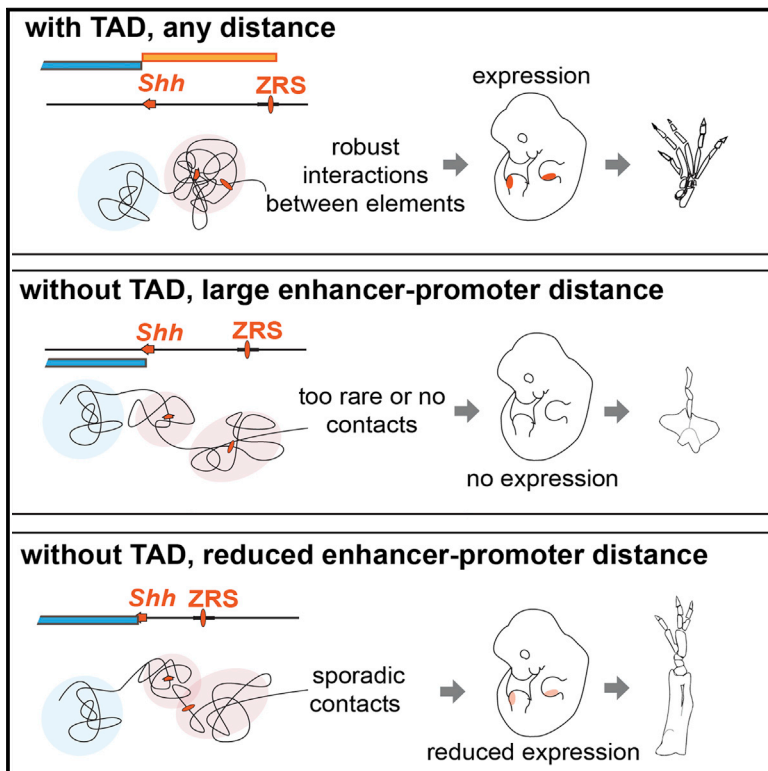


# Developmental Cell

## The *Shh* Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances

### Graphical Abstract



### Authors

Orsolya Symmons, Leslie Pan, Silvia Remeseiro, Tugce Aktas, Felix Klein, Wolfgang Huber, François Spitz

### Correspondence

francois.spitz@pasteur.fr

### In Brief

Using systematic genomic rearrangements at the *Shh* locus, Symmons et al. demonstrate that the topologically associating domain overlapping the locus acts as a functional regulatory mold by promoting contacts between elements that would otherwise be too rare to yield a functional outcome due to the dampening effects of large genomic distances.

### Highlights

- Potential for responsiveness to *Shh* enhancers correlates with contact frequency
- Changing distances within *Shh* TAD has little effect on *Shh* expression
- Inversions bridging the *Shh* TAD boundaries disrupt folding and regulatory contacts
- Decreasing genomic distance can rescue disrupted long-range regulatory interactions



# The *Shh* Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances

Orsolya Symmons,<sup>1,3</sup> Leslie Pan,<sup>1</sup> Silvia Remeseiro,<sup>1</sup> Tugce Aktas,<sup>1,4</sup> Felix Klein,<sup>2</sup> Wolfgang Huber,<sup>2</sup> and François Spitz<sup>1,2,5,6,\*</sup>

<sup>1</sup>Developmental Biology Unit

<sup>2</sup>Genome Biology Unit

European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

<sup>3</sup>Present address: Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>4</sup>Present address: Department of Chromatin Regulation, Max-Planck-Institut für Immunbiologie, 79108 Freiburg, Germany

<sup>5</sup>Present address: Department of Developmental and Stem Cell Biology, Institut Pasteur, 75015 Paris, France

<sup>6</sup>Lead Contact

\*Correspondence: [francois.spitz@pasteur.fr](mailto:francois.spitz@pasteur.fr)

<http://dx.doi.org/10.1016/j.devcel.2016.10.015>

## SUMMARY

Gene expression often requires interaction between promoters and distant enhancers, which occur within the context of highly organized topologically associating domains (TADs). Using a series of engineered chromosomal rearrangements at the *Shh* locus, we carried out an extensive fine-scale characterization of the factors that govern the long-range regulatory interactions controlling *Shh* expression. We show that *Shh* enhancers act pervasively, yet not uniformly, throughout the TAD. Importantly, changing intra-TAD distances had no impact on *Shh* expression. In contrast, inversions disrupting the TAD altered global folding of the region and prevented regulatory contacts in a distance-dependent manner. Our data indicate that the *Shh* TAD promotes distance-independent contacts between distant regions that would otherwise interact only sporadically, enabling functional communication between them. In large genomes where genomic distances per se can limit regulatory interactions, this function of TADs could be as essential for gene expression as the formation of insulated neighborhoods.

## INTRODUCTION

A substantial fraction of gene regulatory elements lie at considerable distance from the nearest promoters (ENCODE Project Consortium et al., 2012; Shen et al., 2012; Visel et al., 2007). While the contribution of these elements to gene expression is generally difficult to estimate, enhancers located hundreds of kilobases from their target genes but essential to their expression are increasingly identified (Sagai et al., 2009, 2005; Spitz et al., 2003; Uslu et al., 2014; Wunderle et al., 1998; Zuniga et al., 2004) (reviewed in de Laat and Duboule, 2013; Visel et al., 2009). Accordingly, mutations or genetic variants in

distant enhancers are a significant cause of genetic diseases (Benko et al., 2009; Bhatia et al., 2013; D'haene et al., 2009; Lettice et al., 2003) and contribute to intra-species (Bauer et al., 2013; Smemo et al., 2014; Sur et al., 2012; Wasserman et al., 2010) and inter-species (Prescott et al., 2015; Prud'homme et al., 2007) phenotypic variability. Although our understanding of regulatory elements has improved tremendously in recent years, it remains unclear how enhancers find a specific target located several hundred kilobases away. There is strong evidence that such interactions require physical proximity (Deng et al., 2012). Yet, how this proximity is established and regulated and how it influences target gene expression is still poorly understood.

Concomitant with the growing appreciation of distant regulatory sequences, improved chromosome conformation capture techniques have provided insights into the three-dimensional organization of the genome and *cis*-interaction networks between genes and surrounding elements (Hughes et al., 2014; Kieffer-Kwon et al., 2013; Li et al., 2012; Mifsud et al., 2015). These approaches have revealed not only loops between distant elements but also that mammalian genomes are partitioned into sub-megabase-sized domains referred to as topologically associating domains or TADs (Dixon et al., 2012; Nora et al., 2012). Several indirect lines of evidence suggest that these self-interacting regions may represent the core units of genome regulatory architecture (Gibcus and Dekker, 2013); a large proportion of TAD boundaries are shared between cell types (Dixon et al., 2015) and largely preserved during evolution (Vietri Rudan et al., 2015). Coordinately regulated tissue-specific enhancer-promoter pairs (Shen et al., 2012) and associated long-range looping interactions (Downen et al., 2014; Jin et al., 2013) are usually comprised within TADs. The regulatory domains defined by enhancers' range of action coincide also largely with TADs (Symmons et al., 2014). Although internal interactions within TADs can be cell-type specific and activity dependent (Phillips-Cremins et al., 2013; Rao et al., 2014), these different findings support the role of TADs as basic structural and functional units.

