extended Data Fig. 1a). Notably, the first published zebrafish Hi-C study [32] reported that, in contrast to other animals [6,33–36], zebrafish embryos have TADs and

compartments before ZGA. Neither the second zebrafish Hi-C study [4], nor our results support this statement. We did not observe any detectable chromosome organization features before the major ZGA wave (Fig. 1a, Extended Data Fig. 1a).

For comparison of our datasets with the published Hi-C data on *Danio rerio* development [4,32], we constructed "developmental trajectories". We used two approaches: (i) we performed pairwise comparisons of all time points in all datasets and represented them in 2D space, and (ii) we compared global compartmentalization and local insulation at boundaries. According to these representations, our dataset is well-aligned with the previous data. Moreover, this analysis shows that changes in global (compartmentalization) and local (insulation) chromosome organization are continuous and gradual (Fig. 1d).

Emergence of loop extrusion activity upon ZGA

Three observations indicate that loop extrusion activity starts shortly after the major ZGA wave and increases with further embryonic development. (1) Extrusion-associated features of insulation and TADs are rather weak at 5.3 hpf (Extended Data Fig.1a). (2) Analysis of the contact probability, P(s), plots can reveal the presence of loops, irrespective of boundary elements [37] [38]. Loops impact the slopes of P(s) curves, as can be seen from their log derivatives by a hump followed by a drop. Analysis of the time-resolved data shows dramatically different slopes of P(s) before and after ZGA, with signatures of ~50kb loops starting to emerge as early as 5.3hpf, yet becoming pronounced and larger (~100-150kb) at 11 hpf (Extended Data Fig. 1c). (3) Consistently, protein abundance measurements show that components of cohesin are present in the chromatin fraction as early as 5.3 hpf [24]. CTCF protein abundance, however, grows steeper than cohesin protein abundance, consistent with the lack of TADs and insulation [24]. Taken together, the evidence suggests that loop extrusion is initiated at ZGA. This loading of cohesin could potentially impact the activated fraction of the genome and facilitate promoter-enhancer contacts at early activating sites.

Fountains are transient structures emerging after ZGA

Local features that emerge at 5.3 hpf are unique, likely representing early stages of chromosome folding. On this basis, we examined this critical stage of development. Characteristic features emerging at the 5.3 hpf stage constitute *patterns of contacts that emanate from a single genomic locus and broaden with distance from the diagonal.* We refer to these features as fountains and to the point where a fountain emanates from the diagonal as the fountain base (Fig. 2a-c). To further characterize fountains, we developed an automatic *cooltools*-based [39] utility, *fontanka,* that is able to distinguish fountains from potential genomic misassemblies (Fig. 2b, Extended Data Fig. 1b, see Methods). Automatic detection allowed us to see both the characteristic average shape of fountains and the shapes of individual fountains that contribute to the average (Fig. 2a). Using *fontanka,* we detected 1460 fountains consistent across two Hi-C replicates at 5.3 hpf (Supplementary Dataset 1). Hi-C features resembling fountains have been recently reported in Wapl-CTCF depleted mouse cells ("plumes") [40], quiescent mouse thymocytes ("jets") [41], and in a recent study in *C.elegans* ("fountains" [42], see also Supplementary Information). The "fountain score" implemented in *fontanka* correlated with the "protractor" score introduced by Guo et al. to detect "jets" (Extended Data Fig. 1e, f), suggesting that fountains are structurally similar to jets. However, fountains were much more numerous, compared with only 38 jets seen across the mouse genome [41].

As we concluded from average and individual fountain maps, fountains represent enrichments of contacts in an approximately 200 kb range from the base. Fountains are comparable in size yet very different in shape from TADs. In contrast to TADs, fountains are pinched near the diagonal, have a distinct base, and lack sharp boundaries (Fig. 2c). Fountains do not colocalize with TAD centers (Extended Data Fig. 1g). Nor do they resemble CTCF-rich TAD boundaries because they do not insulate flanking regions from each other (Extended Data Fig. 1d). Fountains also look distinct from stripes that emanate from CTCF sites 45 degrees from the diagonal [43] [44]. In fact, fountains show no association with or enrichment of CTCF (Extended Data Fig. 4, and see below).

Together, this analysis shows that fountains constitute patterns distinct from known hallmarks of chromosome organization: TADs, stripes, and compartments. We argue that, while distinct from TADs and stripes, fountains reflect early chromosome folding by cohesin-mediated loop extrusion. First, we found that extruded loops of ~50 kb emerge as early as 5.3 hpf (Extended Data Fig. 1c). Second, we observe cohesin accumulation at fountains at this stage (Extended Data Fig. 2a, and Extended Data Fig. 9c, d). Moreover, fountains with high cohesin levels show a more pronounced pattern of contacts (Extended Data Fig. 9d).

Such accumulation of cohesin and similarity in size to other cohesin-mediated features suggests the role of cohesin in creating fountain patterns. While cohesin accumulation is typically observed at CTCF binding sites and active promoters due to their ability to stop extrusion, the lack of these factors suggests that cohesin is accumulated at fountains through a different mechanism that we explore below.

Fountains form at early developmental enhancers

Aiming to understand the functional role of fountains and their relationship to genome activation, we examined ChIP-seq data for histone modifications and binding of pioneer TFs at the bases of fountains. To differentiate fountains from other extrusion features, we compared them to TAD boundaries. However, since TADs become visible only at 11 hpf, we used these “future” boundaries when considering 5.3 hpf data (compare the top and the bottom panels in Fig. 2d). We found that both fountains and “future” TAD boundaries were enriched in regions of open chromatin, active chromatin marks, Pol II and pioneer TF binding as early as 4-5.3 hpf (Fig. 3a, Extended Data Fig. 2a,b). However, fountains showed strong enrichment of enhancer-specific marks H3K4me1 and H3K27ac and rather weak enrichment of the promoter-specific H3K4me3 mark (see Supplemental Information). “Future” TAD boundaries, in contrast, did not display this preference for enhancer marks and were about equally enriched in H3K4me1 and H3K4me3 (Fig. 3a). Together, this suggests that fountains are enriched in enhancer marks and depleted in promoter marks showing patterns distinct from that of TAD boundaries.

To further characterize the functional states of fountains, we examined the enrichment of functional chromatin states (ChromHMM [45]) at 4.3 hpf, focusing on accessible regions (ATAC-seq peaks overlapping with fountain bases). Surprisingly, we found that among ten ChromHMM states, only the enhancer state shows significant enrichment at fountains, as compared to a random control (p-value<<1e-4, Fig. 3). Neither transcriptionally active states, such as transcription start sites, nor inactive or repressed states are enriched at fountains

(Fig. 3b). Interestingly, fountains at 5.3 hpf correspond not only to the early enhancers (4.3 hpf) but also to the enhancers that become active at a later stage (12 hpf, Extended Data Fig. 2e).

As enhancers are mostly located close to their targets, we asked whether fountains were proximal to actively transcribed genes. We found that fountains were indeed significantly enriched in the proximity of zygotic genes that began transcription between 3 and 5.5 hpf (Extended Data Fig. 2f, g).

To further explore the connection between fountains and enhancers, we performed a complementary analysis that does not rely on *a priori* identification of fountains. We stratified all ATAC-seq peaks by their enhancer (H3K4me1) and promoter (H3K4me3) marks and compared average Hi-C pile-ups centered at ATAC-seq peaks from different categories (Fig. 3c, Extended Data Fig. 3a, d). We observed striking differences in Hi-C patterns on promoters and enhancers. Enhancers showed a pronounced fountain pattern that became even stronger when active enhancers were selected using a combination of high H3K4me1 and high H3K27ac (Fig. 3c, Extended Data Fig. 3b, e). On the contrary, Hi-C pile-ups centered at promoters showed an average insulation border (two domains, one upstream and one downstream from a promoter, with depletion of contacts between the two), consistent with the role of transcription as an extrusion barrier [46] (Fig. 3c, see detailed analysis in Extended Data Fig. 3). This analysis, which is independent of fountain calling, complements the observation of the enrichment of enhancers at fountains and demonstrates that fountains form preferentially on enhancers but not on promoters.

Targeted cohesin loading as a mechanism of fountains formation

Known extrusion-mediated patterns are formed by stopping loop extrusion at specific positions, resulting in the accumulation of cohesin at these positions [43,46]. Fountains are distinct from TADs and stripes yet also accumulate cohesin at their bases (Extended Data Fig. 2a, Extended Data Fig. 9c,d). Since we see no enrichment of CTCF at fountains, (Extended Data Fig. 4) cohesin accumulation there can be due to another mechanism. Thus we hypothesize that cohesin accumulation at fountain bases, and fountains themselves, emerge due to targeted cohesin loading.

To test the targeted loading mechanism, we developed polymer simulations in which cohesin loads preferentially at specific sites (Fig. 4a). We defined the site of cohesin loading as a 1-kb region (one simulation monomer), consistent with recent characterization of regulatory elements [47] and the average length of ATAC-seq peaks at the fountain bases (606 bp) [45]. This contrasts a broad 100-kb loading platform used to explain the formation of “jets” in thymocytes; a 100-kb platform is clearly inconsistent with the narrow (~10 kb) bases of fountains that we observe. Upon loading (targeted or uniform), cohesins perform two-sided loop extrusion, blocking each other upon encountering one another[15].

Next, we identified conditions and mechanisms that lead to the formation of fountain shapes in simulated chromosomes. We observe that the loading of extruders at the loading site and extrusion from there does not produce a fountain, but rather yields a narrow hairpin pattern similar to that observed in bacteria (Extended Data Fig. 5). Gradual broadening of fountains (Fig. 3a) as they emanate from the experimental fountain bases suggests that the extrusion

path of cohesin is not perfectly symmetric, i.e., reeling of DNA on one side of cohesin is not perfectly synchronized with another side (Fig. 5a, b). Asymmetric extrusion by cohesin, when averaged, could produce symmetric fountains with increasing spread with distance from the loading site (Fig. 5b). We suggest three mechanisms that can lead to such asymmetry in extrusion dynamics (Fig. 5c). The first mechanism is the intrinsic asymmetry of extrusion by cohesin, i.e. one-sided extrusion with direction switching, as was long anticipated [43,48] and recently observed in vitro [49]. The second mechanism is external to the motor and constitutes pausing of extrusion on randomly positioned boundaries: while stochastically pausing on a boundary, cohesin can continue reeling chromatin on the other side. Such randomly positioned boundaries have been observed, e.g., MCMs [50], and genes can serve as multiple weak boundaries[46]. The third mechanism is desynchronized extrusion due to the pausing of targeted-loaded cohesin upon collisions with other extruders (e.g., uniformly loaded cohesin). While we cannot distinguish between these mechanisms, we demonstrate below that the last mechanism can also explain the weakening of fountains later in development.

Thus, in addition to targeted loading, we introduce background loading of extruders uniformly along the genome. We then sweep the parameters of the model, including the background and targeted loading rates and cohesin processivity. Our parameter sweeps (Extended Data Fig. 5a-b, Extended Data Fig. 6) revealed that two conditions should be met to reproduce the shapes and length of fountains: (1) targeted loading at a 1 kb loading site should be ~10 times greater than the background loading (per 1 kb); (2) cohesin is loaded throughout the genome at approximately 1 per 150 kb (which is consistent with cohesin density in other vertebrate systems [15] and aligns well with a peak on the contact frequency P(s) curve at 5.3 hpf (Extended Data Fig. 1c). Formed through this mechanism, fountains broaden due to interactions with background-loaded cohesins and thus resemble experimental fountains at 5.3 hpf (0.77 Spearman correlation, Extended Data Fig. 5b). Simulations of this mechanism also reproduce a broad (~50-100 kb) peak of cohesin accumulation seen in ChIP-Seq (Fig. 4, Extended Data Fig. 5,6).

Importantly, all these mechanisms do not rely on CTCF for the formation or broadening of fountains, consistent with the absence of CTCF near fountains in zebrafish (Extended Data Fig. 4. Consistently, a recent study by Isiaka et al. [42] observed prominent fountains in *C. elegans* that lack CTCF (lost in nematode evolution). This contrasts the formation of “stripes” at CTCF-bound sites across vertebrate genomes [15 44] and a putative mechanism of formation of ”jets” [41] where CTCF was suggested to be important. Interestingly, in simulations, each fountain contains, on average, ~1 cohesin in the steady state. This indicates that they are different from bacterial hairpins, where multiple (~15-30) extruders of bacterial condensin were simultaneously moving along a hairpin from the loading to the unloading sites [51,52]. Together, our modeling suggests that fountains can result from facilitated (~10-fold) cohesin loading at a narrow site and extruding from there; the fountain shape is a result of interactions between targeted-loaded cohesins with the background-loaded ones or potentially other barriers or the internal properties of cohesin motors.

We also considered and ruled out an alternative mechanism by which fountains are formed by affinities of fountain-proximal regions, i.e., compartment-like mechanisms. If two regions proximal to the fountain origin had higher affinity for each other, they could form such a structure. In this case, however, these “sticky” regions of different fountains would have

affinities for each other, generating enrichment of interactions between fountains [53]. We tested this by examining Hi-C contacts between the fountains and found no such enrichment, ruling out the affinity-mediated mechanism (Extended Data Fig 5c-d).

Changes of fountains through development

We next examined how fountains change through development. We observe that fountains gradually transform into other patterns at 25 hpf, such as dots and TADs, which are known to be formed by loop extrusion (Fig. 2a, c). Fountains are nevertheless visible at 25 hpf as an average pattern but not individually (Fig. 2c-d). possibly masked by TADs, interrupted by CTCFs or suppressed by other mechanisms (see below).

We next examined cohesin and CTCF localization at early (4.5 hpf) and late (12 hpf) time points on the fountains detected at 5.3 hpf, comparing them to “future” TAD boundaries detected at 11 hpf. Cohesin was enriched at the fountains only at 4.5 hpf and depleted at 12 hpf, in contrast to boundaries, where cohesin occupancy increased with time (Extended Data Fig. 2c, Extended Data Fig. 9a, b). As a proxy of CTCF occupancy, we used ATAC-seq signal on CTCF motifs at 4.5 and 12 hpf, and ChIP-seq data at 10 hpf (Extended Data Fig. 2c). CTCF was not enriched at fountains at either time point (Extended Data Fig. 9a, b). Consistently, there was no insulation on the 5.3 hpf fountains at 11 hpf, in contrast to the strong insulation observed on TAD boundaries, which were fully formed at this stage (Extended Data Fig. 2d). Thus, we observed the global redistribution of cohesin from fountains to TAD borders as development progressed (Extended Data Fig. 9a-b). These observations render fountains as transient structures emerging at ZGA and later morphing into known extrusion-mediated features.

Interestingly, the mechanism that combines targeted and background loading can explain the weakening of fountains later in development (Fig. 2c-d). Simulations show that three mechanisms can contribute to the disappearance of fountains: (1) a reduction of targeted loading (Extended Data Fig. 5a); (2) an increase in the background loading (Extended Data Fig. 5b), and (3) the appearance of strong CTCF boundaries that can stop and sequester cohesin. The role of CTCF in contracting fountains is consistent with the persistence of fountains in adult *C. elegans* that lacks CTCF [42]. Distinguishing these mechanisms would require experiments in vertebrate systems where rapid and complete cohesin and CTCF depletions are feasible, other than rapidly dividing cells of the zebrafish embryo. Zebrafish embryo, however, allows to examine a connection between fountains and activity of ZGA enhancers controlled by pioneer transcription factors.

Knocking out pioneer transcription factors leads to reorganization of fountain structure

Early deposition of enhancer genome marks in zebrafish is controlled by the binding of the zygotic genome activators, transcription factors Pou5f3, Sox19b, and Nanog (PSN). These factors act as pioneers by creating open chromatin regions and recruiting histone modifiers, such as p300, to their binding sites [22–28]. We found that fountains were indeed enriched in ChIP-seq signals for Pou5f3, Sox19b, and Nanog (Fig. 3a), suggesting that the pioneer TF binding could play a role in early cohesin loading and facilitate the formation of fountains. Our model of targeted loop extrusion predicted that the removal of the cohesin loading site would lead to the disappearance of the fountain. If such loading sites require chromatin

accessibility and enhancer activity, we expect the removal of Pou5f3, Sox19b or Nanog to disrupt fountain formation on the regions where these factors bind and open chromatin.

To test this mechanism, we performed Hi-C on the 5.3 hpf embryos, null-mutant for Pou5f3 [54], Sox19b [23], Nanog [55], and also on double or triple mutant embryos bearing combinations of two or three of these mutations. We performed two types of analysis: the one that uses called fountains and the one independent of the fountain calling.

Using our approach that stratified ATAC-seq peaks by enhancer (H3K4me1) and promoter (H3K4me3) marks and doesn't rely on fountain calling, we found that the pronounced fountain signature on enhancers faded in the triple mutant, MZ*triple* (Extended Data Fig. 3). In a different analysis, we examined Hi-C patterns at ATAC-seq peaks that changed in MZ*triple*, irrespective of their histone modifications. We found that genomic sites with strong ATAC-seq peaks in the wild-type that were suppressed in MZ*triple* exhibited a similar suppression of the fountain pattern in MZ*triple* (Fig. 6a), but not other categories of ATAC-seq peaks did not exhibit similar behavior. We found that sites where H3K27ac or H3K4me1 were strong in the wild-type and went down in MZ*triple*, also showed a fountain pattern in the wild-type that vanished in MZ*triple* (Fig. 6b-c, also see Supplemental Information). In summary, the loss of accessible regions upon deletion of pioneer TFs caused the loss of fountain structures.

A complementary analysis of fountains called in the wild-type revealed that fountains that were weakened in MZ*triple* had the strongest enhancer signature and were most enriched in PSN binding, compared with strengthened and unchanged fountains (Fig. 7). Consistently, fountains weakened in MZ*triple* corresponded to early enhancers, i.e., had ATAC-seq peaks with enhancer ChromHMM signatures at 4.3 hpf (Extended Data Fig. 7a). Fountains strengthened in MZ*triple* corresponded to late enhancers, i.e., they had enhancer signatures and load RNA polymerase II at 12 hpf (Extended Data Fig. 7b-e). The last finding echoed our recent observation [56] that late enhancers become aberrantly open at ZGA (Fig. 7b), and activate the premature expression of late target genes in the absence of PSN (Extended Data Fig. 7f-g).

It was recently shown that PSN binding results in increased chromatin accessibility and p300-dependent histone acetylation of more than half of genome enhancers [22]. Accordingly, upon triple knockout of PSN, chromatin accessibility, H3K27ac, and p300 occupancy selectively diminished on the weakened fountains while exhibiting either increases or small effects at the strengthened fountains (Fig. 7). Interestingly, the loss of H3K27ac upon p300 inhibition by SGC[4] was associated with weakening of fountains but did not result in the selective collapse of fountains at the sites of PSN binding (Extended Data Fig. 7h, i). This observation suggests that H3K27 acetylation makes fountains stronger but is not necessary for their formation, in line with our finding of the key role of H3K4me1 in the formation of fountains. Thus, fountains that were weakened in MZ*triple* correspond to PSN-dependent early enhancers, while those that were strengthened represent PSN-independent late enhancers.

Pou5f3 and Nanog binding specifically contribute to fountain structure formation by creating open chromatin regions

We next asked if the pioneer TFs' activity selectively led to fountains formation. To address this question, we used our recent finding that Pou5f3 and Nanog promoted chromatin accessibility in different genomic locations [56]. We excluded individual effects of Sox19b from the analysis because of the small number of fountains that contain ATAC-Seq peaks dependent on Sox19b only [56].

In the embryos mutant by Pou5f3, or Nanog, or by all three TFs, chromatin accessibility was significantly reduced on 441 out of 1460 fountain bases (30% of all fountains with FDR <5%). We classified these 441 fountain bases into four groups based on the requirements of the TFs for chromatin accessibility. Namely, in the group (I. P), Pou5f3 but not Nanog was required; in the group (II. N), Nanog but not Pou5f3 was required; in the group (III. T), one of the two factors was sufficient; in the group (IV. All), both Pou5f3 and Nanog were required to open chromatin. In the group (V. the rest), fountain bases PSN were dispensable for chromatin accessibility (Fig. 8a).

