Demarcation of Topologically Associating Domains Is Uncoupled from Enriched CTCF Binding in Developing Zebrafish

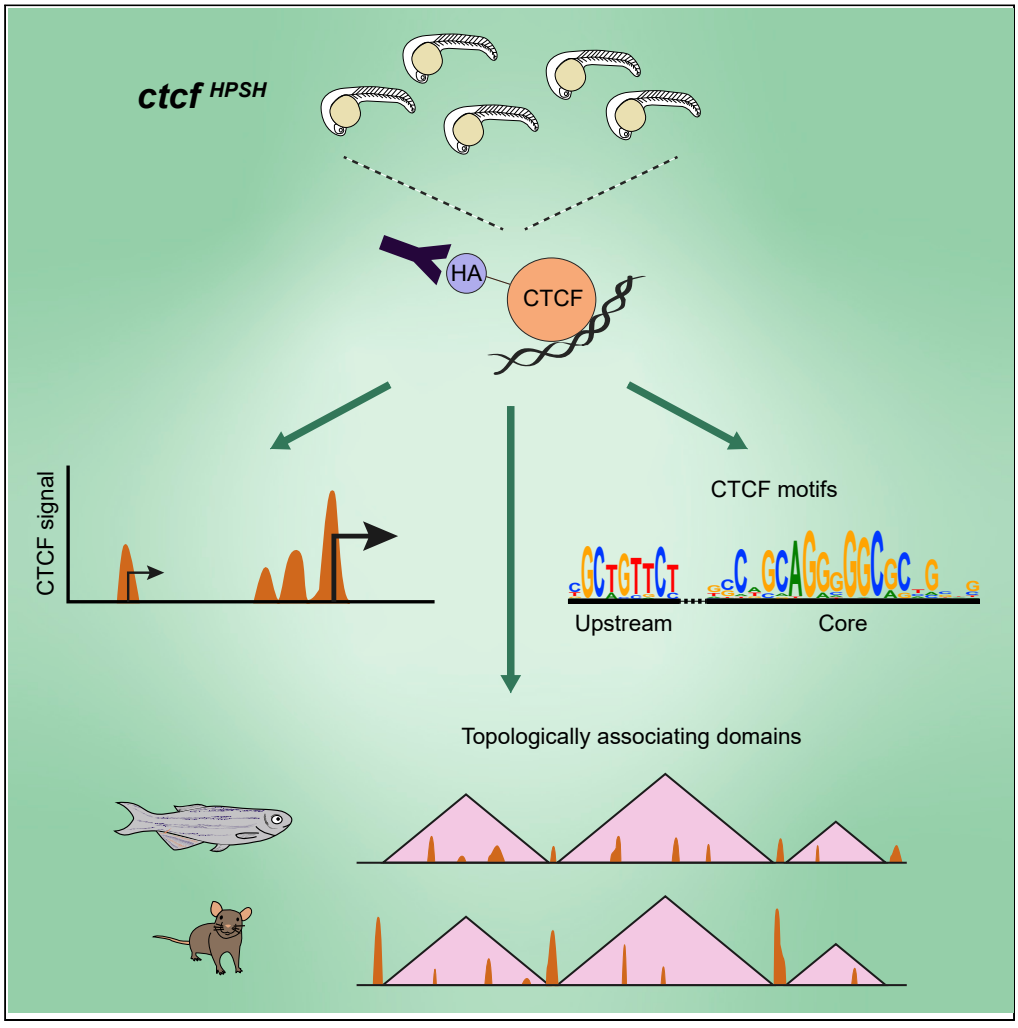
Yuvia A Perèz Rico, Emmanuel Barillot, Alena Shkumatava

iScience

CellPress
OPEN ACCESS

Article

Demarcation of Topologically Associating Domains Is Uncoupled from Enriched CTCF Binding in Developing Zebrafish



Yuvia A. Pérez-Rico, Emmanuel Barillot, Alena Shkumatava

alena.shkumatava@curie.fr

HIGHLIGHTS

Identification of CTCF occupancy in zebrafish embryos using a tagged *ctcf* allele

CTCF binding at promoters correlates with gene expression levels

No general CTCF enrichment at topological domain boundaries in zebrafish embryos

Pérez-Rico et al., iScience *23*, 101046
May 22, 2020
https://doi.org/10.1016/j.isci.2020.101046

iScience

CellPress
OPEN ACCESS

Article

Demarcation of Topologically Associating Domains Is Uncoupled from Enriched CTCF Binding in Developing Zebrafish

Yuvia A. Pérez-Rico,[1,2,3,4] Emmanuel Barillot,[2] and Alena Shkumatava[1,5,*]

SUMMARY

CCCTC-binding factor (CTCF) is a conserved architectural protein that plays crucial roles in gene regulation and three-dimensional (3D) chromatin organization. To better understand mechanisms and evolution of vertebrate genome organization, we analyzed genome occupancy of CTCF in zebrafish utilizing an endogenously epitope-tagged CTCF knock-in allele. Zebrafish CTCF shares similar facets with its mammalian counterparts, including binding to enhancers, active promoters and repeat elements, and bipartite sequence motifs of its binding sites. However, we found that *in vivo* CTCF binding is not enriched at boundaries of topologically associating domains (TADs) in developing zebrafish, whereas TAD demarcation by chromatin marks did not differ from mammals. Our data suggest that general mechanisms underlying 3D chromatin organization, and in particular the involvement of CTCF in this process, differ between distant vertebrate species.

INTRODUCTION

CCCTC-binding factor (CTCF) is a key regulator of gene expression and plays a central role in 3D organization of mammalian genomes (Dixon et al., 2012; Guo et al., 2015; Nora et al., 2017). In mammals, CTCF demarcates topologically associating domain (TAD) boundaries, and disruption of the CTCF sites in these regions results in the formation of ectopic contacts between neighboring domains (Despang et al., 2019; Dowen et al., 2014; Lupiáñez et al., 2015). CTCF and its role in gene regulation are conserved throughout bilaterians (Heger et al., 2012), whereas CTCF functions in 3D genome organization have diverged between invertebrates and vertebrates. In contrast to mammals, *Drosophila* CTCF is not essential for embryogenesis and its binding is not enriched at TAD boundaries (Gambetta and Furlong, 2018; Rowley et al., 2017). Given the functional divergence of CTCF in *Drosophila* and mammals, CTCF analyses in other non-mammalian vertebrate species are key for understanding the evolution and regulation of the 3D chromatin organization. Although the functions of CTCF in zebrafish development have been previously explored (Carmona-Aldana et al., 2018; Delgado-Olguín et al., 2011; Meier et al., 2018; Rhodes et al., 2010), no genome-wide CTCF *in vivo* binding data have been achieved in zebrafish so far. Similar to mammals, predicted CTCF binding motifs are distributed in divergent orientation at TAD boundaries in zebrafish (Gómez-Marín et al., 2015; Kaaij et al., 2018), suggesting the conserved role of CTCF in TAD demarcation. Here, we identified and characterize CTCF occupancy in developing zebrafish embryos using an epitope-tagged allele of *ctcf*. Although several gene regulatory features of zebrafish CTCF are similar to mammals, no enrichment of the *in vivo* CTCF occupancy was detected at TAD boundaries in zebrafish embryos, suggesting functional differences of CTCF in 3D genome architecture between vertebrates.

RESULTS AND DISCUSSION

Identification of *In Vivo* CTCF Binding Sites Using the *ctcf* ^{HPSH} Zebrafish Allele

To determine CTCF occupancy in the zebrafish genome, we generated a tagged allele of *ctcf*, where a tripartite HA-PreScission-His tag was inserted in frame after the start codon of *ctcf* resulting in N-terminally endogenous CTCF tagged by HPSH (*ctcf* ^{HPSH} allele) (Figures 1A and S1A and Transparent Methods). We confirmed the expression of the tagged protein in *ctcf* $^{HPSH/HPSH}$ zebrafish (Figure 1B). Homozygous *ctcf* $^{HPSH/HPSH}$ zebrafish developed normally and were viable and fertile, indicating that the function of CTCF was not affected by the tag (Figures S1B and S1C). Chromatin immunoprecitation sequencing (ChIP-seq) analyses of CTCF binding in 24 hours postfertilization (hpf) *ctcf* ^{HPSH} embryos showed high correlation between biological replicates (Figures 1C and S1D) and confirmed a previously reported autoregulatory binding of CTCF to its promoter (Figure S1E)

[1]Institut Curie PSL Research University, INSERM U934, CNRS UMR 3215 75005, Paris, France

[2]Institut Curie, Mines ParisTech, PSL Research University, INSERM, U900, 75005, Paris, France

[3]Sorbonne Université, Complexité du Vivant, 75005, Paris, France

[4]Present address: Directors' Research, European Molecular Biology Laboratory, 69117 Heidelberg, Germany

[5]Lead Contact

*Correspondence: alena.shkumatava@curie.fr

https://doi.org/10.1016/j.isci.2020.101046

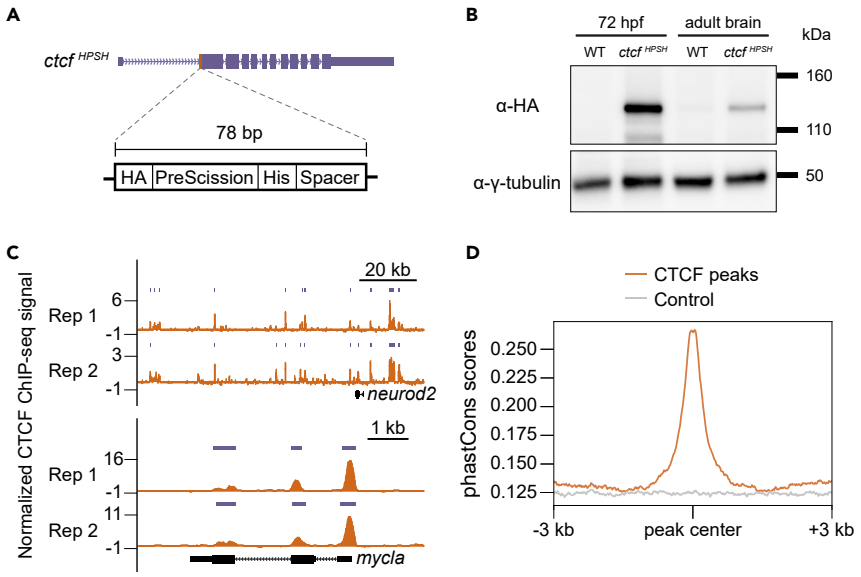


Figure 1. Identification of CTCF Binding in Zebrafish

(A) Schematic representation of the *ctcf* ^{HPSH} zebrafish allele. Orange and purple boxes represent the inserted sequence and exons, respectively.

(B) Western blot using anti-hemagglutinin (HA) antibody on extracts from wild-type (WT) and *ctcf* $^{HPSH/HPSH}$ whole embryos and adult brains. Molecular weights are indicated on the right. γ-Tubulin served as a loading control.

(C) Tracks showing examples of CTCF peaks (purple bars) at the *neurod2* and *mycla* loci (both located on the reverse strand). Displayed signal distributions and peaks correspond to biological replicates (Rep 1, Rep 2). Signal is represented on the *y* axis as $-\log_{10}$ (p value) of the CTCF ChIP-seq signal.

(D) Distribution of the average sequence conservation of CTCF peaks and control regions using peak centers as reference point.

See also Figure S1 and Tables S1 and S2.

(Pugacheva et al., 2006). In the ChIP-seq data merged from both replicates, we identified 36,540 CTCF peaks that showed higher phastCons sequence conservation than random control regions (Figure 1D and Transparent Methods); the same trend was observed when considering only CTCF peaks that do not overlap exons (Figure S1F). Notably, the number of CTCF peaks identified in the zebrafish genome roughly corresponds to the number of CTCF sites in mammalian genomes (Pugacheva et al., 2020). Therefore, the *ctcf* ^{HPSH} zebrafish allele enables reliable and reproducible detection of the *in vivo* CTCF occupancy in the zebrafish genome.