Correlations between regulatory and structural subdivisions of the genome suggest that TADs may constrain the range of



action of enhancers, with TAD boundaries acting as functional “insulators” (Chetverina et al., 2013; Yang and Corces, 2012). TAD boundaries are indeed enriched for elements shown to have insulator activity (such as CTCF binding sites and transcriptional start sites) (Dixon et al., 2012), and insertions of a sensor gene on opposing sides of TAD boundaries show distinct expression patterns (Symmons et al., 2014; Tsujimura et al., 2015). Recent experiments deleting or altering these boundaries showed expansion of chromosomal contacts across the former boundaries, leading to ectopic activation of neighboring genes (Downen et al., 2014; Lupiáñez et al., 2015; Narendra et al., 2015; Tsujimura et al., 2015). Similarly, the consequences of multiple human pathological chromosomal rearrangements can be explained by modification of TAD boundary positions and subsequent enhancer adoption by non-target genes (Flavahan et al., 2015; Hnisz et al., 2016). Together, these experiments clearly established that TAD boundaries are essential for generating isolated domains of regulatory activities. However, other features and potential roles of TADs remain poorly studied.

The *Shh* locus constitutes an ideal system to study long-range enhancer-promoter regulation. *Shh* expression is regulated by a series of tissue-specific enhancers distributed across a region spanning over 900 kb, which also comprises other unrelated genes (Jeong et al., 2006; Lettice et al., 2003; Sagai et al., 2009) and which corresponds to a tissue-invariant and evolutionary conserved TAD (Dixon et al., 2012; Jin et al., 2013). In particular, the specific expression of *Shh* in the zone of polarizing activity (ZPA), which establishes antero-posterior patterning of the developing limbs, is fully determined by the activity of a single *cis*-acting enhancer (Jeong et al., 2006; Lettice et al., 2003; Sagai et al., 2005). This element, the ZRS, lies 850 kilobases away from the *Shh* promoter, in an intron of an unrelated gene, *Lmbr1* (Lettice et al., 2003). In the present work, we took advantage of this prototypic enhancer-promoter pair to study the relationship between distant enhancer-promoter interactions, 3D conformation, and gene expression. We generated a series of mouse strains carrying tagged and structurally rearranged alleles of this locus. We analyzed them *in vivo*, when the mechanisms associated with its regulation are functional and biologically relevant, and *in situ*, in the genomic context where they evolved and normally operate. Our results showed that enhancer-promoter loops occurred within the framework of much more promiscuous contacts, where enhancers scan the entire topological domain they are part of. Remarkably, altering enhancer-promoter distances in the context of the *Shh* TAD did not appear to affect *Shh* expression. In contrast, disruption of the TAD prevented physical and regulatory interactions between *Shh* and its limb enhancer, unless the genomic distance between the two was significantly reduced. Our observations provide evidence that TADs ensure high contact frequency between distant elements by counteracting the effect of genomic distances. TADs do not simply restrict enhancer activity to a specific region to prevent ectopic interactions. They also provide the spatial proximity that is essential for efficient action of remote enhancers on genes located within the same TAD. This regulatory role of TADs can be particularly important in large genomes and may have enabled

expansion of the genomic space available for regulatory innovation during evolution.

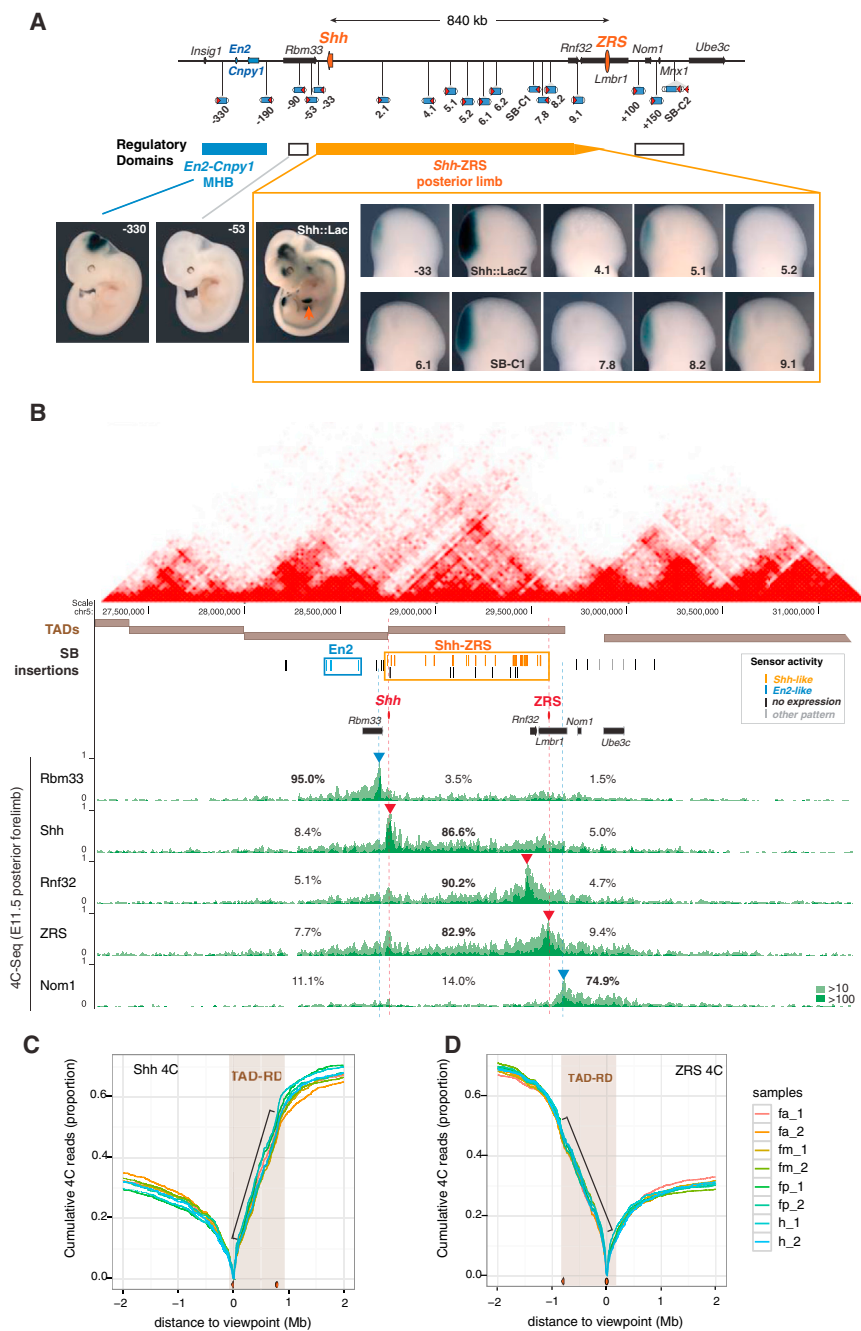
## RESULTS

### The *Shh* Regulatory Domain: Extended but Variable Responsiveness to Enhancers

We had previously shown that insertions of a regulatory sensor (Ruf et al., 2011) at the *Shh* locus show a *Shh*-like expression pattern, and reveal a large regulatory domain that overlaps with the TAD at this locus (Symmons et al., 2014). To get further information on how *Shh* enhancers act within this domain, we generated additional insertions of the regulatory sensor and analyzed its expression in mouse embryos at stages E10–12 (Figures 1A and S1, Table S1). The comparison of the patterns observed with 59 different insertions across the *Shh* genomic region provides a fine-scale view of its regulatory architecture, extending our first observations (Symmons et al., 2014) and those performed with a different promoter (Anderson et al., 2014). Noteworthy, like at other loci (Ruf et al., 2011), the insertion of this naive sensor did not alter the expression of *Shh* or surrounding genes (Figures S1C and S1D), indicating that the activity of the sensor does not trap enhancers away from *Shh*. Instead, the sensor reveals the pre-existing potential of surrounding enhancers to act on a given genomic position.