We asked then whether the same factor that was required to open chromatin on the fountain base was also required for fountain formation. This was indeed the case (Fig. 8b-e, see Extended Data Fig. 8 for the additional analysis of the double mutant embryos). The bases of fountains, that depend on Pou5f3 or Nanog, have higher enrichment of respective factors than the bases of PSN-independent fountains (Extended Data Fig. 7j). This suggests that the fountain formation is a consequence of TF binding on the fountain base. We concluded that chromatin opening by pioneer activity of Pou5f3 and Nanog is key to fountain formation, likely via facilitating targeted cohesin loading.

Altogether, our analysis established a causal relationship between the binding of pioneer transcription factors, the establishment of enhancers, and the formation of fountains: loss of transcription factors leads to failure to activate enhancers and establish fountains. These results underscore multiple and not necessarily interdependent functions of the pioneer factors at ZGA, including their role in promoting early stages of genome folding.

Fountain structures emerge in embryogenesis of other species, in cell lines at enhancers, and are cohesin-dependent

Finally, we asked whether fountains are present in other species and in systems beyond early development. First, we analyzed Hi-C data for development in medaka fish [35] and in *Xenopus tropicalis* [6]. With our fountain caller, we detected individual fountains at stages close to ZGA in these species that provide a clear fountain signal similar to that in zebrafish (Fig. 9a, b). This confirms that fountain structures are not specific to development of zebrafish and are widespread among vertebrates during development.

Next, we examined Mirco-C from mouse embryonic stem cells (mESC) [57]. Although individual fountains were difficult to call automatically as they were overshadowed by TADs, dots and stripes, we piled Micro-C maps at the genomic positions of active enhancers distant from genes (cis-regulatory element of mESC defined recently by ENCODE [58]) and observed distinct fountain pattern (Fig. 9c). This argues that targeted cohesin loading at enhancers is not restricted by early stages of embryonic development.

Finally, we found that enhancer-associated fountain pattern vanishes up cohesin depletion in mECS cells (Fig. 9d). Together these results suggest that fountains are universal cohesin-dependent elements of chromosome architecture emerging at enhancers.

Discussion

By analyzing chromosome organization in early zebrafish development, we found a new class of chromosome structures - fountains, with more than a thousand emerging across the genome around the time of genome activation. We demonstrated that fountains are distinct from known hallmarks of local chromosome folding such as TADs, dots, and stripes -- all of which are associated with the pausing of extruding cohesin by CTCFs, and, weakly, by transcribed genes. Fountains, in contrast, are not associated with CTCF or positions of TAD borders, and do not directly depend on transcription. Importantly, we demonstrated that fountains are not just a peculiarity of a Hi-C map, but rather, they show a striking association and causal relationship to enhancers that are activated at these early stages of development. In synergy with our efforts, a study of Isiaka et al. [42] found enhancer-associated fountains in *C. elegans*, the organism which lacks CTCF, supporting our finding that fountains are formed by a CTCF-independent mechanism.

Others have seen similar patterns in mutants with altered extrusion dynamics ("plumes" in CTCF/Wapl double depletion [40]) or in specific cellular states ("jets" in quiescent thymocytes [41] and "flares" in sperm cells [4]). These features weren't nearly as abundant as fountains (38 "jets", 105 "plumes", 333 "flares": vs. 1460 fountains, see Supplementary Information for discussion), which mark such a fundamental process as ZGA. While different in sizes, shapes, biological contexts, and, possibly, underlying mechanisms, these jets/plumes/fountains may nonetheless constitute a new pattern of chromosome organization.

What kind of mechanisms can generate such features? The resemblance to other local features (~100-200 kb in size) and the accumulation of cohesin at fountains suggest that they can result from cohesin-mediated extrusion. Cohesin-mediated mechanism of fountain formation is strongly supported by a recent study in *C.elegans* [42], which report fountains disappearance upon cleavage of COH-1 cohesin. We, here, and others[41], argued that targeted loading of cohesin followed by bidirectional extrusion could underlie the formation of such a pattern. However, we demonstrate that targeted loading is insufficient since it produces narrow hairpins, akin to patterns seen in bacteria, rather than fountains (Extended Data Fig. 5). We propose that the characteristic broadening shapes of fountains can emerge if the reeling of DNA into an extruded loop from the two sides of cohesin is desynchronized. Such desynchronization can be an inherent property of cohesins, as was recently observed *in vitro* [49], or it can result from collisions with randomly positioned boundaries and/or with other cohesins loaded at other places (through uniform background loading) (Fig. 4).

Among mechanisms of fountain formation, targeted loading acting alongside uniform loading can explain several phenomena. This mechanism can also explain why fountains are more difficult to see as individual patterns in adult cells. Two factors can be at play: (i) abundant loop-extruding cohesins immediately collide with the targeted-loaded cohesin, preventing its further propagation and the formation of a fountain; (ii) patterns formed by other cohesins at CTCF can overshadow fountains. Consistently with this mechanism, we see that at early stages, cohesin is accumulated at fountains and less so at CTCF sites. In contrast, at later

stages, accumulation at fountains diminishes while accumulation at CTCF sites becomes more pronounced (Fig. 3b).

Our simulations also show that fountains formed by targeted loading are extremely dynamic, with around one cohesin present at a fountain at any moment and cohesin traveling through the fountain in mere minutes. We expect fountains to be even more transient and short-lived than CTCF-CTCF loops, which were recently demonstrated to remain for mere ~20 minutes [12,13].

What can be the functional role of fountains? We found that fountains are strongly associated with active enhancers, defined as ATAC-seq peaks enriched in H3K4me1 and H3K27ac (Fig. 3, Extended Data Fig. 3). This association was two-way specific, i.e., among all types of functional annotations, active enhancers were the only ones enriched at locations of fountains (Fig. 3). And conversely, an average Hi-C map at ATAC-seq peaks that also have an H3K4me1 enhancer mark shows a pronounced fountain pattern (Extended Data Fig. 3). Importantly, the latter observation does not rely on definitions or calls of the fountains. Thus, the fountain pattern emerges as a Hi-C feature formed specifically at enhancers and can shed new light on mechanisms of enhancer-promoter interactions.

While it is natural to expect cohesin loading at accessible regions, such as enhancers, surprisingly, other accessible regions, such as promoters or gene bodies, do not show any association with fountains. In fact, an average Hi-C map at active promoters doesn't show the fountain pattern, instead yielding an “anti-fountain” insulation pattern (akin to TAD-boundary) (Fig. 3c). Moreover, preferential cohesin loading at promoters has been recently examined and ruled out [46]. Given the overall similarity between enhancers and promoters, such a selective loading of cohesin on enhancers is surprising, even though it has been suggested in various systems [59,60]. This distinction in cohesin loading between equally accessible enhancers and promoters indicates that cohesin loading is mediated by an enhancer-specific mechanism rather than mere DNA accessibility. Such mechanisms may include interactions with pioneer transcription factors, H3K4me1 and H3K27ac-marked nucleosomes, and/or chromatin modifiers or remodelers [61].

Knockout of pioneer transcription factors (Pou5f3, Sox19b, and Nanog) allows us to establish a causal link between enhancers and fountains. The loss of enhancers (loss of ATAC-seq peaks, H3K27ac, or H3K4me1) in MZ*triple* led to the loss of fountains at these sites (Fig. 6). The effect of pioneer knockouts is specific since only the regions that lost enhancer features lost fountains, while unaffected enhancers continued showing the fountain pattern. This result does not depend on the calling of fountains, demonstrating that fountains emerge as a feature of early enhancers created by pioneer transcription factors. Interestingly, fountains that become stronger upon deletion of pioneer factors correspond to enhancers activated later in the development (~11hpf) (Fig. 6, Fig. 7, Extended Data Fig. 7b). A parallel study in adult *C.elegans* (Isiaka et al., 2023) demonstrated that elimination of fountains (by cohesin COH-1 cleavage) did not result in enhancer weakening, supporting directional causal relationship between enhancer activity and fountain formation. Together, our knockout experiments reveal a cause-consequence relationship between the binding of transcription factors and the establishment of fountains.

While fountains facilitate contacts between regions flanking an enhancer, fountains do not facilitate contacts of the enhancer itself. This prompts the question: how can fountains help enhancers to perform their functions? One possibility is that targeted cohesin-loading at enhancers facilitates contacts between other regulatory regions on one side of an enhancer with genes on the other. Another possibility is that relatively low-resolution Hi-C maps that show fountains may miss less prominent contacts between the enhancer and nearby promoters. Such interactions could be facilitated by some fraction of cohesin loaded at enhancers that remain stuck there, thus performing one-sided extrusion. Seeing such a pattern of enhancer-promoter contacts alongside individual fountains would require much deeper contact mapping data [62].

Mutants also help to reveal the functional roles that fountains might play in mediating enhancer-promoter interactions. In the wild-type, we observe a broad (~50 kb) RNAPol II accumulation pattern around fountains (Extended Data Fig. 2a). In the MZ*triple* mutant, we observe the selective loss of RNA Pol II around fountains that weakened, but not those that were not affected or strengthened compared to the wild-type (Extended Data Fig. 7e). Although we cannot delineate the roles of fountains from that of the enhancers that they localize on, broad and PSN-dependent activation of transcription around such enhancers suggests that fountains could be involved in the spreading enhancer activity over ~50 kb regions.

Broadly, our analysis shows that loop extrusion activity initiates upon ZGA. Yet structures formed by extrusion change through development due to different activity of cohesin-regulating factors. Cohesin-loading activity appears to be the primary factor folding chromosomes (upon ZGA), while cohesin-stopping factors shape chromosomes at later stages. Surprisingly, key regulatory loci that drive transcription, early enhancers, also appear to be the elements driving early folding events by recruiting cohesins. Extruding from there and interacting with other cohesins, they form the new elements of chromosome architecture - fountains. This new role of enhancers in actively folding chromosomes can shed new light on the mechanisms of longer-range regulatory activity of enhancers in development and beyond.

Data availability

Raw and processed Hi-C data for zebrafish embryogenesis is available at GEO at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE195609. The list of fountains is available in Supplementary Table 1, and fountains group assignment based on chromatin accessibility change in mutants is available in Supplementary Table 2. The detailed list of the sequencing datasets is available in Supplementary Table 3.

Code availability

Cooltools[39]-based tool *fontanka* for fountain calling with an example is available at https://github.com/agalitsyna/fontanka with the code of fountain calling under *examples* directory. Simulations of targeted loop extrusion are available at https://github.com/agalitsyna/polychrom_workbench under *targeted_extrusion* directory.

Author contributions:

SU: Hi-C libraries; AxG: Hi-C and genomic analysis, polymer simulations; NB, KP, MiG: bioinformatics; SU, MV, MeG, DO: zebrafish mutants; DO: study design; SR, DO: supervision of the wet part, funding acquisition; MiG, LM: supervision of Hi-C genomic analysis, simulations. AxG wrote the first draft, and SU, SR, LM, and DO developed and edited the manuscript.

Acknowledgments:

We are grateful to Edward Banigan for proofreading the text and to all members of the Mirny lab for many productive discussions. We thank Damir Baranasic for sharing the data on chromatin states before publication. AxG thanks Max Imakaev, Nezar Abdennur, and Anton Goloborodko for early ideas for the polymer simulations; Henrik Pinholt, Simon Grosse-Holz, and Emily Navarrete for ideas on the model and figures; Antoine Coulon, Mark Pownall, and Christopher Bohrer for the thorough discussions of the results and critical assessment of the hypotheses; students Alexey Shkolikov (FBB MSU) and Eli Rybnikova (PRIMES MIT) for their preliminary analysis of fountains. DO is grateful to Mariia Bazarevich and Prof. Lev Yampolsky for the consultation about statistical analysis. This work was supported by DFG-ON86/5-2 and DFG-EXC2189 for DO, by GM114190 to LM, and by Russian Science Foundation (grant #21-64-00001) to SR.

Materials and methods

Experimental model and subject details

Wild-type fish of AB/TL and mutant strains were raised, maintained, and crossed under standard conditions as described by Westerfield [63]. Embryos were obtained by natural crossing (4 males and 4 females in 1,7 l breeding tanks, Techniplast). Wild-type and mutant embryos from natural crosses were collected in parallel in 10-15 minute intervals and raised in egg water at 28.5°C until the desired stage. The staging was performed following the Kimmel staging series [29]. Stages of the mutant embryos were indirectly determined by observation of wild-type embryos born at the same time and incubated under identical conditions. All experiments were performed in accordance with German Animal Protection Law (TierSchG) and European Convention on the Protection of Vertebrate Animals Used for Experimental and Other Scientific Purposes (Strasburg, 1986, https://rm.coe.int/168007a67b). The generation of double and triple mutants was approved by the Ethics Committee for Animal Research of the Koltzov Institute of Developmental Biology RAS, protocol 26 from 14.02.2019.

Generation of MZ*sn*, MZ*pn* double mutant and MZ*triple* triple mutant embryos and maintenance of the mutant fish lines

Maternal-Zygotic (MZ) homozygous mutant embryos MZ*sn* (MZ*sox19b* ^{m1434}*nanog* $^{m1435)}$ were obtained in three subsequent crossings. First, MZ*nanog* ^{m1435} [55] homozygous null-mutant males were crossed with MZ*sox19b* ^{m1434} [23] homozygous null-mutant females. The double heterozygous fish were raised to sexual maturity and incrossed. The progeny developed into phenotypically normal and fertile adults. Genomic DNA from tail fin biopsies was isolated and used for genotyping. We first selected *sox19b-/-* homozygous mutants by PCR-with Sox19b-f1/Sox19br1 primers followed by restriction digest with BbsI, as described in [23]. To select the double homozygous fish, we used the genomic DNA from *sox19b-/-* homozygous mutants by PCR with Nanog-f2/Nanog-r2 primers followed by restriction digest with NdeI, as described in [55]. MZ*sn* embryos for experiments were obtained from incrosses of double homozygous *sox19b*-/-; *nanog*-/- fish. The line was maintained by crossing *sox19b*-/-, *nanog* -/- males with *sox19b*-/-, *nanog* +/- females.

MZ*sn* homozygous mutant embryos were obtained in three subsequent crossings. First, MZ*nanog* ^{m1435} homozygous males [55] were crossed with MZ*spg* ^{m793} [54] homozygous females. *Spg*^{m793} allele carries an A->G point mutation in the splice acceptor site of the first intron of Pou5f3 gene, which results in the frameshift starting at the beginning of the second exon, before the DNA-binding domain. Spg^{m793} is considered to be null allele. The double heterozygous fish were raised to sexual maturity and incrossed. To bypass the early requirement for Pou5f3 in the *spg*^{m793} homozygous mutants, one-cell stage embryos were microinjected with 50-100 pg synthetic Pou5f3 mRNA as previously described [54]. The fish were raised to sexual maturity (3 months), and genomic DNA from tail fin biopsies was isolated and used for genotyping. We first selected *nanog*-/- homozygous mutants, by PCR with Nanog-f2/Nanog-r2 primers followed by restriction digest with NdeI [55]. To select the double homozygous fish, we used the genomic DNA from *nanog*-/- homozygous mutants to

PCR-amplify the region flanking the *spg*^{m793} allele. We used the following PCR primers: spg-f1 5'-` GTCGTCTGACTGAACATTTTGC -3' and spg-r1 5'-` GCAGTGATTCTGAGGAAGAGGT -3'. Sanger sequencing of the PCR products was performed using a commercial service (Sigma). The sequencing traces were examined and the fish carrying A to G mutation were selected. MZ*pn* embryos for experiments were obtained from incrosses of double homozygous *spg*-/-;*nanog*-/- fish. The line was maintained by crossing *spg*-/-, *nanog* -/- males with *spg*-/-, *nanog* +/- females and microinjecting Pou5f3 mRNA in each generation.

To obtain the triple Maternal-Zygotic homozygous mutant embryos MZ*triple*, MZ*sn* double homozygous males were crossed with MZ*ps* (MZ*sox19b*^{m1434}*spg*^{m793} double homozygous females [23]. The *sox19b-/-*; *spg+/-*; *nanog* +/- progeny was raised to sexual maturity and incrossed. The progeny was microinjected with 50-100 pg synthetic Pou5f3 mRNA at one cell stage, raised to sexual maturity and genotyped as described above. MZ*triple* embryos for experiments were obtained from incrosses of triple homozygous fish. The line was maintained by crossing *sox19b*-/-;*spg*-/-, *nanog* -/- males with *sox19b*-/-;*spg*-/-, *nanog* +/- females and microinjecting Pou5f3 mRNA in each generation.

Genomic DNA isolation and PCR for genotyping

Genomic DNA was isolated from individual tail fin biopsies of 3 months old fish. Tail fin biopsies or embryos were lysed in 50µl lysis buffer (10 mM Tris pH 8, 50 mM KCl, 0.3% Tween20, 0.3% NP-40, 1mM EDTA) and incubated at 98°C for 10 min. After cooling down Proteinase K solution (20 mg/ml, A3830, AppliChem) was added and incubated overnight at 55°C. The Proteinase K was destroyed by heating up to 98°C for 10 min. The tail fin biopsies material was diluted 20x with sterile water. 2µl of was used as a template for PCR. PCR was performed in 25-50 µl volume, using MyTag polymerase (Bioline GmbH, Germany) according to the manufacturer's instructions, with 30-35 amplification cycles.

Zebrafish sperm collection

Adult male fish were anesthetized with Tricain (4% 3-Aminobenzoic acid ethyl ester, pH 6.7) and then positioned with the anal area above. The sperm was taken using a capillary, mixed with 5 µl E400 buffer (9.7 g KCl, 2.92 g NaCl, 0.29 g $CaCl_2$ -2H_2O, 0.25 g $MgSO_4$ -7H_2O, 1.8 g D-(+)-Glucose, 7.15 g HEPES in 1L dH_2O, pH 7.9) and then with 150µl SS300 buffer (0.37 g KCl, 8.2 g NaCl, 0.15 g $CaCl_2$ -2H_2O, 0.25 g $MgSO_4$ -7H_2O, 1.8 g D-(+)-Glucose 20ml 1M Tris-Cl, pH 8.0, in 1L dH_2O). The collected sperm was used for nuclei isolation.

Hi-C library preparation

The embryos were obtained from natural crossings in mass-crossing cages (4 males + 4 females). 5-10 cages were set up per genotype, and the eggs from different cages were pooled. The freshly laid eggs were collected in 10-15 min intervals. Embryos were incubated at 28.5°C and dechorionated with pronase E (0.3 mg/ml) shortly before the desired stage. 400-600 embryos were homogenized in 2 ml 0.5 % Danieau's with protease inhibitor cocktail (PIC, Roche) and 1 % (v/v) Methanol-free Formaldehyde (Pierce) and fixed for 10 min on a rotating platform. The fixation was stopped with 0.125 M Glycine by shaking for 5 min on a

rotating platform. Cells were pelleted on the tabletop centrifuge for 5 min, 500 g, and washed three times with PBST (16mM Na_2HPO_4, 4mM NaH_2PO_4,0.08% NaCl(w/v), 0.002% KCl (w/v), 0.1% Tween 20, pH 7.5), with protease inhibitors. The cells were lysed for 1 min in 1 ml lysis buffer (10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 0.5 % NP-40) on ice. The pellet was washed 2 times with 1 ml ice-cold 1x PBST. In order to count the obtained nuclei, the pellet was resolved in 1 ml ice-cold 1x PBST, of which 10 µl were diluted 1:1 with 12 µM Sytox® green. The nuclei were scored under a fluorescence microscope using the Neubauer counting chamber. The residual nuclei were snap-frozen in liquid nitrogen and stored at -80 °C. 2.5-3 million nuclei were used for one Hi-C experiment according to the published protocol [64].

