Preformed chromatin topology assists transcriptional robustness of *Shh* during limb development

Christina Paliou[a,b], Philine Guckelberger[a], Robert Schöpflin[a,c], Verena Heinrich[d], Andrea Esposito[e,f,g], Andrea M. Chiariello[e,f], Simona Bianco[e,f], Carlo Annunziatella[e,f], Johannes Helmuth[h,1], Stefan Haas[d], Ivana Jerković[a,2], Norbert Brieske[a], Lars Wittler[i], Bernd Timmermann[j], Mario Nicodemi[e,f,g], Martin Vingron[d], Stefan Mundlos[a,b,c,3], and Guillaume Andrey[a,b,3,4]

[a]Research Group Development & Disease, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany; [b]Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité Universitätsmedizin Berlin, 13353 Berlin, Germany; [c]Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin, 13353 Berlin, Germany; [d]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany; [e]Dipartimento di Fisica, Università di Napoli Federico II, 80126 Naples, Italy; [f]Istituto Nazionale di Fisica Nucleare (INFN) Napoli, Complesso Universitario di Monte Sant'Angelo, 80126 Naples, Italy; [g]Berlin Institute of Health (BIH), Max Delbrück Center-Berlin, 13125 Berlin, Germany; [h]Otto-Warburg-Laboratory: Epigenomics, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany; [i]Department of Developmental Genetics, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany; and [j]Sequencing Core Facility, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany

Edited by Robb Krumlauf, Stowers Institute for Medical Research, Kansas City, MO, and approved May 2, 2019 (received for review January 26, 2019)

Long-range gene regulation involves physical proximity between enhancers and promoters to generate precise patterns of gene expression in space and time. However, in some cases, proximity coincides with gene activation, whereas, in others, preformed topologies already exist before activation. In this study, we investigate the preformed configuration underlying the regulation of the *Shh* gene by its unique limb enhancer, the *ZRS*, in vivo during mouse development. Abrogating the constitutive transcription covering the *ZRS* region led to a shift within the *Shh–ZRS* contacts and a moderate reduction in *Shh* transcription. Deletion of the CTCF binding sites around the *ZRS* resulted in the loss of the *Shh–ZRS* preformed interaction and a 50% decrease in *Shh* expression but no phenotype, suggesting an additional, CTCF-independent mechanism of promoter–enhancer communication. This residual activity, however, was diminished by combining the loss of CTCF binding with a hypomorphic *ZRS* allele, resulting in severe *Shh* loss of function and digit agenesis. Our results indicate that the preformed chromatin structure of the *Shh* locus is sustained by multiple components and acts to reinforce enhancer–promoter communication for robust transcription.

genetics | 3D genome | gene regulation | development | limb

During development, precise spatiotemporal gene-expression patterns are established by regulatory regions called enhancers. The specificity of enhancers is instructed by the combination of bound transcription factors and their transcriptional activities are transmitted to associated gene promoters. The communication between enhancers and promoters is ensured by physical proximity in the nucleus, even when separated by a large genomic distance. This communication is delimited by domains of preferential interactions called topologically associating domains (TADs) (1, 2). TADs are separated by boundary elements, which interact together and have been associated with convergent CTCF/Cohesin binding and constitutive transcription (1–6). Depletion of CTCF or Cohesin induces a genome-wide loss of TADs and boundary interactions but results in only modest gene expression changes, bringing into question the functional relevance of these structures (7–10). Moreover, the role of transcription at boundary regions is yet to be elucidated.

Although TADs are mostly invariant, intra-TAD interactions between regulatory elements can be dynamic or preformed (11, 12). In particular, dynamic enhancer–promoter interactions occur in a tissue- or time-specific manner and participate in the regulation of gene transcription (13–17). By contrast, preformed interactions are detected before gene transcriptional activation and are tissue-invariant. These interactions can occur between regulatory regions located within TADs, but also in the direct vicinity of boundary elements. Preformed interactions have been associated with various *trans*-acting factors including CTCF, paused polymerase, constitutive transcription, and the Polycomb complex (11, 18–20). Functionally, preformed topologies have been postulated to enable more efficient and robust gene activation by ensuring rapid communication between enhancers and their target promoters (21, 22). However, as the activities of enhancers and their capacity to interact with their target promoter often overlap, the sole contribution of preformed topologies to gene activation remains speculative.

In this work, by using the *Shh* locus as testbed, we set out to study the function of a stable, preformed topology. *Shh* is expressed in the posterior part of the developing limb, within the zone of polarizing activity (ZPA). This highly specific expression pattern is critical to ensure the development of limb extremities

Significance

Enhancers control the activation of genes in space and time through physical proximity in the nucleus. Recent studies have shown that the control of this proximity can modulate gene expression. On the contrary, it is unclear how constant proximity participates in transcriptional regulation. Here, by using the *Shh* locus as a testbed, we set out to investigate how the preformed interaction with its enhancer, the *ZRS*, contributes to its regulation. We find that the alteration of CTCF binding or constitutive transcription around the *ZRS* alters this interaction as well as the expression of *Shh*. Overall, we demonstrate the significance of preformed promoter–enhancer interactions to achieve maximal expression level and therefore transcriptional robustness.

Author contributions: C.P., S.M., and G.A. designed research; C.P., P.G., I.J., N.B., L.W., B.T., and G.A. performed research; A.E., A.M.C., S.B., C.A., and M.N. contributed new reagents/analytic tools; C.P., R.S., V.H., J.H., S.H., M.V., and G.A. analyzed data; and C.P., S.M., and G.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE123388).

[1]Present address: Humangenetik, Labor Berlin–Charité Vivantes, 13353 Berlin, Germany.

[2]Present address: Institute of Human Genetics, Department of Genome Dynamics, Group Chromatin and Cell Biology, UMR 9002 CNRS, Université de Montpellier 34396 Montpellier, France.

[3]To whom correspondence may be addressed. Email: mundlos@molgen.mpg.de or guillaume.andrey@unige.ch.

[4]Present address: Department of Genetic Medicine and Development, Medical School, University of Geneva, 1211 Geneva, Switzerland.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1900672116/-/DCSupplemental.

Published online May 30, 2019.

and, in particular, the number and identity of digits (23). In the limb bud, *Shh* is regulated by a single enhancer, the *ZRS*, the deletion of which results in a complete *Shh* loss of function in the limb, leading to digit aplasia (24). The *ZRS* is located almost 1 Mb away from the *Shh* promoter within the intron 5 of the constitutively expressed gene, *Lmbr1* (24, 25). Despite this large genomic separation, FISH experiments have demonstrated complete colocalization of the *Shh* promoter and the *ZRS* in posterior limb buds, where *Shh* is expressed. Moreover, in contradiction to many enhancer–promoter interactions that are tissue- and time-specific, the two elements are found in close proximity even when inactive, suggesting a preformed mode of interaction (26, 27). However, how this preformed topology is established or how it relates to the expression of *Shh* in developing limb buds remains unclear. In this work, we address these questions by interrogating the *Shh* locus structure with high-resolution capture-HiC (cHi-C) using targeted genetic disruption of predicted CTCF/transcriptional architectural features.

Results

The *Shh* Regulatory Domain Architecture Is Tissue-Invariant. To investigate whether the 3D architecture of the *Shh* locus changes depending on ongoing tissue-specific regulation, we produced cHi-C maps of three different tissues/cell types: embryonic stem cells (ESCs), E10.5 midbrain, and E10.5 embryonic limb buds. In ESCs, the *Shh* locus is in a poised transcriptional state, whereas, in limbs and midbrain, it is found in an active state under the control of different enhancers (28). Independently of the locus' transcriptional state, we observed a conserved 1-Mb-sized TAD, defined by a centromeric boundary located close to the *Shh* gene and a telomeric boundary positioned around the TSS of the *Lmbr1* gene in the vicinity of the *ZRS* enhancer (Fig. 1 *A*–*C*). Moreover, the interaction between *Shh* and the *ZRS* remains unchanged among the three tissues (Fig. 1*D*).

Next, we investigated the chromatin accessibility and the H3K27ac active enhancer mark at the locus in all three tissues by using ATAC-seq and ChIP-seq, respectively (11, 29–31) (Fig. 1*E*). In agreement with enhancer reporter assays demonstrating