Common Features of CTCF Binding Sites in Vertebrates

Because of the function of mammalian CTCF in establishing enhancer-promoter interactions (Sanyal et al., 2012), we analyzed zebrafish CTCF binding with respect to histone modifications and DNA accessibility in 24-hpf zebrafish embryos (Aday et al., 2011; Bogdanović et al., 2012; Gehrke et al., 2015; Irimia et al., 2012; Ulitsky et al., 2011) (Tables S1 and S2 and Transparent Methods). We found that zebrafish CTCF peaks were enriched for poised (H3K4me1), active (H3K4me3, H3K27ac), and accessible chromatin (ATAC-seq), but not for inactive chromatin (H3K27me3) (Figure 2A). Furthermore, *de novo* motif discovery identified a 20-bp core motif that was present in 78% of CTCF peaks and showed more similarity to the human CTCF motif than to the human CTCFL or *Drosophila* CTCF motifs (Figure 2B and Data S1). As reported for other vertebrate species (Boyle et al., 2011; Filippova et al., 1996; Kadota et al., 2017; Rhee and Pugh, 2011; Schmidt et al., 2012), we also identified enriched CTCF upstream motifs, separated from the core motif by 8- or 12-bp spacers (Figures 2C and S2A). Similar to mammalian CTCF binding sites, a fraction of which propagated in the genome through retrotransposition of repeat elements (Schmidt et al., 2012), non-autonomous DNA transposons were significantly enriched on CTCF binding sites (Figures 2D, S2B, S2C, and Table S3). Taken together, our analyses show that zebrafish and mammalian CTCF binding sites share similar features.

CTCF Abundance at Promoters Correlates with the Gene Expression Levels

Although CTCF peaks were mainly located in intronic and intergenic regions, ~6% of the peaks were found within promoters (Figure 3A). In human, a fraction of CTCF sites in promoter and intragenic regions is engaged in loops

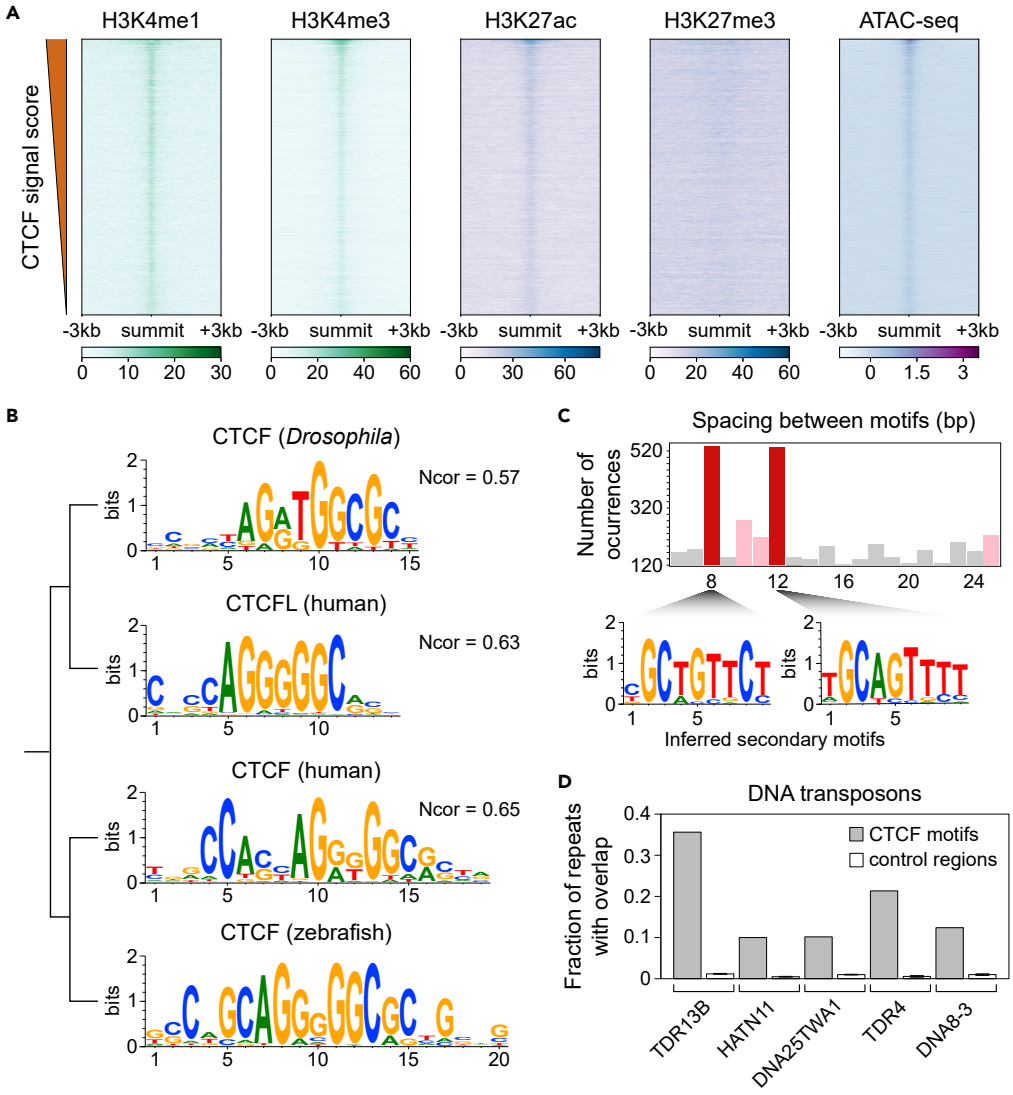


Figure 2. Characterization of CTCF Peaks and Binding Sites

(A) Heatmap profiles of four histone marks and ATAC-seq data at CTCF peaks ranked by decreasing CTCF ChIP-seq signal over the displayed region. Normalized signal is shown in FPM (fragments per million mapped fragments) for ATAC-seq and in RPKM (reads per kilobase million) for histone marks. All datasets correspond to the 24-hpf zebrafish embryonic stage.

(B) Dendrogram representing hierarchical clustering results of CTCF and CTCFL motifs. In total, 28,538 of 36,540 CTCF ChIP-seq peaks contained at least one matching site to the zebrafish motif, compared with 7,360 of 36,540 control sequences. Information content of each position on the x axis is expressed in bits on the y axis. Ncor, normalized Pearson correlation.

(C) Histogram showing the number of co-occurrences of the CTCF core and upstream motifs at different spacing distances (6–25 bp). Non-significant enriched spacing distances are shown in gray, enrichments are shown in pink, and the highest enrichments are shown in red. Bottom, inferred upstream motifs using sequences matching to the reference motif at the indicated distances from the CTCF core motif. Information content of each position on the x axis is expressed in bits on the y axis.

(D) Top five DNA transposon types enriched for CTCF binding sites. The fraction of repeats overlapping at least one CTCF motif is shown on the y axis. For control regions, the mean and the standard deviation (error bars) calculated by bootstrap analyses are shown.

See also Figure S2, Tables S1–S3 and Data S1.

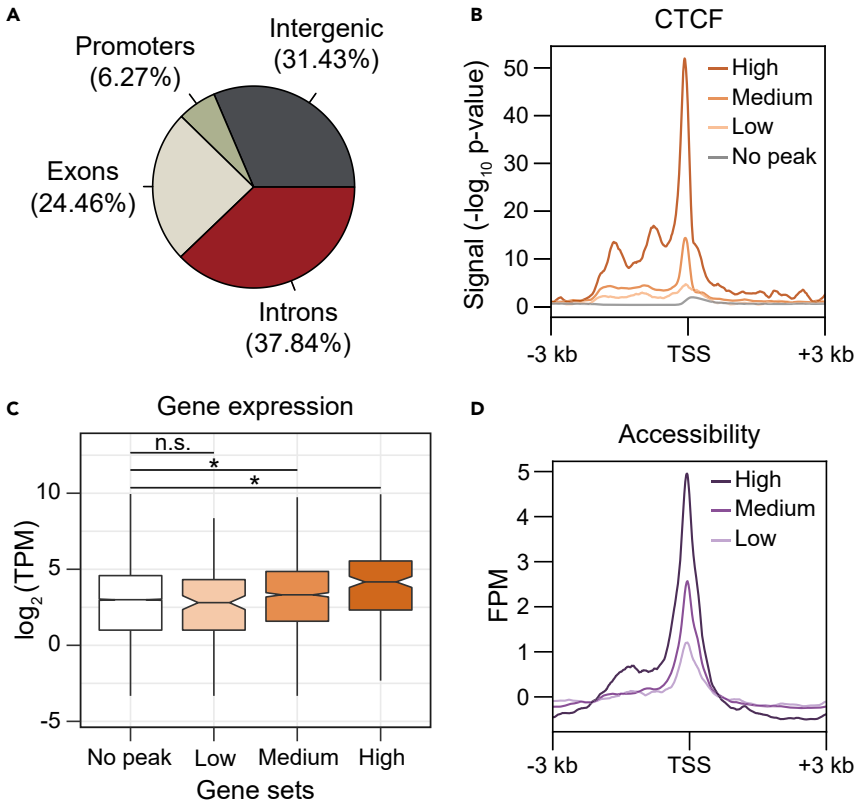


Figure 3. High Abundance of CTCF Binding at Promoters Associates with High Gene Expression Levels

(A) Distribution of CTCF peaks across different zebrafish genomic regions. Percentages represent the number of CTCF peaks for each category.

(B) Average CTCF ChIP-seq signal profiles over promoters. Each line represents one of the three gene categories defined by CTCF abundance at promoters (low, medium, high) or promoters with no CTCF peaks (no peak).

(C) Expression of the stratified gene categories and genes without CTCF peaks at promoters. Differences in distribution are denoted as significant (*) and non-significant (n.s.) according to two-sided Wilcoxon rank-sum test (p value $\leq 1.5 \times 10^{-5}$).

(D) Average ATAC-seq signal profiles over promoters of gene categories defined by CTCF abundance as explained in (B). See also Figure S3.

that impact exon inclusion (Ruiz-Velasco et al., 2017), raising the possibility of an equivalent mechanism in zebrafish. Because the resolution of the available zebrafish high-throughput chromosome conformation capture (Hi-C) data (~20 kb) (Kaaij et al., 2018) does not allow investigation of this type of looping interactions, we sought to determine distinctive features of genes with CTCF-bound promoters. All genes that showed CTCF binding at their promoters were classified in three categories based on the signal strength of CTCF peaks ranging from high (top 10% percentile) to low (bottom 10% percentile) (Figure 3B and Transparent Methods). We found a positive correlation between the presence of CTCF motifs and CTCF occupancy regardless of the site orientation relative to transcription (Figure S3A) (χ^2 tests of independence, p value $\leq 7.9 \times 10^{-10}$). The increased CTCF occupancy at promoter also positively correlated with increased gene expression (White et al., 2017) (Figure 3C) and DNA accessibility (Figure 3D). Notably, CTCF binding at promoters is not a mere reflection of permissive chromatin, as we also identified promoters with high ATAC-seq signal but no CTCF binding (Figure S3B). Genes with high CTCF promoter occupancy had high signals for histone marks associated with enhancers (H3K4me1, H3K27ac), active promoters (H3K4me3, H3K27ac), and transcriptional elongation (Figures S3C–S3F). By contrast, no correlation between repressive chromatin (H3K27me3) and CTCF abundance was found (Figure S3G), whereas enrichment of the H3K27me3 repressive mark was overall higher at CTCF-bound promoters than at promoters without CTCF peaks (Figure S3H). In summary, our findings suggest that CTCF binding at promoters generally correlates with chromatin states that favor transcription. This observation could be explained by CTCF playing a role in the generation of nucleosome-depleted regions (Nora et al., 2017) or reflect CTCF binding in specific cell types to prevent ectopic gene expression, as previously reported for *cis*-regulatory elements of *runx1* in zebrafish (Marsman et al., 2014).

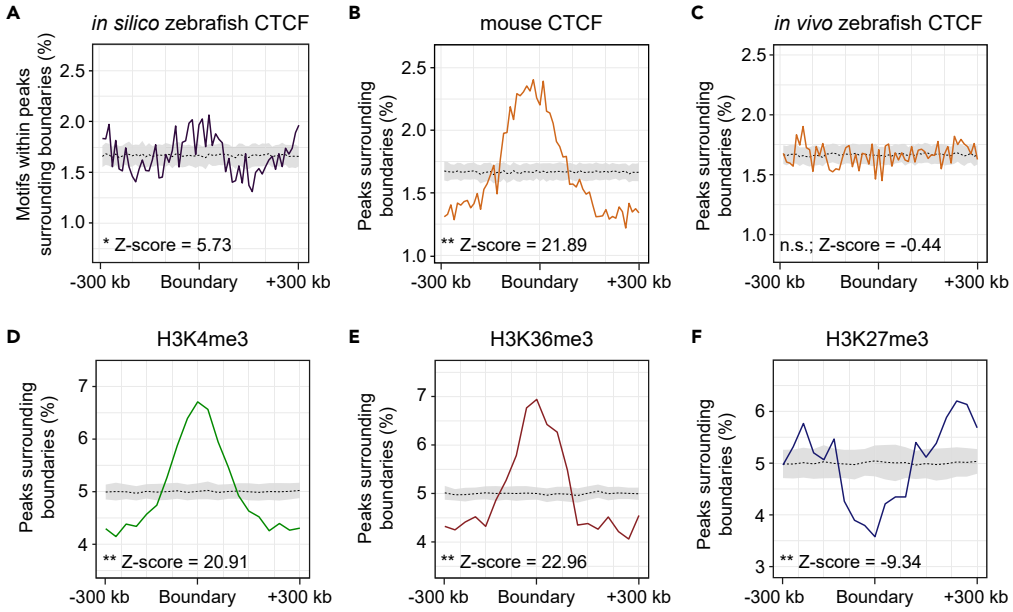


Figure 4. Active Chromatin Marks but Not CTCF Are Enriched at TAD Boundaries

(A) Distribution of predicted CTCF motifs within ATAC-seq peaks (purple) along 600-kb regions centered on TAD boundaries (x axis) in 24-hpf zebrafish. The y axis shows the percentage of total peak counts in the 600-kb region located at each genomic position. The dashed line represents the mean background distribution, and the gray ribbon depicts the ±1 standard deviation range from the mean. Differences in mean percentages (central 60 kb) were assessed by *Z* scores. Non-significant, n.s.; *$p < 1 \times 10^{-5}$; **$p < 1 \times 10^{-20}$.