Chromatin accessibility changes on fountains in MZ*triple*, MZ*spg* and MZ*nanog* mutants compared to the wild-type

To evaluate the changes in chromatin accessibility on fountains in the absence of Pou5f3, Nanog, or all three zygotic genome activators, we analyzed ATAC-seq data in the respective mutants and the wild-type [56] using the European galaxy server [65].

The coverage of ATAC-seq reads on 1460 fountains was scored using MultiCovBed (Bedtools) in four ATAC-seq replicates of the wild-type, three replicates of MZ*spg* at 4.3 hpf (GEO accession number: GSE188364), two replicates of MZ*nanog* and four replicates of MZ*triple* at 4.3 hpf (GEO accession number: GSE215956). Deseq2 [66] was used to normalize the reads and compare chromatin accessibility in each mutant to the wild-type. The fountain was scored as weakened in the mutant if the log2 fold change to the wild-type was negative with FDR < 5%. As a result of this analysis, 1460 fountains were divided to five groups, as shown in Fig. 8a: I. P - downregulated in MZ*spg*, II. N - downregulated in MZ*nanog*, III. T - downregulated in MZ*triple*, not in MZ*spg* not in MZ*nanog*, IV. All - downregulated in MZ*triple*, MZ*spg* and MZ*nanog*, V. the rest - not downregulated in three mutants. Group assignments, Deseq2-normalised ATAC-seq values per replicate, average ATAC-seq values, and the values of log2 fold change in each of the mutants compared to the wild-type are listed in the Supplementary Dataset 2 for all fountains and were used for Figure 8d.

Hi-C data mapping

Hi-C reads were mapped with Open2C *bwa-mem* [67] and *pairtools*[68]-based pipeline *distiller-nextflow* version 0.3.3. For mapping, we used UCSC's danRer11 assembly with removed unplaced contigs [69]. For parsing, we used the walks policy "all" for parsing pairs [68]. This increased the library pairs yield by parsing all possible pairs from each sequenced read pair. Cool files were created by *cooler,* and replicates were merged by summation [70].
Normalization was done by iterative correction [71] with default parameters (minimal count per genomic bin 0, minimum number of non-zero pixels per bin 10, maximum median absolute deviation of pixel values per bin 5, both cis- and trans-contacts included, tolerance 1e-05 and 200 maximum iterations).
Processed coolers from this work are available at GEO [72] https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE195609. We also re-processed data from previously published zebrafish embryogenesis works [4,32], and made the output available through the Open Science Foundation [73] website: https://osf.io/mt4vf/.

Hi-C data processing

Hi-C data visualization

Normalized Hi-C signal was visualized with HiGlass genomic browser [74] (Figure 1a, 2a), and resgen [75] (Extended Data Fig. 2b) in log-scales and “fall” colormap.

Centromere positioning in *Danio rerio*

We visualized Hi-C maps for *Danio rerio* in HiGlass genomic browser [74] and manually marked the locations of centromeres following criteria: (i) centromeres must be located far from chromosome ends; (ii) regions around centromeres should have enriched interactions with regions around other centromeres in trans; (iii) centromeres should have unmapped DNA as a sign of centromeric repeats. The beginning and the end of the unmapped DNA served as centromere start and end, respectively. We later use this annotation for plotting the Rabl configuration.

P(s) curves and derivatives

To calculate P(s) curves, also named scalings, for Extended Data Fig. 1c, we used *cooltools* [39] v0.5.4 *expected_cis* at 1 kb resolution. We removed all Hi-C contacts shorter than 2 kb, binned contacts in log space to produce evenly sized points in double log coordinates, then smoothed contacts with Gaussian kernel with sigma 0.05 to even out noisy interactions at the ends of chromosomes and applied aggregation per chromosome arm (defined as in “Centromere positioning”). P(s) derivatives were calculated by taking the gradient at each point of the resulting P(s) plot. For each dataset, the approximate distance ranges of peak locations were manually marked and then validated by searching for local maxima of the derivative plots at distances under 120 kb.
Note that for plotting saddle plots and snipping Hi-C data, we also calculated expected P(s) (see below), but did not smooth and did not aggregate contacts for these separate analyses.

Removal of poorly mapped and surrounding genomic regions

While analyzing centromere positions, we noticed that the sequences Hi-C data frequently has rearrangements and poorly mapped regions. Further analysis, thus, required the exclusion of poorly mapped regions. For that, we created a track of unmappable genomic bins (or “bad bins”).
Firstly, we defined bin as “bad” if it was impossible to normalize through the default cooler balancing (see criteria in the “Hi-C data mapping”) in at least one of the replicate or merged datasets. Secondly, we marked all genomic regions located at least 50 Kb away from these bins as bad because they might be affected by balancing problems due to the absence of short-range interactions. Moreover, bad bins frequently had genomic rearrangements around them which might affect TAD and fountain calling.
These two criteria resulted in the extended list of 49843 genomic bins subject to removal from most of the Hi-C analyses (out of the total 134526 bins, with 4799 bad bins located at chr4).

Building developmental trajectories

To build developmental trajectories, we calculated pairwise SCC (stratum-adjusted correlation coefficient) [76] between all zebrafish Hi-C datasets from previous studies and this work (including replicates and merged datasets). For that, we used the R library HiCRep [76] with a minimum genomic separation of 0, maximum separation of 5 Mb, window size of 3, and data resolution of 100 Kb. The resulting SCC was input to PCA analysis with *sklearn* [77]. The first two PCs jointly explained around 79% of the variance, with a substantial drop at PC3: PC1 47.7%, PC2: 31.4%, PC3: 9.9%, PC4: 6.2% (Figure 1d).

Average Rabl configuration

To assess Rabl configuration (Figure 1b), we plotted average Rabl pileups. We extracted the trans interactions for each pair of chromosome arms at 250 Kb resolution (per-arm snippets). We reshaped per-arm snippets of different sizes to the 200x200 squares with *cooltools* [78], which is a common strategy for averaging Hi-C pileups of different sizes [79]. We then averaged all snippets and displayed log2 over expected (average number of trans-interactions).

The centromeric score measures the enrichment of contacts between centromeres of different chromosomes relative to all trans interactions. To calculate the centromeric score, we started with the average Rabl pileup (reshaped and normalized), took a 40x40 window around its center, and calculated the median value.

Compartment calling and saddle plots

For compartment calling (Figure 1c), we used *cooltools* cis eigendecomposition with *eigs_cis* [39]. We used cis decomposition and not trans to avoid the influence of prominent Rabl configuration.

Phasing of the eigenvector is an important step of compartmental calling that defines the sign of the eigenvector with A compartment [8,71]. Phasing of the mammalian first eigenvector is based on correlating the eigenvector with wither number of genes or GC content [8,71,80], because both are enriched in A compartment of mammals [39]. For example, in mouse, the first eigenvector at the late 2-cell stage of embryo development [33] highly correlates with both GC content (Pearson corr. 0.64 at 25 kb bin size) and coverage by genes (Pearson corr. 0.28).

Surprisingly, the first eigenvector of zebrafish Hi-C at 5.3 hpf does not correlate with GC content (Pearson corr. of -0.037 at 25 kb bin size) and poorly correlates with coverage by genes (P.corr. 0.137). However, the first eigenvector at 5.3 hpf correlates well with replication timing at shield stage [81] (P. corr. 0.59). Thus, we phased the first eigenvector of Hi-C maps with replication timing instead of conventional GC content. 11 hpf first eigenvector correlated the best with bud RT (P. corr 0.65), and 25 hpf first eigenvector correlated the best with RT measured at 28 hpf (P. corr 0.65) [81].

Saddle plots (Figure 1c) were constructed by aggregating observed over expected cis Hi-C signal at 100 kb (calculated by normalizing balanced Hi-C signal by expected with no smoothing), based on grouping into 20 groups by first eigenvector for corresponding stage (see details in [39]).

Compartment score (Figure 1d) was defined as the mean log2 number of AA and BB interactions in saddle plots minus the mean log2 number of AB and BA interactions. The transition point between A and B compartments was defined by zero of the first eigenvector.

For the distribution of compartmental scores, we calculated the compartment scores for each chromosome and plotted the resulting values as distributions.

Insulation score

To capture the general properties of the scalings in embryogenesis, we first calculated insulation scores [82] with *cooltools* [39] at 5 Kb resolutions for each replicate of each developmental stage of zebrafish. We set the window size to 200 Kb. We next found local peaks in the insulation score profiles and considered peaks reproducible between replicates as boundaries and used them for average insulation score plotting per experiment (insulation strength in Figure 1d, Extended Data Fig. 1a).
For pileups, we first constructed *bigwig* files with insulation scores with *bioframe* [83] and then aggregated the signal in the genomic windows by *pybbi* [84], as done in [85].

TAD calling

We next performed more rigorous TAD boundary calling. We processed each replicate of each developmental stage with *cooltools*-based and insulation score-based *HiChew* utility [86] at 10 kb, as explained below. We set the expected TAD size parameter of 200 Kb [32] and automatically assessed the optimal parameters of TAD calling to match the expectation. We next removed the TAD boundaries that overlapped bad bins. All code for this analysis is deposited in the GitHub repository: https://github.com/encent/danio-2022.
Function *hichew.calling.boundaries* produces TAD boundaries calling given a certain resolution and the expected TAD size parameter. By iterating over the insulation window size grid, it finds for each chromosome the insulation window size parameter that minimizes the difference between the user-defined expected TAD size parameter and the median TAD sizes obtained while iterating the grid. The resulting TAD boundaries annotation is calculated with *cooltools* functions *calculate_insulation_score* and *find_boundaries* using the insulation window size parameters found for each chromosome, as explained before. All code for TAD boundaries calling using *HiChew* utility is deposited in the Jupyter notebook: https://github.com/encent/danio-2022/blob/main/src/TAD_boundaries_calling.ipynb.
This set of boundaries was used to compare with the positions of fountains (Extended Data Fig. 1g). For the confident set of TAD boundaries at 11 hpf in Figure 2d, 3a, Extended Data Fig. 2d, we further selected only those boundaries that contain at least one ATAC-seq peak supported by CTCF motif (see “CTCF binding inference from ATAC-seq”).

Fountain calling with *fontanka*

Reference fountains
We first arbitrarily picked regions of chromosomes 1 and 2 and manually marked fountains there. In total, we collected 34 fountains from chromosome 1 and 18 fountains from chromosome 2. This fountains set was used as a set of reference fountains for fountain calling control.

Fontanka protocol
We designed *fontanka*, *cooltools*-based [39] utility for fountains calling in Hi-C maps (Figure 2b). Although *fontanka* is designed for calling any on-diagonal pattern in Hi-C map, here we applied it to call fountains.Fontanka has four principal steps:

(i) **Snippets extraction.** *Fontanka* rolls a square window of specified size along the main diagonal of the Hi-C matrix and extracts observed over the expected signal. Each window is assigned to the genomic point of its center.

(ii) **Convolution.** Instead of insulation diamond window implemented in *cooltools* insulation [39], *fontanka* performs two types of convolution:

(a) Fountain score calculation. Each snippet is convolved with the **fountain mask** producing a **fountain score** for the corresponding genomic region. In this work, we used average pileup for all reference fountains as a fountain mask. The convolution score can be based on the summed product of the snippet with a mask, mean square error, and Pearson or Spearman correlation. Here, we used Pearson correlation as a measure for fountain score, which proved to be the best in preliminary tests of the library. Thus, the fountain score can have values between (-1, 1), where 1 is the best correlation with the fountain pattern, and 0 is the absence of correlation.

(b) Noise score calculation. Since Hi-C maps frequently displayed rearrangements relative to the default danRer11 genome, we aimed to filter out potential rearrangements as a noise source in our data. For that, we applied another convolution round with the Scharr operator, approximating the gradient of the Hi-C map [87] [88]. We averaged the Scharr operator by two orientations (vertical and horizontal) for each window pixel, resulting in a **Scharr score**. The Scharr score measures the noise and sharpness of the signal in the corresponding window. Too noisy or sharp patterns in Hi-C snippets result in high noise scores, while smooth Hi-C maps will have Scharr scores closer to zero.

As a result, each fountain candidate thus has two characteristics: fountain score and noise score (Extended Data Fig. 1b). The more the fountain score, the more pronounced the fountain-like structure. The smaller the Scharr score, the less noisy the local Hi-C map and the less probable the rearrangements at the corresponding genomic locus.

(iii) **Peak calling.** *Fontanka* detects local maxima in the fountain score by *cooltools find_peak_prominence* function [39]. This function detects peaks in the 1D genomic signal of the fountain score and assigns **peak prominence** to each one. The resulting peaks were considered candidate fountains.

(iv) **Thresholding and filtration.** Candidate fountains contain potentially very weak and noisy false calls. In order to minimize the contribution of false calls, we set strict requirements for fountains and select the most confident set of fountains.

(a) We first remove the fountains with weak prominence. We noticed that the distribution of the fountain peaks prominence is a mixture of two distributions, small noisy calls and long tail of high values of potential true peak calls. To separate those distributions, we applied Li's iterative method for finding the separation point by using the slope of the cross-entropy [89], as implemented in *scikit-learn* [90].

(b) Next, we remove fountains that fall too close (<50 Kb) to the nearest bad bin (see "Removal of poorly mapped and surrounding genomic regions" section of Methods).

(c) We remove all the fountains that have a negative correlation with the fountain mask.

(d) We next require that the fountains detected in the merged dataset is detected in individual replicates with an offset of 20 kb (on either side from the fountain).

(e) We apply a filter for potential genomic rearrangements and misassembly problems. Genomic rearrangements should be present at all the stages, including 2.75 hpf

that does not have visible non-artifactual on-diagonal patterns in Hi-C. Thus, any fountain detected at 2.75 hpf is probably a result of the erroneous call of genomic rearrangement as a fountain. We removed those fountains at 5.3 hpf that fall within 20 kb from the nearest constant “fountain” at 2.75 hpf (reproducible between replicates at this time point).

(f) Another filter for potential genomic misassembly problems filters out the top 25% of candidate fountains based on their Scharr score.

Out of 3391 candidate fountains at 5.3 hpf passing filters iv.a-iv.c, 2199 were confirmed in replicates (64.8%, filter iv.d). 1110 genomic bins were found as constant “fountains” at 2.75 hpf. Filter (iv.e) resulted in 1947 candidates, out of which 1460 were finally selected for the analysis (iv.f). The final list of fountains with fountain scores and fountain peak scores is provided as Supplementary Dataset 1.

15 out of 34 reference fountains at chr1 (44.1%), and 10 out of 18 reference fountains at chr2 (55.5%) were successfully found by *fontanka* with all filtration steps (Extended Data Fig. 1b).

Since Open2C software and fontanka are under ongoing development, we fixed the versions of the software to fontanka v0.1, bioframe v0.4.1 [83], cooler v0.8.11 [70], and cooltools v0.5.2 [39] in this work.

Comparison of fontanka and Chromosight

We noted that the fountain structures are similar to hairpins [91], and the hairpin calling algorithm (*Chromosight*) can be applied to call fountains in our dataset. With *Chromosight* v1.6.3 and default parameters, we identified 9 out of 34 reference fountains at chr1 (26.5%) and 7 out of 18 reference fountains at chr2 (38.9%). Notably, *Chromosight* output had an average score of 0.14 and an average pileup that did not resemble the average reference fountain.

Comparison of fontanka and protractor tool

Another algorithm for fountain-like structure calling was proposed for jets [41]. It relies on calculating the contacts at certain distances away from the base (protractor tool). To replicate this approach, we calculated the number of Hi-C interactions normalized by expected in a line perpendicular to the main diagonal of the Hi-C map for the whole genome. Then, we compared the fountain average protractor versus all non-fountain bins. Indeed, the protractor tool showed enrichment of the contacts at distances up to 150 kb (Ext. Data Fig. 1e). We summed up the protractor tool average over all distances up to 200 kb and correlated it with the fountain score (Ext. Data Fig. 1a). Both scores correlated with 0.32 Pearson correlation, 0.44 Spearman correlation.

Enrichment of chromatin states

For analysis of chromatin states at the fountains (Figure 3b, Ext. Data Fig. 2e, 7a-b), we first split the genome into non-overlapping 10 kb bins. Then, we calculate the coverage of each genomic bin by (1) ATAC-Seq peaks and (2) PADREs state (predicted ATAC-seq-supported developmental regulatory elements, from [45]) with bioframe coverage [83]. Then, for each bin, we calculate the relative **enrichment of the chromatin state** (the ratio of open chromatin per bin covered by the given state). We then average these enrichments for all bins at fountain bases.

For control, we randomly pick the fountains from all genomic bins and calculate the average

enrichments again for each state. This procedure preserves the distribution of ATAC-Seq peak sizes and the relative contribution of each state to PADRE annotation.

For FDR calculation, we approximated the control distribution of Gaussian and calculated the probability of obtaining an observed value or larger.

Hi-C snipping and average pileup

To snip the Hi-C maps, we used *cooltools pileup* function [39]. Pileup takes one or multiple genomic locations and returns a square window of specified size around these locations. To build average pileups over genomic loci (e.g., fountains), we took the mean of corresponding pixels of all windows snipped around fountains.

ChIP-Seq data analysis

We re-analyzed ChIP-Seq data from [4,22,24,26,92] with *chipseq-nf* 1.2.1 with default parameters [93]. The mode of processing was set either to single-end or paired-end based on raw sequencing annotation in the SRA database (see Supplementary Dataset 3).

CTCF binding inference from ATAC-seq

CTCF ChIP-Seq data is unavailable for most stages of development of zebrafish that we study (except one dataset for 10 hpf from [24]). However, multiple ATAC-Seq profiles are available from the Danio Code database [94]. This allowed us to use CTCF binding inference from ATAC-Seq.
We first called CTCF motifs in the danRer11 genome. Generic vertebrates CTCF motif (JASPAR MA0139.1) instances were called by JASPAR whole-genome *PWMScan*-based [95] motif scanner [96], as described in [97]. The background GC content was set to 0.317 (A and T frequencies) and 0.183 (G and C). We used a p-value threshold of 5e-02 and an r threshold of 0.8.
Next, we downloaded ATAC-Seq peaks from DCC database [94] in narrow peak format for 4.5 hpf (DCD019127DT) and 12 hpf (DCD019077DT). Each ATAC-Seq peak was marked as CTCF-bound if it had at least one CTCF motif.
The selection of parameters for motif calling was made by benchmarking against CTCF ChIP-Seq peaks obtained at 10 hpf from [24] (re-processed with *chipseq-nf* with default parameters [98]).
For coverage pileups in Extended Data Fig. 2c, 9a-b, we stored the CTCF-containing peaks with their abundance as bigwig files with *bioframe* [83] and then aggregated the signal in the genomic windows by *pybbi* [84].

Simulations of loop extrusion

Simulations of loop extrusion were based on *OpenMM*-based [99] Python library *Polychrom* [100]. We split it into five steps: (1) 1D simulation of extruders movement on the genome, (2) 3D simulation of polymer, (3) *in silico* reconstruction of interaction probabilities, (4) assessment of goodness of fit of the simulations to real data, (5) parameter sweep of the model.
The implementation is available at https://github.com/agalitsyna/polychrom_workbench under *targeted_extrusion* directory.

1D simulation

We simulated the translocation of extruders on a 1D lattice [101], where each point represented a genomic position. Each position can either (1) have a regular role or (2) serve as a loading platform for the targeted extruder. For regular positions, we assumed uniform probabilities of loading and unloading. For loading platforms, we set up the probability of cohesin landing increased by enrichment factor.

To simplify the model, we assumed (1) immediate reloading of unloaded cohesin, and (2) separate pools of extruders that can load at regular positions versus loading platforms. Before running the simulations, we ensured that cohesins are fully loaded on a lattice.

As in our previous works [101], simulations were organized in rounds, with all cohesins performing translation with both legs in the opposite directions at each round. Each simulation lasted for 10 010 000 translocation rounds, and we sampled each 10th step as a readout of 1D simulation. We removed the first 10000 steps to ensure that we sample the steady state of loop formation and not the earlier stages of loop formation and maturation [102].