We found that in the region beginning 33 kb downstream of *Shh* and extending to the ZRS, most insertions showed expression patterns that closely matched *Shh* expression in the limb (Figure 1A) as well as in other tissues (Figure S1). Outside of this *Shh* regulatory domain, insertions showed no expression or a divergent one. The *Shh* expression patterns detected by insertions in the *Shh* regulatory domain included domains for which enhancers have been mapped (Jeong et al., 2006; Sagai et al., 2009) (Figure S1), as well as domains for which no enhancers have been identified to date (e.g., choroid plexus; Figure S1G). This widespread responsiveness indicated that most *Shh* enhancers can act long range and not only in their vicinity or close to the *Shh* promoter. We found that expression of the sensor at a given position was highly reproducible, both when comparing littermate embryos and in independent replicate experiments (Table S1). Yet, lacZ stainings of neighboring insertions can sometimes differ extensively, even when only a few kilobases apart (Figures 1A and S1). A small number of insertions within the regulatory domain, such as insertion 5.2, showed no expression in any tissue. But more typically, variation was quantitative and differed depending on the tissue. For example, at position 5.1, we observed robust expression in the notochord and floor plate, but only weak staining in the limb and in the genital bud; at position C1, we observed the reverse relative intensities (Figures S1B, S1E and S1F). The reporter insertion at position –33 showed high responsiveness to the ZRS but not to other enhancers (Figure S1). We also observed this quantitative variability at stages other than E11.5 (Figures S1G and S1H). Our data show that, for the same promoter, the responsiveness to enhancer(s) can vary extensively within an otherwise largely permissive regulatory domain.

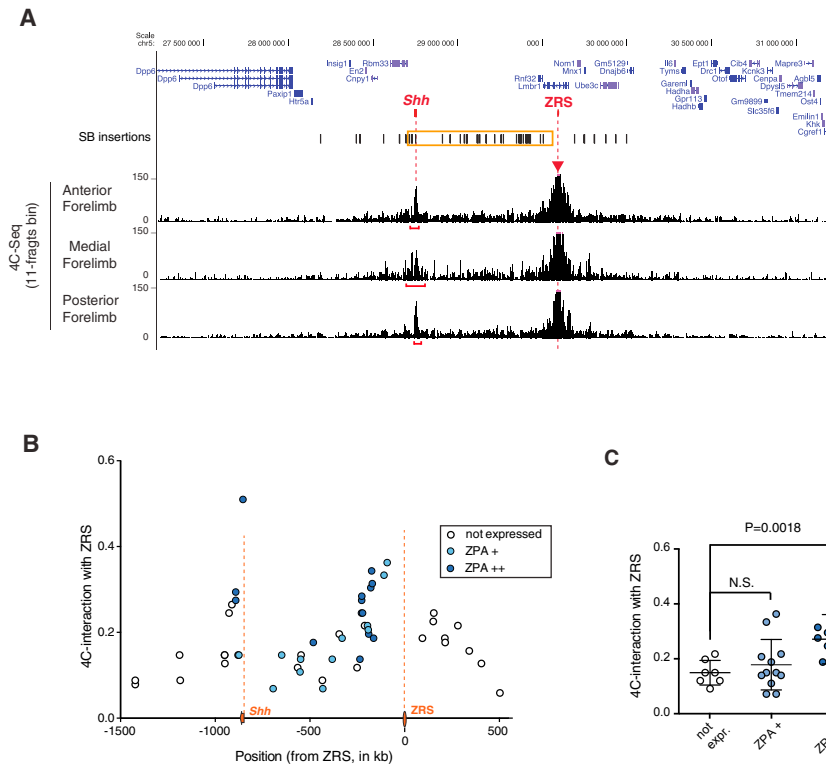
To understand what factors modulate responsiveness to regulatory inputs, we focused on the limb where *Shh* expression is



### Figure 1. Topological and Regulatory Organization of the *Shh* Locus

(A) Schematic representation of the *Shh* regulatory domain, as defined by the collection of 60 transposon insertions obtained with GROMIT. The location of representative insertions and their expression patterns is shown. Bars represent regulatory domains (*En2-Cnpy1*, *Shh*), as outlined by expression patterns reminiscent of the ones of the associated endogenous genes. White bars indicate that insertions in those regions have no expression. Orange arrowhead indicates the ZPA.

(B) The *Shh* regulatory domain compared with the 3D conformation of the locus. Hi-C map of the locus from CH12 cells (Rao et al., 2014) (red contact maps, image generated with 3Dgenome browser, <http://www.3dgenome.org>) and TADs identified in ESCs (Dixon et al., 2012) (brown bars) are shown. Position and activity of insertions are indicated by colored lines (orange, *Shh*-like expression; blue, *En2*-like expression; black, no expression; gray, other/non-attributed expression). Corresponding regulatory domains are boxed. Shown beneath are 4C-interaction profiles (hit percentage with 10 and 100 count thresholds in light and dark green, respectively) of three viewpoints (*Shh*, *Rnf32*, ZRS, red arrowheads and lines) located in the regulatory domain and of two viewpoints (*Rbm33*, *Nom1*, blue arrowheads and lines) flanking it. For each viewpoint, we indicate the percentage of reads from regions in the *Shh* domain or from the 1 Mb flanking regions. (C and D) Cumulative 4C read counts as a function of distance from the *Shh* viewpoint (C) or the ZRS viewpoint (D). Data from different microdissected limb compartments is shown in different colors (fa, anterior forelimb; fm, medial forelimb; fp, posterior forelimb; h, hindlimb, 1 and 2 indicate biological replicates), the TAD/regulatory domain is highlighted in brown and the black bar indicates the constant slope of the curve.



## Figure 2. Organization of the *Shh* Locus and Responsiveness to ZRS

(A) 4C-seq interaction profile (read counts, binned in 11-fragment sliding windows) of the ZRS (viewpoint indicated by red triangle) in the anterior, medial, and posterior compartments of E11.5 forelimbs. A red bar underlines the peak contact region around *Shh*. For further comparisons, see also Figure S2.