(B) Distribution of CTCF peaks (orange) relative to TAD boundaries identified in mouse embryonic stem cells as described in (A).

(C) Distribution of CTCF peaks (orange) relative to TAD boundaries identified in 24-hpf zebrafish embryos as described in (A).

(D) Distribution of H3K4me3-enriched peaks along TAD boundaries as described in (A).

(E) Distribution of H3K36me3-enriched peaks along TAD boundaries as described in (A).

(F) Distribution of H3K27me3-enriched peaks along TAD boundaries as described in (A).

See also Figures S4, S5, and Table S4.

No Enrichment of CTCF Binding at TAD Boundaries in Zebrafish Embryos

Next, we sought to investigate the role of CTCF in zebrafish 3D genome organization by analyzing its distribution at TAD boundaries. Visualization of the available 24-hpf Hi-C data (Kaaij et al., 2018) showed enriched interactions of centromeres and telomeres and an uneven distribution of the signal along chromosomes (Figure S4A), which was even more pronounced at earlier developmental stages (Figure S4B). Although this signal distribution reflects the Rabl organization of chromosomes with continuous arm pairing over the cell cycle characteristic for dividing cells (Stadler et al., 2017) and the cell cycle heterogeneity of 24-hpf embryos (Figures S4B and S4C), we, nevertheless, identified 1,307 TADs (median size = 580 kb) using insulation scores (Crane et al., 2015) (Figure S4D and Table S4 and Transparent Methods). The latter was possible given that the 24-hpf Hi-C maps are a composite of the interactions occurring in dividing and interphase cells, which are characterized by lack and presence of TADs, respectively.

Although the annotated CTCF motif was enriched within accessible chromatin sites at TAD boundaries, this enrichment was not boundary specific, as it was also found enriched at accessible sites within TADs (enrichment p values $< 1 \times 10^{-20}$) (Figure S5A). Our analysis showed a moderate enrichment of *in silico* predicted CTCF sites and their divergent orientation bias within accessible chromatin at TAD boundaries, consistent with previous reports (Gómez-Marín et al., 2015; Kaaij et al., 2018) (Figures 4A and S5B). Similar to mouse CTCF (Dixon et al., 2012), only a small fraction of zebrafish CTCF peaks was located at TAD boundaries (Figure S5C). However, unlike enriched CTCF binding at TAD boundaries in mammals (Figure 4B), neither CTCF peaks nor the *in vivo* identified CTCF motifs were enriched at TAD boundaries in 24-hpf zebrafish embryos (Figures 4C and S5D). To exclude analysis bias, we applied the reciprocal insulation method

and identified hierarchical domains (Zhan et al., 2017). In contrast to mammalian CTCF and in agreement with our zebrafish above-mentioned results, no reciprocal insulation value at which zebrafish CTCF enrichment was clearly maximized was identified (Figures S5E and S5F and Transparent Methods), potentially reflecting low variability in sequence composition of the genome (Costantini et al., 2007). Likewise, we found no zebrafish CTCF enrichment at boundaries of domains identified at 72.5% reciprocal insulation (i.e., the value at which the highest percentage of boundaries overlaps with CTCF peak summits) (Figure S5G). In agreement with the previous report (Kaaij et al., 2018), we found that TAD boundaries in zebrafish were enriched for chromatin marks associated with active transcription (Figures 4D and 4E), depleted for the H3K27me3 repressive mark (Figure 4F), and showed no enrichment for H3K27ac (Figure S5H) suggesting that the biochemical features of zebrafish TAD boundaries are similar to mammals. Our results do not imply that CTCF is dispensable for TAD establishment in the zebrafish genome, as we found moderate enrichment of predicted CTCF motifs within accessible chromatin at TAD boundaries and motif orientation biases, but they rather indicate that, in contrast to mammals, there is no strong correlation between TAD boundaries and high enrichment of CTCF. Importantly, it will require further investigations to determine if this moderate enrichment of CTCF is sufficient to establish TAD boundaries in zebrafish. It is also reasonable to propose that additional architectural proteins or the active chromatin state may play a role in TAD establishment in zebrafish, similar to *Drosophila* and other eukaryotes lacking this architectural protein. Interestingly, replication timing that is tightly associated with TAD distribution correlates with transcriptional status in zebrafish (Siefert et al., 2017) supporting the latter hypothesis. Moreover, it will be important to investigate colocalization of CTCF and cohesin binding in zebrafish, as the N-terminal CTCF region mediating the interaction with cohesin in mammals differs in its amino acid composition in zebrafish (Li et al., 2020; Pugacheva et al., 2006, 2020). Indeed, this N-terminal region is highly conserved in organisms, in which CTCF is enriched at TAD boundaries including mammals and chicken (Fishman et al., 2018), but it is not conserved in *Drosophila*, in which CTCF does not delineate TAD boundaries (Moon et al., 2005; Rowley et al., 2017). Therefore, possible differences in the interaction between CTCF and cohesin in zebrafish may explain lack of CTCF enrichment at TAD boundaries. CTCF/cohesin ChIP-seq and Hi-C analyses in specific cell types and using single-cell approaches will be required to further investigate the functions of CTCF in the higher-order organization of the zebrafish genome.

Limitations of the Study

Although we found a positive correlation between CTCF binding at promoters and elevated gene expression, the cellular heterogeneity of 24-hpf zebrafish embryos does not allow to distinguish between CTCF-facilitating gene expression in specific cell types while acting as an insulator in other cells. Future cell-type-specific depletion of CTCF followed by purification of these cells will be required to interrogate CTCF binding and gene expression changes. In addition, it will also be important to analyze the relationship between CTCF enrichment and TAD boundaries in specific cell types to establish whether lack of strong CTCF enrichment at boundaries is maintained across different cell types or is cell specific.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Alena Shkumatava (alena.shkumatava@curie.fr).

Materials Availability

The *ctcf* $^{HPSH/HPSH}$ zebrafish line generated in this study is available from the Lead Contact without restriction.

Data and Code Availability

CTCF ChIP-seq sequencing data generated in this study are available in the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/). The accession number for the sequencing data reported in this study is NCBI GEO: GSE133437. No previously unreported algorithms were used to generate the results.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101046.

ACKNOWLEDGMENTS

We thank members of the Shkumatava laboratory for discussions. ChIP-seq libraries generated in this study were sequenced at the ICGex NGS platform of Institut Curie funded by grants from "L'Agence Nationale de la Recherche", France (ANR-10-EQPX-03 (Equipex) and ANR-10-INBS-09-08 (France Génomique Consortium) from the "Investissements d'Avenir" program) and by the Cancéropole Île-de-France, France. This research was funded by grants from the European Research Council (FLAME-337440), "Fondation pour la Recherche Médicale", France (DBI201312285578), and LabEx DEEP, France (ANR-11-LABX-0044, ANR-10-IDEX-0001-02).

AUTHOR CONTRIBUTIONS

Conceptualization, Y.A.P.-R.; Methodology, Y.A.P.-R.; Formal Analysis, Y.A.P.-R.; Investigation, Y.A.P.-R.; Resources, E.B. and A.S.; Data Curation, Y.A.P.-R.; Writing – Original Draft, Y.A.P.-R.; Writing, Review & Editing, Y.A.P.-R and A.S.; Visualization, Y.A.P.-R.; Supervision, E.B. and A.S.; Funding Acquisition, E.B. and A.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 24, 2019
Revised: March 19, 2020
Accepted: April 3, 2020
Published: May 22, 2020

REFERENCES

Aday, A.W., Zhu, L.J., Lakshmanan, A., Wang, J., and Lawson, N.D. (2011). Identification of cis regulatory features in the embryonic zebrafish genome through large-scale profiling of H3K4me1 and H3K4me3 binding sites. Dev. Biol. *357*, 450–462.

Bogdanović, O., Fernandez-Miñán, A., Tena, J.J., De La Calle-Mustienes, E., Hidalgo, C., Van Kruysbergen, I., Van Heeringen, S.J., Veenstra, G.J.C., and Gómez-Skarmeta, J.L. (2012). Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res. *22*, 2043–2053.

Boyle, A.P., Song, L., Lee, B.K., London, D., Keefe, D., Birney, E., Iyer, V.R., Crawford, G.E., and Furey, T.S. (2011). High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. Genome Res. *21*, 456–464.

Carmona-Aldana, F., Zampedri, C., Suaste-Olmos, F., Murillo-de-Ozores, A., Guerrero, G., Arzate-Mejía, R., Maldonado, E., Navarro, R.E., Chimal-Monroy, J., and Recillas-Targa, F. (2018). CTCF knockout reveals an essential role for this protein during the zebrafish development. Mech. Dev. *154*, 51–59.

Costantini, M., Auletta, F., and Bernardi, G. (2007). Isochore patterns and gene distributions in fish genomes. Genomics *90*, 364–371.

Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J., and Meyer, B.J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature *523*, 240–244.

Delgado-Olguín, P., Brand-Arzamendi, K., Scott, I.C., Jungblut, B., Stainier, D.Y., Bruneau, B.G., and Recillas-Targa, F. (2011). CTCF promotes muscle differentiation by modulating the activity of myogenic regulatory factors. J. Biol. Chem. *286*, 12483–12494.

Despang, A., Schöpflin, R., Franke, M., Ali, S., Jerković, I., Paliou, C., Chan, W.-L., Timmermann, B., Wittler, L., Vingron, M., et al. (2019). Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. Nat. Genet. *51*, 1263–1271.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374–387.

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., and Lobanenkov, V.V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol. Cell. Biol. *16*, 2802–2813.

Fishman, V., Battulin, N., Nuriddinov, M., Maslova, A., Zlotina, A., Strunov, A., Chervyakova, D., Korablev, A., Serov, O., and Krasikova, A. (2018). 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. Nucleic Acids Res. *45*, 846–860.

Gambetta, M.C., and Furlong, E.E.M. (2018). The insulator protein CTCF is required for correct hox gene expression, but not for embryonic development in Drosophila. Genetics *210*, 129–136.

Gehrke, A.R., Schneider, I., de la Calle-Mustienes, E., Tena, J.J., Gomez-Marin, C., Chandran, M., Nakamura, T., Braasch, I., Postlethwait, J.H., Gómez-Skarmeta, J.L., et al. (2015). Deep conservation of wrist and digit enhancers in fish. Proc. Natl. Acad. Sci. U S A *112*, 803–808.

Gómez-Marín, C., Tena, J.J., Acemel, R.D., López-Mayorga, M., Naranjo, S., de la Calle-Mustienes, E., Maeso, I., Beccari, L., Aneas, I., Vielmas, E., et al. (2015). Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. Proc. Natl. Acad. Sci. U S A *112*, 7542–7547.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. Cell *162*, 900–910.

Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., and Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. Proc. Natl. Acad. Sci. U S A *109*, 17507–17512.

Irimia, M., Tena, J.J., Alexis, M.S., Fernandez-Miñan, A., Maeso, I., Bogdanovic, O., de la Calle-Mustienes, E., Roy, S.W., Gómez-Skarmeta, J.L., and Fraser, H.B. (2012). Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. Genome Res. *22*, 2356–2367.

Kaaij, L.J.T., van der Weide, R.H., Ketting, R.F., and de Wit, E. (2018). Systemic loss and gain of chromatin architecture throughout zebrafish development. Cell Rep. *24*, 1–10.e4.