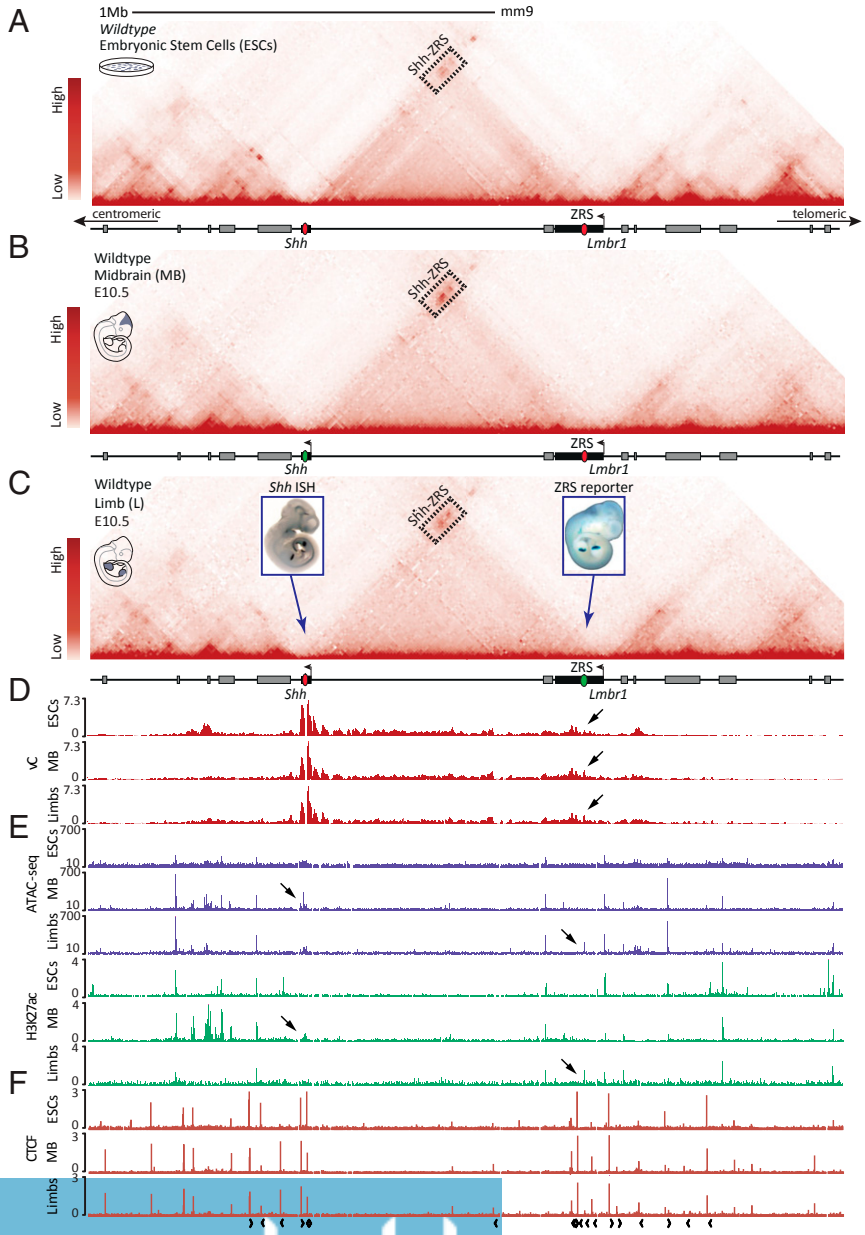

Fig. 1. *Shh* and the *ZRS* enhancer form a tissue-invariant chromatin interaction. (*A*) cHi-C map of the extended *Shh* locus in ESCs. Midbrain and limb inactive enhancers are indicated by red ovals on the lower gene track. (*B*) cHi-C map in E10.5 midbrain. The inactive *ZRS* enhancer is indicated by a red oval, and the active midbrain enhancer (*SBE1*) in the transcribed *Shh* gene is indicated by a green oval. (*C*) cHi-C map in E10.5 limb buds (forelimbs and hindlimbs). The active limb enhancer (*ZRS*) is indicated by a green oval, and the inactive midbrain enhancer is indicated by a red oval. The left embryo staining shows a wildtype WISH of *Shh* at E10.5. The right embryo displays LacZ staining for the *ZRS* enhancer activity. Reprinted from ref. 65. The black boxes in *A*–*C* indicate the domain of high interaction between *Shh* and the *ZRS* region. *Lmbr1* and *Shh* genes in *A*–*C* are indicated as black bars, whereas neighboring genes are colored gray. (*D*) Virtual capture-C (vC) profiles from the *Shh* promoter in ESCs, midbrain, and limbs show similar interactions with the *ZRS* (black arrows). (*E*) ATAC-seq (purple) and H3K27ac ChIP-seq (green) tracks in ESCs, midbrain, and limbs. Black arrows indicate the active limb and midbrain enhancers. (*F*) CTCF ChIP-seq tracks of the extended *Shh* locus in ESCs, midbrain, and limbs. Note the absence of changes between the three tissues. Black arrows under the ChIP-seq track indicate the orientation of CTCF sites.

a limb-restricted activity, the *ZRS* is accessible and decorated with H3K27ac exclusively in the limb tissue (25). By contrast, several other regions, including the known enhancer *SBE1* in intron 2 of the *Shh* gene, were accessible and marked by H3K27ac specifically in the midbrain, suggesting that they drive *Shh* expression only in this tissue (32) (Fig. 1*E*). Similarly to previous studies in the limb (26, 27), we concluded that the 3D chromatin topology of the *Shh* locus is preformed in ESCs and invariable among the tested tissues, despite the changes in the enhancer repertoires that are being used.

Preformed stable 3D chromatin interactions and TAD boundaries have been proposed to rely on the presence of convergent CTCF binding sites and the Cohesin complex (3, 11). At the *Shh* locus, the centromeric and telomeric TAD boundaries are bound by two and three major tissue-invariant convergent CTCF binding sites, respectively (Fig. 1*F*) (29, 31). Additionally, constitutive transcription has also been observed to correlate with stable chromatin topologies (1, 22). Here, the *Lmbr1* gene, which overlaps the *ZRS* and colocalizes with the telomeric TAD boundary, is expressed across most tissues and cell types. We thereby hypothesized that the disruption of the CTCF sites or of the *Lmbr1* transcription would impact the preformed chromatin interaction between *Shh* and the *ZRS* as well as the overall TAD architecture of the locus.

Loss of *Lmbr1* Transcription Results in Down-Regulation of *Shh*. The preformed chromatin interaction between *Shh* and the *ZRS* involves the constitutively transcribed *Lmbr1* gene (Fig. 1). To test whether the transcription of *Lmbr1* participates in the establishment of this long-range interaction as well as to the *ZRS* activity, we engineered a homozygous 2.3-kb deletion of its promoter ($Lmbr1^{\Delta prom/\Delta prom}$) by using the CRISPR-Cas9 system (33). By using RNA-seq and ChIP-seq for the transcriptional elongation mark H3K36me3, we could observe a complete loss of *Lmbr1* transcription in mutant tissues (Fig. 2*A*). By using qRT-PCR, from somite-staged E10.5 embryos, we then quantified the effect of the mutation on *Shh* transcription and found a 20% decrease in *Shh* expression ($P = 0.0063$; Fig. 2*B*). However, this mild loss of expression did not result in any obvious skeletal phenotypes in E18.5 limb buds.

To determine whether the loss of *Lmbr1* transcription disrupts the *ZRS* activity, we compared ATAC-seq and H3K27ac ChIP-seq profiles between wildtype and $Lmbr1^{\Delta prom/\Delta prom}$ limb buds and could not observe any significant changes. Thus, we concluded that the activity of the *ZRS* appears unaffected by the loss of *Lmbr1* transcription (Fig. 2*B*). Next, we examined whether the loss of *Lmbr1* transcription results in altered 3D chromatin architecture of the locus. CHi-C maps were generated for wildtype and $Lmbr1^{\Delta prom/\Delta prom}$ limb buds, showing no global changes in the domain of $Lmbr1^{\Delta prom/\Delta prom}$ limbs (Fig. 2*C*). Nevertheless, subtraction of the two heatmaps revealed a slight increase in the interaction between *Shh* and the centromeric part of the *Lmbr1* gene (*SI Appendix*, Fig. S1). This specific increase was also detected in the subtraction of wildtype and $Lmbr1^{\Delta prom/\Delta prom}$ 4C