The 1D lattice size was 50000 monomers, with each monomer corresponding to 1 Kb of DNA (thus, the whole simulated molecule was 50 Mb).
The size of the loading platform was set to 1 monomer (1 kb), and the platforms were distributed each 500 kb to correspond to the mean distance between fountains in Hi-C data. Technically, we generated 20 groups of the size 2500 to allow averaging over multiple groups in a single conformation when calculating the average fountain, as we proposed previously [15].

For the basic simulations of targeted loading, we assumed that there are (1) no boundaries that stall cohesin legs, (2) only other cohesins can stall the legs of cohesins, (3) legs of cohesin (if not stalled) are moving simultaneously, (4) free leg of cohesin with a stalled leg can move.

For simulations with random barriers, we assumed that (1) cohesins load only at the loading platforms, (2) random barriers are placed at each genomic position in a 100 kb window around the platform, (3) random barriers stall the cohesin leg with some probability (best fit to real data obtained for the stalling probability of 0.005, see Fig. 5c).

For simulations with decoupled legs of cohesin, we assumed that (1) cohesins load only at the loading platforms, (2) the probability of stepping is set for each leg independently (best fit to real data obtained for the probability of stepping set to 0.99, see Fig. 5c).

3D simulation

As in our previous works [102] [43] [15], we represented chromatin as a polymer with spherical monomers connected by harmonic bonds with stiffness and soft-core repulsive potential. Simulations were performed with variable Langevin integrator in the periodic boundary conditions [15] with volume density 0.1. For simulations and setting up the bonds, we used the OpenMM-based [99] Python library Polychrom [100]. A harmonic bond connected the two monomers held by the extruder. The number of 3D-simulation time steps per 1D simulation step was set to 200 [103].

In silico reconstruction of interaction probabilities

To calculate contact maps, we recorded all the pairs of monomers located closer than 5 monomer radii. We then averaged the interactions over multiple equivalent groups to obtain

a simulated 2.5 Mb region with fountains. To obtain a simulated average fountain at 10 kb resolution, we snipped 200 kb windows around the loading platforms and coarse-grained them by a factor of 10. We then normalized the snippets by expected.

Goodness of fit of the simulations to real data

The goodness of fit was assessed by the correlation of the simulated average fountain with the Hi-C average fountain. For that, we compared each *in silico* Hi-C of the simulated fountain with an average reference fountain. As a quality measure, we calculated the Spearman correlation between matrices of average fountains. Both matrices were taken at 10 kb resolution, and *in silico* Hi-C matrices were coarse-grained for this purpose (for example, see Figures 4-5).
For additional validation, we checked the interactions between loading platforms (fountain bases) by snipping off-diagonal in silico Hi-C matrices and verified that there was no enrichment of interactions (Ext. Data Fig. 5c-d).
We also assessed the shape of the fountains by plotting the cross-section and protractor slides of average fountains (Ext. Data Fig. 6).

Parameter sweep

The simulations were organized as a set of Python scripts: 1D extrusion, 3D simulations, and 2D contact map construction. For the 1D extrusion, we could vary lifetime, separation, and the level of targeted loading. For 3D simulations and 2D contact map construction, we fixed the parameters. To automate the parameters sweep, we implemented a snakemake [104] pipeline combining all three steps. Tested parameters and the resulting goodness of fit are presented in Ext. Data Fig. 5b and 6.

Figure 1

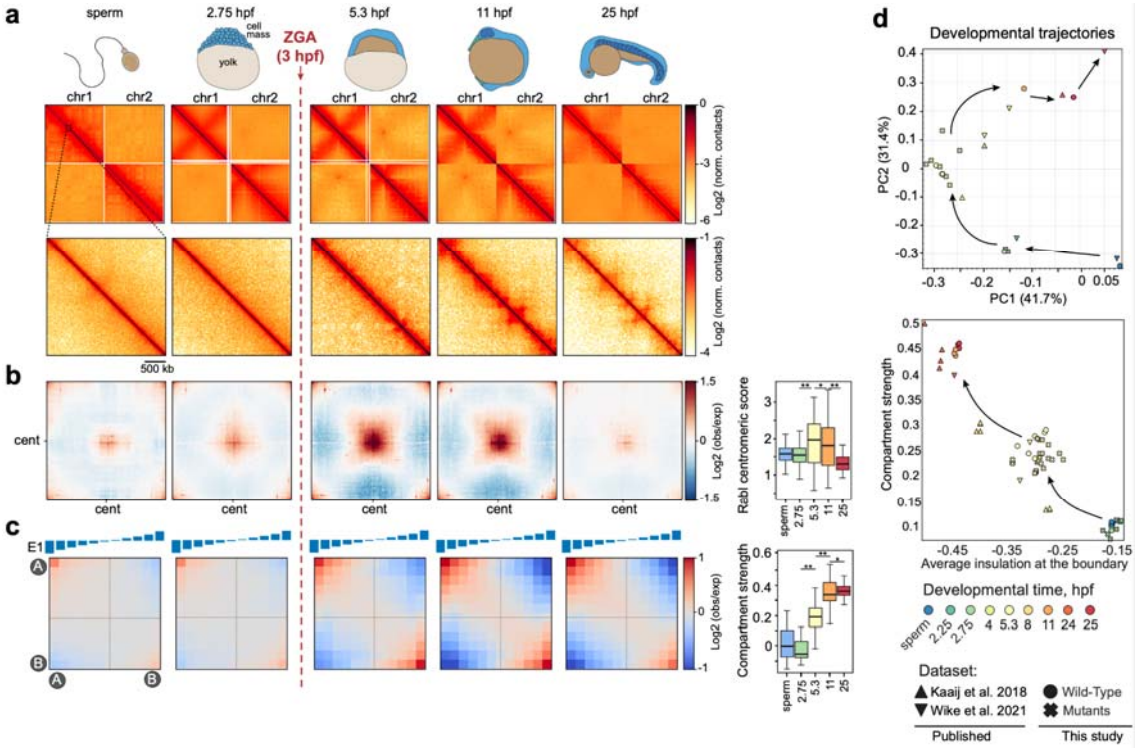


Figure 1. Dynamics of large-scale chromatin spatial organization in zebrafish early embryogenesis. ZGA – zygotic genome activation; hpf – hours post fertilization.

a. Representative examples of contact heatmaps showing dynamics of Rabl configuration (top) and local contact patterns (bottom).

b. Centromere-centromere interactions measured by average Rabl plots (right) and Rabl centromeric scores (left).

c. Compartmentalization measured by saddle plots (right) and compartment strength (left, see Methods).

d. Developmental trajectories based on stratum-adjusted correlation coefficient (SCC, left) and average insulation plotted against compartmental strength (right). Plots include datasets from three studies ([4] [32] and this work). Mutants for the transcription factors (cross marks) will be discussed in Fig. 6-8.

Figure 2

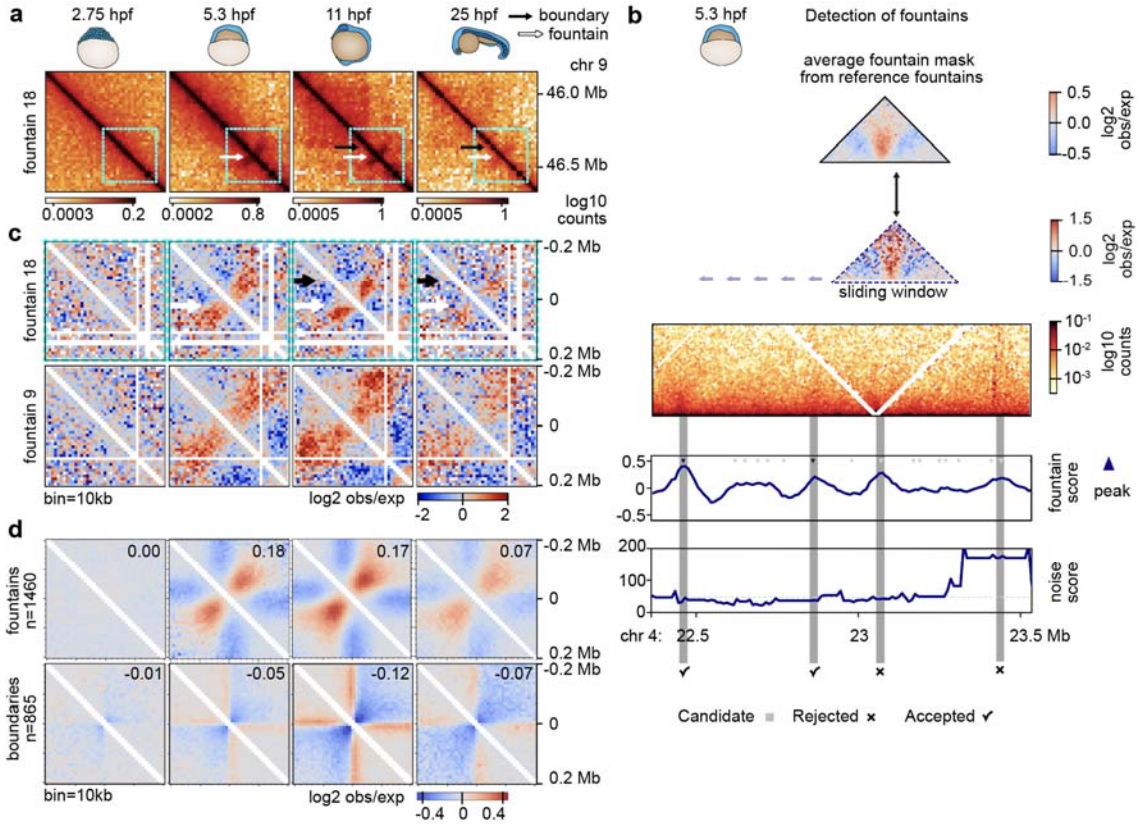


Figure 2. Emergence of fountains and boundaries after ZGA.

a. A fountain (white arrow) is detectable on the contact heatmap after ZGA (5.3 hpf). A TAD boundary (black arrow) is detectable at 11 hpf. Bin size=20 kb. The region in the blue frame is also shown at the top of Fig. 2b.

b. Detection of fountains with the fountain-calling algorithm *fontanka*. From top to bottom: log2 observed/expected ratios of manually picked fountains at 5.3 hpf were averaged to obtain a fountain mask (top). The fountain mask was compared to a 0.4 Mb sliding window to obtain the fountain score for each 10 kb genomic bin. The peaks in the fountain score (gray triangles) were considered fountains if their noise score was smaller than the threshold (bottom, see Methods for full fountains selection criteria). The 1460 fountains found on the 5.3 hpf contact map using this procedure are referred to as “fountains” and are further analyzed in this work.

c. Two representative examples of fountains called at 5.3 hpf (see Supplementary Dataset 1 for precise genomic positions). Heatmaps show observed over expected Hi-C signal at each time point.

d. Top row: Average log2 observed/expected ratios of 1460 fountains at 5.3 hpf. Developmental stages are the same as in (a). The average fountain score is shown at the top right corner of each panel.

Bottom row: Average log2 observed/expected ratios of strong insulating TAD boundaries at 11 hpf. The average fountain score is shown at the top right corner of each panel.

Figure 3

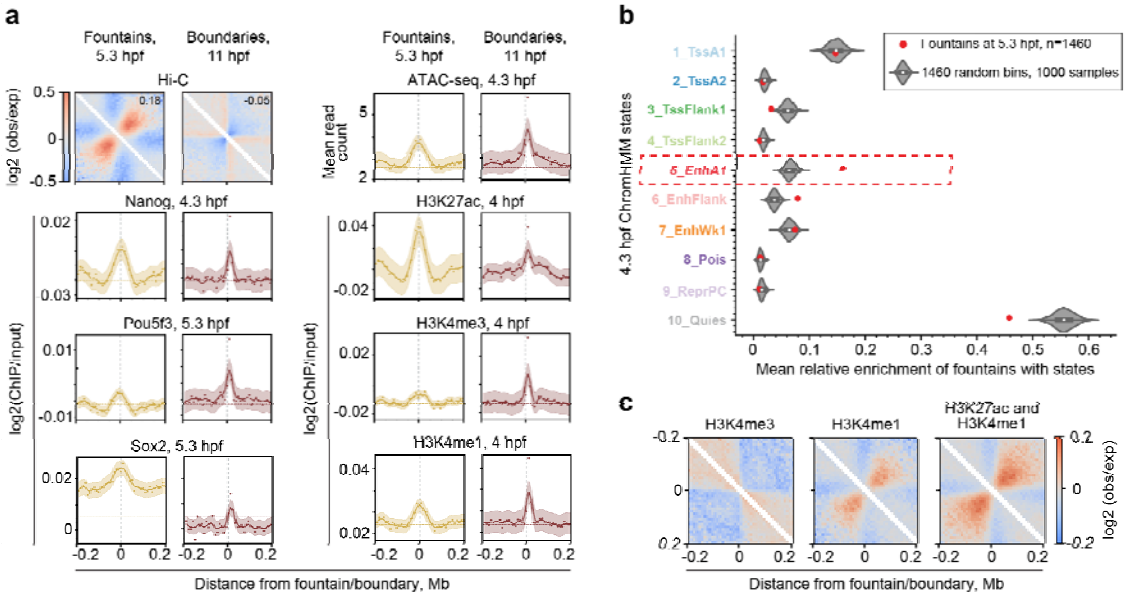


Figure 3. Fountains emerge on enhancers

a. Genomic positions of fountains (5.3 hpf) and TAD boundaries (11 hpf) are comparably enriched in H2K27ac, open chromatin, H3K4me1, and the transcription factors Pou5f3, Nanog, and Sox2. Boundaries at 11 hpf are more enriched in H3K4me3.

b. Fountains colocalize with enhancers at 4.3 hpf. ChromHMM data from [45], EnhA1 FDR<< 1e-4, obs. ~0.158, exp. mean ~0.065, exp. std ~0.009, EnhFlank FDR << 1e-4, obs. ~0.079, exp. mean ~0.038, exp. std ~0.007 (see Methods).

c. Average pileups of Hi-C maps at 5.3 hpf for regions enriched in promoter and enhancer histone marks (also see Extended Data Fig. 3).

Figure 4

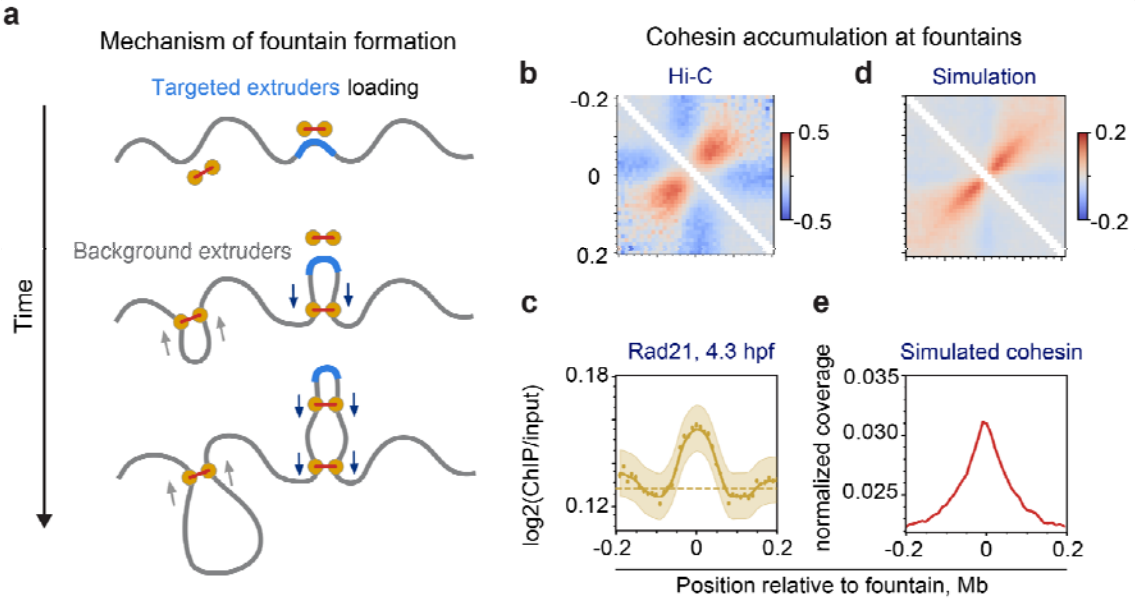


Figure 4. Structural basis for fountains formation.

a. Proposed mechanism of fountain formation. The extruder molecule (cohesin complex) loads preferentially at the target region (highlighted in blue) and starts extruding DNA on two sides. Multiple extruders can load at the target sites, resulting in the formation of a hairpin-like structure. Extruder molecules can also load at other genomic loci (background extrusion with lower rates of extruder loading).

b. Average Hi-C pileup based on experimental data (observed-over-expected signal).

c. Distribution of cohesin subunit Rad21 (ChIP-Seq) around 5.3 hpf boundaries. Pearson correlation between fountain strength and cohesin abundance is 0.373 (two-sided p-value << 0.01, Beta-test implemented in *scipy* v1.11.1).

d. Average simulated Hi-C pileup based on simulations with the best parameters according to Extended Data Fig 5 (see Methods).

e. Cohesin coverage profile in simulations.

Figure 5

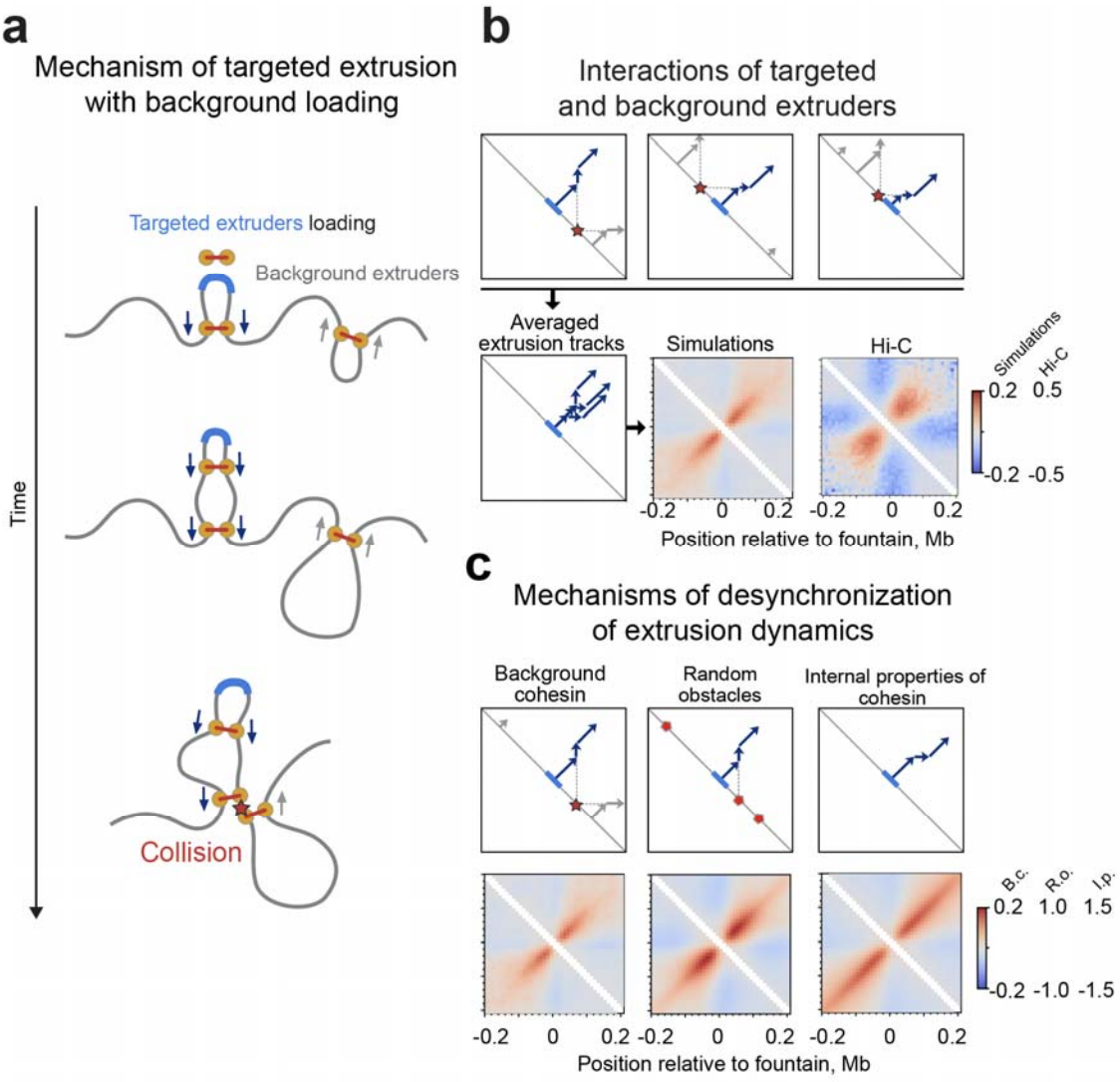


Figure 5. Emergence of fountains. Random barriers are required for decoupling of extrusion by the two legs and forming disperse fountains.

a. Mechanism of targeted extrusion with background loading with collisions.

b. Interactions of targeted and background extruders that can lead to the decoupling of the legs. Cohesins loaded at target sites can randomly bump into background cohesins (top three trajectories), which deflects the hairpin pattern and forms a dispersive fountain structure on average (bottom left).

c. Other mechanisms of desynchronization of extrusion dynamics: background cohesin (left), random obstacles (center), and internal properties of cohesin (right).