(B) ZRS-interaction values at the insertion points of the transposon (in the absence of the transposon). x axis, distance to the ZRS; y axis, 4C-interaction score; dot color represents intensity of LacZ staining in the ZPA.

(C) Comparison of interaction scores with responsiveness to the ZRS for positions within the *Shh* TAD (not expressed versus strongly expressed in ZPA;  $p = 0.0018$ , two-sided Mann-Whitney test).

determined by a single enhancer, in contrast to many other tissues where it is associated with several enhancers with overlapping activities (Jeong et al., 2006; Sagai et al., 2009; Tsukiji et al., 2014). Critically, similarly to other tissues, the sensor showed significant variability in ZPA expression at different positions within the *Shh* TAD (Table S2). This indicated that variability in responsiveness is not limited to complex situations involving multiple enhancers. Responsiveness to the ZRS showed no correlation with linear distance to the ZRS nor did it appear to be influenced by the orientation of the sensor (Fisher exact test  $p = 0.387$ ) or local chromatin features (proximity to repeat elements such as LINES or SINEs; accessibility measured by DNaseI hypersensitivity or chromatin acetylation/methylation) (Table S2, data not shown) signifying that it is determined by other factors.

### Regulatory and Topological Domains Coincide at the *Shh* Locus

As noted before (Anderson et al., 2014; Symmons et al., 2014), the *Shh* regulatory domain shows strong overlap with an underlying TAD, conserved in different cell lines (Dixon et al., 2012; Jin et al., 2013; Rao et al., 2014) (Figure 1B). Since 3D conformations can vary between cell types, we performed chromosome conformation capture sequencing experiments (4C-seq) on the posterior compartment of E11.5 microdissected limb buds (Figures 1B and S2). We chose viewpoints within the TAD (*Shh* promoter, ZRS, and *Rnf32*) and outside (*Rbm33*, *Nom1*). In the posterior limb, the three viewpoints located in the *Shh* TAD showed prominent contacts along the entire TAD, while contact frequency with regions outside the TAD decreased quickly (Figure 1B). Reciprocally, the

viewpoints located immediately outside the TAD, on either side, showed limited interactions with sequences in the *Shh* TAD (Figure 1B). Both the ZRS and *Shh* had contact frequencies that remained constantly high throughout the TAD, with limited effect of genomic distance (Figures 1C and 1D). Together, these data show that the *Shh*-ZRS region forms a self-interacting chromatin domain in

the posterior limb bud, which corresponds well to the TAD described in other cell types.

### The Potential for Responsiveness Is Influenced by the 3D Organization

Even among the generally robust interactions detected along the TAD, the ZRS showed a particularly stronger interaction with the *Shh* promoter in the posterior limb bud (Figure 2A), in agreement with previous 3C and FISH data (Amano et al., 2009). Interestingly, based on 4C-seq, the compartmentalization of interactions and the fine-scale interactions of the ZRS did not appear very different between E11.5 posterior limbs (where *Shh* and the ZRS are active) and E11.5 anterior and medial forelimb samples (where *Shh* and the ZRS are inactive) (Figures 1C, 1D, and S2); the ZRS showed stronger contact with *Shh* in all limb compartments (Figure 2A), although the interaction peaks appeared more diffuse in the inactive situations than in the ZPA. Hence, similarly to other loci (Ghavi-Helm et al., 2014; Montavon et al., 2011), conformation and enhancer-target gene contacts appear to be in part constitutive and independent of transcriptional and regulatory activity.

To understand if responsiveness of the regulatory sensor to the ZRS was related to the native folding of the locus, we compared the interaction profile of the ZRS (from wild-type limb, without sensor insertions) to the expression of the sensor in the ZPA at the different insertion sites (Figures 2B and Table S2). We found that, within the *Shh* TAD, positions with ZPA expression had overall stronger contacts with the ZRS than weakly or non-expressed ones (Figure 2C). This correlation indicates a relationship between the distribution of enhancer activity and the native structural folding of the locus. It should be noted,

however, that some positions contacted by the ZRS with similar efficiency (as measured by 4C, at a resolution of 5–10 kb) displayed different activation potential, indicating that average contact frequency is not the sole determinant for regulatory activation.

### TADs Buffer the Effect of Genomic Distances between Promoters and Enhancers

To further identify the mechanisms that govern distant interactions, we decided to systematically change different genomic parameters of the locus. First, we modulated the genomic distance separating the ZRS from *Shh*, while respecting the TAD boundaries. To this end, we engineered mice carrying either intra-TAD deletions or duplications, using *Cre-loxP* in vivo recombineering (Hérault et al., 1998) (Figures 3A and 3B). We then assessed limb morphology and *Shh* expression in animals carrying these rearrangements over a *Shh* null allele (*Shh<sup>del</sup>*) (Gonzalez-Reyes et al., 2012).

In agreement with ZRS deletion alleles (Lettice et al., 2014; Sagai et al., 2005), deletions that included the ZRS completely abolished limb expression of *Shh* (Figures 3C and S3A) and led to fore- and hindlimb monodactyly (Figures 3D and S3). We also observed loss of limb expression of the inserted sensor gene retained in DEL(C1-Z), showing that the remaining region comprised no limb enhancer. In contrast, compound embryos carrying either the DEL(5–8) deletion, which reduced the distance between *Shh* and the ZRS by 260 kb, or duplications that increased the distance to 1.1 Mb, DUP(5–8), DUP(C1-Z), showed normal limb morphology (Figures 3E and S3G). We did not detect major changes in *Shh* expression in E10.5 forelimbs as assessed by in situ hybridization (Figure 3F) and RT-qPCR (Figures 3G–3I, S3D, and S3F). We observed a slight reduction of *Shh* expression with the DUP(C1-Z) allele (*Shh*-ZRS distance = 1.08 Mb), but since the other duplication of similar size, DUP(5–8) (*Shh*-ZRS distance = 1.11 Mb), showed normal levels of *Shh*, this effect cannot be due solely to the increased distance. The difference between the two duplications may stem from the extra copy of *Rnf32* in DUP(C1-Z), which could act as a competitor for ZRS activity. However, previous reports have shown that *Rnf32* is not regulated by the ZRS (Amano et al., 2009), and we did not detect upregulation of *Rnf32* (Figure 3I) beyond the 1.5-fold increase that corresponds to the increase in *Rnf32* copy number from 2 in *Shh<sup>del/+</sup>* to 3 in *Shh<sup>del</sup>/DUP(1C-Z)*.

Altogether, *Shh* expression appeared largely resilient to changes in enhancer-promoter distances when TAD boundaries were left unchanged and when no element normally external to the *Shh* TAD was introduced.