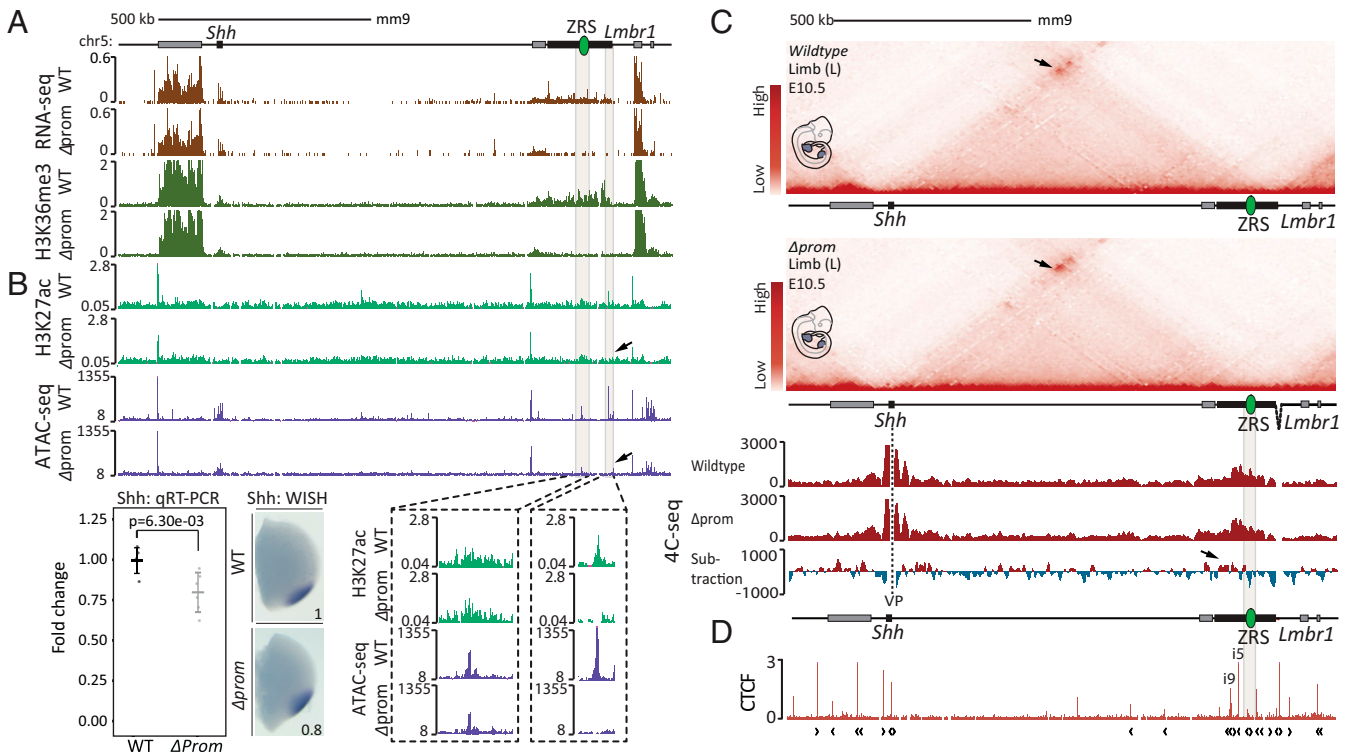

Fig. 2. Abrogation of *Lmbr1* transcription results in down-regulation of *Shh*. (*A*) RNA-seq (brown) and H3K36me3 ChIP-seq (green) of wildtype and $Lmbr1^{\Delta prom/\Delta prom}$ E10.5 limb buds. *Lmbr1* and *Shh* genes in *A*–*D* are indicated as black bars, whereas neighboring genes are colored gray. (*B*) H3K27ac ChIP-seq (light green) and ATAC-seq (purple) of wildtype and $Lmbr1^{\Delta prom/\Delta prom}$ E10.5 limb buds. In the zoom-in of the H3K27ac and ATAC-seq tracks (*Bottom*), note that the *ZRS* enhancer remains open and active (*Bottom Left*) and that a complete loss of signal at the region of the promoter is visible as a result of its deletion (*Bottom Right*). qRT-PCR of *Shh* in wildtype ($n = 5$) and $Lmbr1^{\Delta prom/\Delta prom}$ E10.5 limb buds ($n = 3$). The *P* value was calculated by one-sided Student *t* test ($P = 6.3e\text{-}03$). Error bars represent SD. (*C*) cHi-C maps of wildtype and $Lmbr1^{\Delta prom/\Delta prom}$ E10.5 limb buds. The black arrows indicate the differential interaction of the *Shh-ZRS* region between wildtype and $Lmbr1^{\Delta prom/\Delta prom}$ E10.5 mutants. Note that the overall structure does not change between the wildtype and mutant limb buds. 4C-seq tracks from the *Shh* promoter as viewpoint (VP) in wildtype and $Lmbr1^{\Delta prom/\Delta prom}$ E10.5 limb buds. The lowest panel shows a subtraction of both tracks (blue and red indicate loss and gain of contact, respectively, in the mutant). The black arrow indicates the increase of interaction between *Shh* and the centromeric part of the *Lmbr1* gene in the $Lmbr1^{\Delta prom/\Delta prom}$ mutants. (*D*) CTCF ChIP-seq tracks of the extended *Shh* locus. The CTCF sites at introns 5 and 9 are indicated as i5 and i9, respectively. Black arrows under the ChIP-seq track indicate the orientation of CTCF sites.

profiles using the *Shh* promoter as a viewpoint (Fig. 2*C*). The precise genomic location of the increased contacts indicates a centromeric shift of the interaction, excluding the *ZRS* region, which itself displays a decreased contact with the *Shh* promoter. Thus, it is likely that the *ZRS*, located in the telomeric part of *Lmbr1*, becomes more isolated from the *Shh* TAD and gene. In fact, the centromeric part of *Lmbr1* is bound by prominent CTCF sites within intron 5 (i5) and intron 9 (i9) of the gene, the binding frequency of which could be affected by the *Lmbr1* transcriptional loss (Fig. 2*D*) (34). Consequently, our findings support that transcription of *Lmbr1* modulates the distribution of *Shh* interactions within *Lmbr1*, which in turn defines the *Shh-ZRS* contact and ultimately *Shh* expression.

Deletions of CTCF Sites Result in Ectopic CTCF and Cohesin Binding at Neighboring Sites. In all cell types we investigated, three major CTCF binding events occur on either side of the *ZRS*, which we termed i4 (intron 4), i5, and i9, respectively (Fig. 3*A*). These binding sites are oriented in the direction of *Shh*, which is itself flanked by two CTCF sites in a convergent orientation, and could account for the preformed interaction between the *ZRS* and *Shh*, ultimately leading to a decreased *Shh* expression. By using CRISPR-Cas9, we engineered homozygous deletions specifically targeting the CTCF binding motifs individually or sequentially in combination (Fig. 3*B* and *SI Appendix*, Table S1). First, the deletion of the i4 binding site (*ΔCTCF i4*) resulted in the disruption of CTCF binding solely at the i4 site. In contrast, the deletion of the i5 site (*ΔCTCF i5*) eliminated the CTCF binding at the targeted site and, surprisingly, induced ectopic CTCF binding at two neighboring sites; one on the centromeric side of the *ZRS* and the other adjacent to the *Lmbr1* promoter. Finally, we retargeted the *ΔCTCF i5* allele to delete the i4 CTCF binding site and thus obtain a combined deletion allele (*ΔCTCF i4:i5*). *ΔCTCF i4:i5* limb buds displayed loss of i4 and i5 CTCF binding as well as increased CTCF binding at both the *ZRS* and *Lmbr1* promoter ectopic binding sites, as seen in the *ΔCTCF i5* mutant (Fig. 3*B*).

The increased binding of CTCF at the *ZRS* indicates cooperativity in the occupancy between neighboring binding sites. This ectopic binding precisely occurred on the centromeric edge of the previously characterized long-range regulatory region of the *ZRS*, suggesting a potential compensating redundancy that

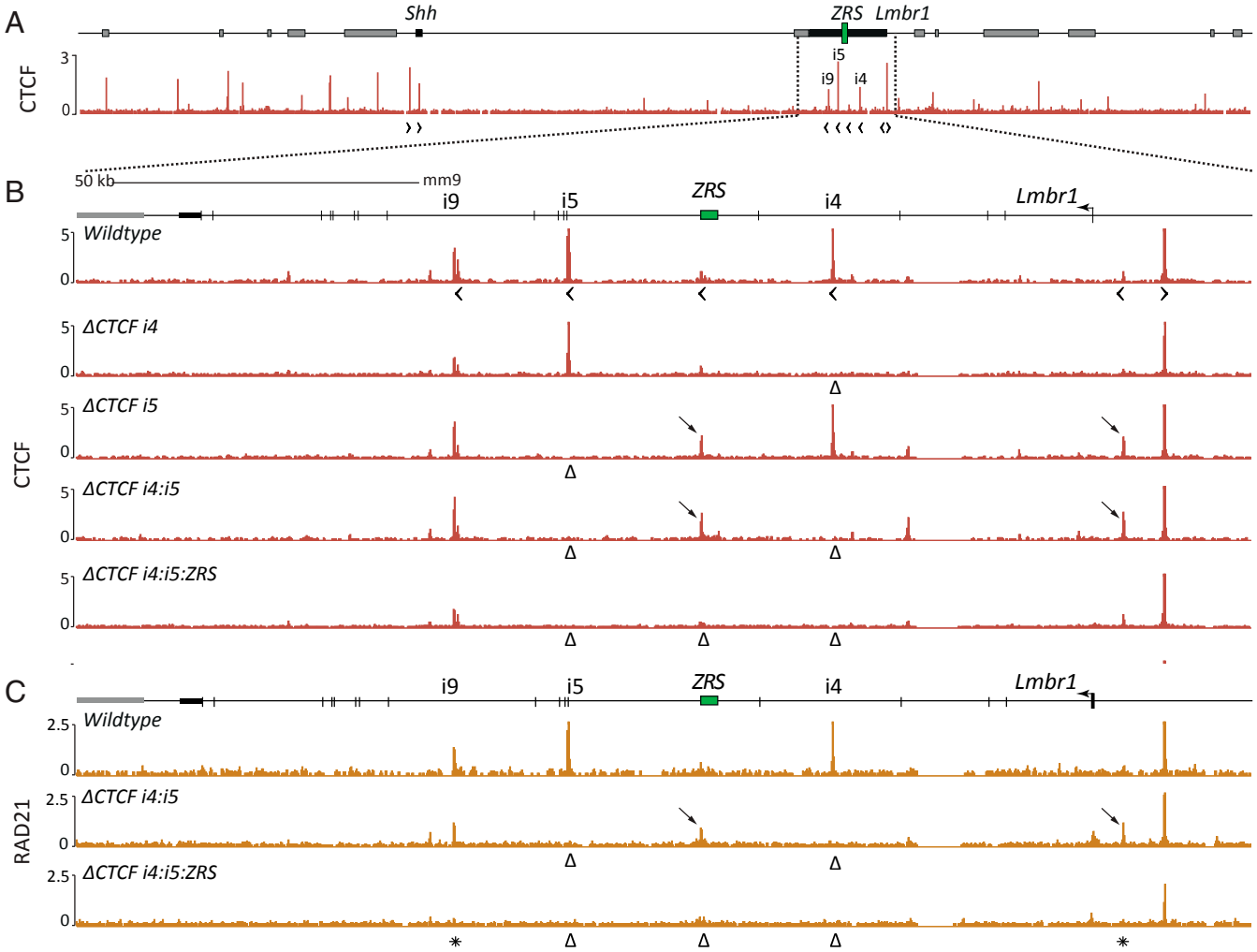

Fig. 3. Deletions of CTCF binding sites result in ectopic CTCF and Cohesin binding at neighboring sites. (*A*) CTCF ChIP-seq enrichment in wildtype E10.5 limb buds at the *Shh* locus. Note the i4, i5, and i9 CTCF binding sites around the *ZRS*. (*B*) Zoom-in at the *Lmbr1* gene of CTCF ChIP-seq enrichment tracks in wildtype, *ΔCTCF i4*, *ΔCTCF i5*, *ΔCTCF i4:i5*, and *ΔCTCF i4:i5:ZRS* homozygote mutants. Black arrowheads below the wildtype track indicate the CTCF motif orientation. The Δ signs indicate the CTCF motif deletions leading to loss of CTCF binding, and the arrows indicate the increased binding on ectopic neighboring sites. (*C*) RAD21 ChIP-seq in wildtype, *ΔCTCF i4:i5*, and *ΔCTCF i4:i5:ZRS* homozygote mutants. The Δ signs indicate the CTCF motif deletions. Black arrows show the increased binding at the *ZRS* and *Lmbr1* promoter ectopic sites. Black asterisks indicate the cooperative RAD21 loss at the i9 and *Lmbr1* promoter in triple *ΔCTCF i4:i5:ZRS* mutants.

ensures formation of the 3D architecture (35). Therefore, within the *ΔCTCF i4:i5* background, we additionally deleted the *ZRS* CTCF binding site (*ΔCTCF i4:i5:ZRS*) while leaving intact any other characterized transcription factor binding sites required for the *ZRS* function (Fig. 3*B* and *SI Appendix*, Fig. S5*C*). The *ΔCTCF i4:i5:ZRS* allele resulted in a complete loss of CTCF binding at the *ZRS* site and up to the i9 binding site, 40 kb away.