Figure 6

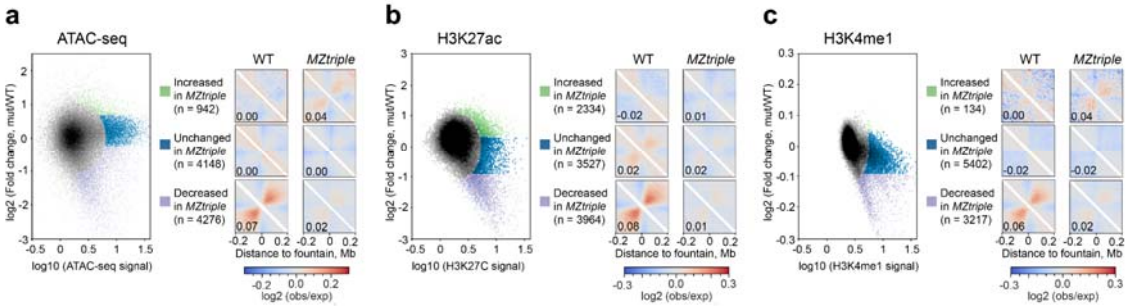


Figure 6. Fountain formation is defined by chromatin openness, H3K27ac, and H3K4me1.

a. Chromatin openness in MZ*triple* mutant is coupled with fountain pattern formation on Hi-C. **(left)** Distribution of ATAC-seq responses in MZ*triple* and definition of the groups of genomic bins based on their ATAC-seq responses. **(center)** An average pileup of different groups from the left panel in wild type Hi-C 5.3 hpf. **(right)** An average pileup of different groups from the left panel in MZ*triple* Hi-C 5.3 hpf.

b. H3K27ac enrichment in MZ*triple* mutant is coupled with fountain pattern formation on Hi-C. **(left)** Distribution of H3K27ac responses in MZ*triple* and definition of the groups of genomic bins based on their ATAC-seq responses. **(center)** An average pileup of different groups from the left panel in wild type Hi-C 5.3 hpf. **(right)** An average pileup of different groups from the left panel in MZ*triple* Hi-C 5.3 hpf.

c. H3K4me1 enrichment in MZ*triple* mutant is coupled with fountain pattern formation on Hi-C. **(left)** Distribution of H3K4me1 responses in MZ*triple* and definition of the groups of genomic bins based on their ATAC-seq responses. **(center)** An average pileup of different groups from the left panel in wild type Hi-C 5.3 hpf. **(right)** An average pileup of different groups from the left panel in MZ*triple* Hi-C 5.3 hpf.

Figure 7

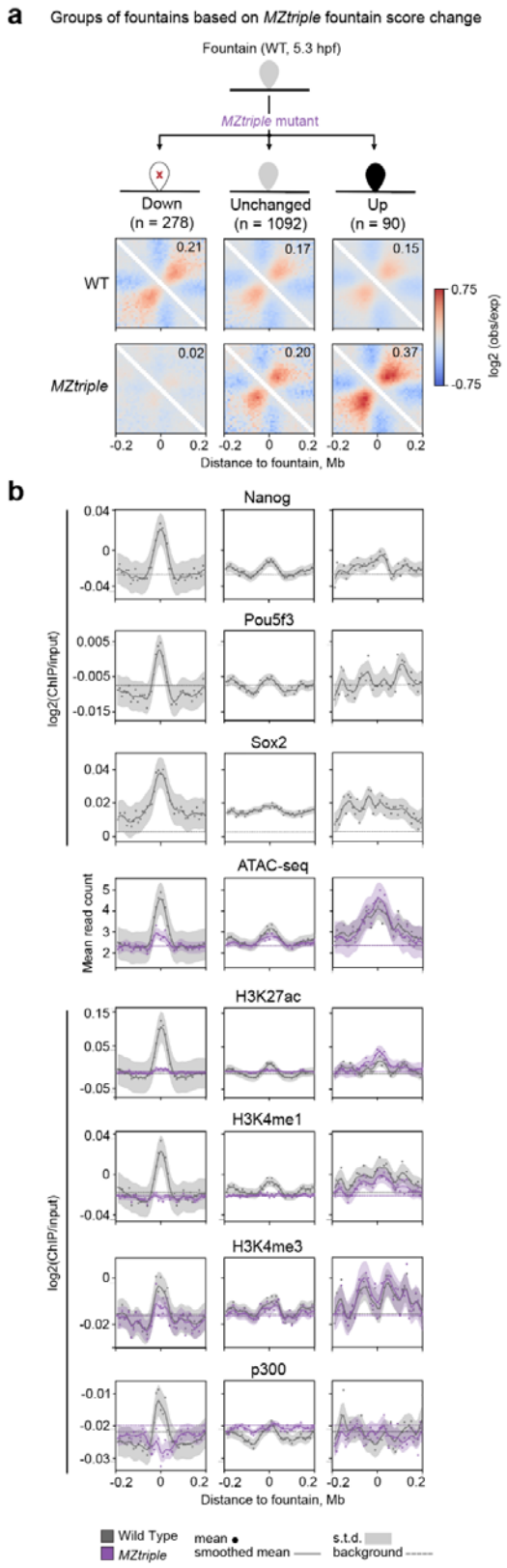

a
Groups of fountains based on MZtriple fountain score change
Fountain (WT, 5.3 hpf)
MZtriple mutant
Down
(n = 278)
Unchanged
(n = 1092)
Up
(n = 90)
WT
MZtriple
0.21
0.17
0.15
0.02
0.20
0.37
0.75
-0.75
log2 (obs/exp)
-0.2
0
0.2
Distance to fountain, Mb
b
Nanog
Pou5f3
Sox2
ATAC-seq
H3K27ac
H3K4me1
H3K4me3
p300
log2(ChIP/input)
Mean read count
Distance to fountain, Mb
Wild Type
MZtriple
mean
smoothed mean
s.t.d.
background

Figure 7. Weakened fountains in MZ*triple* mutant bear the strongest early enhancer mark.

a. Three groups of fountains by strength change in the triple Pou5f3/Nanog/Sox19b mutant (MZ*triple,* this work). Out of 1460 fountains, 278 weakened, 90 strengthen, and 1091 remained unchanged (see Methods for definition).

b. The binding of transcription factors Pou5f3, Sox19b, and Nanog is enriched at the weakened fountains (ChIP-seq for Nanog at 4.3 hpf from [92] and for Pou5f3 and Sox19b at 5.3 hpf from [22]). Chromatin accessibility decreases in the weakened fountains in the MZ*triple* mutant and increases in the strengthened fountains (ATAC-seq data from this work).

(pileups below, right) H3K27ac and p300 are lost from the weakened fountains in the Pou5f3/Nanog/Sox19b mutant (MZ*nps*, ChIP-seq data for 4 hpf from [22])

Figure 8

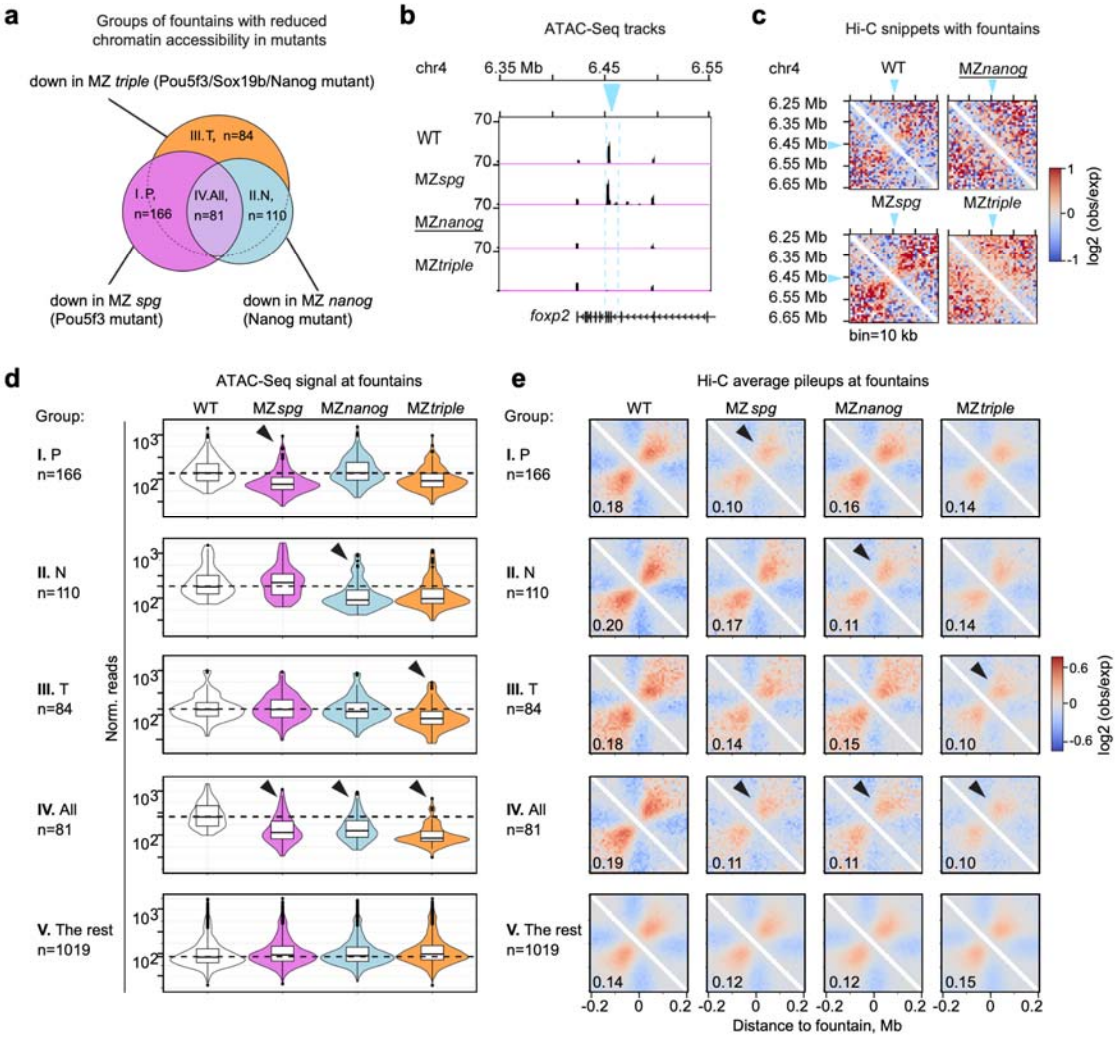


Figure 8. Pioneer activities of Pou5f3 and Nanog separately contribute to the formation of the fountain structures.

a. The Venn diagram shows the groups of fountains selected by the requirement for Pou5f3, Nanog, or their combination for chromatin accessibility at the fountain bases (groups I-IV). The groups of fountains were assigned by the differential analysis of normalized ATAC-seq reads per 10 kb fountain base. ATAC-seq data for the indicated mutant genotypes and the wild-type at 4.3 hpf from [56]. **b, c**. Example of a fountain where both chromatin accessibility (b) and fountain structure (c) are regulated by Nanog but not by Pou5f3.

b. The 10 kb fountain base is shown by blue arrows and dotted lines. ATAC-seq signals (rkpm) in the wild-type, Nanog mutant MZ*nanog*, Pou5f3 mutant MZ*spg* and triple mutant MZ*triple*, 200 kb window, 4.3 hpf ATAC-seq data from [56]. Note that only Nanog is required for chromatin accessibility.

c. Hi-C signal in 400 kb window centered at the fountain in the indicated genotypes. The fountain base is represented by the blue arrow. Note that only Nanog is required for fountain formation: the fountain is present in the wild-type and MZ*spg* but disappears in MZ*nanog* and MZ*triple.*

d. Violin plots show ATAC-seq signals in groups of fountains I-IV, selected by the requirement for Pou5f3, Nanog, or their combination for chromatin accessibility (as in a), and in group V (the rest), where Pou5f3 and Nanog were not required for chromatin accessibility. Dashed lines show the median wild-type ATAC-seq signal. Arrowheads show the genotypes used to assign the groups (where the ATAC-seq signal was significantly reduced in all fountains). 4.3 hpf ATAC-seq data from [56]. Group assignment and the corresponding ATAC-Seq signal are listed in Supplementary Dataset 2.

e. Average on-diagonal pileup plots with log2 observed/expected ratios of the groups shown in (d). Note that the chromatin accessibility and fountain structure strength are regulated in parallel by the same factors (black arrowheads). The average fountain score is shown at the bottom left corner of each panel.

Figure 9

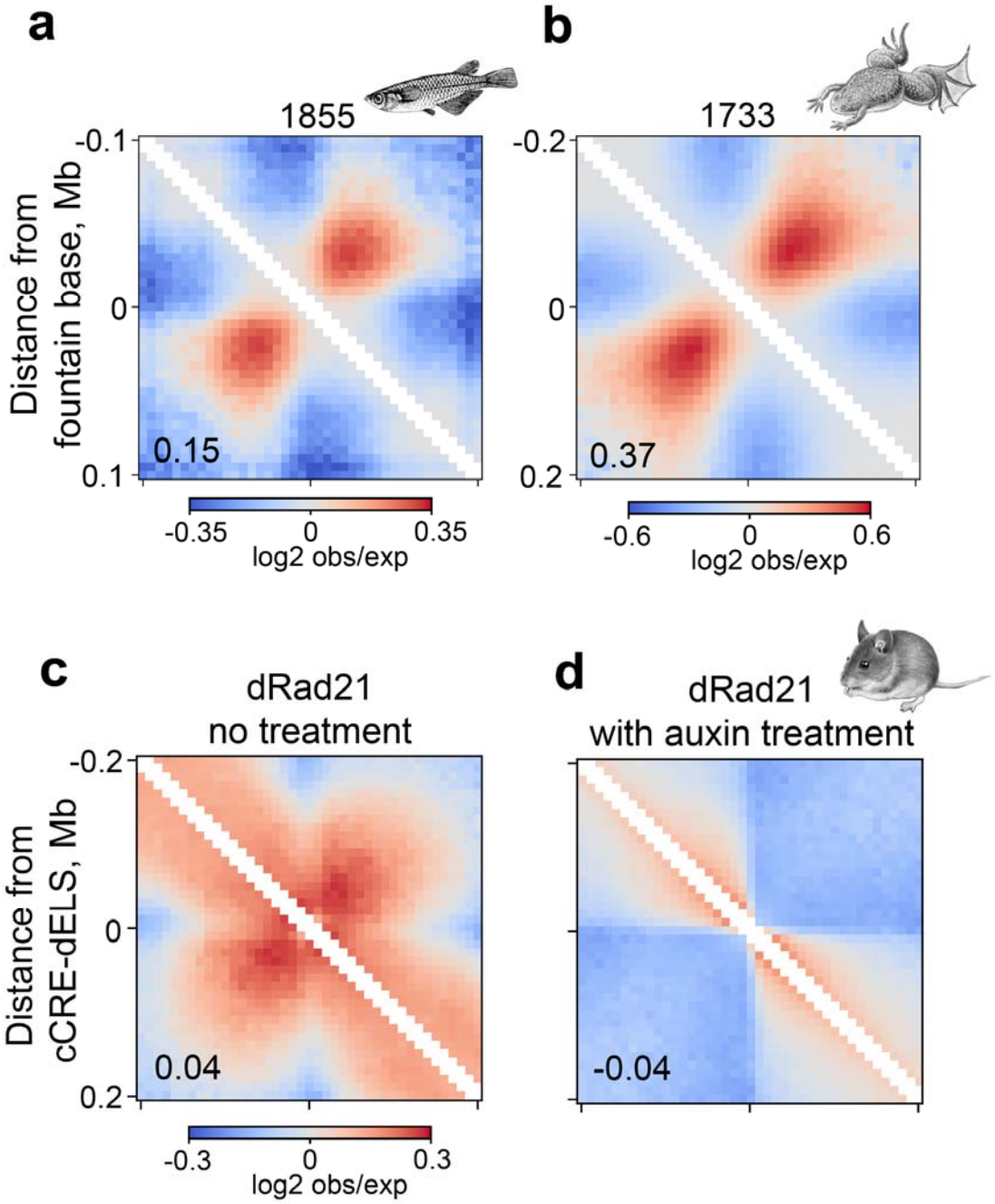


Figure 9. Fountain structure is a universal cohesin-dependent element of chromosome architecture.