### TAD Content Influences the Distribution of Enhancer Responsiveness

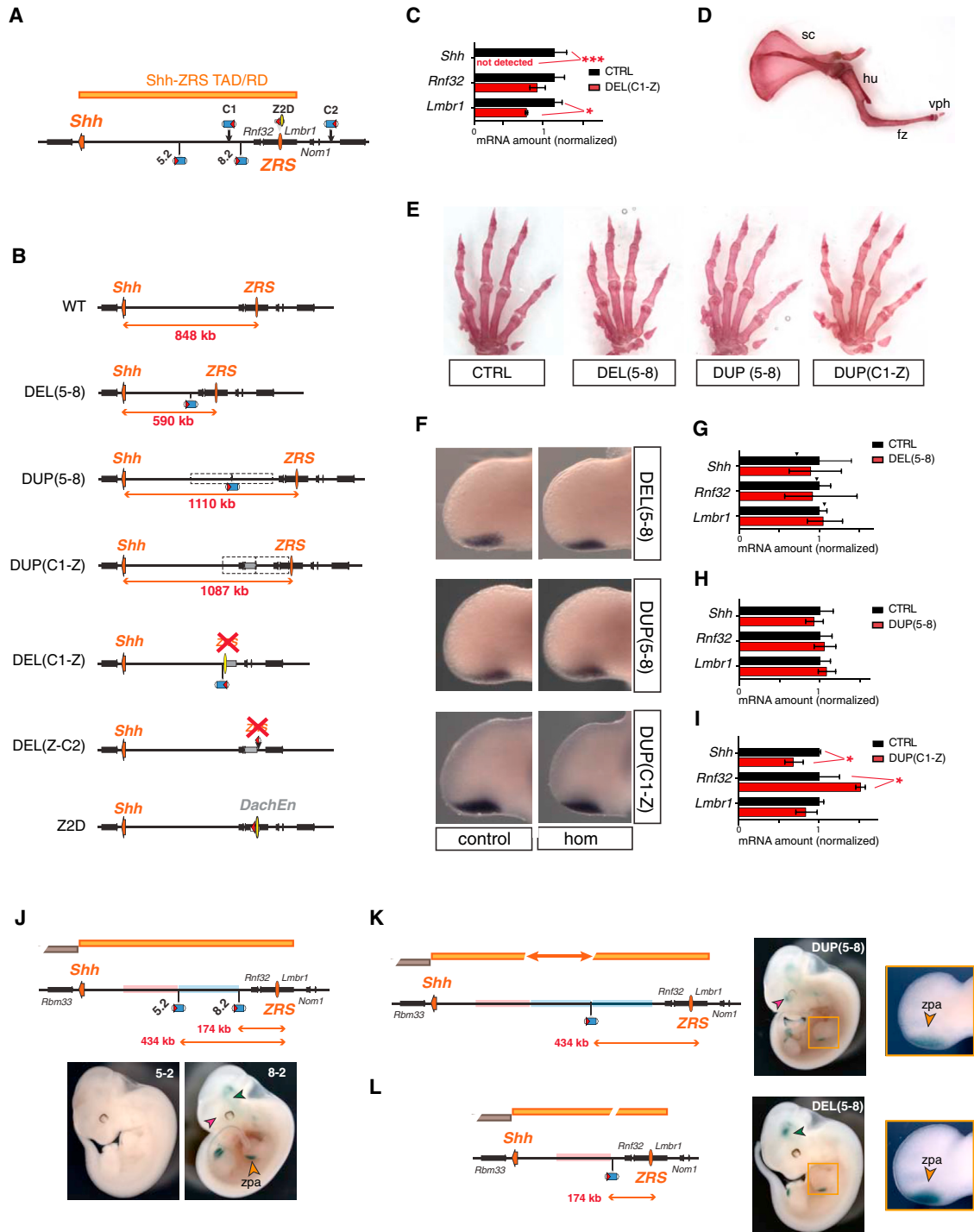
We next examined if these intra-TAD rearrangements, which showed no major impact on *Shh* expression, could nonetheless alter the distribution of ZRS responsiveness. Genes outside the *Shh* TAD (*Lmbr1*, *Rbm33* or *Nom1*) showed no significant expression changes in any of these genomic configurations (Figure S3), showing that confinement of enhancer activity is maintained. To look at the responsiveness within the TAD, we compared the expression of the regulatory sensor in the native context and in the context of the genomic rearrangements

(Figure 3B). Prior to rearrangement, position 5.2 is refractory to activation by *Shh* enhancers, constituting one of the rare “dead spots” present in the domain, while 8.2 responds to multiple *Shh* enhancers (Figure 3J). Surprisingly, in the context of DUP(5–8), the sensor showed robust expression in the ZPA (Figure 3K), even though its position is identical to 5.2 with respect to the ZRS (same distance, same intervening sequences). Moreover, additional *Shh* expression domains (not observed at positions 5.2 and 8.2, but detected with insertions elsewhere in the locus), were also un-masked in the context of DUP(5–8) and DEL(5–8) (Figures 3K and 3L, pink and green arrowheads). Some of the new expression domains in DUP(5–8) may be associated with duplicated enhancers. But as the ZRS, the only limb enhancer active in the region, is located far outside the duplicated region, the gained expression in DUP(5–8) and in DEL(5–8) requires another explanation. We considered first that expression at position 5.2 could be locally repressed. If this is done by a centromeric repressor element, the reporter in DEL(5–8) should also be repressed. If this putative repressor was telomeric to position 5.2, then the reporter in DUP(5–8) should be repressed. Since both DEL(5–8) and DUP(5–8) show activity, the hypothesis of a local repressor at 5.2 is unlikely, as it would imply the existence of a cryptic de-repressor next to 8.2 that can counteract the repressor at 5.2. Even if we cannot fully rule out the existence of such a series of local elements, we propose that the rearrangements modulate the relative 3D folding of the TAD, and therefore change which regions are functionally exposed to the influence of the enhancers dispersed throughout this domain. This model is not only more parsimonious, but also fits well with the wide distribution of sensor cold spots, which correlates with 3D conformation.

### TAD-Breaking Inversions Disrupt Regulatory Interactions between *Shh* and the ZRS

The resilience of *Shh* expression to changes in enhancer-promoter distance can be interpreted as evidence for classical looping models, where *Shh* is directed and tethered to the ZRS. Such looping interactions could be driven by a combination of elements present at the enhancer (Lettice et al., 2014) or the promoter (Calhoun and Levine, 2003; Kwon et al., 2009; Williamson et al., 2011; Zabidi et al., 2015). To test these models, we engineered balanced inversions that should split the *Shh*-ZRS TAD, while keeping *Shh* within the range of action of the ZRS defined by the previous experiments (Figures 4A, 4B, and S4). INV(-500-C1) exchanged sequences between the *Shh* TAD and the centromeric *En2-Rbm33* TAD, while INV(6-C2) interspersed a region telomeric to the *Shh* TAD between the two halves of the original *Shh* TAD. In both cases, *Shh*-ZRS distances remained below 850 kb.