We then performed ChIP-seq for RAD21 to assess whether the loss of CTCF binding altered the recruitment of the Cohesin complex. In wildtype limb buds, RAD21 accumulates exactly at the same positions as CTCF around the *ZRS* (Fig. 3*C*). In double-deleted *ΔCTCF i4:i5* limb buds, RAD21 is lost at the i4 and i5 sites and bound at the *ZRS* and *Lmbr1* promoter ectopic CTCF sites. In triple-deleted *ΔCTCF i4:i5:ZRS* limb buds, RAD21 was also lost from the ectopic ZRS site, as expected from the deletion of the underlying CTCF site. Surprisingly, in these latter mutants, we further observed a complete loss of RAD21 binding over the ectopic *Lmbr1* promoter site and at the neighboring i9 site, even though neither were mutated and were thus still bound by CTCF (Fig. 3 *B* and *C*). We concluded that the genetic alteration of these three CTCF sites results in a cooperative loss of Cohesin binding within the *Lmbr1* gene and without the appearance of ectopic binding sites.

CTCF Binding Sites Enable a Strong *Shh–ZRS* Interaction and Ensure Normal *Shh* Transcription. To assess the effect of impaired CTCF binding on the 3D architecture and transcription of the *Shh* locus, we produced cHi-C maps of *ΔCTCF i4:i5* and *ΔCTCF i4:i5:ZRS* E10.5 limb buds and compared them versus wildtype controls. First, in *ΔCTCF i4:i5*, a strong loss of interaction was observed between *Shh* and the *Lmbr1* boundary (Fig. 4 *A*–*C*). In the triple *ΔCTCF i4:i5:ZRS* mutant limb buds, the loss of contact was as strong as in the double *ΔCTCF i4:i5* mutants (*SI Appendix*, Fig. S2 *A* and *B*). In both mutants, we also observed a partial deinsulation of the *Shh* TAD denoted by decreased inner TAD interactions and increased interactions with the telomeric neighboring *Mnx1*-containing TAD (Fig. 4 *B* and *C* and in *SI Appendix*, Fig. S2*B*). A similar change in the locus structure was observed in *ΔCTCF i4:i5:ZRS* midbrain, indicating that this effect was not restricted to the limb (*SI Appendix*, Fig. S2 *D* and *E*).

To visualize the difference in structure, we modeled the 3D architecture of the *Shh* locus using the wildtype and *ΔCTCF i4:i5* cHi-C datasets and a conformation prediction approach based on polymer physics (36–38) (Fig. 4*D*, Movies S1 and S2, and *SI Appendix*, Fig. S3). Specifically, in the wildtype model, *Shh* and the *ZRS* are found in close proximity and separated from *Mnx1*. In contrast, in *ΔCTCF i4:i5* mutant limb buds, the distance between *Shh* and the *ZRS* is increased and *Mnx1* is found closer to both *ZRS* and *Shh*, in agreement with the observed deinsulation between the *Mnx1* and *Shh* TADs (Fig. 4*B*). The increased *Shh–ZRS* distance was further confirmed by a shift in the distribution of distances across all of the polymer-based models derived from wildtype and mutant limb buds (Fig. 4*E*).

We next investigated whether these alterations of the chromatin structure have an effect on *Shh* regulation. By using qRT-PCR, we detected a 51% ($P = 1.01e-04$) loss of *Shh* expression in *ΔCTCF i4:i5* limb buds and a 52% ($P = 2.65e-06$) reduction in *ΔCTCF i4:i5:ZRS* limb buds compared to wildtype (Fig. 4*F* and *SI Appendix*, Fig. S2*C*, respectively). The absence of transcriptional and structural differences between *ΔCTCF i4:i5* and *ΔCTCF i4:i5:ZRS* alleles suggests that the *ZRS* CTCF binding site, despite being bound by RAD21, cannot rescue the loss of the i4 and i5 binding sites. We next examined the limb skeleton of E18.5 embryos and could not observe any obvious limb skeletal phenotype. To further test how the structural changes influence *Shh* transcription when regulated by a different enhancer, e.g., *SBE1*, qRT-PCR was performed in *ΔCTCF i4:i5:ZRS* midbrain. In this tissue, we did not observe a significant *Shh* reduction (*SI Appendix*, Fig. S2*F*). Altogether, our results underline that the removal of CTCF binding sites around the *ZRS* and the subsequent alterations in the long-range *Shh–ZRS* interaction have a significant effect on *Shh* transcription only in the limb. Moreover, our findings suggest that a CTCF-independent mode of communication between *Shh* and the *ZRS* sustains the residual *Shh* transcription in *ΔCTCF i4:i5* and *ΔCTCF i4:i5:ZRS* embryos.

CTCF Supports Residual *Shh* Expression in a *ZRS* Hypomorphic Allele Background. The loss of *Shh* expression in CTCF mutant animals suggests that CTCF and the preformed topology of the locus provide robustness to *Shh* regulation and, as a result, maximize its transcriptional levels. To test this hypothesis, we engineered a hypomorphic allele of the ZRS by deleting 400 bp that contains several ETS binding sites and so contributes to the *ZRS* enhancer strength (*ΔZRSreg*; Fig. 5*A* and *SI Appendix*, Fig. S5*C*) (35, 39, 40). To characterize the effect of this allele on the locus 3D structure, we produced cHi-C maps of *ΔZRSreg* limbs. We observed an increased interaction between *Shh* and the i5 CTCF site, but the overall structure of the locus remained unchanged (Fig. 5*B* and *SI Appendix*, Fig. S4). Moreover, the preformed interaction between *Shh* and the region of the *ZRS* appeared unaffected by the mutation. We then assessed the transcriptional outcome of the *ΔZRSreg* allele by using whole-mount in situ hybridization (WISH) and qRT-PCR and found a 75% reduction in *Shh* transcription ($P = 2e-07$) in mutant developing limb buds (Fig. 5*C*). This significant loss of expression resulted in a fully penetrant forelimb oligodactyly, a typical *Shh* loss-of-function phenotype (24), and normal hindlimbs (Fig. 5 *D*, *Middle*).

We then introduced the same hypomorphic *ΔZRSreg* mutation on the *ΔCTCF i4:i5* background (*ΔCTCF i4:i5:ZRSreg*; Fig. 5*A*). CHi-C in *ΔCTCF i4:i5:ZRSreg* limb buds showed, similarly to *ΔCTCF i4:i5* mutants, a strong loss of interaction between *Shh* and the *Lmbr1* boundary as well as a partial deinsulation of the *Shh* TAD (Fig. 5*B*). Moreover, the interaction loss between *Shh* and the *ZRS* in *ΔCTCF i4:i5:ZRSreg* is similar to what is observed in *ΔCTCF i4:i5:ZRS* mutants (*SI Appendix*, Fig. S5 *A* and *B*). Accordingly, it appears that the removal of this 400-bp segment of the *ZRS*, which includes the ZRS CTCF binding site as well as several ETS binding sites, does not further affect the contact between *Shh* and the *ZRS* (*SI Appendix*, Fig. S5 *B* and *C*).

In *ΔCTCF i4:i5:ZRSreg*, WISH experiments could not detect any expression of *Shh*, and qRT-PCR experiments confirmed that *Shh* expression is reduced by more than 98% ($P = 5.8e-07$). The appendages of these animals were strongly malformed; in forelimbs, we could observe a fully penetrant digit agenesis, as expected from a complete *Shh* loss of function. In contrast, hindlimbs were slightly less affected, displaying oligodactyly and syndactyly (Fig. 5 *D*, *Lower*). Accordingly, in comparison with the *ΔZRSreg* animals, we observed a near-complete loss of *Shh* expression and a more severe digit phenotype in *ΔCTCF i4:i5:ZRSreg* mutants. These results demonstrate that, in the context of the *ZRS* hypomorphic allele, the 3D structure mediated by CTCF and Cohesin is the only backup supporting the residual *Shh* expression. Thus, the preformed chromatin structure at the locus confers robustness in *Shh* regulation and ultimately enables the formation of five digits.

Discussion

In this study, we focused on the role of the preformed interaction established between *Shh* and its limb enhancer, the *ZRS*, both located in the vicinity of two TAD boundaries. By using CRISPR-Cas9, we eliminate constitutive transcription or CTCF binding motifs around the *ZRS* and assess the contribution of these factors on the locus structure and regulation in vivo.