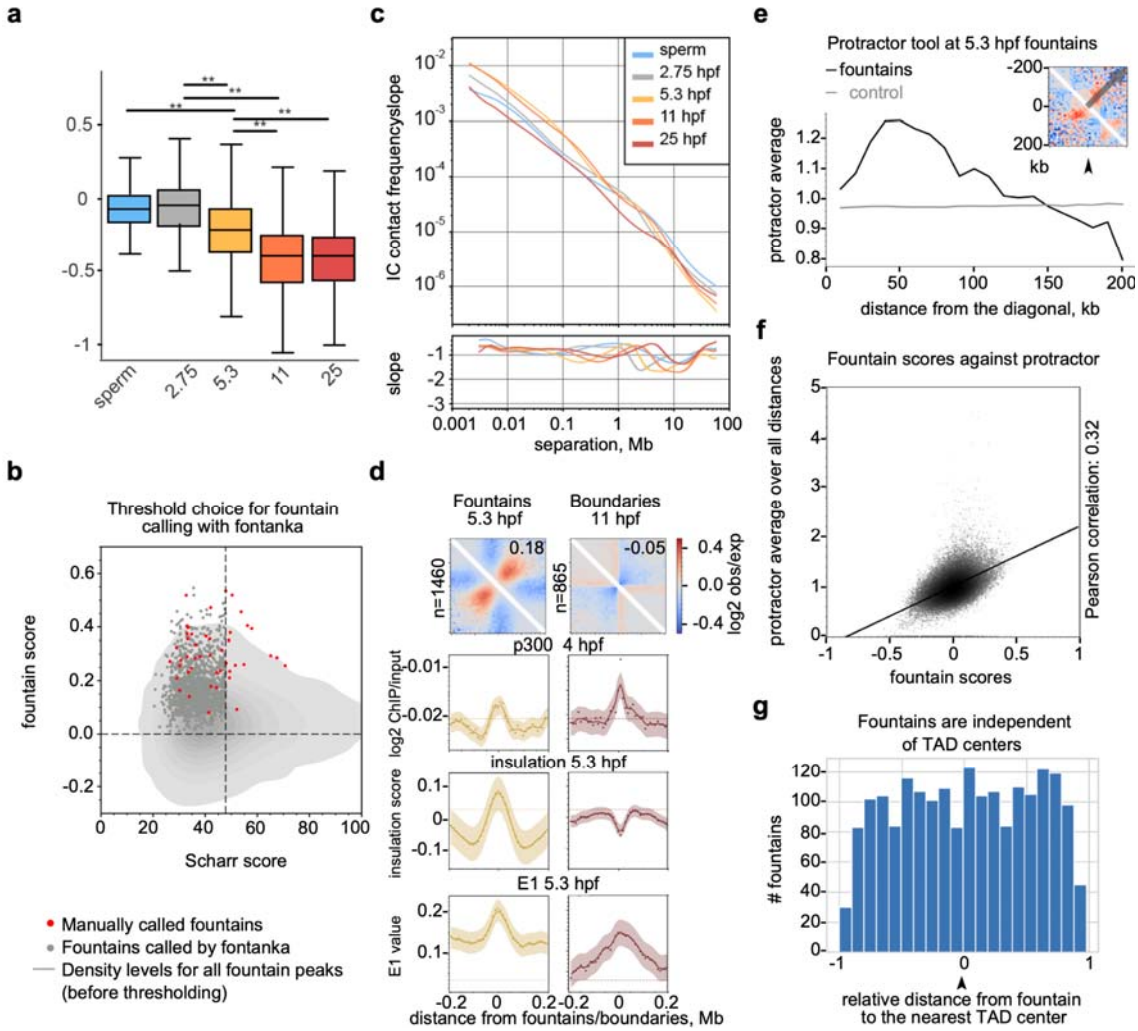
a. Average fountain observed in medaka fish based on 1855 fountains called at 5 Kb resolution in 100 Kb windows with fontanka. Hi-C for developmental stage 10 from [6].

b. Average fountain observed in *Xenopus tropicalis* based on 1733 fountains called at 10 Kb resolution in 200 Kb windows with fontanka. Hi-C for developmental stage 11 from [35].

c. Pileup of candidate cis-Regulatory Elements (cCREs) with TSS-distal enhancer-like signatures (dELS) for mouse embryonic stem cells [58]. Micro-C from [57], control experiment with no auxin treatment.

c. Pileup of candidate cis-Regulatory Elements (cCREs) with TSS-distal enhancer-like signatures (dELS) for mouse embryonic stem cells with auxin-inducible degradation of Rad21 [57]. Average fountain disappears upon Rad21 degradation.

Extended Data Figure 1



Extended Data Figure 1. Establishment of 3D chromatin in development, finding of fountain structures. (Supporting material for main Fig. 1 and Fig. 2)

a. Insulation score boxplots at different stages of embryo development. 2.75 hpf boxplots are based on boundaries found at the same stage; 5.3, 11, and 25 hpf insulation are based on boundaries found at 25 hpf. Note that the insulation score is negative (more negative scores for more insulation)

b. Setting up the thresholds of fountain peak score and Scharr score for selecting fountains.

c. Top: dependence of the contact probability $P_c(s)$ on the genomic distance s for sperm and the four indicated developmental stages. Bottom: the first derivative indicating the slope of $P_c(s)$.

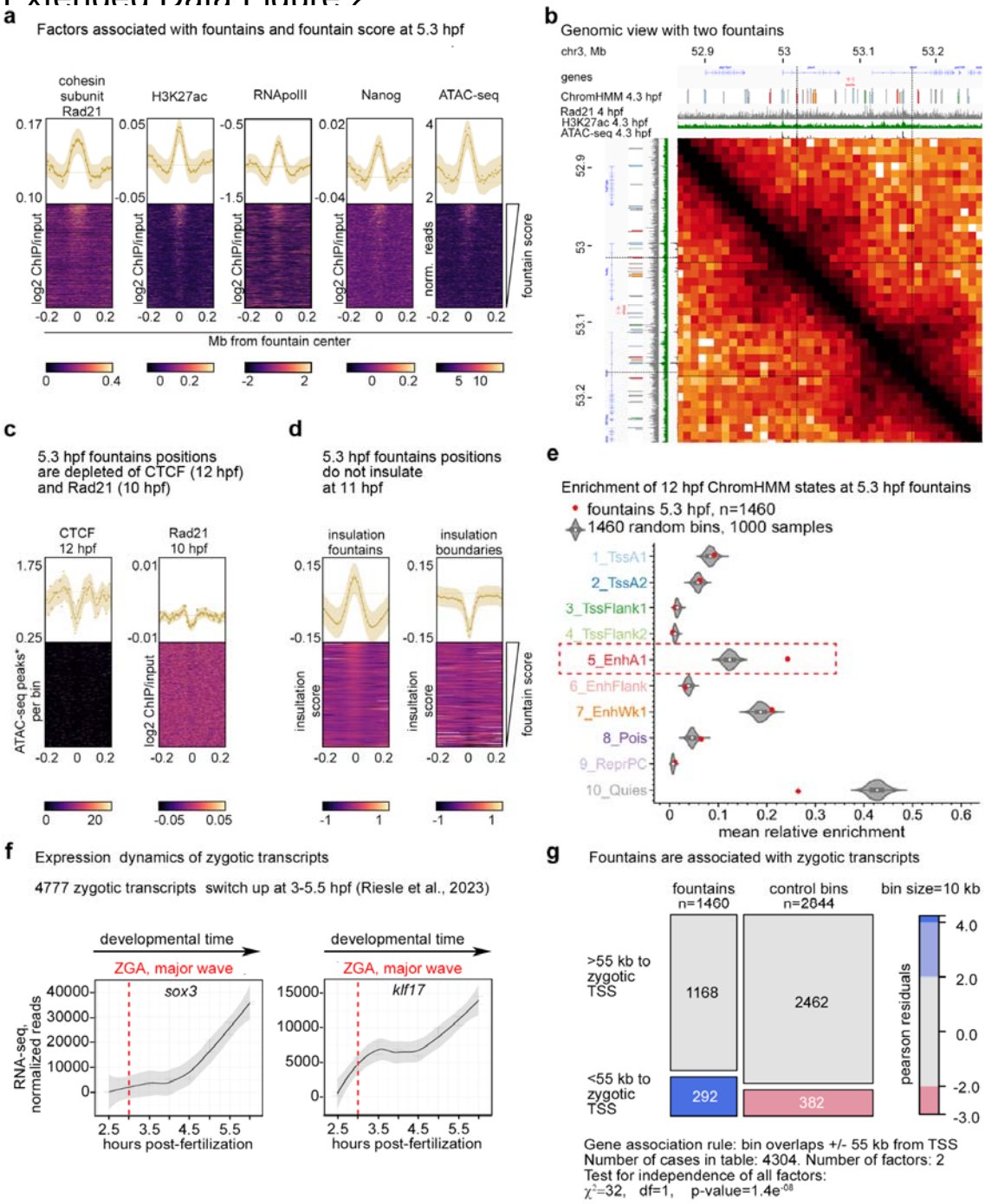
d. Application of protractor tool [41] for validation of fountains. The grey arrow in the inset (fountain) represents the perpendicular to the main diagonal (ideal hairpin). The grey line is a control based on non-fountain genomic regions.

e. The scatterplot shows the correlation of the protractor tool score ([41], see Methods) with the fountain score as defined in this paper.

f. Distribution of fountains around centers of TADs: no correlation was found.

g. Profiles of p300 ChIP-seq (4 hpf), insulation score (5.3 hpf), and E1 (5.3 hpf) centered on the genomic positions of the fountains at 5.3 hpf (left) or on the genomic positions of the TAD boundaries at 11 hpf (right). Insulation is negative for strong TAD boundaries (meaning fewer average interactions) and positive for strong fountains (more average interactions).

Extended Data Figure 2



Extended Data Figure 2. Properties of the fountains (Supporting material for the main Fig. 3)

a. Profiles and heatmaps of the factors associated with fountains are shown for ChIP-seq data for Rad21 [4], RNA polymerase II [22], Nanog [92] and H3K27ac [22], and ATAC-seq data [23] at the indicated stages. The heat maps were sorted by descending fountain score, as indicated on the right. Note that all factors positively correlate with the fountain score. Bottom: scales for the heatmaps.

b. Single fountain with gene track and epigenetic annotations.

c. CTCF binding at 12 hpf and Rad21 binding at 10 hpf are depleted at the genomic positions of 5.3 hpf fountains. *- only CTCF-motif-containing ATAC-seq peaks at 12 hpf [105] were scored.

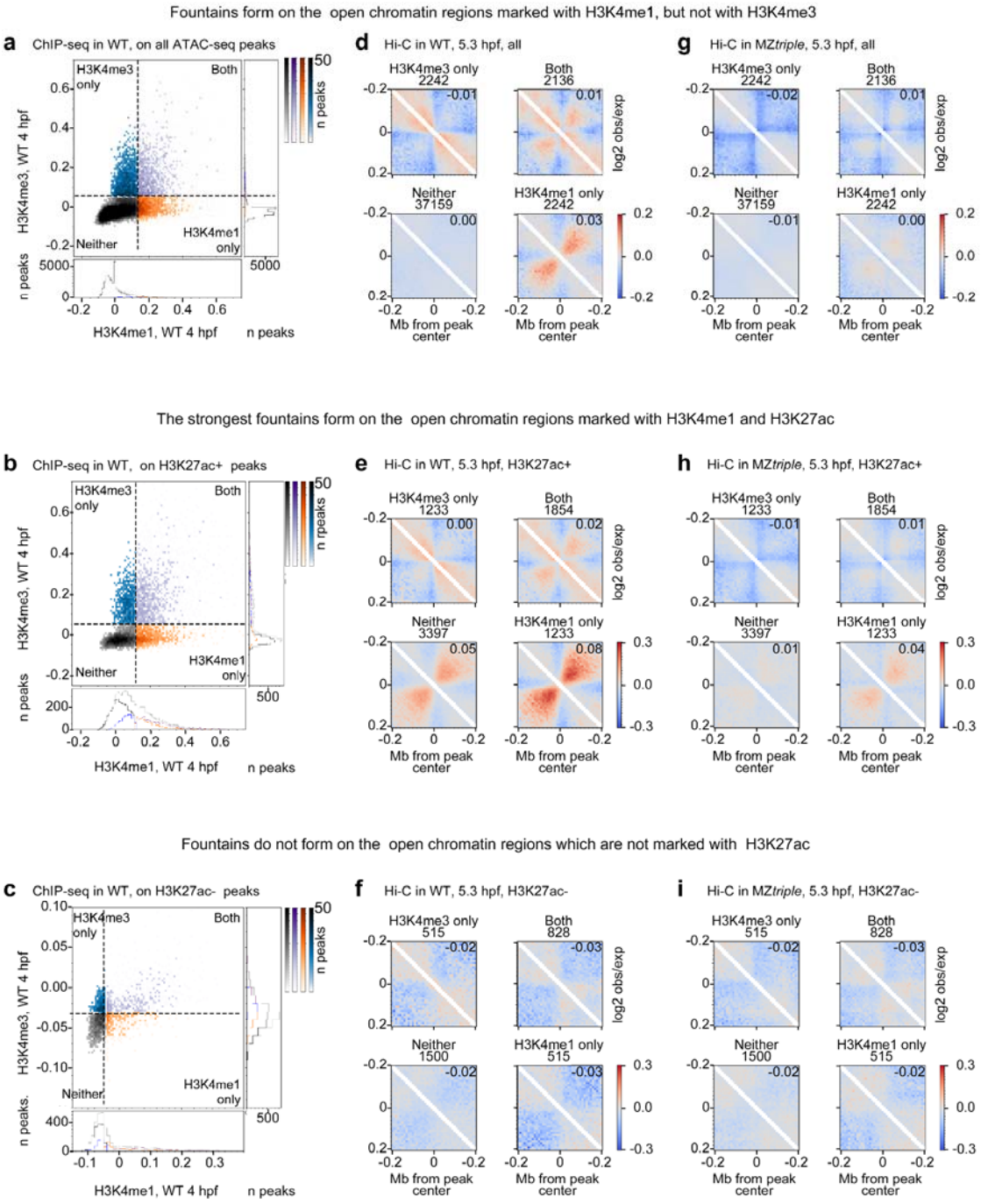
d. Insulation at 11 hpf is below the background at the genomic positions of 5.3 hpf fountains, in contrast to strong insulation at 11 hpf TAD boundaries (note that the insulation score is a negative value). The heatmaps were sorted by descending fountain score (5.3 hpf).

e. Fountains colocalize with enhancers at 12 hpf, but to less extent than with enhancers at 4.3 hpf (compare to Fig. 3b). ChromHMM data from [45].

f. Dynamic increase of mRNAs over time for actively transcribed (zygotic) *sox3* and *klf17* genes. The time point of the major ZGA wave (3 hpf) is highlighted in red. The number of zygotic genes in the embryo (4777) and RNA-seq data are from [56].

g. Fountains preferentially localize near zygotic genes. Chi-squared test. Control 10 kb bins were taken at the distance +/- 1 Mb from 1460 fountains, the overlapping fountain bins were excluded.

Extended Data Figure 3



Extended Data Figure 3. Fountains form on accessible chromatin regions enriched in H3K4me1 and H3K27ac and depend on the pioneer transcription factors Pou5f3, Sox19b, and Nanog. (Supporting material for the main Fig. 3 and the main Fig. 6)**.**

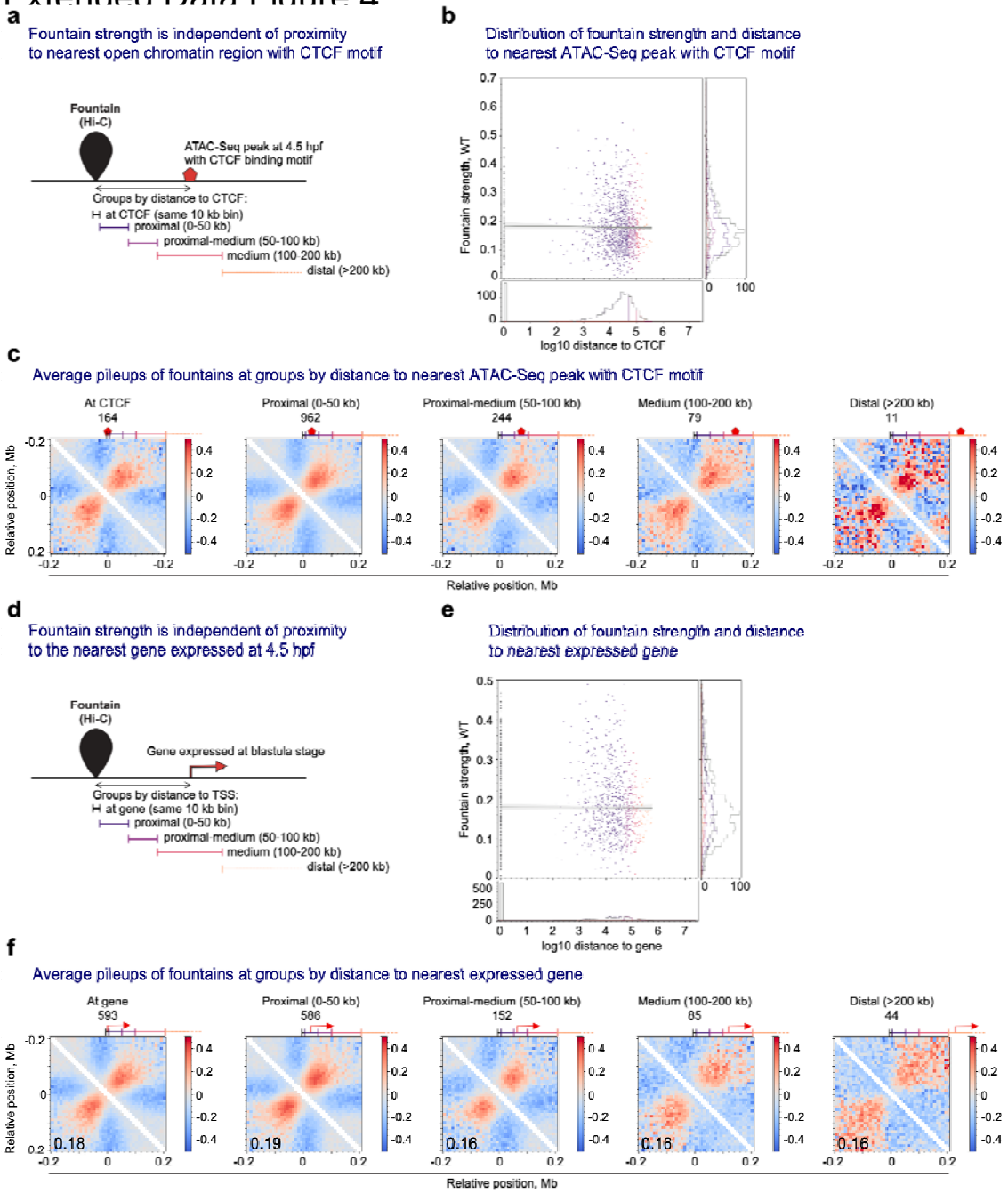
43779 accessible chromatin regions in the wild-type embryos at 4.3 hpf (ATAC-seq peaks from [23]), overlapping with 10 kb fountain bases, were sorted by enrichment in H3K4me1, H3K4me3, and H3K27ac histone modifications at 4 hpf [22]

(**a-c**) Scatterplots for the histone modifications and definition of four groups (“Neither”, “Both”, “H3K4me1 only”, “H3K4me3 only”). Colors represent the division of the scatterplot by 90% percentile for each axis. **a.** Distribution of ChIP-Seq signals for H3K4me1 and H3K4me3, in all 43779 ATAC-seq peaks. **b.** Distribution of ChIP-Seq signals for H3K4me1 and H3K4me3 in 7717 accessible regions, which were also enriched in H3K27ac (the top 10% genomic bins by H3K27ac signal). **c.** Distribution of ChIP-Seq signals for H3K4me1 and H3K4me3 in 3358 accessible regions which were not enriched in H3K27ac (the bottom 10% genomic bins by H3K27ac signal).

(**d-f**) Average on-diagonal pileup plots for genomic regions falling into specific groups defined in a-c, respectively, in 5.3 hpf wild-type embryos. **d.** On all the accessible regions, fountains are detectable in the “H3K4me1 only” and “Both” groups. **e**. On the accessible regions with high H3K27ac, the strongest fountain signal is seen on the “H3K4me1 only” group, enriched in of H3K27ac and H3K4me1 but not H3K4me3, i.e., on the active enhancers. Fountain strength increases in the order “Both” < “Neither” < “H3K4me1 only” groups. **f**. In the absence of the H3K27ac mark, fountains do not form at the accessible regions. The average fountain score is shown at the top right corner of each graph.

(**g-i**) Average on-diagonal pileup plots for genomic regions falling into specific groups defined in a-c, respectively, in the MZ*triple* 5.3 hpf embryos. Note that the fountain formation is absent or severely reduced in all groups, compared to the wild-type (compare the panels g-i to panels d-f).

Extended Data Figure 4



Extended Data Figure 4. Fountains formation is independent of chromatin openness at CTCF motifs and active transcription in the vicinity of fountains.

a. Fountains formation is independent of proximity to the nearest open chromatin region with CTCF motif.

b. Distribution of fountain strength and distance to the nearest ATAC-Seq peak with CTCF motif.