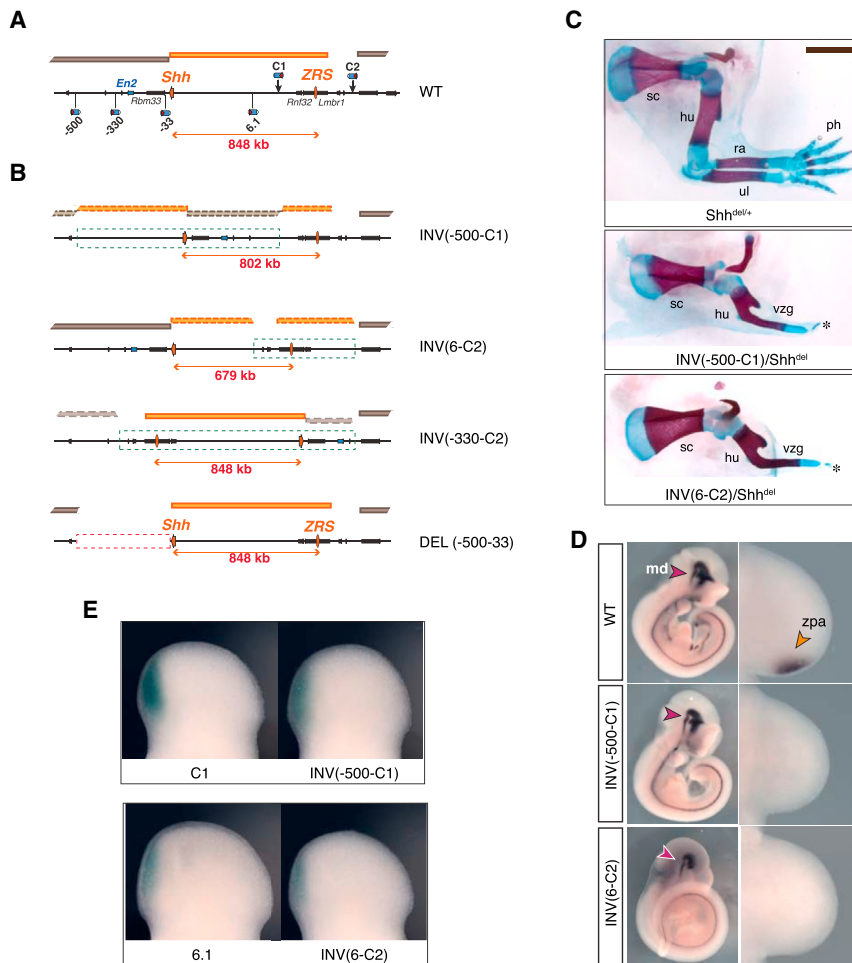
Animals carrying these inversions over a deletion of *Shh* (Figure 4C) or an inactivating substitution of the ZRS (Figures S4B and S4C) showed monodactyly on both fore- and hindlimbs. Expression of *Shh* was lost in the limb of E10 embryos homozygous for either inversion (Figure 4D). Importantly, *Shh* expression was detected in other tissues (Figure 4D), showing that the gene was not globally repressed. Furthermore, in both configurations, the associated regulatory sensor remained at the same position relative to the ZRS before and after inversion and maintained expression in the ZPA (Figures 4E, S4D, and S4F), indicating



**Figure 3. Changing Distances within the *Shh* TAD**

(A) Schematic representation of the region, including the different insertion points and *loxP* sites used.  
 (B) Schematic representation of the rearranged alleles. The distance separating the ZRS (orange oval) from *Shh* is indicated. The transposon at the junction point (when retained) is indicated, and dashed rectangles mark the duplicated regions. The Z2D allele is a replacement of the ZRS by another limb enhancer (yellow oval, *DachEn*/hs126; [Visel et al., 2007](#)), which appeared to be essentially inactive when inserted at this position (Figure S3G).  
 (C) Gene expression by RT-qPCR in DEL(C1-Z) versus WT E11 forelimb buds (for each gene, reference value in WT set as 1, the error bars correspond to SEM. Statistical significance done with *t* tests: \**p* < 0.05; \*\*\**p* < 0.001).  
 (D) Forelimb skeleton of a DEL(Z-C2)/*Shh*<sup>del</sup> mouse showing monodactyly and fused zeugopod. sc, scapula; hu, humerus; fz, fused zeugopod; vph, vestigial phalanges.

(legend continued on next page)



**Figure 4. Consequences of TAD Disrupting Alleles**

(A) Representation of the insertions used to produce the inversions and del(-500-33). The *Shh*-ZRS TAD is in orange, the flanking ones in brown. (B) Representations of the rearranged alleles, with the inverted and deleted regions outlined by dashed green and red boxes, respectively. The linear distance between *Shh* and the ZRS is indicated. Dashed orange and brown blocs indicate the segment corresponding to former TADs. (C) Forelimb skeletons of E18.5 embryos for *Shh<sup>del/+</sup>* (control), *INV(-500-C1)/Shh<sup>del</sup>*, and *INV(6-C2)/Shh<sup>del</sup>*, with the latter two showing the typical *Shh* loss of function limb phenotype. sc, scapula; hu, humerus; ra, radius; ul, ulna; vzg, fused zeugopod; \*, single vestigial digit. (D) Expression of *Shh* in E10 mouse embryos by in situ hybridization. The orange arrowhead indicates expression in the ZPA, the purple one expression in the ventral midbrain (md). (E) LacZ expression in E11.5 autopods of embryos carrying the starting insertions in normal (C1, 6.1) or inverted configurations, *INV(-500-C1)*, *INV(6-C2)*. See also Figure S4.

Overall, our experiments argue against the presence of a strong specific recognition system that will suffice to bring together *Shh* and the ZRS, as shown by the lack of *Shh* limb expression in *INV(6-C2)* and *INV(-500-C1)*, despite shorter genomic distances than normal. These inversions, which reshuffled sections of different TADs, were the only ones from our series of rearrangements

that the endogenous activity of the ZRS is unaltered. *Shh* loss of expression is also unlikely to result from the disruption of a specific accessory element, since the two inversions used different breakpoints. As further controls, we produced two additional rearrangements, *INV(-330-C2)* and *DEL(-500-33)*, this time changing the sequences flanking the *Shh* TAD without modifying the TAD itself (Figure 4B). In both configurations, *Shh* expression and function appeared unaffected (Figures S4G and S4H), which led us to conclude that disruption of the neighboring domains had minimal effect on *Shh* regulation.

that affected *Shh*-ZRS communication, further strengthening the importance of TADs as regulatory units.

#### TAD-Breaking Balanced Inversions Affect the Global Topology of the Locus

To assess the consequences of the TAD-reshuffling inversions on the topology of the locus, we repeated the 4C analysis in E11.5 limbs for the *INV(6-C2)* allele (Figures 5 and S5). To account for the loss of *Shh* expression following inversion, we compared *INV(6-C2)* forelimb 4C profiles with the ones