Transcriptionally active gene promoters have been shown to interact over a large distance independently of CTCF and thereby suggested to be drivers of genome architecture (5). In fact, strong gene transcriptional activation was associated with loop and boundary formation at the *Zfp608* locus in neuronal progenitors (5). However, the ectopic activation of the *Zfp608* gene alone did not result in the formation of a new TAD

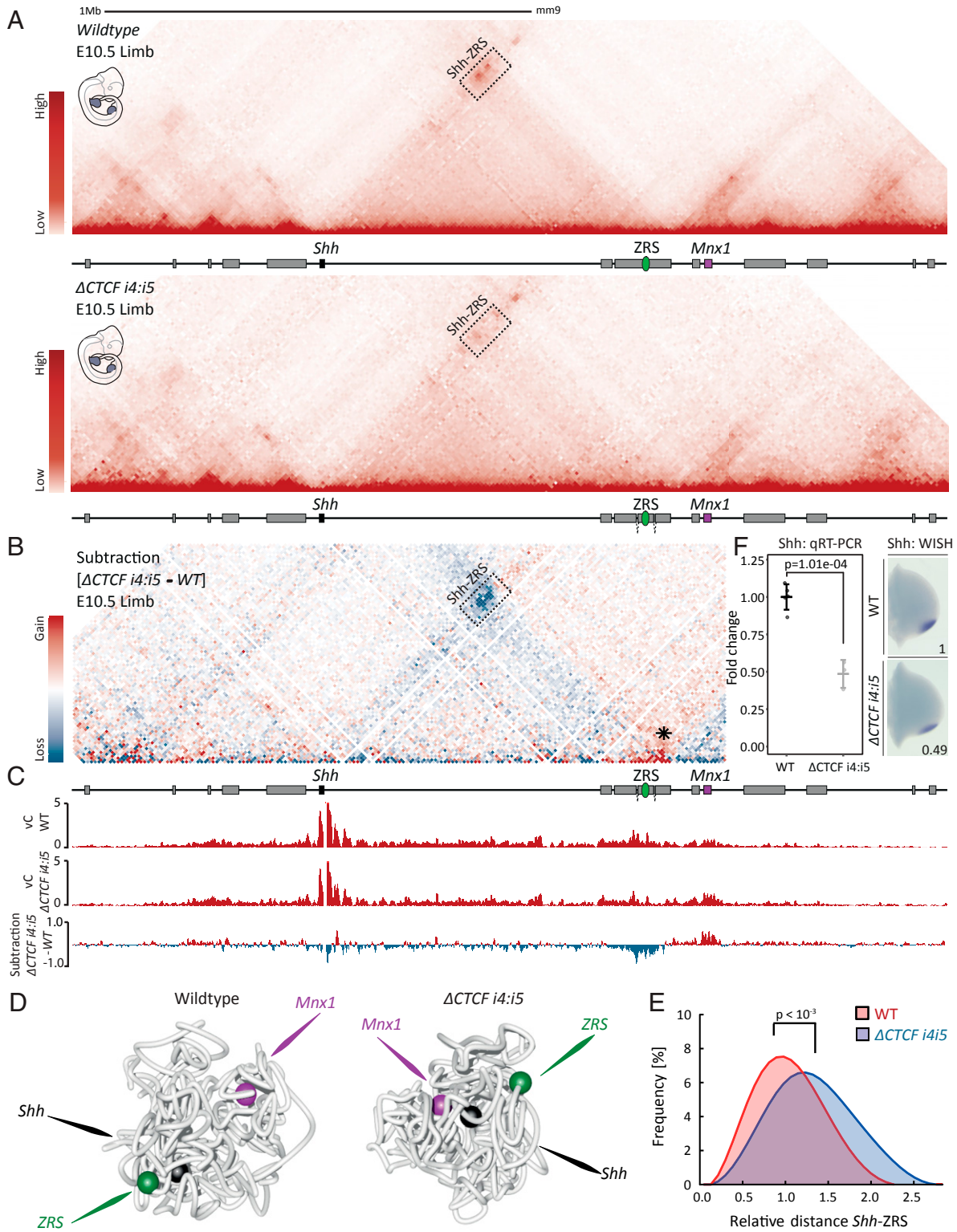

Fig. 4. CTCF enables the long-range interaction between *Shh* and *ZRS* and ensures proper *Shh* expression level. (*A*) cHi-C maps of wildtype (*Upper*) and *ΔCTCF i4:i5* (*Lower*) E10.5 limb buds. Black box indicates the domain of high interaction between *Shh* and the centromeric side of *Lmbr1*, which contains the *ZRS* enhancer. Note the decreased interaction within the box in *ΔCTCF i4:i5* compared with wildtype tissue. (*B*) Subtraction map between wildtype and *ΔCTCF i4:i5* maps, where red and blue indicate gain and loss of contact, respectively, in *ΔCTCF i4:i5 limbs*. The black asterisk indicates loss of insulation between *Shh* and *Mnx1* TADs. (*C*) Virtual capture-C (vC) profiles using *Shh* promoter as viewpoint derived from cHi-C in E10.5 wildtype and *ΔCTCF i4:i5* limb buds. Lower track: the subtraction of the two upper tracks, where positive values indicate gain (red) and negative values indicate loss (blue) of interaction in *ΔCTCF i4:i5* limb buds. Note the loss of chromatin interactions with the *ZRS* region and the gain of interaction with the neighboring *Mnx1* gene. (*D*) A 3D polymer model of wildtype and *ΔCTCF i4:i5* limb cHi-C data. Note the changes in proximity between *Shh* and *ZRS* and between *Shh* and *Mnx1*. (*E*) Frequency plot of the distance distribution between *Shh* and the *ZRS* in wildtype and *ΔCTCF i4:i5* limb buds. Note the increase in relative distance in the mutant limbs. *P* value was calculated by Mann–Whitney test. (*F*) qRT-PCR and WISH of *Shh* in wildtype ($n = 5$) and *ΔCTCF i4:i5* limb buds ($n = 3$). The *P* value was calculated by one-sided Student *t* test ($P = 1.1e-04$). Error bars represent SD.