Kadota, M., Hara, Y., Tanaka, K., Takagi, W., Tanegashima, C., Nishimura, O., and Kuraku, S. (2017). CTCF binding landscape in jawless fish with reference to Hox cluster evolution. Sci. Rep. *7*, 4957.

Li, Y., Haarhuis, J.H.I., Sedeño Cacciatore, Á., Oldenkamp, R., van Ruiten, M.S., Willems, L., Teunissen, H., Muir, K.W., de Wit, E., Rowland, B.D., et al. (2020). The structural basis for cohesin-CTCF-anchored loops. Nature *578*, 472–476.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025.

Marsman, J., O'Neill, A.C., Kao, B.R.Y., Rhodes, J.M., Meier, M., Antony, J., Mönnich, M., and Horsfield, J.A. (2014). Cohesin and CTCF differentially regulate spatiotemporal runx1 expression during zebrafish development. Biochim. Biophys. Acta *1839*, 50–61.

Meier, M., Grant, J., Dowdle, A., Thomas, A., Gerton, J., Collas, P., O'Sullivan, J.M., and Horsfield, J.A. (2018). Cohesin facilitates zygotic genome activation in zebrafish. Development *145*, dev156521.

Moon, H., Filippova, G., Loukinov, D., Pugacheva, E., Chen, Q., Smith, S.T., Munhall, A., Grewe, B., Bartkuhn, M., Arnold, R., et al. (2005). CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. EMBO Rep. *6*, 165–170.

Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. Cell *169*, 930–944.e22.

Pugacheva, E.M., Kwon, Y.W., Hukriede, N.A., Pack, S., Flanagan, P.T., Ahn, J.C., Park, J.A., Choi, K.S., Kim, K.W., Loukinov, D., et al. (2006). Cloning and characterization of zebrafish CTCF: developmental expression patterns, regulation of the promoter region, and evolutionary aspects of gene organization. Gene *375*, 26–36.

Pugacheva, E.M., Kubo, N., Loukinov, D., Tajmul, M., Kang, S., Kovalchuk, A.L., Strunnikov, A.V., Zentner, G.E., Ren, B., and Lobanenkov, V.V. (2020). CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. Proc. Natl. Acad. Sci. U S A *117*, 2020–2031.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408–1419.

Rhodes, J.M., Bentley, F.K., Print, C.G., Dorsett, D., Misulovin, Z., Dickinson, E.J., Crosier, K.E., Crosier, P.S., and Horsfield, J.A. (2010). Positive regulation of c-Myc by cohesin is direct, and evolutionarily conserved. Dev. Biol. *344*, 637–649.

Rowley, M.J., Nichols, M.H., Lyu, X., Ando-Kuri, M., Rivera, I.S.M., Hermetz, K., Wang, P., Ruan, Y., and Corces, V.G. (2017). Evolutionarily conserved principles predict 3D chromatin organization. Mol. Cell *67*, 837–852.e7.

Ruiz-Velasco, M., Kumar, M., Lai, M.C., Bhat, P., Solis-Pinson, A.B., Reyes, A., Kleinsorg, S., Noh, K.M., Gibson, T.J., and Zaugg, J.B. (2017). CTCF-mediated chromatin loops between promoter and gene body regulate alternative splicing across individuals. Cell Syst. *5*, 628–637.e6.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109–113.

Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Gonalves, Â., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell *148*, 335–348.

Siefert, J.C., Georgescu, C., Wren, J.D., Koren, A., and Sansam, C.L. (2017). DNA replication timing during development anticipates transcriptional programs and parallels enhancer activation. Genome Res. *27*, 1406–1416.

Stadler, M.R., Haines, J.E., and Eisen, M.B. (2017). Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early Drosophila melanogaster embryo. Elife *6*, 1–29.

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell *147*, 1537–1550.

White, R.J., Collins, J.E., Sealy, I.M., Wali, N., Dooley, C.M., Digby, Z., Stemple, D.L., Murphy, D.N., Billis, K., Hourlier, T., et al. (2017). A high-resolution mRNA expression time course of embryonic development in zebrafish. Elife *6*, 1–32.

Zhan, Y., Mariani, L., Barozzi, I., Schulz, E.G., Blüthgen, N., Stadler, M., Tiana, G., and Giorgetti, L. (2017). Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. Genome Res. *27*, 479–490.

iScience, Volume *23*

Supplemental Information

Demarcation of Topologically Associating Domains Is Uncoupled from Enriched CTCF Binding in Developing Zebrafish

Yuvia A. Pérez-Rico, Emmanuel Barillot, and Alena Shkumatava

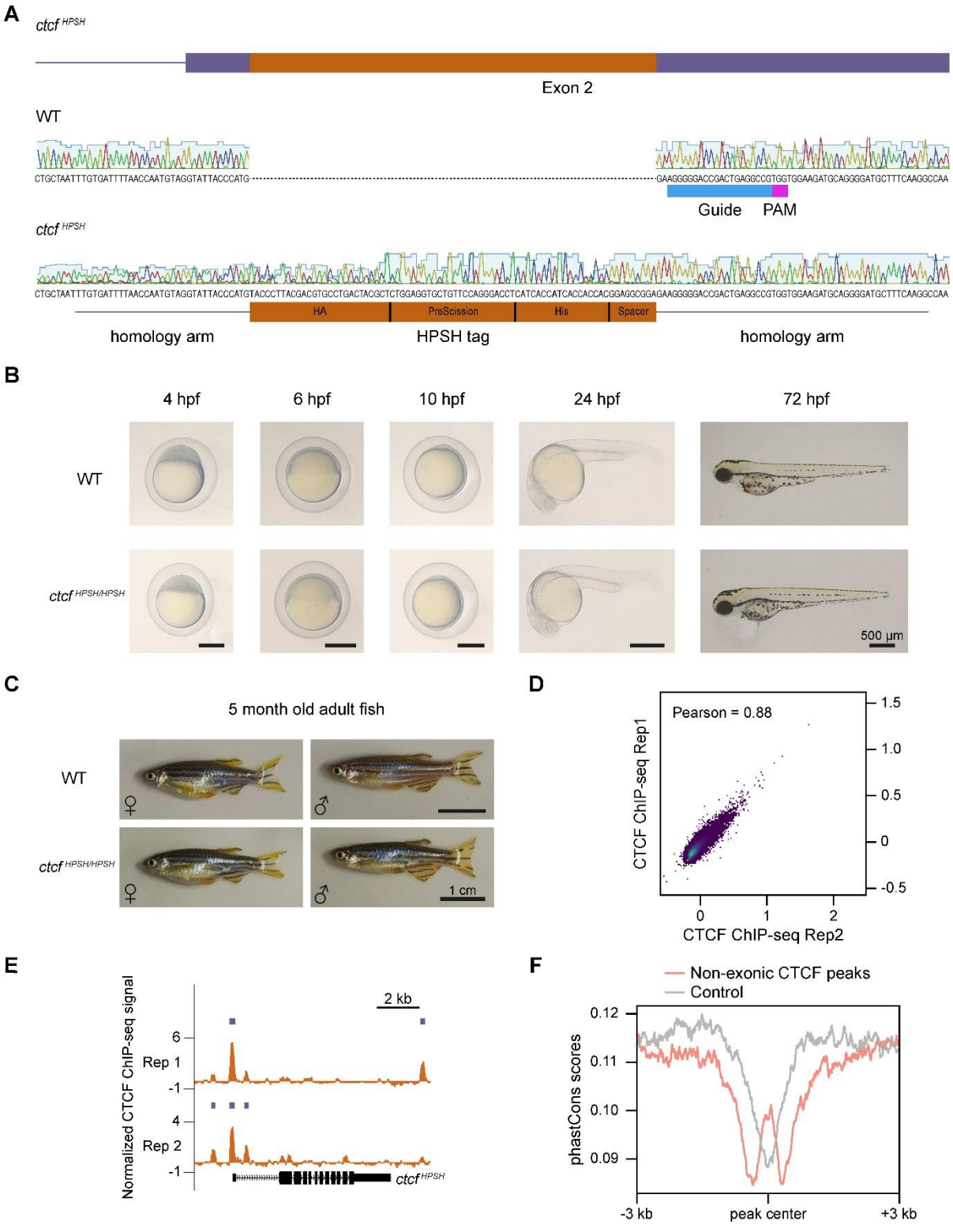


Figure S1. Characterization of the *ctcf* ^{HPSH} allele and identification of CTCF binding in zebrafish, Related to Figure 1.

(A) Schematic representation of the *ctcf*^{HPSH} zebrafish locus showing exon 2 that was targeted to generate the tagged allele (top panel). Middle and bottom panels show DNA sequencing chromatographs and nucleotide composition of wild type (WT) *ctcf* and *ctcf*^{HPSH} zebrafish. The positions of the short guide RNA and PAM sequences are indicated with blue and magenta blocks, respectively. In the bottom panel, the HPSH tag is indicated with orange blocks and homology arms are indicated with black lines. (B) Wild type (WT) and *ctcf*$^{HPSH/HPSH}$ embryos during early development. Scale bars for each stage are displayed. Hfp, hours post-fertilization. (C) Wild type (WT) and *ctcf*$^{HPSH/HPSH}$ adult fish. Representative fish of each sex and genotype are shown. Scale bars for each genotype are displayed. (D) Correlation between CTCF ChIP-seq biological replicates. In the scatter plot, each dot represents one 1 Mb bin and the axes correspond to the average signals in replicate 1 (y axis) and replicate 2 (x axis). (E) Genome track showing CTCF signal distribution at the *ctcf* locus in two biological replicates. (F) Distribution of average sequence conservation, measured by phastCons scores on the y axis, of non-exonic CTCF peaks and control regions using as reference point the center of peaks and controls.

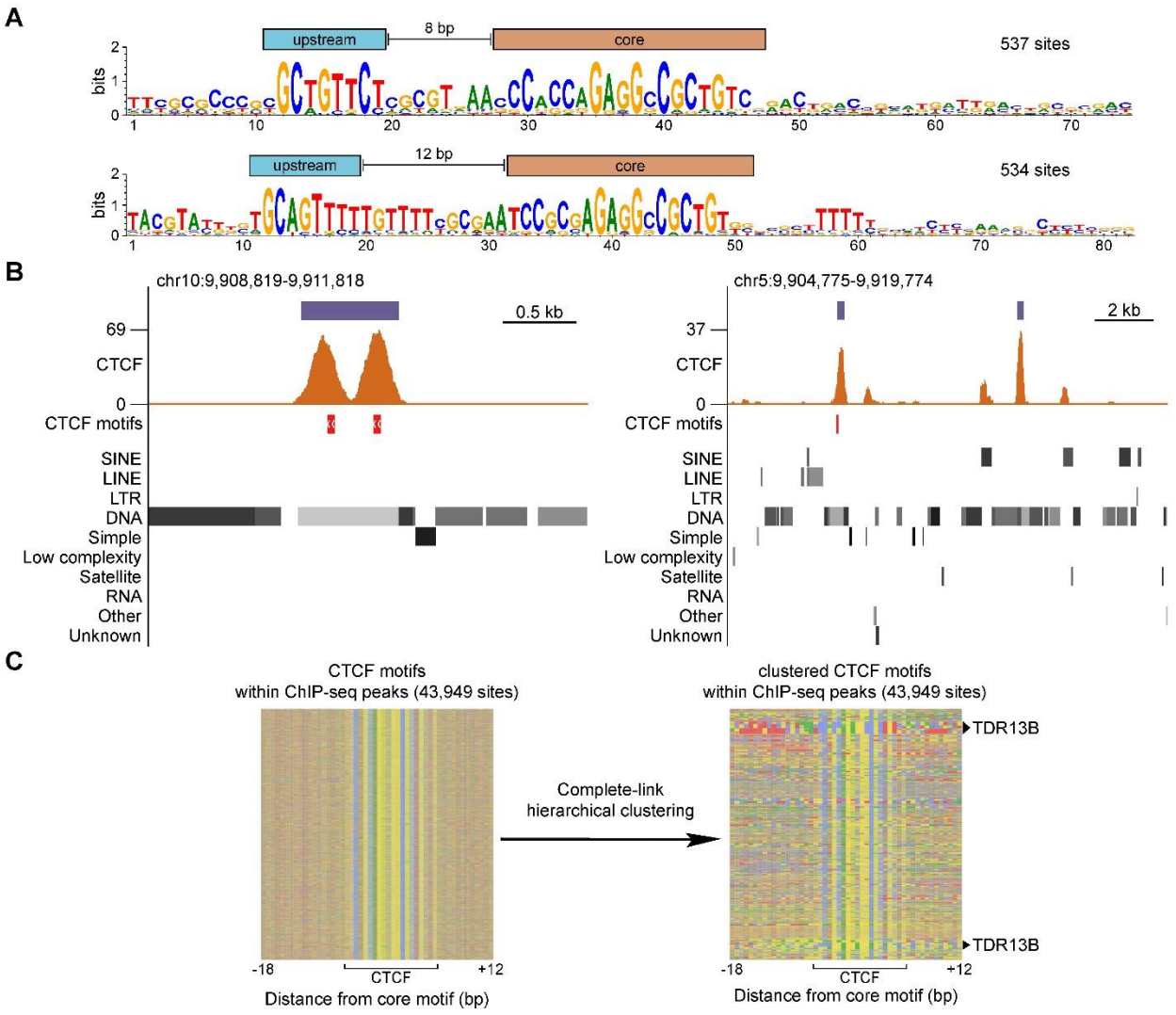


Figure S2. Analysis of CTCF binding sites, Related to Figure 2.