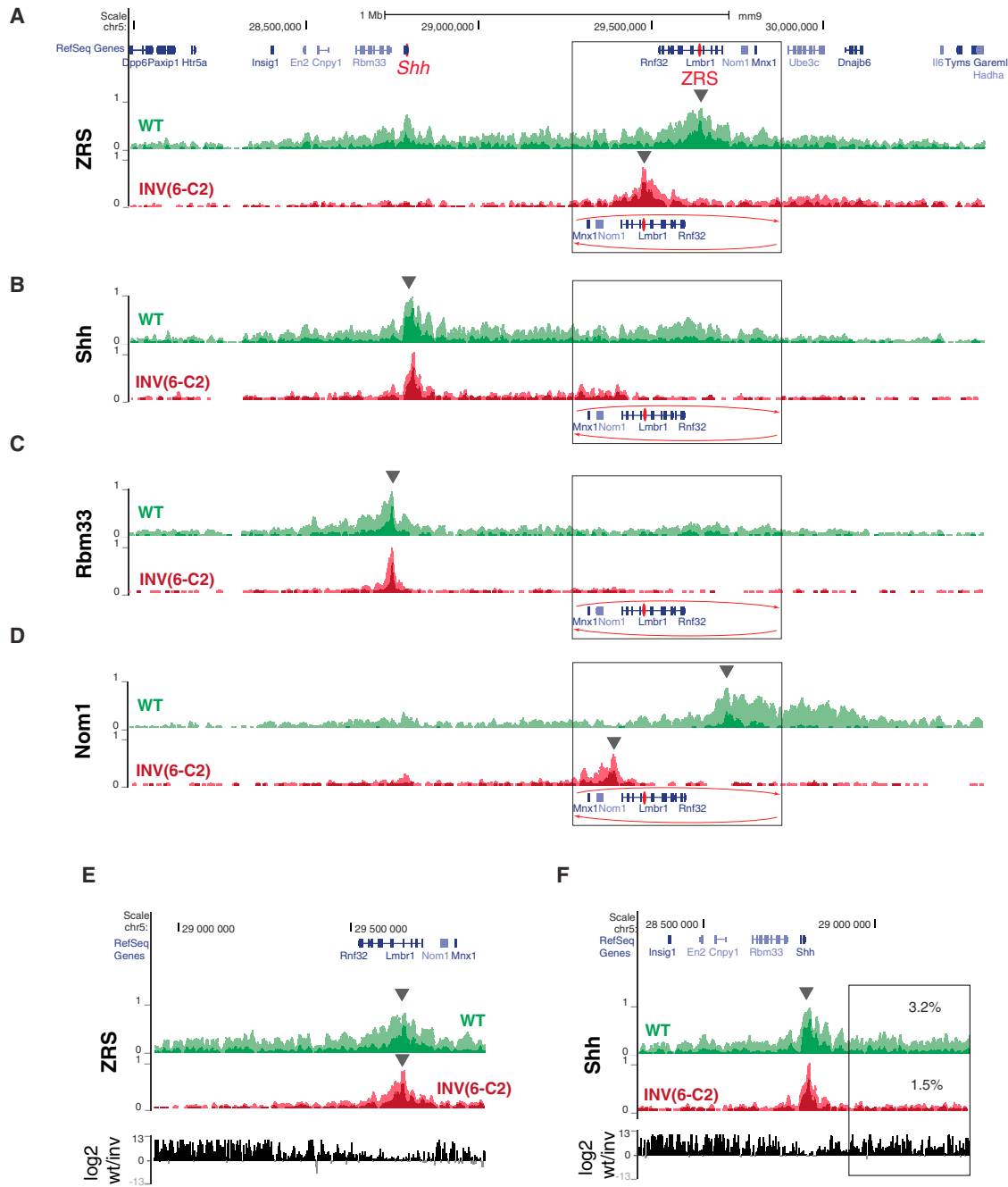
(E) Hand skeletons of adult mice with different rearranged alleles. Alleles are in *trans* of either *Shh<sup>del</sup>* (for DUP(5-8) and DUP(C1-Z)) or of a ZRS replacement (Z2D allele, for CTRL and DEL(5-8)), because DEL(5-8) homozygous or compound mutants with *Shh<sup>del</sup>* die at birth due to holoprosencephaly and cranial defect (not shown).

(F) Expression of *Shh* in E10.5 forelimbs in the different alleles. For each line, in situ hybridization was performed on wild-type control littermates and homozygous mutants ( $n = 3$ ).

(G-I) RT-qPCR data in DEL(5-8) (G), DUP(5-8) (H), and DUP(C1-Z) (I) E11 forelimb buds. Homozygous mutant samples are in red ( $n = 3$ ), stage-matched wild-type samples from the same litters ( $n = 3$ ) are used as control, except for (G), where wild-type samples include embryos from separate litters (the arrowheads indicate the expression level in wild-type littermates of the mutants). The error bars correspond to SEM. \* $p < 0.05$  (t test).

(J-L) LacZ staining of E11.5 embryos with insertions of the sensor at positions 5.2, 8.2, and in the context of DEL(5-8) and DUP(5-8). On the schematic representation of the alleles, the *Shh*-ZRS TAD is in orange, and red and blue rectangles label the centromeric and telomeric flanking regions of 5.2, respectively. The sensor showed expression in the ZPA in 8.2 DUP(5-8) and DEL(5-8) embryos (orange arrowheads and insets) but not in 5.2 embryos. Expression domains observed in DUP(5-8) or DEL(5-8), but in none of the starting insertions, are labeled with pink and green arrowheads, respectively. For further gene expression and phenotypic data. See also Figure S3.





**Figure 5. 4C Profiles in INV(6-C2) Alleles**

(A–D) For each viewpoint, the hit percent profiles (with 10 and 100 count thresholds in light and dark color, respectively) obtained from WT (green) and INV (red) samples are plotted on their respective genomic configurations (i.e., with an inversion of the [6-C2] genomic segment for INV). The inverted region is boxed, and the new position of the genes in the INV allele is depicted. The viewpoints are indicated by black arrowheads. To take into account the loss of *Shh* expression and monoactivity in INV(6-C2), we compared INV whole forelimbs with WT anterior forelimb compartments.

(E) Comparison of the interaction profile of the ZRS between WT and INV in the inverted region (plotted with the same orientation).

(F) Same comparison as in (E) for the interaction profile of *Shh* between WT and INV. The box delimits the intra-TAD segment not affected by the inversion and the percentage of counts contained within it.

See also Figure S5.

obtained from the anterior compartment of E11.5 WT limbs. We found that the reciprocal interaction peaks between *Shh* and the ZRS found on the WT allele were lost in INV(6-C2) (Figures

5A and 5B). In addition, *Shh* and ZRS contacts became mostly local, focused in ~100 kb around each viewpoint, replacing the broadly distributed contacts throughout the *Shh*-ZRS TAD

characteristic of the wild-type chromosome (Figures 5E and 5F). This was particularly striking for the region between *Shh* and the inversion breakpoint, since the linear organization of this segment is not directly changed by the inversion (Figure 5B). In INV(6-C2), *Shh* showed some interactions with the *Mnx1*, *Nom1*, and *Lmbr1* promoters, but only marginally above what was observed in WT (Figures 5B and 5D), particularly if the reduced distance is considered. We did not observe broad reciprocal contacts between the *Lmbr1-Nom1* viewpoint and the *Shh* region, which could have indicated the reformation of a new TAD, as described at other loci (Lupiáñez et al., 2015; Tsujimura et al., 2015). Instead, in this case, the global reduction of contact frequency suggests that the overall conformation of the locus is disrupted, changing from a dense network of interactions along the TAD, either into a diffuse structure or into small, independent domains. Interestingly, the flanking viewpoints *Nom1* and *Rbm33* still displayed an asymmetric distribution of their contacts, avoiding the *Shh*-ZRS interval, indicating that some “insulating” aspects were retained (Figures 5C and 5D).