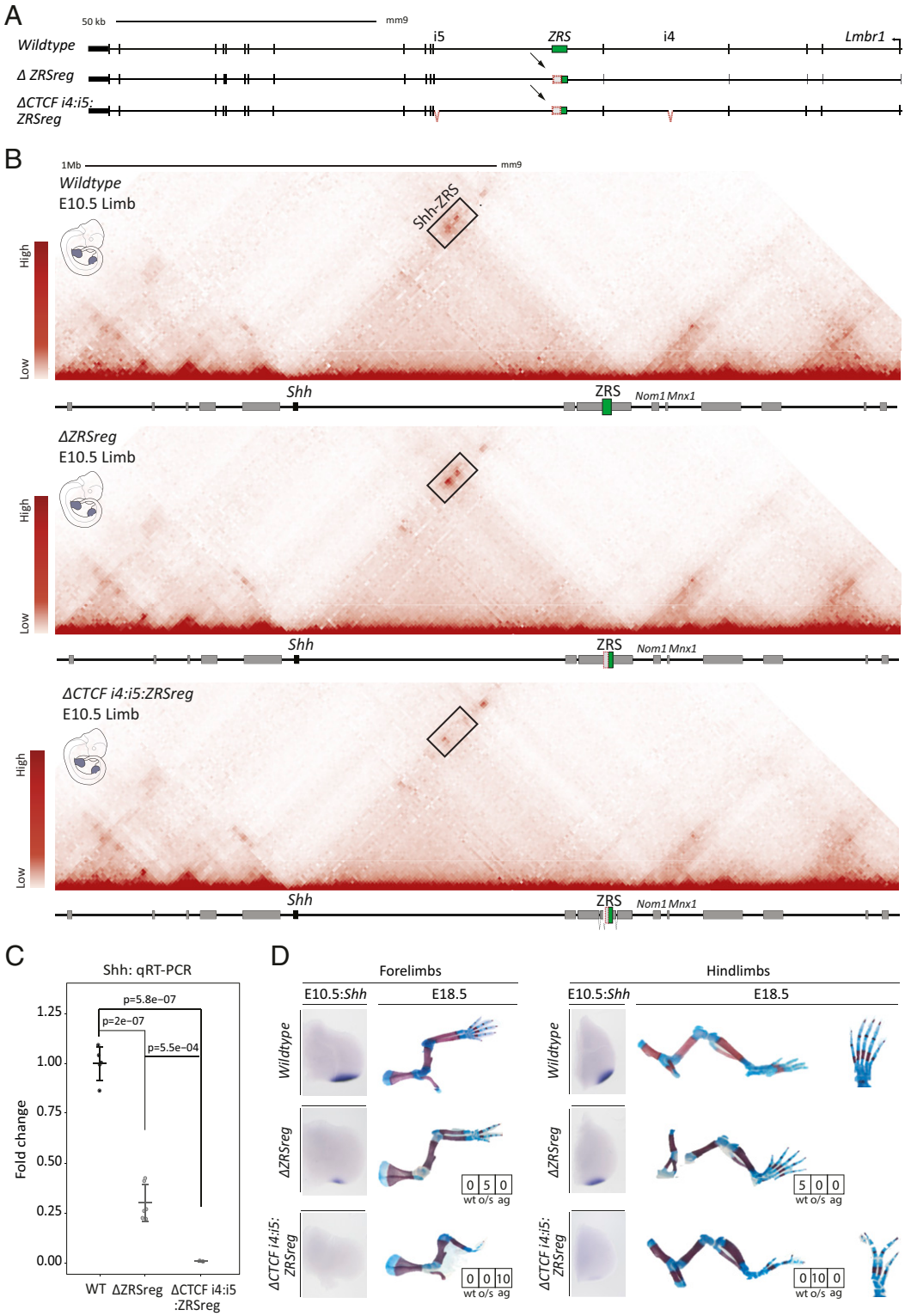

Fig. 5. CTCF binding site deletions on a hypomorphic background lead to *Shh* loss-of-function phenotype. (*A*) Schematic representation of the alleles. *ΔZRSreg* bears a 400-bp deletion encompassing the long-range control domain at the centromeric side of the *ZRS*. *ΔCTCF i4:i5:ZRSreg* bears the *ΔZRSreg* deletion on top of the mutated CTCF binding sites i4 and i5. (*B*) cHi-C maps of wildtype (*Upper*), *ΔZRSreg* (*Middle*), and *ΔCTCF i4:i5:ZRSreg* (*Lower*) E10.5 limb buds. Black box indicates the domain of high interaction between *Shh* and the centromeric side of *Lmbr1*, which contains the *ZRS* enhancer. Note that wildtype and *ΔZRSreg* overall structures are similar, whereas interactions within the box are decreased in *ΔCTCF i4:i5:ZRSreg* limb buds. (*C*) qRT-PCR of *Shh* in wildtype ($n = 5$), *ΔZRSreg* ($n = 6$), and *ΔCTCF i4:i5:ZRSreg* ($n = 3$) E10.5 limb buds. The *P* value was calculated by one-sided Student *t* test. Error bars represent SD. (*D*) *Shh* WISHs in E10.5 limb buds and E18.5 limb skeletal preparation of wildtype and mutants. The boxes represent the number of animals with wildtype (wt), oligodactyly/syndactyly (o/s), and agenesis of digits (ag). Note the differences in expression and skeletal defects between *ΔZRSreg* and *ΔCTCF i4:i5:ZRSreg* mutants.

boundary or chromatin interactions. Here, the abrogation of *Lmbr1* constitutive transcription did not result in an alteration of the TAD structure, but in an interaction shift toward the centromeric end of the gene body. This specific gain of interactions occurs in a region centromeric to the *ZRS*, bearing CTCF sites, located in intron 5 (i5) and 9 (i9) of *Lmbr1* (Fig. 2*D*). The observed gain of interactions is likely linked to the loss of the elongating polymerase as recently reported (34). In this latter work, the blocking of transcription leads to increased CTCF and Cohesin binding and stronger chromatin loops (34). The loss of *Lmbr1* transcription might thus increase the binding of CTCF and Cohesin within its gene body, which could result in a preferential interaction between the centromeric i5 and i9 sites and the opposing *Shh* CTCF sites. As a result, the distribution of interactions shifts toward the centromeric side of the *Lmbr1* boundary, thereby increasing the insulation between *ZRS* and *Shh*, decreasing *Shh* transcriptional activation.

Within the *Lmbr1* gene, disruption of only three CTCF sites is sufficient to abolish the binding of the Cohesin complex at five different binding sites. According to the loop extrusion biophysical model of chromatin organization, Cohesin rings bind and extrude DNA until they reach convergent CTCFs acting as natural barriers (4, 6). Thus, we hypothesized that the deletion of these CTCF sites would lead to the inability of Cohesin to be halted at the mutated sites. Additionally, the loss of Cohesin binding over the i9 and over the ectopic *Lmbr1* sites, which remain intact, suggests an interdependent binding mechanism between neighboring CTCF sites and Cohesin loading as previously observed at the *Hoxa* locus (41, 42). As a result of this binding loss, the strong *Shh–ZRS* interaction is significantly weakened and *Shh* expression is decreased by 50%. This observation suggests that the normal expression of *Shh* depends on the presence of CTCF and on the preformed 3D structure. These data also indicate that, in the absence of the CTCF-driven chromatin interaction, *Shh* and the *ZRS* can still communicate in the 3D space of the nucleus, probably via an alternative mechanism such as molecular bridging of phase separation (43). Furthermore, we noticed in these mutants a partial deinsulation between the *Shh* TAD and its neighboring telomeric *Mnx1* TAD. In this genetic configuration, a complete TAD deinsulation is likely prevented by the intact convergent CTCF/Cohesin binding sites in the *Mnx1* TAD, which provide a telomeric border to the *Shh* interaction domain. Overall, our results demonstrate that the CTCF/Cohesin-mediated preformed topology of the *Shh* locus ensures maximal gene expression in vivo, but is not essential to achieve long-range gene activation.

Despite the dramatic loss of *Shh* expression associated with the reduced *Shh–ZRS* interaction, we did not detect a digit phenotype in the CTCF mutant animals. The lack of limb malformations in these animals is in agreement with findings in heterozygote *ZRS* loss-of-function animals, which bear normal appendages despite having a 50% lower *Shh* expression (44). Such a phenotype arises only when the function of the *ZRS* has been partially impaired in a hypomorphic allele, thus providing a "sensitized" genetic background. This particular mutation alone leads to partial digit phenotype as a result of a number of ETS binding site deletions (*SI Appendix*, Fig. S5*C*), which define the spatial expression of *Shh* in the limb and determine the enhancer "strength" (39). In this allele, further disruption of the CTCF binding around the *ZRS* results in *Shh* loss of function with complete digit aplasia in the forelimbs and oligodactyly in the hindlimbs. Accordingly, we concluded that this stable, preformed, CTCF/Cohesin-based chromatin topology provides robustness in gene regulation that can buffer variations in enhancer activity by maintaining *Shh* mRNA in high excess. Finally, our data suggest that the modest transcriptional alterations upon global CTCF/Cohesin depletion in nondifferentiating cells can be crucial to the process of differentiation and morphogenesis (7, 8, 10).

Dynamic and preformed chromatin interactions were shown to occur between enhancers and promoters at various loci during development and lineage commitment (11, 15). In this work, we show that the unique and specific *Shh* limb enhancer, the *ZRS*, contacts the *Shh* promoter in a preformed manner that enables a constant proximity and ultimately drives the promoter to mirror the activity of its enhancer. This finding suggests that the regulatory role of stable chromatin structure differs greatly from dynamic, tissue-specific ones. Indeed, even though stable chromatin organization might ensure permanent enhancer–promoter communication, dynamic interactions were found to refine unspecific enhancer activities into more specific gene transcriptional output, and, consequently, their alterations were shown to cause gene misregulation (17, 45). Ectopic looping between the *β-globin* promoter and its *LCR* enhancer in primary erythroid progenitor cells induces *β-globin* misexpression (45). Furthermore, at the *Pitx1* locus, our previous work showed that dynamic changes in chromatin structure restrict the activity of a fore- and hindlimb enhancer to hindlimb only (17). The stable mode of interaction found at the *Shh* locus does not restrict the activity of the *ZRS*. Instead, it promotes gene transcriptional robustness from a single specific enhancer. This contrasts other loci, where robustness is mainly ensured via the additive effect of multiple, partially redundant enhancers. At the *Hoxd* and *Ihh* loci, for example, several enhancers have been shown to globally contribute to the final transcription pattern (46–49). At these loci, a preformed interaction may be unnecessary, as a promoter samples the cumulative activities of all its enhancers, each of which possesses an overlapping and redundant activity. Nevertheless, both types of chromatin interactions, dynamic and preformed, seem to coexist at many developmental loci. As a consequence, it will be important to determine if the dynamic and preformed chromatin topologies maintain their regulatory functions by restricting enhancer activities and by enforcing constitutive enhancer–promoter communication, respectively.

Methods

Cell Culture and Mice.

CRISPR-Cas9 engineered allelic series. All deletion alleles of this study were generated by using the CRISPR-Cas9 editing system according to previous works (33, 50). To target the CTCF motif sites specifically, only one gRNA was used. All CTCF deletions are summarized in *SI Appendix*, Table S1. For longer deletions, two gRNAS were used on each side of the target DNA. A detailed protocol for CRISPR-Cas9 engineering is provided in *SI Appendix, Supplementary Methods*.

Aggregation of mESC. Embryos and live animals were generated from ESCs, which were thawed, seeded on CD1 feeders, and grown for 2 d, by diploid or tetraploid complementation (51). CD1 female mice were used as foster mothers. The mouse lines were maintained when necessary by crossing them with C57BL6/J mice.

Animal procedures. All animal procedures were in accordance with institutional, state, and government regulations (Berlin: LAGeSo G0247/13 and G0346/13).

RNA isolation, cDNA synthesis, and qRT-PCR. Forelimb and hindlimb buds of 34–35 somite stage embryos (E10.5) were microdissected in cold PBS solution, snap-frozen in liquid N_2, and immediately stored at −80 °C. To isolate the RNA, tissues were thawed on ice and the RNA extraction was performed by using the RNeasy Mini kit according to the manufacturer's instructions (no. 74106; Qiagen). cDNA was generated by using the SuperScript II First-Strand Synthesis System (Thermo Fisher Scientific) whereby 500 ng of RNA was reverse-transcribed using random hexamer primers. To quantify the mRNA, qRT-PCR analysis of three to six biological replicates (two fore- and two hindlimb buds per biological replicate) in three technical triplicates was performed by using GoTaq qPCR Master Mix (Promega). qRT-PCR of midbrain included three biological wildtype replicates and two biological mutant replicates each divided in two technical replicates. qPCR primers used were as follows: ShhFw, 5′-ACCCCGACATCATATTTAAGGA-3′; ShhRev, 5′-TTAACTTGTCTTTGCACCTCTGA-3′; Rps9Fw, 5′-GACCAGGAGCTAAAGTTGATTGGA-3′; and Rps9Rev, 5′- TCTTGGCCAGGGTAAACTTGA-3′

RNA-seq. Forelimb and hindlimb buds of 34–35 somite stage embryos (E10.5) were microdissected in cold PBS solution, snap-frozen in liquid N_2, and immediately stored at −80 °C. To isolate the RNA, tissues were thawed on ice, and 500 μL of TRIzol was added. Homogenization of the tissue was achieved by filtering the samples using a 0.4-mm syringe until cell clumps were dissolved. Then, 200 μL of chloroform was added and the samples were mixed vigorously and centrifuged at 12,000 × *g* for 15 min at 4 °C. Then, the upper phase was transferred to new Eppendorf tubes, and 500 μL of isopropanol

was added. After 10 min incubation at RT, another centrifugation step at maximum speed for 10 min at 4 °C followed. The pellet was washed two times with 100% and 70% EtOH accordingly, centrifuged, air-dried for 10 min, and eluted with nuclease-free water. Samples were poly-A–enriched and sequenced (paired-end 50 bp) by using Illumina technology following standard protocols. Two biological replicates were used for each experiment.