(A) Extended motif logos for the sequences containing matches to the CTCF core and upstream motifs at spacing distances of 8 (top) and 12 bp (bottom). The number of sequences used to generate the logos is shown. Information content of each position on the x axis is expressed in bits on the y axis. (B) Tracks showing examples of CTCF peaks (purple bars) and CTCF motifs (red bars) at DNA transposons. Displayed signal distributions and peaks correspond to the combined analysis of biological replicates. Signal is represented on the y axis as $-\log_{10}$ (p-value) of the CTCF ChIP-seq signal. (C) Left, color chart representing all sequences with CTCF motifs centered on the region matching to the zebrafish CTCF motif. Right, color chart with the same sequences

clustered by edit distances. Clustered sequences showing enrichment on TDR13B repeats (the DNA transposon with the highest enrichment of sites) are indicated with triangles. Nucleotides are represented as follows: A = green, T = red, C = blue, G = yellow.

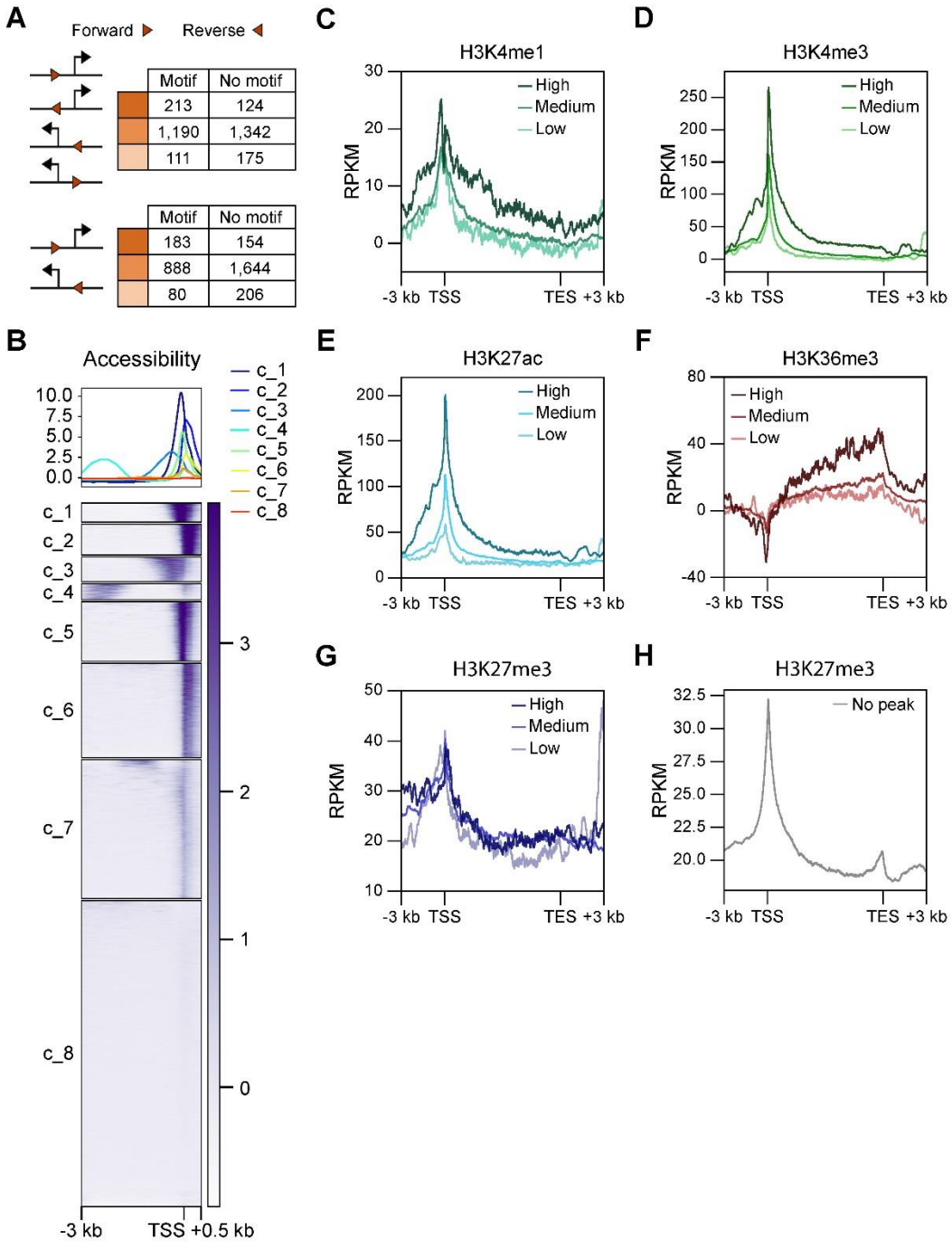


Figure S3. Analysis of promoters with CTCF peaks, Related to Figure 3.

(A) Contingency tables showing the number of gene promoters for each of the three defined categories (color-coded according to ChIP-seq signals, lighter colors represent lower abundances) that contain a CTCF motif irrespective of its orientation (top) and in

the same orientation as transcription (bottom). (B) Profiles and heat maps of ATAC-seq signal in promoters that do not contain a CTCF peak. Promoters were assigned to eight clusters (c_1-8) using k-means. (C-G) Average ChIP-seq signal profiles of (C) H3K4me1, (D) H3K4me3, (E) H3K27ac, (F) H3K36me3 and (G) H3K27me3 over gene bodies and flanking sequence of genes with CTCF bound at promoters. Normalized signal is shown in RPKM (reads per kilobase million). Genes were classified based on CTCF signal strength at promoters as High (strongest peaks), Medium and Low (weakest peaks). (H) Average H3K27me3 ChIP-seq signal profiles over gene bodies and flanking sequence of genes without CTCF peaks at promoters.

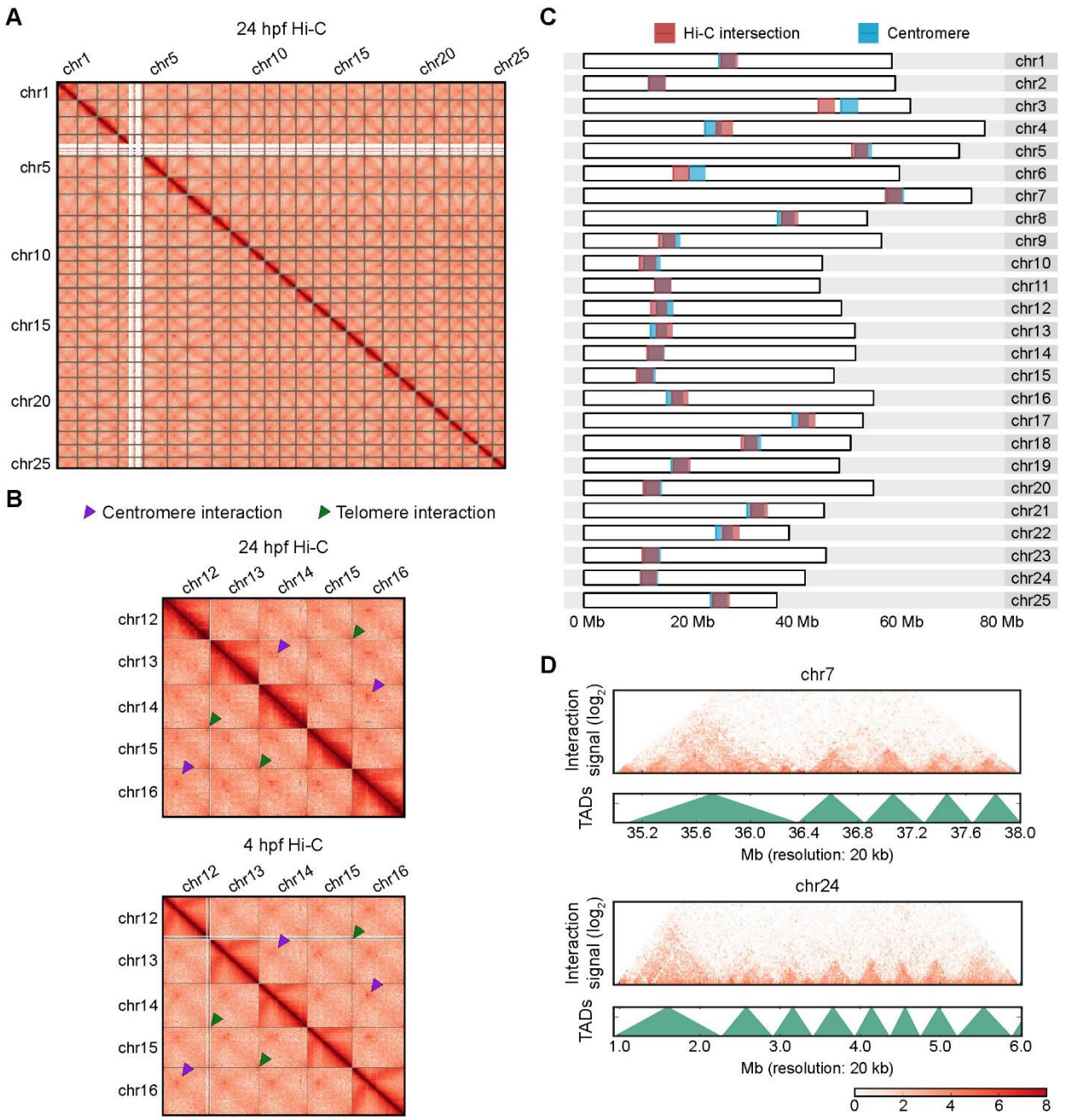


Figure S4. Visualization of zebrafish Hi-C maps shows Rabl configuration of chromosomes, Related to Figure 4.

(A) Hi-C maps of zebrafish at 24 hpf showing intra-chromosomal and inter-chromosomal normalized interaction signal for all chromosomes. (B) Zoom into Hi-C maps showing intra-chromosomal and inter-chromosomal interaction signal of 5 zebrafish chromosomes at 24 hpf (top) and 4 hpf (bottom). Arrowheads point to representative enriched

interactions between centromeres (purple) and telomeres (green). (C) Karyotype of zebrafish chromosomes showing the locations of centromeres (blue) and the intersections between the two strong diagonals visualized on the Hi-C maps (red). (D) Hi-C maps of zebrafish at 24 hpf displaying examples of TADs annotated using the insulation score approach (green) at genomic regions in chromosomes 7 (top) and 24 (bottom).