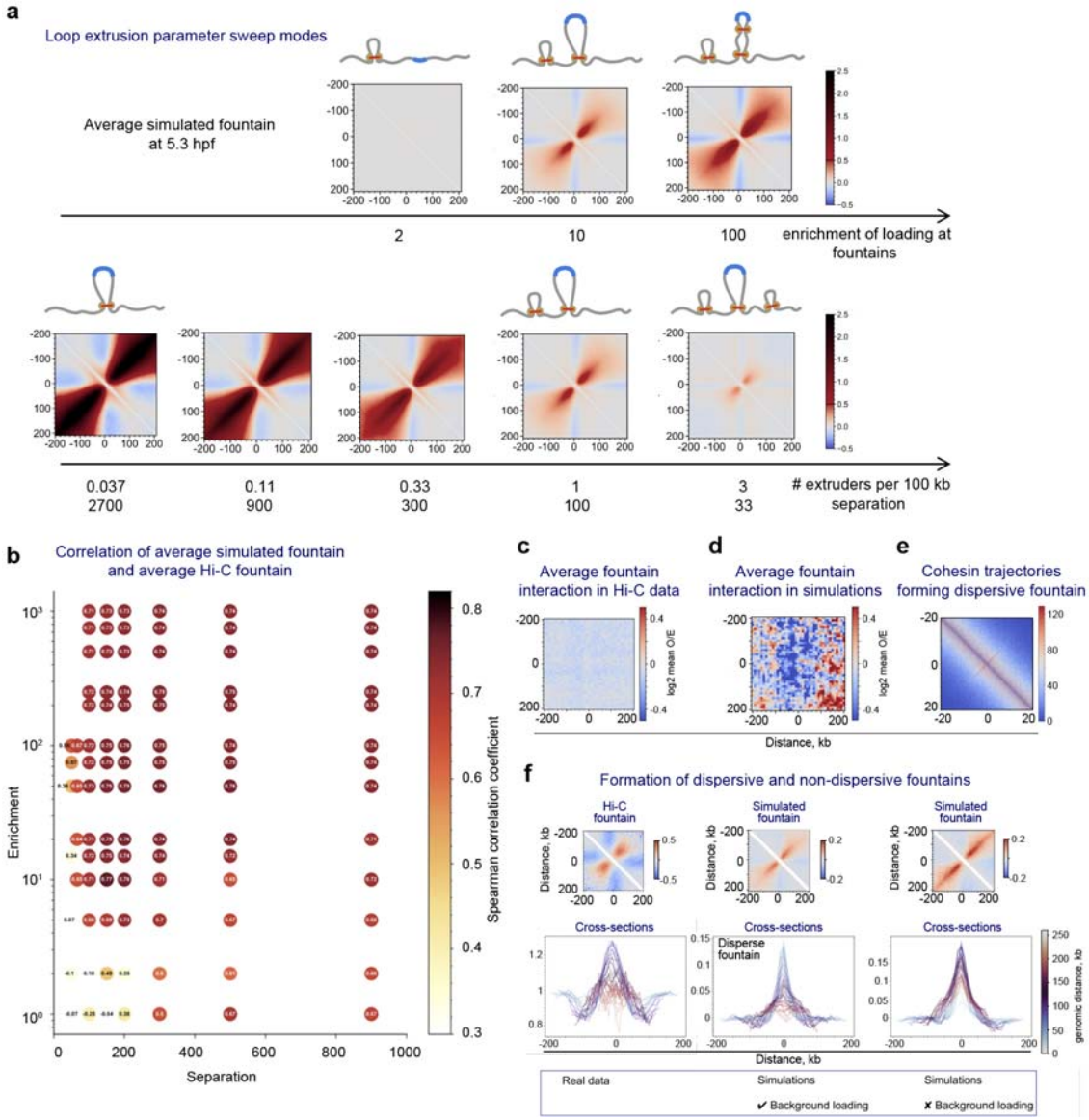
c. Average pileups of fountains by groups defined by distance to the nearest ATAC-Seq peak with CTCF motif.

d. Fountains formation is independent of proximity to the nearest expressed TSS.

e. Distribution of fountain strength and distance to the nearest expressed TSS.

f. Average pileups of fountains by groups defined by distance to the nearest expressed TSS.

Extended Data Figure 5



Extended Data Figure 5. Simulations of targeted cohesin loading.

a. Loop extrusion parameter sweep modes. Average fountains for simulations with different parameters of targeted extrusion.

b. Parameter sweep for the model of targeted extruder loading. The quality of fit is measured by the Spearman correlation coefficient between Hi-C and simulated average fountains.

c. Average fountain interaction in Hi-C data.

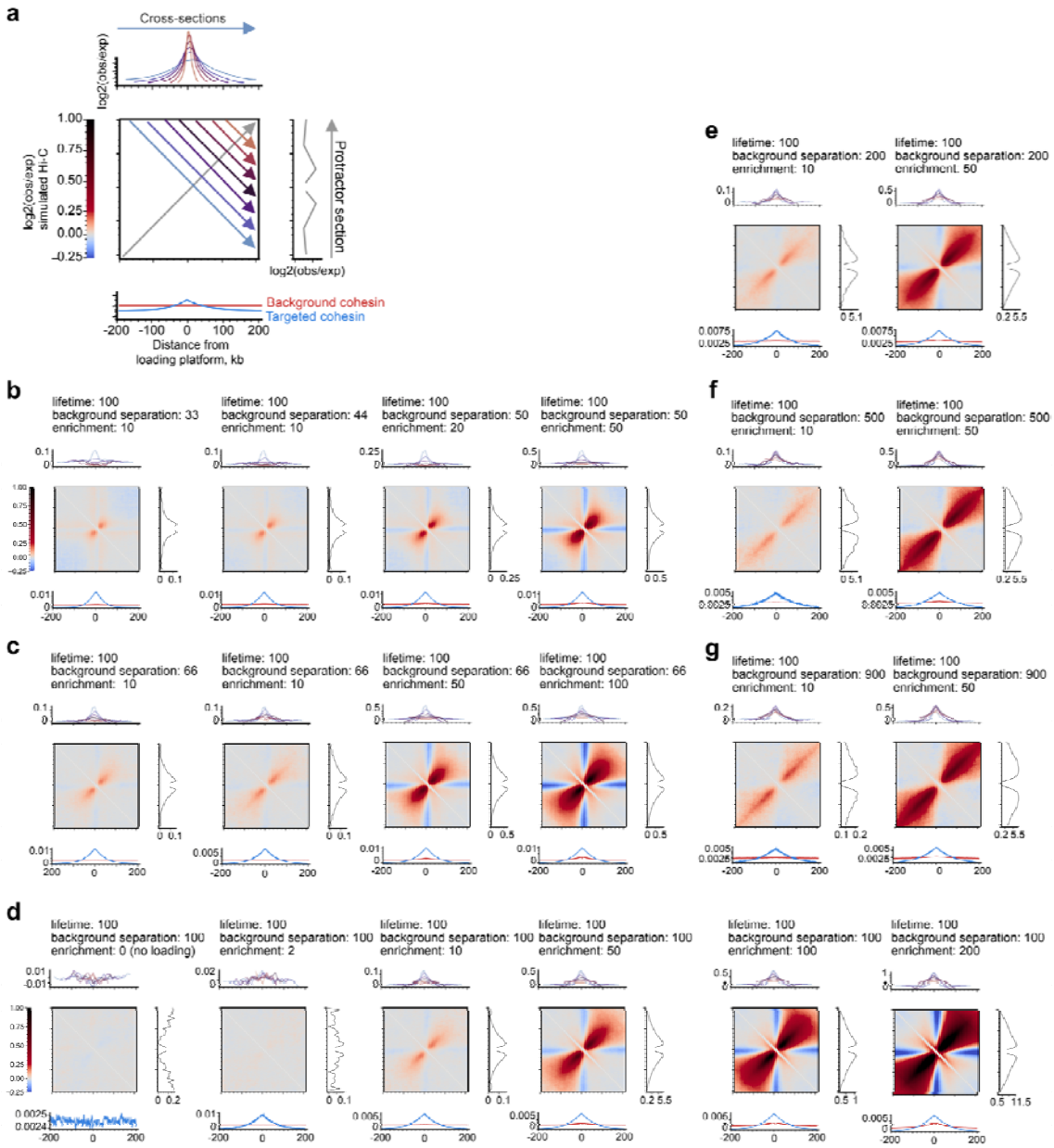
d. Average fountain interaction in simulated maps.

e. Cohesin trajectories forming dispersive fountains.

f. Experimental Hi-C data (left, top) resembles disperse fountains formed in the presence of the background loading (center, top) but is less similar to the non-disperse model of

fountains (right, top). Cross-section plots indicate that the difference between models can be explained as the growth of dispersion of Hi-C signal enrichment with genomic distance (bottom).

Extended Data Figure 6

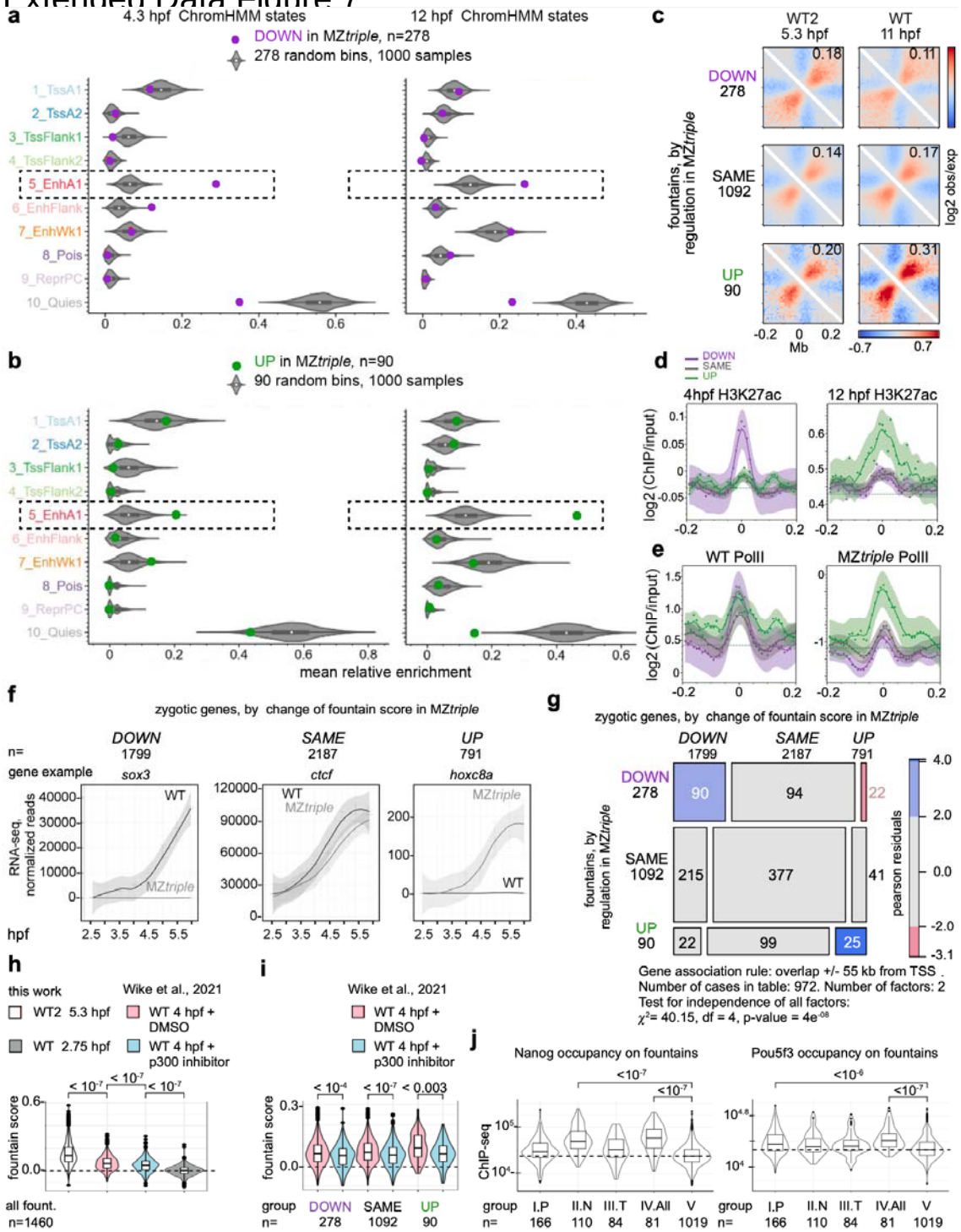


Extended Data Figure 6. Fountain shapes in different modes of polymer simulations of fountains.

a. Scheme of the average pileup plot of the simulated fountain with protractor and cross-sections, targeted and background cohesin coverages.

b-g. Parameter sweep varying background separation and enrichment of targeted loading resulting in the formation of fountains of various shapes.

Extended Data Figure 7



Extended Data Figure 7. Changes in fountain formation in the pioneer factor mutants are due to the changes in the active enhancer repertoire in these mutants (Supporting material for the main Fig. 7 (panels a-h) and for the main Fig. 8 (panel i).

a. Fountains weakened in MZ*triple* mutant (DOWN) colocalize with early (4.3 hpf) enhancers.

b. Fountains strengthened in MZ*triple* mutant (UP) colocalize with late (12 hpf) enhancers (a, b) ChromHMM data from [45].

c. The strength of fountains that weaken in the MZ*triple* mutant also decreases with time. Conversely, the strength of fountains that strengthen in the MZ*triple* mutant also increases with normal developmental time. Average on-diagonal pileup plots for “DOWN”, “SAME”, and “UP” fountains in the wild-type embryos at 5.3 hpf and 11 hpf. The number of reads was the same for the two “WT2” Hi-C replicates at 5.3 hpf and for the two wild-type Hi-C replicates at 11 hpf; the replicates were merged.

d. H3K27ac profiles on “DOWN”, “SAME”, and “UP” fountains at 4.3 hpf (ChIP-seq data from [23]) and 12 hpf (ChIP-seq data from [106]). Genomic positions with fountains that strengthen in MZ*triple* accumulate H3K27ac later in development.

e. RNA polymerase II profiles on “DOWN”, “SAME”, and “UP” fountains at 4 hpf in wild-type and MZ*triple* mutants (ChIP-seq data from [22]). Genomic positions with fountains that strengthen are enriched in MZ*triple* PolII.

f. Three groups of zygotic transcripts, by expression change in MZ*triple* mutants: “*DOWN*”, “*SAME*” and “*UP*”. Expression time curves in the wild-type and MZ*triple* are shown for three representative genes. Data from [56].

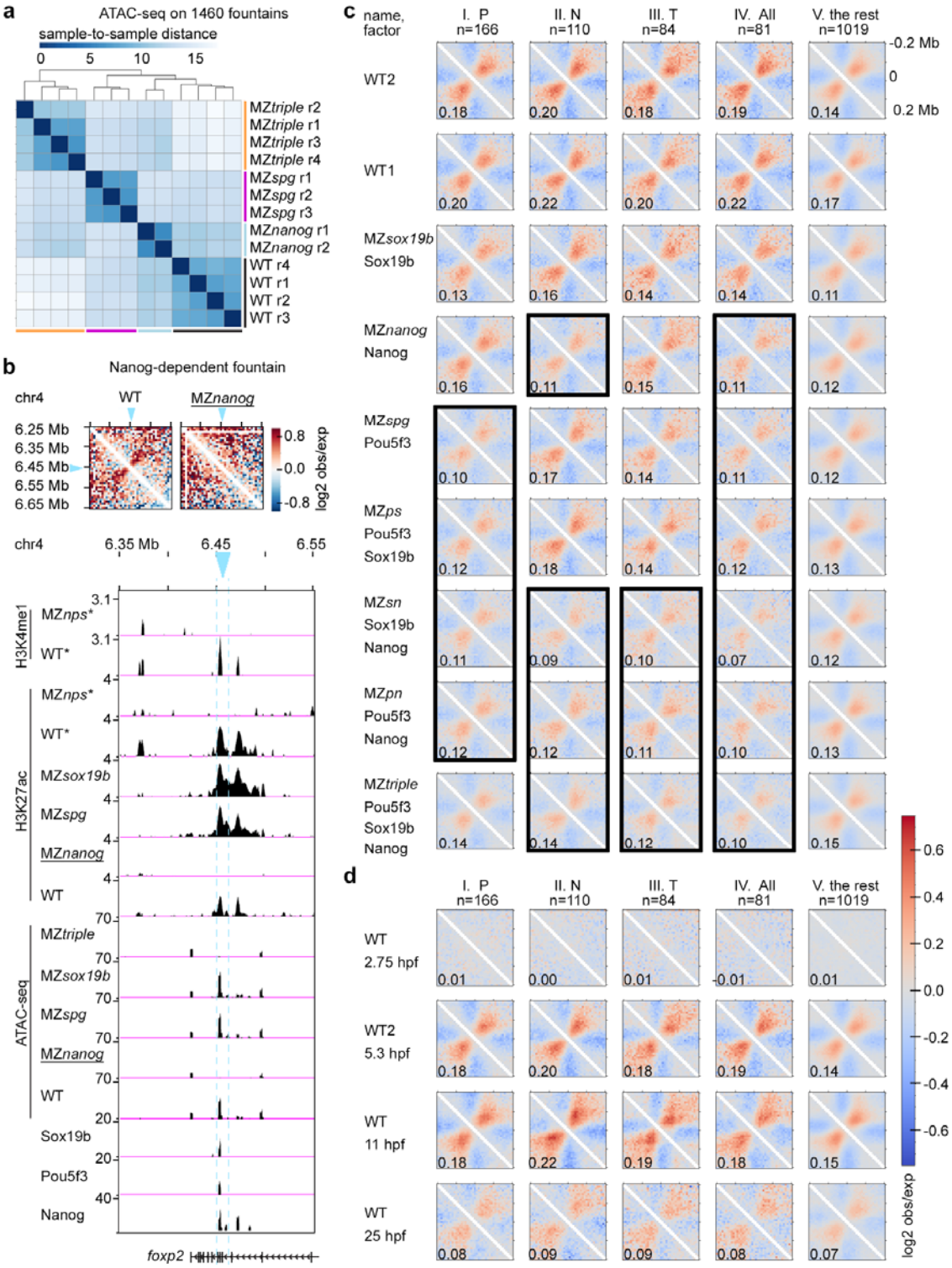
g. Weakened (DOWN) and strengthened (UP) fountains are located near down- and up-regulated transcripts in MZ*triple*. The chi-squared test was done for the fountains overlapping +/- 55 kb from TSS (transcription start sites) of the indicated gene groups (shown with the numbers on the graph).

h. Fountains are present at 4 hpf and are reduced upon the inhibition of p300. Fountain scores for 1460 fountains (Supplementary Dataset 1) for 5.3 hpf and 2.75 hpf Hi-C data from this work compared to the Hi-C data at 4 hpf (control-treated or p300 inhibitor-treated) from [4]. The p-values in the Tukey-Kramer test are shown, p-value in 1-way ANOVA < 2e-16. The median fountain value at 2.75 hpf is zero (shown by the dashed line).

i. Fountain strength of all fountain groups (“DOWN”, “SAME”, “UP” by the change of the fountain score in MZ*triple*) is equally reduced when p300 is inhibited. The dashed line shows a zero baseline of the fountain score.

j. Pou5f3 and Nanog bind to the fountains they regulate. ChIP-seq signals for Pou5f3 and Nanog were scored per 10kb fountain in the groups I-V (see the main Fig. 8 and Supplementary Dataset 2 for the assignment of the groups). Note that each TF is mostly bound on the fountains it regulates Pou5f3 on “I. P” and “III. T” groups, Nanog on “II. N” and “III. T” groups. Dashed lines show the median TF occupancy in the non-regulated group V (the lowest in both cases). P-values in the Tukey-Kramer test are shown on the graphs; p-value in 1-way ANOVA < 2e-16 in both cases.