WISH. *Shh* WISH was performed on 34–35 somite stage mouse embryos (E10.5) by using a digoxigenin-labeled *Shh* antisense riboprobe transcribed from a cloned Shh probe (PCR DIG Probe Synthesis Kit; Roche). A detailed protocol is provided in *SI Appendix, Supplementary Methods*.

Skeletal preparation. E18.5 fetuses were processed and stained for bone and cartilage with standard Alcian blue/Alizarin A staining procedures (more detailed methodology is provided in *SI Appendix, Supplementary Methods*).

ATAC-seq. ATAC-seq experiments were performed according to a previous work (52) by using primary E10.5 limb buds dissected in 1× PBS solution. A detailed protocol is provided in *SI Appendix, Methods*.

Tissue collection and fixation for ChIP-seq, Capture-HiC, and 4C-seq. Limbs from homozygous E10.5 embryos were microdissected in 1× PBS solution and pooled together. Limbs were washed once with 1× PBS solution and homogenized in 500 μL collagenase solution [0.1% collagenase type 1a (C9891; Sigma), 0.1% (wt/vol) trypsin, 5% FCS or chicken serum in DMEM:Ham's F-12, 1:1] for ∼15 min in a Thermomixer. Additional disruption of cell clumps was achieved by using a 0.4-mm needle. Then, samples were transferred in a 50-mL Falcon tube through a 40-μm cell strainer and complemented with 10% FCS/PBS solution. Formaldehyde 37% diluted to a final 1% for ChIP experiment and 2% for Capture Hi-C and 4C-seq was used to fix the samples for 10 min at RT. To quench the fixation, 1.425 M glycine was used. Formaldehyde solution was removed by centrifugation (300 × *g*, 8 min), and fresh lysis buffer (10 mM Tris, pH 7.5, 10 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA complemented with Protease Inhibitor) was added to isolate the nuclei. The samples were incubated for 10 min on ice, centrifuged for 5 min at 480 × *g*, washed with 1× PBS solution, and snap-frozen in liquid N_2.

ChIP-seq. Chromatin from at least 16 pairs of E10.5 limb buds was sonicated in a size range of 200–500 bp by using the Bioruptor UCD-300 system (Diagenode). CTCF and RAD21-bound chromatin was immunoprecipitated by using the iDeal Kit for Transcription Factors (C01010055; Diagenode) according to the manufacturer's instructions. Chromatin 20 μg was used for transcription factor ChIP-seq, and the experiments were performed in duplicates (CTCF for *Wildtype*, *ΔCTCF i4i5*, *ΔCTCF i4i5ZRS*) or in singletons. For histone modifications, the immunoprecipitation was performed with 10–15 μg of chromatin as described previously (53, 54). The experiments were performed in duplicates. Libraries were prepared by using the Nextera adaptors and were sequenced as single-end 50- or 75-bp reads. Antibodies used were H3K27ac (C15410174; Diagenode), K3K36me3 (ab9050; Abcam), CTCF (C15410210; Diagenode), and RAD21 (ab992; Abcam).

3C-library for Capture Hi-C and 4C-seq. 3C-libraries were prepared from at least 10–12 pairs of homozygous E10.5 forelimb and hindlimb buds as described previously (17, 55, 56). A detailed protocol is provided in *SI Appendix, Supplementary Methods*.

Three-dimensional polymer modeling. To investigate the spatial conformations of the *Shh* chromatin region, we employed the *Strings & Binders Switch* (SBS) model (36, 37), whereby a chromatin segment is modeled as a self-avoiding string made of consecutive beads interacting with diffusing Brownian molecular binders. Because of the specific interaction between beads and their cognate binders, distant loci can bridge with each other, forming loops, enabling the spontaneous folding of the polymer. To estimate the minimal SBS polymer model that best reproduces the folding of the *Shh* region, i.e., the best distribution of the different binding sites along the polymer chain, we used the PRISMR method (38). Briefly, PRISMR is a machine learning-based approach that takes experimental contact data as its input and, by simulated annealing Monte Carlo, finds the minimal number of different types of binding sites and their arrangement along the polymer that results in the best agreement between the input data (cHi-C here) and the equilibrium contact matrix derived by the model. Next, by molecular dynamics simulations of the best SBS polymer, an ensemble of single-molecule spatial conformations is derived for the studied loci. Simulation details are provided in *SI Appendix, Supplementary Methods*.

Statistical Analysis and Computational Analysis.

qRT-PCR analysis. The fold change between wildtype and mutant samples was calculated by using the delta delta Ct method (ΔΔCt) (57). For statistical analysis, one-sided Student's *t* tests were used. Error bars represent SD between at least three biological replicates.

RNA-seq. Paired-end reads (50 bp) were mapped to the mouse reference genome (mm9) by using the STAR mapper version 2.4.2a with default settings besides the following options: outFilterMultimapNmax = 5; outFilterMismatchNoverLmax = 0.1; alignIntronMin = 20; alignIntronMax = 500000; and chimSegmentMin = 10. Reads per gene were counted based on the UCSC annotation tracks "known genes" and "RefSeq" combined via shared exon boundaries. The counting was implemented by applying the R function "summarizeOverlaps" with "mode = Union" and "fragments = TRUE". Finally, differential expression analysis was performed with the DEseq2.

ATAC-seq. ATAC-seq data were processed as described in the ENCODE (29) guidelines for mouse ATAC-seq samples. First, ATAC-seq paired-end reads were trimmed to 30 bp to allow fragments with close-by transposition events (<50 bp) to map, i.e., increase read coverage at nucleosome-free regions. Second, trimmed reads were mapped with Bowtie2 (58) and duplicated fragments removed with Picard RemoveDuplicates (https://broadinstitute.github.io/picard/). Last, bigWig files for display were generated with deepTools2 (59) for properly mapped read pairs (FLAG 0x2) with mapping quality ≥20.

ChIP-seq. Single-end reads (50 or 75 bp) were mapped to the reference NCBI37/mm9 genome by using Bowtie 2.2.6 (58) and filtered for mapping quality ≥10, and duplicates were removed by using SAMtools (https://github.com/samtools/samtools). Reads were extended (chromatin modifications, 300 bp; transcription factors, 200 bp) and scaled (1 million/total of unique reads) to produce coverage tracks. For figure display purposes, some replicate ChIP-seq BigWig files were merged by using the bigWigMerge from UCSC tools. BigWig files were visualized in the UCSC browser. ESC H3K27ac and CTCF ChIP-seq were downloaded from the ENCODE portal (https://www.encodeproject.org/) produced by the laboratory of Bing Ren with the following identifiers: ENCSR000CDE, ENCSR000CCB (29, 31).

4C-seq. Biological replicates were merged on the raw read level. Reads were filtered for the primer sequence, including the first restriction enzyme DpnII. After preprocessing, clipped reads were mapped to the reference NCBI37/mm9 genome by using BWA-MEM (v0.7.12-r1044) (60) with default settings. 4C-seq contacts were analyzed in the murine region chr5:28,000,000–30,000,000. To calculate read count profiles, the viewpoint and adjacent fragments 2 kb up- and downstream were removed. A sliding window of five fragments was chosen to smooth the data, and data were normalized to reads per million mapped reads (RPM). To compare interaction profiles of different samples, subtraction of normalized reads was applied.

Capture-HiC

cHi-C processing. Raw sequencing reads were preprocessed with cutadapt v1.15 (61) to trim potential low-quality bases (-q 20 -m 25) and Illumina sequencing adapters (-a and -A option with Illumina TruSeq adapter sequences according to the cutadapt documentation) at the 3′ end of reads. Next, sequencing reads were mapped to reference genome mm9 and filtered and deduplicated using the HiCUP pipeline v6.1.0 (62) (no size selection, Nofill: 1, format: Sanger). The pipeline was set up with Bowtie2 v2.3.4.1 (58) for short read mapping. The trimming of sequencing adapters in the first step would not have been necessary because potentially remaining sequencing adapter in reads from valid ligation products should also be removed by the truncation after ligation sites implemented in the HiCUP pipeline. In case replicates were available, they were combined after the processing with the HiCUP pipeline. Juicer command line tools v1.7.6 (63) was used to generate binned contact maps from valid and unique reads pairs with MAPQ ≥ 30 and to normalize maps by Knights and Ruiz matrix balancing (3, 63, 64). For binning and normalization, only the genomic region chr5:27,800,001–30,600,000 enriched in the DNA-capturing step was considered. Therefore, only read pairs mapping to this region were kept, shifted by the offset of 27,800,000 bp and imported with Juicer tools by using a custom chrom.sizes file, which contained only the length of the enriched region (2.8 Mb). Afterward, KR normalized maps were exported for 5-kb and 10-kb resolution, and coordinates were shifted back to their original values. Subtraction maps were generated from KR normalized maps, which were normalized in a pairwise manner before subtraction. To account for differences between two maps in their distance-dependent signal decay, maps were scaled jointly across their subdiagonals. Therefore, the values of each subdiagonal of one map were divided by the sum of this subdiagonal and multiplied by the average of these sums from both maps. Afterward, the maps were scaled by 10^6/total sum. cHiC maps of count values, as well as subtraction maps, were visualized as heatmaps in which absolute values greater than the 97th percentile were truncated for visualization purposes.