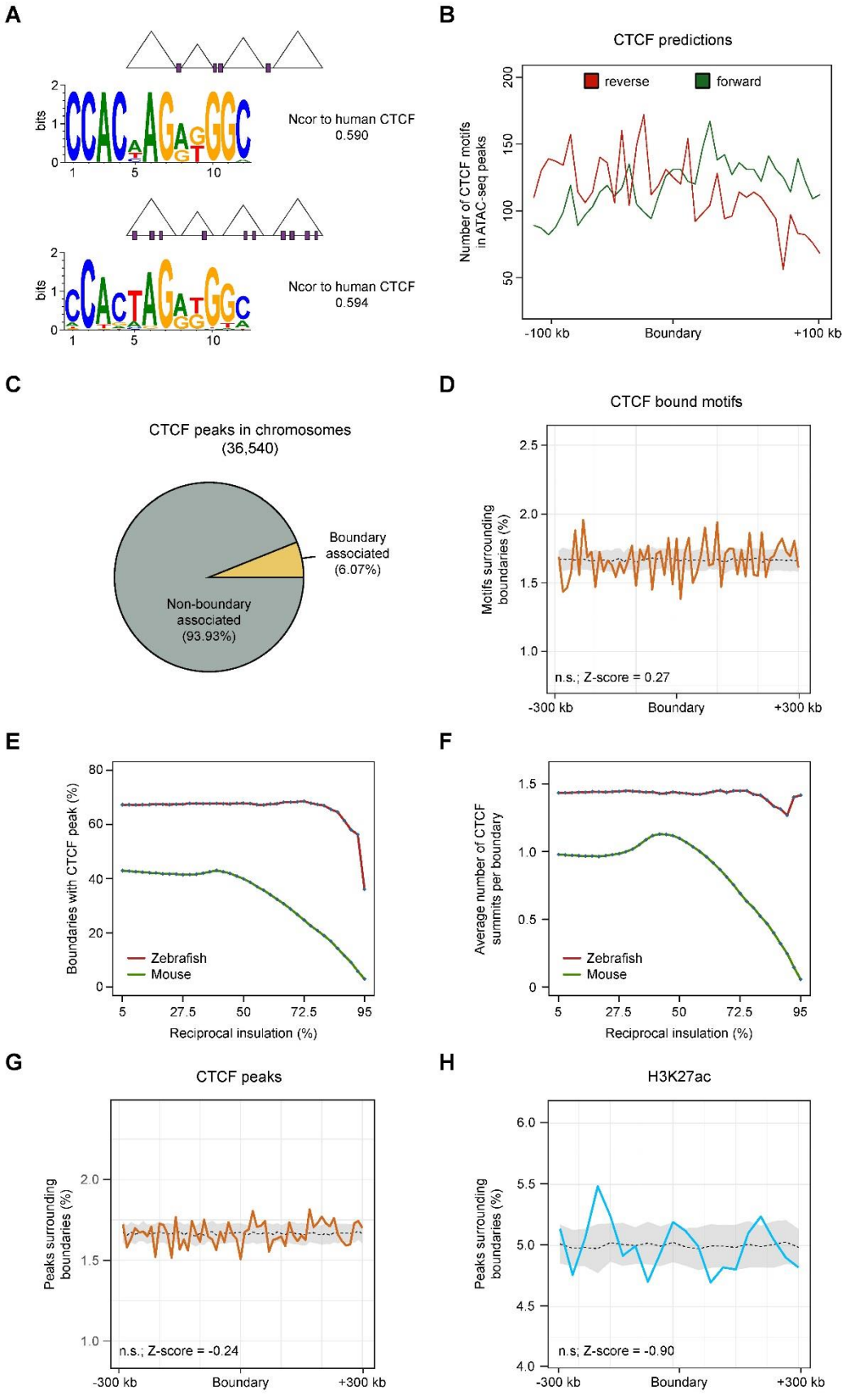

A
bits
Ncor to human CTCF
0.590
Ncor to human CTCF
0.594
B
CTCF predictions
reverse
forward
Number of CTCF motifs
in ATAC-seq peaks
-100 kb
Boundary
+100 kb
C
CTCF peaks in chromosomes
(36,540)
Boundary
associated
(6.07%)
Non-boundary
associated
(93.93%)
D
CTCF bound motifs
Motifs surrounding
boundaries (%)
n.s.; Z-score = 0.27
-300 kb
Boundary
+300 kb
E
Boundaries with CTCF peak (%)
Zebrafish
Mouse
Reciprocal insulation (%)
F
Average number of CTCF
summits per boundary
Zebrafish
Mouse
Reciprocal insulation (%)
G
CTCF peaks
Peaks surrounding
boundaries (%)
n.s.; Z-score = -0.24
-300 kb
Boundary
+300 kb
H
H3K27ac
Peaks surrounding
boundaries (%)
n.s; Z-score = -0.90
-300 kb
Boundary
+300 kb

Figure S5. Relationship between CTCF peaks and domain boundaries, Related to Figure 4.

(A) CTCF motifs enriched at TAD boundaries (top) and within TADs (bottom). Ncor, normalized Pearson correlation. (B) Distribution of predicted CTCF motifs within ATAC-seq peaks present in forward (green) and reverse (red) orientation along 200 kb regions centered on TAD boundaries (x axis). The y axis shows the number of motifs overlapping each bin. (C) Classification of CTCF peaks relative to TAD boundaries. (D) Distribution of CTCF motifs within ChIP-seq peaks (orange) along 600 kb regions centered on TAD boundaries (x axis). The y axis shows the percentage of total peak counts in the 600 kb region located at each genomic position. The dash line represents the mean background distribution and the grey ribbon depicts the ± 1 standard deviation range from the mean. Differences in mean percentages (central 60 kb) were assessed by Z-scores, n.s.: non-significant. (E) Percentage of domain boundaries identified at different reciprocal insulation scores (x axis) that overlap at least one CTCF peak (y axis) in zebrafish (red) and mouse (green). (F) Average number of CTCF peaks (y axis) within boundaries of domains identified at different reciprocal insulation scores (x axis) in zebrafish (red) and mouse (green). (G) Distribution of CTCF peaks (orange) along 600 kb regions centered on domain boundaries identified at 72.5 % reciprocal insulation (x axis) as described for D. (H) Distribution of H3K27ac-enriched peaks along TAD boundaries as described for D.

Data S1: Zebrafish CTCF motif (homer format) and upstream inferred motifs (meme format), Related to Figure 2.

>GCCWGCAGGGGGCGCTGSDG CTCF_zebrafish 3.643980 -616.688268 0
T:1993.0(42.57%),B:7136.0(20.02%),P:1e-267
0.125 0.078 0.550 0.247
0.047 0.576 0.309 0.068
0.046 0.908 0.015 0.031
0.453 0.046 0.171 0.330
0.015 0.108 0.876 0.001
0.061 0.892 0.046 0.001
0.953 0.001 0.015 0.031
0.001 0.001 0.997 0.001
0.123 0.001 0.861 0.015
0.078 0.123 0.732 0.067
0.001 0.001 0.997 0.001
0.015 0.001 0.969 0.015
0.001 0.997 0.001 0.001
0.202 0.015 0.782 0.001
0.001 0.860 0.124 0.015
0.217 0.281 0.062 0.440
0.092 0.092 0.815 0.001
0.123 0.396 0.357 0.124
0.283 0.157 0.219 0.341
0.062 0.219 0.611 0.108

MEME version 4

ALPHABET= ACGT

strands: + -

Background letter frequencies (from unknown source):
A 0.322 C 0.178 G 0.178 T 0.322

MOTIF upstream_near_CTCF_gap_8_orientation_0

letter-probability matrix: alength= 4 w= 9 nsites= 20 E= 0e+0
0.039106 0.584730 0.106145 0.270019
0.007449 0.011173 0.981378 0.000000
0.001862 0.960894 0.014898 0.022346
0.204842 0.011173 0.001862 0.782123
0.013035 0.063315 0.918063 0.005587
0.011173 0.109870 0.011173 0.867784
0.033520 0.011173 0.096834 0.858473
0.020484 0.953445 0.009311 0.016760
0.024209 0.186220 0.007449 0.782123

MEME version 4

ALPHABET= ACGT

strands: + -

Background letter frequencies (from unknown source):
A 0.322 C 0.178 G 0.178 T 0.322

MOTIF upstream_near_CTCF_gap_12_orientation_0

letter-probability matrix: alength= 4 w= 9 nsites= 20 E= 0e+0

0.161049 0.071161 0.031835 0.735955
0.009363 0.018727 0.970037 0.001873
0.000000 0.971910 0.009363 0.018727
0.898876 0.003745 0.009363 0.088015
0.016854 0.076779 0.902622 0.003745
0.014981 0.059925 0.007491 0.917603
0.043071 0.011236 0.095506 0.850187
0.018727 0.207865 0.009363 0.764045
0.031835 0.142322 0.005618 0.820225

Table S1: Data sets generated or analyzed in this study, Related to Figures 1 and 2.

Data	NCBI GEO serie	Sample	Input sample	Approximate fragment size (bp)	Reference
ATAC-seq (24 hpf)	GSE61065	GSM1496130	-	-	Gehrke et al., 2015
H3K4me1 ChIP-seq (24 hpf)	GSE20600	GSM520620, GSM686660	GSM520622, GSM686662	150	Aday et al., 2011
H3K4me3 ChIP-seq (24 hpf)	GSE20600	GSM520621, GSM686661	GSM520622, GSM686662	150	Aday et al., 2011
H3K27ac ChIP-seq (24 hpf)	GSE32483	GSM803832	NA	200	Bogdanović et al., 2012
H3K27me3 ChIP-seq (24 hpf)	GSE35050	GSM861348	NA	200	Irimia et al., 2012
H3K36me3 ChIP-seq (24 hpf)	GSE32880	GSM813752	GSM813756	150	Ulitsky et al., 2011
CTCF ChIP-seq (24 hpf)	GSE133437	GSM3908626, GSM3908627	GSM3908628, GSM3908629	185	This report
Hi-C (4 and 24 hpf)	GSE105013	allValidPairs files	-	-	Kaaij et al., 2018
mouse CTCF ChIP-seq (mESCs)	GSE29184	GSM723015	GSM723020	300	Shen et al., 2012
mouse Hi-C (mESCs)	GSE35156	GSM862720, GSM862721	-	-	Dixon et al., 2012

Table S2: ATAC-seq and ChIP-seq mapping statistics, Related to Figures 1 and 2.

Data	Species	Total reads	Mapped reads	% of mapped reads	Clean reads	% of clean reads
ATAC-seq	zebrafish	179831298	144502903	80.4	77697950	43.2
CTCF ChIP-seq 1	zebrafish	15793189	12861051	81.4	5732512	36.3
CTCF ChIP-seq 2	zebrafish	23303996	18968756	81.4	8700853	37.3
input (CTCF ChIP-seq 1)	zebrafish	30132254	21301146	70.7	12757021	42.3
input (CTCF ChIP-seq 2)	zebrafish	35265251	25364354	71.9	15361366	43.6
CTCF ChIP-seq	mouse	19433603	15591650	80.2	10343560	53.2
input (CTCF ChIP-seq)	mouse	33137002	29424170	88.8	23494267	70.9
H3K4me1 ChIP-seq	zebrafish	54206950	43782567	80.8	27424455	50.6
H3K4me3 ChIP-seq	zebrafish	41419284	34804137	84.0	19159968	46.3
input (H3K4me1 and H3K4me3 ChIP-seq)	zebrafish	44315679	39178459	88.4	24704265	55.7
H3K27ac ChIP-seq	zebrafish	11440822	10804838	94.4	7679223	67.1
H3K27me3 ChIP-seq	zebrafish	11989124	11236840	93.7	7788005	65.0
H3K36me3 ChIP-seq	zebrafish	15019635	12565631	83.7	7814849	52.0
input (H3K36me3 ChIP-seq)	zebrafish	12541421	10338391	82.4	6456010	51.5

Table S3: Zebrafish repeat types enriched with CTCF sites, Related to Figure 2.

Repeat type	Number of repeats with CTCF site	p-value (Z-scores)
TDR13B	2874	0
HATN11_DR	562	0
DNA25TWA1_DR	1310	0
TDR4	334	0
DNA8-3_DR	308	0
DNA-1-3_DR	305	0
LTR1_DR	290	0
hAT-N45_DR	178	6.36E-283
DNA-1-4_DR	258	7.83E-241
Copia-7-I_DR	107	4.94E-201
DNA-1-5_DR	217	4.45E-194
Copia-6-I_DR	128	4.66E-178
DNA-1-3B_DR	235	1.93E-167
HATN3_DR	153	2.00E-154
DNA-8-1_DR	107	7.62E-148
DIRS1a_DR	538	9.87E-144
hAT-N31_DR	94	9.05E-127
Nimb-1_DR	150	1.69E-120
DIRS1_DR	520	1.11E-101
DNA9NNN1_DR	655	1.42E-75
Looper-N8_DR	96	1.10E-69
HATN5_DR	221	2.78E-68
Harbinger-N11_DR	96	1.43E-65
Gypsy-169-I_DR	76	4.31E-53
DIRS-1_DR	114	2.62E-51
DIRS-10_DR	95	1.52E-48
Harbinger-N13_DR	100	1.13E-45
HATN3B_DR	92	1.08E-41
HarbingerN1_DR	217	1.10E-40
LRS_DR	65	2.10E-36
Kolobok-1N1_DR	100	1.38E-32
hAT-N76_DR	100	3.79E-28
DIRS-8_DR	59	6.50E-22
Harbinger-N9_DR	71	2.56E-20
DNA-8-23_DR	57	1.07E-19
SAT-1_DR	57	2.97E-16
DNA-8-9_DR	123	3.62E-06
HATN9_DR	71	3.62E-05
EXPANDER1_DR	110	3.80E-03
hAT-N25_DR	52	5.13E-03