Extended Data Figure 8



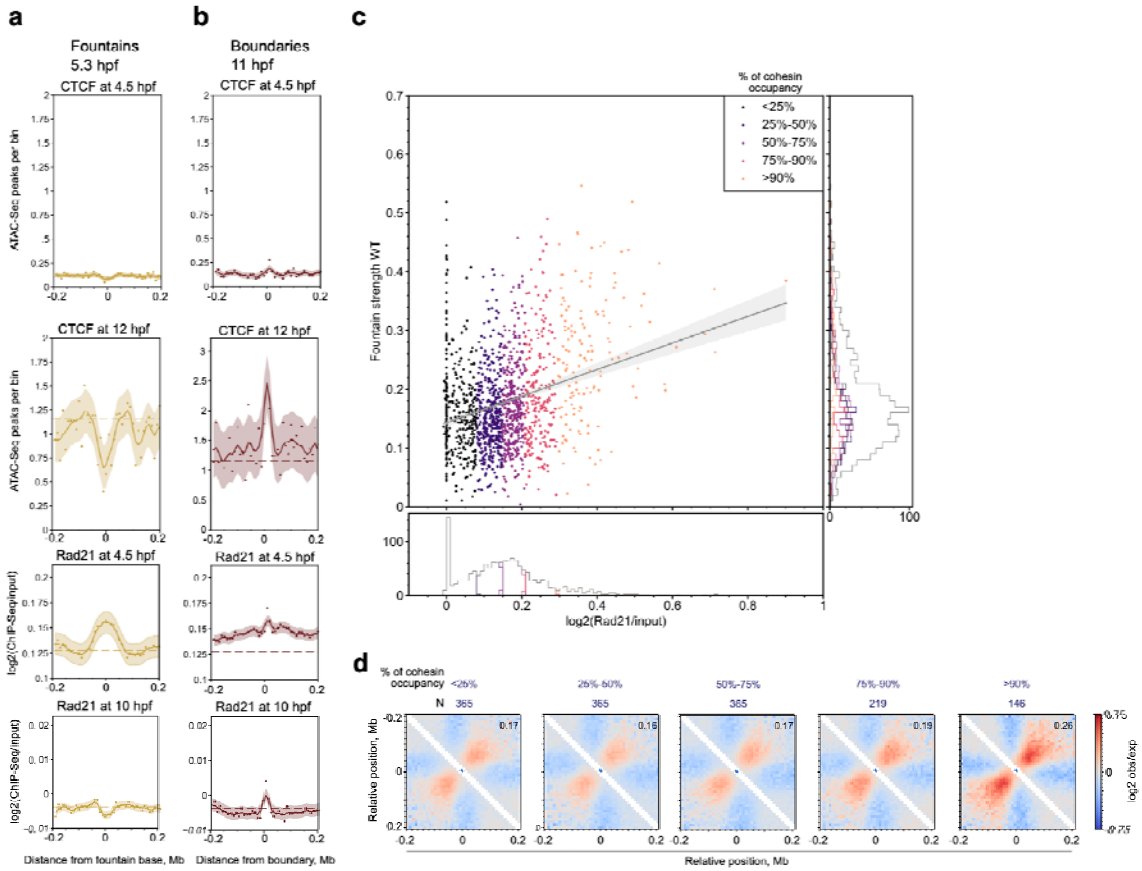
Extended Data Figure 8. Fountains specifically depend on the pioneer activity of the transcription factors (Supporting material to the main Fig. 8).

a. Hierarchical clustering of ATAC-seq replicate experiments[56] on 1460 fountains in the 4.3 hpf embryos of the following genotypes: mutants by pluripotency transcription factors Pou5f3 (MZ*spg*), Nanog (MZ*nanog*), triple Pou5f3/Nanog/Sox19b mutant (MZ*triple*) and the wild-type (WT). ATAC-seq signal was scored per 10 kb (fountain base size).

b. Example of the Nanog - dependent fountain. Top: the 10 kb fountain bin is shown by blue arrows. Bottom: UCSC browser view shows enrichment for pluripotency TFs, chromatin modifications, and open chromatin on the fountain (dotted lines). *-ChIP-seq data from [22]. Nanog ChIP-seq is from [92]. TF occupancy and ATAC-seq are shown in rpkm, chromatin modifications in log2 ChIP to input ratio.

c, d. Mean Hi-C observed/expected signal on five groups of fountains in indicated genotypes at 5.3 hpf (c) or in the wild-type in four developmental time points (d). The groups were selected by requirement for pluripotency factors for chromatin accessibility: Pou5f3 (I. P), Nanog (II. N), Pou5f3 or Nanog or Sox19b (III. T), Pou5f3 and Nanog (IV. All), none (V, the rest), as shown in Venn diagram in the main Fig. 8a. WT1 and WT2 refer to two Hi-C experiments with different read depth: 200 mln reads per duplicate in WT2, 600 mln reads per duplicate in WT1. Single mutants MZ*sox19b*, MZ*nanog,* and MZ*spg*, and the WT embryos at 2.75 hpf, 11 hpf, and 25 hpf were sequenced to 200 mln reads and should be compared to WT2. Double mutants MZ*ps*, MZ*pn*, MZ*sn,* and triple mutant MZ*triple* were sequenced to 600 mln reads per replicate and should be compared to WT1. Merge duplicates of each experiment were used to derive the fountain scores. The mean fountain score is shown at the lower left corner of each plot. (c) The mutant genotypes where the fountain score was lower than the respective control, more than 0.6 (arbitrary selected threshold) are highlighted by black thick frames. Notably, Sox19b mutant has the unifrom drop of fountain strength, which is not specific to group III. T. This is in line with our previous findings that Sox19b does not act alone as a pioneer factor on early active enhancers [23,56].

Extended Data Figure 9



Extended Data Figure 9. Redistribution of cohesin from fountains to TAD boundaries in embryogenesis.

a. Distribution of early CTCF (4.5 hpf ATAC-Seq peaks supported by CTCF motif), late CTCF (12 hpf ATAC-Seq peaks supported by CTCF motif), early Rad21 (4.5 hpf ChIP-Seq) and late Rad21 (10 hpf ChIP-Seq) around fountains called at 5.3. hpf.

b. Distribution of the same signal as in (**a**), but around boundaries called at 11 hpf.

c. Scatterplot of 4.5 hpf Rad21 ChIP-Seq (OX) against fountain score (OY) colored by the percentiles of Rad21 signal distribution. Each dot represents a single fountain in the genome.

d. An average pileup of Hi-C at 5.3 hpf for different groups with Rad21 enrichment at fountains from (**c**).

References

1. Vastenhouw, N. L., Cao, W. X. & Lipshitz, H. D. The maternal-to-zygotic transition revisited. *Development* **146**, (2019).
2. Larson, E. D., Marsh, A. J. & Harrison, M. M. Pioneering the developmental frontier. *Mol. Cell* **81**, 1640–1650 (2021).
3. 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.e19 (2017).
4. Wike, C. L. *et al.* Chromatin architecture transitions from zebrafish sperm through early embryogenesis. *Genome Res.* **31**, 981–994 (2021).
5. Pecori, F. & Torres-Padilla, M.-E. Dynamics of nuclear architecture during early embryonic development and lessons from liveimaging. *Dev. Cell* **58**, 435–449 (2023).
6. Niu, L. *et al.* Three-dimensional folding dynamics of the Xenopus tropicalis genome. *Nat. Genet.* **53**, 1075–1087 (2021).
7. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
8. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
9. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930–944.e22 (2017).
10. Gibcus, J. H. *et al.* A pathway for mitotic chromosome formation. *Science* **359**, (2018).
11. Franke, M. *et al.* CTCF knockout in zebrafish induces alterations in regulatory landscapes and developmental gene expression. *Nat. Commun.* **12**, 5415 (2021).
12. Gabriele, M. *et al.* Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* eabn6583 (2022).
13. Mach, P. *et al.* Cohesin and CTCF control the dynamics of chromosome folding. *Nat.*

Genet. **54**, 1907–1918 (2022).

14. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305–320.e24 (2017).
15. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
16. Symmons, O. *et al.* Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).
17. Karpinska, M. A. & Oudelaar, A. M. The role of loop extrusion in enhancer-mediated gene activation. *Curr. Opin. Genet. Dev.* **79**, 102022 (2023).
18. Kane, L. *et al.* Cohesin is required for long-range enhancer action at the Shh locus. *Nat. Struct. Mol. Biol.* **29**, 891–897 (2022).
19. Morales, C. & Losada, A. Establishing and dissolving cohesion during the vertebrate cell cycle. *Curr. Opin. Cell Biol.* **52**, 51–57 (2018).
20. Arzate-Mejía, R. G., Recillas-Targa, F. & Corces, V. G. Developing in 3D: the role of CTCF in cell differentiation. *Development* **145**, (2018).
21. Hug, C. B. & Vaquerizas, J. M. The Birth of the 3D Genome during Early Embryonic Development. *Trends Genet.* **34**, 903–914 (2018).
22. Miao, L. *et al.* The landscape of pioneer factor activity reveals the mechanisms of chromatin reprogramming and genome activation. *Mol. Cell* **82**, 986–1002.e9 (2022).
23. Gao, M. *et al.* Pluripotency factors determine gene expression repertoire at zygotic genome activation. *Nat. Commun.* **13**, 788 (2022).
24. Meier, M. *et al.* Cohesin facilitates zygotic genome activation in zebrafish. *Development* **145**, (2018).
25. Lee, M. T. *et al.* Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**, 360–364 (2013).
26. Leichsenring, M., Maes, J., Mössner, R., Driever, W. & Onichtchouk, D. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* **341**, 1005–1009 (2013).

27. Veil, M., Yampolsky, L. Y., Grüning, B. & Onichtchouk, D. Pou5f3, SoxB1, and Nanog remodel chromatin on high nucleosome affinity regions at zygotic genome activation. *Genome Res.* **29**, 383–395 (2019).
28. Pálfy, M., Schulze, G., Valen, E. & Vastenhouw, N. L. Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome activation. *PLoS Genet.* **16**, e1008546 (2020).
29. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203**, 253–310 (1995).
30. Rabl, C. Uber zellthilung. *Morphol Jahrb* **10**, 214–330 (1885).
31. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2**, a003889 (2010).
32. Kaaij, L. J. T., van der Weide, R. H., Ketting, R. F. & de Wit, E. Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. *Cell Rep.* **24**, 1–10.e4 (2018).
33. Du, Z. *et al.* Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232–235 (2017).
34. Ke, Y. *et al.* 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell* **170**, 367–381.e20 (2017).
35. Nakamura, R. *et al.* CTCF looping is established during gastrulation in medaka embryos. *Genome Res.* **31**, 968–980 (2021).
36. Chen, X. *et al.* Key role for CTCF in establishing chromatin structure in human embryos. *Nature* **576**, 306–310 (2019).
37. Gassler, J. *et al.* A mechanism of cohesin□dependent loop extrusion organizes zygotic genome architecture. *EMBO J.* **36**, 3600–3618 (2017).
38. Polovnikov, K., Belan, S., Imakaev, M., Brandão, H. B. & Mirny, L. A. Fractal polymer with loops recapitulates key features of chromosome organization. *bioRxiv* 2022.02.01.478588 (2022) doi:10.1101/2022.02.01.478588.
39. Open2C *et al.* Cooltools: enabling high-resolution Hi-C analysis in Python. *bioRxiv*

2022.10.31.514564 (2022) doi:10.1101/2022.10.31.514564.

40. Liu, N. Q. *et al.* Rapid depletion of CTCF and cohesin proteins reveals dynamic features of chromosome architecture. *bioRxiv* 2021.08.27.457977 (2021) doi:10.1101/2021.08.27.457977.
41. Guo, Y. *et al.* Chromatin jets define the properties of cohesin-driven in vivo loop extrusion. *Mol. Cell* (2022) doi:10.1016/j.molcel.2022.09.003.
42. Isiaka, B. N. *et al.* Cohesin forms fountains at active enhancers in C. elegans. *bioRxiv* (2023).
43. Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A. & Mirny, L. A. Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harb. Symp. Quant. Biol.* **82**, 45–55 (2017).
44. Vian, L. *et al.* The Energetics and Physiological Impact of Cohesin Extrusion. *Cell* **175**, 292–294 (2018).
45. Baranasic, D. *et al.* Multiomic atlas with functional stratification and developmental dynamics of zebrafish cis-regulatory elements. *Nat. Genet.* **54**, 1037–1050 (2022).
46. Banigan, E. J. *et al.* Transcription shapes 3D chromatin organization by interacting with loop extrusion. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2210480120 (2023).
47. Vierstra, J. *et al.* Global reference mapping of human transcription factor footprints. *Nature* **583**, 729–736 (2020).
48. Banigan, E. J., van den Berg, A. A., Brandão, H. B., Marko, J. F. & Mirny, L. A. Chromosome organization by one-sided and two-sided loop extrusion. *Elife* **9**, (2020).
49. Davidson, I. F. *et al.* CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature* **616**, 822–827 (2023).
50. Dequeker, B. J. H. *et al.* MCM complexes are barriers that restrict cohesin-mediated loop extrusion. *Nature* **606**, 197–203 (2022).
51. Wang, X., Brandão, H. B., Le, T. B. K., Laub, M. T. & Rudner, D. Z. *Bacillus subtilis* SMC complexes juxtapose chromosome arms as they travel from origin to terminus. *Science* **355**, 524–527 (2017).

52. Brandão, H. B., Ren, Z., Karaboja, X., Mirny, L. A. & Wang, X. DNA-loop-extruding SMC complexes can traverse one another in vivo. *Nat. Struct. Mol. Biol.* **28**, 642–651 (2021).
53. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. Interplay of Loop Extrusion, Compartmentalization and Global Chromosome Dynamics Across Conditions and the Cell Cycle. in vol. 2019 S65.009 (2019).
54. Lunde, K., Belting, H.-G. & Driever, W. Zebrafish pou5f1/pou2, homolog of mammalian Oct4, functions in the endoderm specification cascade. *Curr. Biol.* **14**, 48–55 (2004).
55. Veil, M. *et al.* Maternal Nanog is required for zebrafish embryo architecture and for cell viability during gastrulation. *Development* **145**, (2018).
56. Riesle, A. J. *et al.* Activator-blocker model of transcriptional regulation by pioneer-like factors. *bioRxiv* 2023.01.11.523615 (2023) doi:10.1101/2023.01.11.523615.
57. Hsieh, T.-H. S. *et al.* Enhancer–promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat. Genet.* **54**, 1919–1932 (2022).
58. ENCODE Project Consortium *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
59. Han, R. *et al.* Targeted cohesin loading characterizes the entry and exit sites of loop extrusion trajectories. *bioRxiv* 2023.01.04.522689 (2023) doi:10.1101/2023.01.04.522689.
60. Vos, E. S. M. *et al.* Interplay between CTCF boundaries and a super enhancer controls cohesin extrusion trajectories and gene expression. *Mol. Cell* **81**, 3082–3095.e6 (2021).
61. Muñoz, S., Minamino, M., Casas-Delucchi, C. S., Patel, H. & Uhlmann, F. A Role for Chromatin Remodeling in Cohesin Loading onto Chromosomes. *Mol. Cell* **74**, 664–673.e5 (2019).
62. Goel, V. Y., Huseyin, M. K. & Hansen, A. S. Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. *Nat. Genet.* **55**, 1048–1056 (2023).
63. Westerfield, M. A guide for the laboratory use of zebrafish (Danio rerio). *(No Title)*

(2000).

64. Ulianov, S. V. *et al.* Suppression of liquid-liquid phase separation by 1,6-hexanediol partially compromises the 3D genome organization in living cells. *Nucleic Acids Res.* **49**, 10524–10541 (2021).
65. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46**, W537–W544 (2018).
66. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
67. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013).
68. Open2C *et al.* Pairtools: from sequencing data to chromosome contacts. *bioRxiv* (2023) doi:10.1101/2023.02.13.528389.
69. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
70. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).
71. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
72. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* **41**, D991–5 (2013).
73. Foster, E. D. & Deardorff, A. Open Science Framework (OSF). *J. Med. Libr. Assoc.* **105**, 203 (2017).
74. Kerpedjiev, P. *et al.* HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 125 (2018).
75. Llc, R. G. Reservoir. https://resgen.io/.
76. Yang, T. *et al.* HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* **27**, 1939–1949 (2017).
77. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *arXiv [cs.LG]* (2012).

78. Venev, S. *et al. open2c/cooltools: v0.5.2*. (2022). doi:10.5281/zenodo.7246348.

79. Flyamer, I. M., Illingworth, R. S. & Bickmore, W. A. Coolpup.py: versatile pile-up analysis of Hi-C data. *Bioinformatics* **36**, 2980–2985 (2020).

80. Spracklin, G. *et al.* Diverse silent chromatin states modulate genome compartmentalization and loop extrusion barriers. *Nat. Struct. Mol. Biol.* **30**, 38–51 (2023).

81. Siefert, J. C., Clowdus, E. A., Goins, D., Koren, A. & Sansam, C. L. Profiling DNA Replication Timing Using Zebrafish as an In Vivo Model System. *J. Vis. Exp.* (2018) doi:10.3791/57146.

82. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).

83. Open2C *et al.* Bioframe: Operations on Genomic Intervals in Pandas Dataframes. *bioRxiv* 2022.02.16.480748 (2022) doi:10.1101/2022.02.16.480748.

84. Abdennur, N. *pybbi: Python bindings to UCSC BigWig and BigBed library*. (Github).

85. Ulianov, S. V. *et al.* Order and stochasticity in the folding of individual Drosophila genomes. *Nat. Commun.* **12**, 41 (2021).

86. Bykov, N. S., Sigalova, O. M., Gelfand, M. S. & Galitsyna, A. A. HiChew: a Tool for TAD Clustering in Embryogenesis. in *Bioinformatics Research and Applications* 381–388 (Springer International Publishing, 2020).

87. Scharr, H. Optimal operators in digital image processing. (2000).

88. Scharr, H. Optimal Filters for Extended Optical Flow. in *Complex Motion* 14–29 (Springer Berlin Heidelberg, 2007).

89. Li, C. H. & Tam, P. K. S. An iterative algorithm for minimum cross entropy thresholding. *Pattern Recognit. Lett.* **19**, 771–776 (1998).

90. van der Walt, S. *et al.* scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).

91. Matthey-Doret, C. *et al.* Computer vision for pattern detection in chromosome contact maps. *Nat. Commun.* **11**, 5795 (2020).

92. Xu, C. *et al.* Nanog-like regulates endoderm formation through the Mxtx2-Nodal

pathway. *Dev. Cell* **22**, 625–638 (2012).

93. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
94. Tan, H., Onichtchouk, D. & Winata, C. DANIO-CODE: Toward an Encyclopedia of DNA Elements in Zebrafish. *Zebrafish* **13**, 54–60 (2016).
95. Castro-Mondragon, J. A. *et al.* JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**, D165–D173 (2022).
96. *JASPAR-UCSC-tracks: Code and data used to create the JASPAR UCSC Genome Browser tracks data hub*. (Github).
97. Deforzh, E. *et al.* Promoter and enhancer RNAs regulate chromatin reorganization and activation of miR-10b/HOXD locus, and neoplastic transformation in glioma. *Mol. Cell* **82**, 1894–1908.e5 (2022).
98. Patel, H., Wang, C. & Ewels, P. nf-core/chipseq: nf-core/chipseq v1. 2.2—Rusty Mole. *Zenodo https://doi. org/10.5281/zenodo*.
99. Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).
100. *polychrom: Polymer simulations of chromosomes and generating 'in silico' Hi-C maps*. (Github).
101. Goloborodko, A., Marko, J. F. & Mirny, L. A. Chromosome Compaction by Active Loop Extrusion. *Biophys. J.* **110**, 2162–2168 (2016).
102. Goloborodko, A., Imakaev, M. V., Marko, J. F. & Mirny, L. Compaction and segregation of sister chromatids via active loop extrusion. *Elife* **5**, (2016).
103. Banigan, E. J. *et al.* Transcription shapes 3D chromatin organization by interacting with loop extrusion. *bioRxiv* 2022.01.07.475367 (2022) doi:10.1101/2022.01.07.475367.
104. Mölder, F. *et al.* Sustainable data analysis with Snakemake. *F1000Res.* **10**, 33 (2021).
105. Danio-code dcc. https://danio-code.zfin.org/daniocode/.
106. Bogdanovic, O. *et al.* Dynamics of enhancer chromatin signatures mark the transition

from pluripotency to cell specification during embryogenesis. *Genome Res.* **22**, 2043–2053 (2012).