Virtual Capture-C profiles. To obtain more fine-grained interaction profiles for selected loci, we defined 10-kb viewpoint regions and generated virtual Capture-C–like profiles based on the filtered, unique read pairs that were also used for the cHiC maps. A read pair was considered in a profile when it had an MAPQ ≥ 30 and one read mapped to the defined viewpoint region whereas the other read mapped outside of it. The reads outside of the viewpoint were counted per restriction fragment, and read counts were binned afterward to 1-kb bins. In case a fragment was overlapping more than one bin, the read count was distributed proportionally. Afterward,

each binned profile was smoothed by averaging over a running window of five bins and scaled by 10^3/sum of all its counts on chr5. The viewpoint and a window ±5 kb around it were not considered for the computation of the scaling factor. The profiles were generated with custom Java code by using htsjdk v2.12.0 (https://samtools.github.io/htsjdk/).

Virtual Capture-C viewpoints were as follows. Proximal CTCF site: chr5:28,777,001–28,787,000; Shh promoter: chr5:28,789,001–28,799,000; and ZRS enhancer: chr5:29,637,001–29,647,000.

Data Accessibility. Sequencing data are available at the Gene Expression Omnibus repository with the accession number GSE123388.

ACKNOWLEDGMENTS. We thank Karol Macura and Judith Fiedler from the transgenic facility of the Max Planck Institute for Molecular Genetics; members of the laboratory of S.M.; and Martin Franke, Ivana Jerković, and Darío Lupiáñez for their critical reading of the manuscript. This study was supported by Deutsche Forschungsgemeinschaft Grants SP1532/2-1 and MU 880/14 (to S.M.); a grant from the Max Planck Foundation (to S.M.); Swiss National Science Foundation Grants PP00P3_176802 and P300PA_160964 (to G.A.); European Union Marie Skłodowska-Curie Action European Training network Grant n.813282 (to M.N.); NIH Grant 1U54DK107977-01 (to M.N.); Consorzio Interuniversitario Lombardo per l'Elaborazione Automatica (CINECA) Italian SuperComputing Resource Allocation Grant HP10CRTY8P (to M.N.); the Einstein BIH Fellowship Award (to M.N.); and computer resources from INFN, CINECA, and *Scope* at the University of Naples (to M.N.).

1. J. R. Dixon *et al*., Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
2. E. P. Nora *et al*., Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
3. S. S. P. Rao *et al*., A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
4. A. L. Sanborn *et al*., Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6456–E6465 (2015).
5. B. Bonev *et al*., Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572.e24 (2017).
6. G. Fudenberg *et al*., Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
7. E. P. Nora *et al*., Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944.e22 (2017).
8. W. Schwarzer *et al*., Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
9. B. Bintu *et al*., Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**:eaau1783 (2018).
10. S. S. P. Rao *et al*., Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017).
11. G. Andrey *et al*., Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res.* **27**, 223–233 (2017).
12. J. R. Dixon *et al*., Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
13. B. Tolhuis, R. J. Palstra, E. Splinter, F. Grosveld, W. de Laat, Looping and interaction between hypersensitive sites in the active β-globin locus. *Mol. Cell* **10**, 1453–1465 (2002).
14. M. Simonis *et al*., Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* **38**, 1348–1354 (2006).
15. W. de Laat, D. Duboule, Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506 (2013).
16. P. Freire-Pritchett *et al*., Global reorganisation of *cis*-regulatory units upon lineage commitment of human embryonic stem cells. *eLife* **6**, 1–26 (2017).
17. B. K. Kragesteen *et al*., Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nat. Genet.* **50**, 1463–1473 (2018).
18. Y. Ghavi-Helm *et al*., Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).
19. S. Cruz-Molina *et al*., PRC2 facilitates the regulatory topology required for poised enhancer function during pluripotent stem cell differentiation. *Cell Stem Cell* **20**, 689–705.e9 (2017)
20. M. Barbieri *et al*., Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* **24**, 515–524 (2017).
21. O. Hakim *et al*., Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements. *Genome Res.* **21**, 697–706 (2011).
22. E. E. M. Furlong, M. Levine, Developmental enhancers and chromosome topology. *Science* **361**, 1341–1345 (2018).
23. R. Zeller, J. López-Ríos, A. Zuniga, Vertebrate limb bud development: Moving towards integrative analysis of organogenesis. *Nat. Rev. Genet.* **10**, 845–858 (2009).
24. T. Sagai, M. Hosoya, Y. Mizushina, M. Tamura, T. Shiroishi, Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**, 797–803 (2005).
25. L. A. Lettice *et al*., A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
26. T. Amano *et al*., Chromosomal dynamics at the Shh locus: Limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* **16**, 47–57 (2009).
27. I. Williamson, L. A. Lettice, R. E. Hill, W. A. Bickmore, Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development* **143**, 2994–3001 (2016).
28. E. Anderson, P. S. Devenney, R. E. Hill, L. A. Lettice, Mapping the Shh long-range regulatory domain. *Development* **141**, 3934–3943 (2014).
29. T. E. P. Consortium; ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
30. M. de Dieuleveult *et al*., Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. *Nature* **530**, 113–116 (2016).
31. C. A. Davis *et al*., The encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).
32. D. J. Epstein, A. P. McMahon, A. L. Joyner, Regionalization of Sonic hedgehog transcription along the anteroposterior axis of the mouse central nervous system is regulated by Hnf3 dependent and independent mechanisms. *Development* (1999). http://dev.biologists.org/content/develop/126/2/281.full.pdf. Accessed 17 October 2018.
33. K. Kraft *et al*., Deletions, inversions, duplications: Engineering of structural variants using CRISPR/Cas in mice. *Cell Rep.* **10**, 833–839 (2015).
34. S. Heinz *et al*., Transcription elongation can affect genome 3D structure. *Cell* **174**, 1522–1536.e22 (2018).
35. L. A. Lettice *et al*., Development of five digits is controlled by a bipartite long-range cis-regulator. *Development* **141**, 1715–1725 (2014).
36. M. Barbieri *et al*., Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16173–16178 (2012).
37. A. M. Chiariello, C. Annunziatella, S. Bianco, A. Esposito, M. Nicodemi, Polymer physics of chromosome large-scale 3D organisation. *Sci. Rep.* **6**, 29775 (2016).
38. S. Bianco *et al*., Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat. Genet.* **50**, 662–667 (2018).
39. L. A. Lettice *et al*., Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly. *Dev. Cell* **22**, 459–467 (2012).
40. E. Z. Kvon *et al*., Progressive loss of function in a limb enhancer during snake evolution. *Cell* **167**, 633–642.e11 (2016).
41. V. Narendra *et al*., CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017–1021 (2015).
42. V. Narendra, M. Bulajić, J. Dekker, E. O. Mazzoni, D. Reinberg, CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev.* **30**, 2657–2662 (2016).
43. F. Erdel, K. Rippe, Formation of chromatin subcompartments by phase separation. *Biophys. J.* **114**, 2262–2270 (2018).
44. D. Matsumaru *et al*., Genetic analysis of Hedgehog signaling in ventral body wall development and the onset of omphalocele formation. *PLoS One* **6**, e16260 (2011).
45. W. Deng *et al*., Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**, 1233–1244 (2012).
46. G. Andrey *et al*., A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* **340**, 1234167 (2013).
47. N. Lonfat, T. Montavon, F. Darbellay, S. Gitto, D. Duboule, Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. *Science* **346**, 1004–1006 (2014).
48. T. Montavon *et al*., A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
49. A. J. Will *et al*., Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). *Nat. Genet.* **49**, 1539–1545 (2017).
50. G. Andrey, M. Spielmann, "CRISPR/Cas9 genome editing in embryonic stem cells" in *Enhancer RNAs: Methods and Protocols, Methods in Molecular Biology*, Ørom UA, Ed. (Springer New York, New York, NY, 2017), pp. 221–234.
51. J. Artus, A.-K. Hadjantonakis, Generation of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. *Methods Mol. Biol.* **693**, 37–56 (2011).
52. J. D. Buenrostro, B. Wu, H. Y. Chang, W. J. Greenleaf, ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* **109**, 21.29.1–21.29.9 (2015).
53. T. Lee, S. Johnston, R. Young, Chromatin immunoprecipitation from C. elegans embryos. *Nat. Protoc.* **1**, 729–748 (2006).
54. I. Jerković *et al*., Genome-wide binding of posterior HOXA/D transcription factors reveals subgrouping and association with CTCF. *PLoS Genet.* **13**, e1006567 (2017).
55. H. J. G. van de Werken *et al*., 4C technology: Protocols and data analysis. *Methods Enzymol.* **513**, 89–112 (2012).
56. M. Franke *et al*., Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).
57. K. J. Livak, T. D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Δ Δ C(T)) Method. *Methods* **25**, 402–408 (2001).
58. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
59. F. Ramírez *et al*., deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-5 (2016).
60. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
61. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
62. S. Wingett *et al*., HiCUP: Pipeline for mapping and processing Hi-C data. *F1000 Res.* **4**, 1310 (2015).
63. N. C. Durand *et al*., Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
64. P. A. Knight, D. Ruiz, A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* **33**, 1029–1047 (2013).
65. T. Amano *et al*., Two Types of Etiological Mutation in the Limb-Specific Enhancer of Shh. *G3* **7**, 2991–2998 (2017).