TRANSPARENT METHODS

Generation of the epitope-tagged *ctcf* ^{HPSH} zebrafish allele

The AB zebrafish strain was used to generate the *ctcf* ^{HPSH} allele using a CRISPR-Cas9-based knock-in protocol (Lavalou et al., 2019). The HPSH tag was introduced into the 5'UTR of *ctcf* by using one gRNA (5'-AGGGG GACCG ACTGA GGCCG-3') designed to target 3 bp down-stream of the *ctcf* ATG. A 163-nt single-stranded DNA oligo (ssDNA oligo) thiolated at 5'- and 3'-end nucleotides with 33- and 54-bp homology arms, respectively, flanking both sides of the HPSH tag was designed and manufactured by Ultramer IDT (ssDNA oligo: 5'-TTGTG ATTTT AACCA ATGTA GGTAT TACCC ATGTA CCCTT ACGAC GTGCC TGACT ACGCT CTGGA GGTGC TGTTC CAGGG ACCTC ATCAC CATCA CCACC ACGGA GGCGG AGAAG GGGGA CCGAC TGAGG CCGTG GTGGA AGATG CAGGG GATGC TTTCA AGG-3'; HPSH tag is underlined). One-cell stage embryos were injected with 5 pg GFP-Cas9 protein (kind gift of Jean-Paul Concordet), 166 pg sgRNA, 5 pg ssDNA donor oligo and 0.45 pg morpholino against *xrcc4* (morpholino: 5'-CACTA CTGCT GCGAC ACCTC ATTCC-3'; Gene Tools LLC). All adult fish (female and male) were individually genotyped to verify the integrity of both HPSH-tagged and wild type *ctcf* loci. The presence of the *ctcf* ^{HPSH} sequence was scored by PCR (genotyping primers: forward 5' GGAGA CAGAA AGTGG TCGAG GC 3'; reverse 5' GGCTC CCCAT CTTTA GGCAT GG 3'), the amplified DNA region was subjected to DNA sequencing to confirm tag integration (Figure S1A). Heterozygous *ctcf* ^{HPSH} animals were backcrossed to wild type AB fish for 3 generations before generating homozygous *ctcf* $^{HPSH/HPSH}$ animals; wild type siblings of *ctcf* ^{HPSH} fish with no HPSH-tag insertion were used as controls.

Zebrafish experimental procedures were approved by the ethics committee of the Institut Curie CEEA-IC #118 (project CEEA-IC 2017-017). Zebrafish were maintained according to standard protocols (Westerfield, 2000) that follow the current Directive 2010/63/EU.

Protein extraction and Western blot assays

Protein extracts were isolated from ~45 embryos or 3 male adult brains. Embryos were dechorionated (Pronase, Roche) and deyolked (55 mM NaCl, 1.8 mM KCl, 1.25 mM $NaHCO_3$) prior to cell dissociation. Samples were homogenized in 400 µL of dissociation solution (1 cOmplete tablet, Roche REF 11873580001, in 5 ml of 1X PBS) and spun down at 2,000 rpm for 2 minutes. Protein extracts were obtained by sequential resuspension of pellets using cytoplasmic

(10 mM KCl, 10 mM Hepes (pH 7.9), 3 mM $MgCl_2$, 0.45% Triton X-100, 0.05% Tween-20) and nuclear (400 mM KCl, 10 mM Hepes (pH 7.9), 3 mM $MgCl_2$, 0.45% Triton X-100, 0.05% Tween-20) extract buffers. Proteins were separated on a NuPAGE 4-12% Bis-Tris gel (Life technologies, Lot 17022070). Western blot analysis was performed with antibodies detecting HA epitope (HA.11 Epitope tag antibody, BioLegend, Clone 16B12, Lot B224726) and γ-tubulin (Sigma, Clone GTU-88, Lot #026M4832V).

ChIP-seq library preparation

Two biological replicates were prepared. For each replicate, ~2000 24 hpf *ctcf*^{HPSH} embryos were used. ChIP-seq was performed as previously described in (Pérez-Rico et al., 2017) using anti-HA.11 epitope tag antibody (BioLegend, Clone 16B12, Lot B224726). Purified chromatin was used for single-end library preparation following TruSeq – ChIP-seq Illumina protocol. Libraries were sequenced in a HiSeq 2500 system.

Processing of CTCF ChIP-seq data

Quality of mouse and zebrafish (published and newly generated) ChIP-seq libraries was assessed using *FastQC* (v0.9.3, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), 3' adapters were removed using *cutadapt* v1.3 (Martin, 2011) with the following specifications –O 5 –match-read-wildcards –m 50. Reads were mapped to the danRer10 and mm10 genome versions with *Bowtie 2* version 2.2.5 (Langmead and Salzberg, 2012) using –end-to-end –sensitive parameters. Alignment reports were converted to sorted BAM files using *samtools* version 1.1 (Li et al., 2009) to discard reads with mapping quality lower than 20. Duplicated reads were removed using *MarkDuplicates* from *Picard Tools* (version 1.45, https://broadinstitute.github.io/picard/) with parameters: REMOVE_DUPLICATES = true OPTICAL_DUPLICATE_PIXEL_DISTANCE = 100. Reads from technical replicates were combined with *samtools merge*. Peak calling was performed with *MACS2 callpeak* using the input libraries as control of the ChIP libraries and the following parameters: -g 9.9e8 for zebrafish and 2.3e9 for mouse –keep-dup all –bw [187|183] for zebrafish and 300 for mouse –call-summits (Zhang et al., 2008). Only peaks with q-values < $1x10^{-4}$ within chromosomes were used for further analyses. Pileup tracks were generated using *MACS2* commands *callpeak* (--keep-dup all –B –model –extsize [187|183] –SPMR –g 9.9e8 or 2.3e9) and *bdgcmp* (-m subtract). Bedgraph files were filtered to keep only the signal in chromosomes and converted to bigwig format with *bedGraphToBigWig* version 4 (Kent et al., 2010). Pearson correlation between zebrafish biological duplicates was calculated with *deepTools* version 3.0.1

(Ramírez et al., 2016) *multiBigwigSummary* (--binSize 10000) and *plotCorrelation* (--corMethod pearson –log1p --removeOutliers). Zebrafish biological replicates were combined by first normalizing the signal for each replicate using MACS2 *callpeak* (-B –nomodel –extsize 185 –g 9.9e8 –keep-dup all) and *bdgcmp* (-m ppois). Replicates were combined using the generated bedgraph files as input of *cmbreps* (-m fisher) and peaks identified with *bdgpeakcall* using the bedgraph file with the combined signal (-l 200 –g 100 –c 8). The final bedgraph file was converted to bigwig format with *bedGraphToBigWig*.

Sequence conservation

Peak coordinates were converted from danRer10 to danRer7 using the *liftOver* tool (Karolchik, 2004). Conversion was performed for two peak sets: all peaks and peaks with no overlap to exons or with a fraction of overlap < 40% of the peak size. Ensembl 91 gene annotations of zebrafish (http://dec2017.archive.ensembl.org/Danio_rerio/Info/Index) were used as reference for exon coordinates. Control regions were obtained with *bedtools shuffle* (Quinlan and Hall, 2010) and different exclusion regions for all peaks (-excl danRer7_gaps.bed –chrom -noOverlapping) and the non-exonic peaks (-excl danRer7_gaps_exons.bed –chrom -noOverlapping). Profiles of conservation were generated using the 8 vertebrate NCBI PhastCons track of danRer7 as reference (Siepel et al., 2005), *deepTools computeMatrix* (-a 3000 –b 3000 –averageTypeBins mean –referencePoint center –missingDataAsZero) and *plotProfile* programs. Significance of the observed differences between the distributions of CTCF peaks and control regions were assessed by two-sided Kolmogorov-Smirnov tests performed with *ks.test* in *R* version 3.3.0 (R Development Core Team 2008).

Processing of histone ChIP-seq and ATAC-seq data

All zebrafish ChIP-seq and ATAC-seq data sets analyzed in this study correspond to 24 hpf of zebrafish development; all data sets analyzed in this study and their origin are listed in Table S1. The quality of these libraries was analyzed with *FastQC*. Only the ATAC-seq data showed biases in sequence content and therefore, *fastx_trimmer* (FASTX Toolkit 0.0.13, http://hannonlab.cshl.edu/fastx_toolkit/) was used to remove the first 15 bp of all reads (-f 15 –Q 33). For all libraries, reads were mapped to the danRer10 version of the genome with *Bowtie 2*, low quality (MQ < 20) alignments were filtered out and duplicates were removed with *MarkDuplicates* as described for CTCF ChIP-seq data. The only difference in mapping was for the ATAC-seq library that was mapped in paired-end mode with –very-sensitive parameters.

Mapping statistics for all ChIP-seq and ATAC-seq datasets are shown in Table S2. BAM files of ChIP-seq biological replicates were merged using *samtools merge*. Peak calling of ChIP-seq libraries was performed with *SICER* v1.1 (Zang et al., 2009) using the following parameters: redundancy threshold = 2, window size = 200, fragment size = [150|200], effective genome fraction = 0.7, gap size = 600. For ChIP-seq libraries with input controls, false discovery rate (FDR) was set to $1x10^{-5}$, whereas for libraries without controls, an E-value equal to 10 was used. Pileup tracks of ChIP-seq libraries were generated with *bamCompare* (--binSize 10 –effectiveGenomeSize 938313030 –normalizeUsing RPKM –ignoreForNormalization chrM –skipNonCoveredRegions –extendReads [150|200]). For those libraries with input controls, input signal was subtracted from the ChIP signal (--operation subtract). Peak calling of the ATAC-seq data was done using *MACS2 callpeak* (-f BAMPE –keep-dup all –B –SPMR –g 9.3e8 –call-summits), and only peaks with q-values < $1x10^{-4}$ were used for further analyses. The pileup track of the ATAC-seq library was obtained using the bedgraph files generated by *callpeak*, including the control signal generated, in combination with *bdgcmp* (-m subtract) and *bedGraphToBigWig*.

Heat maps over CTCF peaks

Heat maps were generated using normalized H3K4me1, H3K4me3, H3K27ac, H3K27me3 and ATAC-seq signal over CTCF peak regions centered in the summit position and ranked by decreasing CTCF ChIP-seq signal. Average signal was calculated with *computeMatrix* (-a 3000 –b 3000 –referencePoint center –sortRegions keep –averageTypeBins mean --misingDataAsZero) and plotted with *plotHeatmap* (--zMin 0 0 0 0 -1 –zMax 30 60 80 60 3.5 –sortRegions keep).

Motif analyses

The central core motif of the zebrafish CTCF matrix was first identified using all peak summit positions (summit ± 100 bp) as input of the *HOMER* v4.10.1 (Heinz et al., 2010) program f*indMotifsGenome.pl* (danRer10 -size -100,100 -len 8,10,12 -S 25 -mis 2 -cpg). The motif was then optimized with the same program (danRer10r -opt motif -size given -len 20 -mis 4 -cpg) using as reference sequences identified as centrally enriched on CTCF and CTCFL (HOCOMOCO v11 Full) (Kulakovskiy et al., 2018) by the *Centrimo* program (MEME Suite 5.0.1 patch 1) (Bailey and MacHanick, 2012; Bailey et al., 2009). CTCF matrix models of zebrafish, human and *Drosophila* from JASPAR 2018 (Khan et al., 2018) were clustered using the RSAT program *matrix-clustering* (Castro-Mondragon et al., 2017; Nguyen et al., 2018) with the following clustering options: metric

for similarity = Ncor, agglomeration rule for hierarchical clustering = complete, merge matrices = sum. Binding sites within the CTCF ChIP-seq peaks were identified using *matrix-scan* (Turatsinze et al., 2008) with the zebrafish CTCF matrix (pseudo counts = 1) and an organism specific background model (GRCz10, upstream-noorf, pseudo-frequencies = 0.01). Sequences were scanned on both strands to report individual matches using the end of the sequence as origin. Only the first-rank matches and those with p-value ≤ $1x10^{-5}$ were used for further analyses. *Spamo* (Whitington et al., 2011) was used to identify significant spacings between the zebrafish CTCF motif and a previously identified upstream motif of CTCF in mammals (downloaded from CTCFBSDB (Ziebarth et al., 2013) and converted to meme format with *transfac2meme*). The sequences used as input for *spamo* were the sequences with binding sites identified with *matrix-scan* and extended in both directions to reach 350 bp long sequences (upstream - 168 bp, motif - 20 bp, downstream - 162 bp). *Spamo* was run with the following options: -numgen 1 -minscore 4 –dumpseqs -shared 0.8 -bgfile background_residues.txt (containing the nucleotide probabilities reported by *matrix-scan*; A 0.32028, C 0.17713, G 0.17903, T 0.32356). Clustering of sequences shown in Figure S2C was performed using extended binding site sequences identified by *matrix-scan* too, but restricted to a shorter central region of 50 bp. Distances between sequences were calculated using the Levenshtein method from the function *stringdistmatrix* of the *stringdist* package version 0.9.5.1 (https://cran.r-project.org/web/packages/stringdist) in *R* version (3.4.4). Sequences were clustered using calculated distances and complete linkage hierarchical clustering in *R* with the *hclust* function. All logos were generated using *WebLogo* version 3.6.0 (Crooks et al., 2004).

Enrichment of CTCF sites in repeat regions

The repeatMasker danRer10 track of NCBI was used (Casper et al., 2018) to assess enrichment of CTCF sites in repeats. Only repeat types with more than 50 repeats overlapped by extended CTCF binding site sequences used for clustering were tested for significance of the overlaps. Identification of overlapping regions between motif sequences and each repeat type was done with *bedtools intersect* (-f 0.76 to consider only those repeat annotations that overlapped at least 75% of the sequence) to calculate the fraction of total repeats with overlap. The same analysis was repeated 100 times using control regions (*bedtools shuffle* with the CTCF binding site extended coordinates and -chrom –noOverlapping options) to generate expected values of randomly distributed regions and assess the significance of the overlap with CTCF sites using Z-scores (Table S3).

Distribution of CTCF peaks in the genome

CTCF peaks were assigned to exonic, intronic, intergenic and promoter (2 kb upstream of TSSs) categories based on the location of their summit. Considering that genomic regions can be annotated as promoters, exons or introns of different transcriptional units, peak assignment was performed sequentially to first identify those overlapping promoters, then exons and finally introns. Genomic overlaps were identified with *bedtools intersect* using the Ensembl 90 gene annotations of zebrafish (http://aug2017.archive.ensembl.org/Danio_rerio/Info/Index). Those peaks that do not overlap with gene bodies or promoters were assigned as intergenic. The pie chart shown in Figure 3A was generated with *R pie* function.

Gene expression analyses

Expression levels of genes were obtained from RNA-seq data deposited in the Expression Atlas of zebrafish (White et al., 2017). Normalized counts per gene (TPM) for each developmental stage were downloaded from https://www.ebi.ac.uk/gxa/experiments/E-ERAD-475/Downloads. Only data from the "pharyngula prim-5" developmental stage were used in this study. Promoters overlapping summits of CTCF peaks were classified in three categories based on the *MACS2* scores of the CTCF peak. 'Low', 'medium' and 'high' categories corresponded to the following ppois values (Poison p-values): $x \leq 219$ (10th quantile, relative to all peaks), $219 > x < 948.2$, and $x \geq 948.2$ (90th quantile, relative to all peaks). Gene IDs of the promoters were used to filter the three categories and ensure that a given gene was assigned to only one category based on its CTCF peak with the highest score. TPM values for each category, including those genes with no CTCF peak at the promoter, were retrieved from the Expression Atlas table and used to generate the box plot shown in Figure 3C. Analyses of significance of detected differences were carried out using two-sided Wilcoxon rank-sum tests in *R* (*wilcox.test*).

Histone and ATAC-seq profiles over gene bodies

Gene IDs of the three gene categories based on *MACS2* scores were used to extract genomic coordinates using Ensembl 90 annotations. Average CTCF and ATAC-seq signal over promoters was obtained with *computeMatrix reference-point* (-a 3000 –b 3000 –averageTypeBins mean –binSize 10), while *computeMatrix scale-regions* (-a 3000 –b 3000 –averageTypeBins mean –regionBodyLength 8000 --missingDataAsZero) was used for histone ChIP-seq signal. All profiles were plotted with *plotProfile*. Differences in the distribution were tested for significance with two-

sided Kolmogorov-Smirnov tests. ATAC-seq profile and heat map of genes without CTCF-bound promoters was performed with *computeMatrix reference-point* (-a 500 –b 3000 –averageTypeBins mean) and *plotHeatmap* (--kmeans 8).

Generation of Hi-C maps

Normalized Hi-C maps of 4 and 24 hpf zebrafish embryos were generated using the valid pairs reported by Kaaij et al. 2018 and deposited at the NCBI GEO data base. Files with valid pairs were used as input for *HiC-Pro* version 2.9.0 (Servant et al., 2015) to build contact maps at two different resolutions (1 Mb and 20 kb) and one iteration of ICE normalization was performed on those matrices. Contact maps were visualized with *HiCPlotter* version 0.6.6 (Akdemir and Chin, 2015) using the following parameters for whole genome maps: -tri 1 –wg 1 –r 1000000 –chr chr25 –hmc 1 –dpi 500. Hi-C maps showing examples of annotated TADs were also generated with HiCPlotter, but using the 20 kb normalized matrix of 24 hpf embryos. Mouse Hi-C analyses were done using biological duplicates of embryonic stem cells (Dixon et al., 2012). Raw data sets were processed with *HiC-Pro* to map reads, filter read pairs and generate normalized contact matrices by merging the valid pairs from both replicates at 20 kb resolution.

Annotation of centromeres

Centromeres were annotated based on enrichment of Type I transposable elements (TEs). The danRer10 nestedRepeats track from NCBI was downloaded and coordinates of Type I TEs (LINE|SINE|LTR) saved as a bed file. This bed file was used to generate a coverage track using TE counts per base and then calculating the average counts over 3 kb windows using *igvtools count* (version 2.3.57, http://www.broadinstitute.org/igv). The coverage track was converted to bigwig format with *wigToBigWig* and used to calculate the enrichment of Type I TEs over 3 Mb overlapping windows with a shift of 1 kb by *bigWigAverageOverBed*. Finally, the window with the highest enrichment for each chromosome was annotated as the centromere. Only chromosome 3 and chromosome 4 showed several prominent peaks and the selection of the centromere was performed using an average score including zero or discarding scores higher than 0.38 (values corresponding to the long arm that is known to be enriched on repeats) for chromosome 3 and 4, respectively.

Identification of TADs

Zebrafish and mouse TADs were annotated using perl scripts of the cworld::dekker module version 1.01 (https://github.com/dekkerlab/cworld-dekker) and the contact maps generated with *HiC-Pro* at 20 kb resolution. First, insulation was calculated with *matrix2insulation.pl* indicating the following parameters: --is 400000 --ss 80000 --ids 240000 --bmoe 0 --nt 0.1. Second, to obtain TAD coordinates, the calculated insulation scores and boundaries identified were used as input of *insulation2tads.pl*, indicating a value of 0 for the option –mts. Given that the danRer10 genome has evident misassembled regions, as indicated by the Hi-C maps, all TADs annotated within those regions that could have hampered the annotation (mainly telomeric, Table S4) were discarded. A total of 1,307 TADs were used for further analyses.

Identification of hierarchical domains

Hierarchical domains of zebrafish were annotated at 20 kb resolution using the reciprocal insulation score method (Zhan et al., 2017) implemented in CaTCH version 1.1 in R version 3.4.4. Boundaries of domains identified at 37 reciprocal insulation scores were analyzed to test for enrichment of CTCF. Boundaries for each set of domains were extended to obtain 60 kb regions and calculate the percentage of boundaries overlapping CTCF peak summits and the average number of CTCF summits within each boundary using *bedtools intersect* and *bedtools coverage*.

Enrichments at TAD boundaries

TAD boundaries and boundaries of domains identified at 72.5% reciprocal insulation were defined as the 20 kb regions located at both ends of domains. CTCF ChIP-seq peaks were categorized as associated or non-associated with boundaries according to the location of their summit relative to extended boundaries (± 20 kb) accounting for uncertainty in the exact definition of boundaries. CTCF and histone mark enrichments over boundaries were assessed by using the center of boundaries as reference point to extend regions on both sides (± 300 kb). These extended regions were divided in 10 kb or 30 kb sized bins for CTCF and histone mark analyses, respectively, and significant peaks identified by MACS and SICER were used to compute the number of overlapping peaks per bin for each boundary using *bedtools coverage* (-count option). These analysis generated matrices in which each row represents a boundary region and each column one of the bins. Total counts were obtained from each column to calculate the percentage of peaks overlapping each bin and plot their distribution. The same strategy was followed to assess

enrichment of predicted and *in vivo* identified CTCF motifs at boundaries using 10 kb sized bins. This strategy was also used to calculate the distribution of forward and reverse CTCF motifs along the ± 100 kb region around boundaries using 5 kb sized bins. Background controls were generated for each enrichment analysis by randomly distributing peaks in the same chromosome using *bedtools shuffle* (–chrom –noOverlapping options) and calculating percentage distributions as described above. This process was repeated 100 times to obtain the mean and standard deviation of expected distributions.

Data and Code Availability

CTCF ChIP-seq sequencing data generated in this study are available in the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/). The accession number for the sequencing data reported in this paper is under accession NCBI GEO: GSE133437. No previously unreported algorithms were used to generate the results.

SUPPLEMENTAL REFERENCES

Akdemir, K.C., and Chin, L. (2015). HiCPlotter integrates genomic data with interaction matrices. Genome Biol. *16*, 1–8.

Bailey, T.L., and MacHanick, P. (2012). Inferring direct DNA binding from ChIP-seq. Nucleic Acids Res. *40*, 1–10.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME Suite: Tools for motif discovery and searching. Nucleic Acids Res. *37*, 202–208.

Casper, J., Zweig, A.S., Villarreal, C., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Karolchik, D., et al. (2018). The UCSC Genome Browser database: 2018 update. Nucleic Acids Res. *46*, D762–D769.

Castro-Mondragon, J.A., Jaeger, S., Thieffry, D., Thomas-Chollier, M., and Van Helden, J. (2017). RSAT matrix-clustering: Dynamic exploration and redundancy reduction of transcription factor binding motif collections. Nucleic Acids Res. *45*, 1–13.

Crooks, G., Hon, G., Chandonia, J., and Brenner, S. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188–1190.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589.

Karolchik, D. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, 493D – 496.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: Enabling browsing of large distributed datasets. Bioinformatics *26*, 2204–2207.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. *46*, D260–D266.

Kulakovskiy, I. V., Vorontsov, I.E., Yevshin, I.S., Sharipov, R.N., Fedorova, A.D., Rumynskiy, E.I., Medvedeva, Y.A., Magana-Mora, A., Bajic, V.B., Papatsenko, D.A., et al. (2018). HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. Nucleic Acids Res. *46*, D252–D259.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lavalou, P., Eckert, H., Damy, L., Constanty, F., Majello, S., Bitetti, A., Graindorge, A., and Shkumatava, A. (2019). Strategies for genetic inactivation of long noncoding RNAs in zebrafish. RNA *25*, 897–904.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and

Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10.

Nguyen, N.T.T., Contreras-Moreira, B., Castro-Mondragon, J.A., Santana-Garcia, W., Ossio, R., Robles-Espinoza, C.D., Bahin, M., Collombet, S., Vincens, P., Thieffry, D., et al. (2018). RSAT 2018: Regulatory sequence analysis tools 20th anniversary. Nucleic Acids Res. *46*, W209–W214.

Pérez-Rico, Y.A., Boeva, V., Mallory, A.C., Bitetti, A., Majello, S., Barillot, E., and Shkumatava, A. (2017). Comparative analyses of super-enhancers reveal conserved elements in vertebrate genomes. Genome Res. *27*, 259–268.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. *16*, 259.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.D.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. *15*, 1034–1050.

Turatsinze, J.V., Thomas-Chollier, M., Defrance, M., and van Helden, J. (2008). Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules. Nat. Protoc. *3*, 1578–1588.

Westerfield, M. (2000). The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). University of Oregon Press.

White, R.J., Collins, J.E., Sealy, I.M., Wali, N., Dooley, C.M., Digby, Z., Stemple, D.L., Murphy, D.N., Billis, K., Hourlier, T., et al. (2017). A high-resolution mRNA expression time course of embryonic development in zebrafish. Elife *6*, 1–32.

Whitington, T., Frith, M.C., Johnson, J., and Bailey, T.L. (2011). Inferring transcription factor complexes from ChIP-seq data. Nucleic Acids Res. *39*, e98–e98.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics *25*, 1952–1958.

Zhan, Y., Mariani, L., Barozzi, I., Schulz, E.G., Blüthgen, N., Stadler, M., Tiana, G., and Giorgetti, L. (2017). Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. Genome Res. *27*, 479–490.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Ziebarth, J.D., Bhattacharya, A., and Cui, Y. (2013). CTCFBSDB 2.0: A database for CTCF-binding sites and genome organization. Nucleic Acids Res. *41*, 188–194.