### Insulators versus Distance Effects

Our data from the INV(6-C2) and INV(-550-C1) inversions showed that gross disruption of the *Shh* domain abolished the ability of the ZRS to contact and activate *Shh*. This loss of contact could be explained by the presence of an insulator at the telomeric end of the TAD, since it would be relocated between *Shh* and the ZRS in the inversion. To test this possibility, we generated additional inversions, which utilized the same C2 telomeric breakpoint and different centromeric breakpoints (4.2 and 2.1), located closer to *Shh* (Figure 6A). The resulting rearranged alleles repositioned the same ZRS-*Lmbr1*-C2 intervening sequence between *Shh* and the ZRS as in INV(6-C2) but displaced increasing portions of the *Shh* regulatory domain (Figures 6B–6D). At first glance, as more enhancers were moved away, we observed a progressive increase in the severity of the phenotypes (Figures 6E–6H). This was particularly obvious for the cranio-facial and axial skeletons, which showed a stepwise increase in the extent of malformations. INV(2-C2) mice essentially copied the phenotype of complete *Shh* null lines (Chiang et al., 1996) or of mice where the entire regulatory region is removed (Niedermaier et al., 2005).

Importantly, these inversions also moved the ZRS progressively closer to *Shh*. While INV(6-C2) almost fully recapitulated the ZRS null limb phenotype (Figure 6J), we observed a gradual recovery of the limb structures, especially in the hindlimb. INV(4-C2) embryos still showed severely affected monodactylous limbs (Figure 6K), but INV(2-C2) embryos showed partially restored hindlimb morphology: feet usually comprised three digits, with an anterior big toe with two phalanges and two toes with three phalanges, while the tibia-fibula elements were distinct and only partially fused (Figure 6L). These limb phenotypes imply a gradual restoration of antero-posterior polarity and growth of zeugopod and autopod structures, consistent with a partial rescue of *Shh* activity. While we were unable to detect *Shh* expression in the limb of E10.5 embryos, prior work on other ZRS mutants has shown that reducing *Shh* expression to 10% of wild-type level results in a somewhat less severe hindlimb phenotype than the INV(2-C2) embryos (Lettice et al.,

2014). Therefore, the INV(2-C2) phenotype is consistent with expression that is either extremely low or that occurs only during a very limited time period.

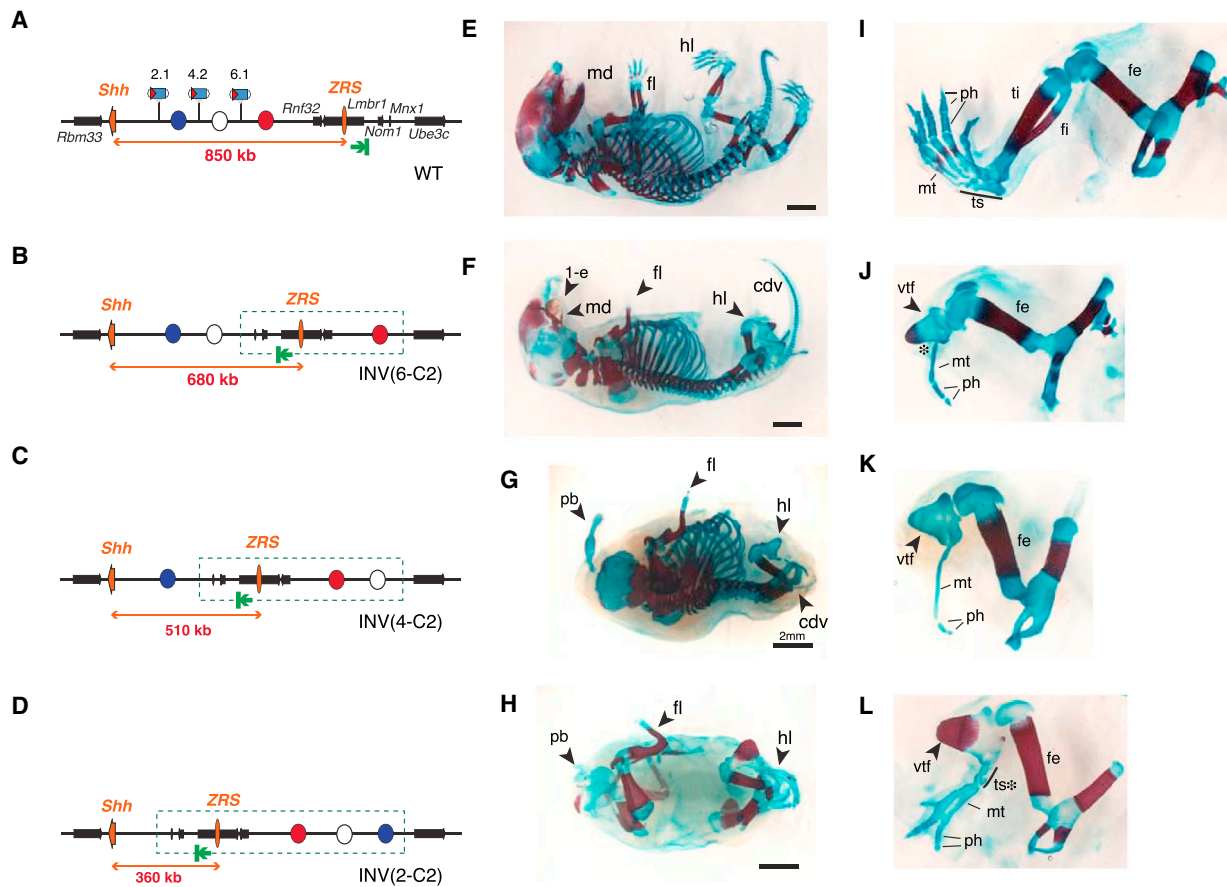
Compound mutants over an inactive ZRS allele (Z2D; Figure S6) also showed the same progressive restoration of limb morphology, indicating allelism to ZRS activity. In brief, this allelic series reveals that reducing *Shh*-ZRS distance can restore functional interactions between these elements, and that the presence of the *Lmbr1*-C2 region is not sufficient to block these interactions.

## DISCUSSION

Although the ability of enhancers to act in a distance-independent manner is part of their original definition (Banerji et al., 1981), this property was established on plasmid assays (i.e., at distances up to 10 kb). In their native genomic environment, enhancers have been shown to select their target gene through mechanisms influenced by proximity (Dillon et al., 1997; Kmita et al., 2002), even though promoter preference (Butler and Kadonaga, 2001; Ohtsuki et al., 1998; Zabidi et al., 2015), occupancy by specific transcription factors (Deng et al., 2012), and/or tethering elements (Calhoun et al., 2002) may modulate these effects. Our present study of the *Shh* locus provides new insights into the organizing principles of long-distance enhancer-promoter interactions.

### Domain-wide but Variable Action of Remote Enhancers

Confirming previous reports (Anderson et al., 2014; Symmons et al., 2014), our data demonstrate that enhancers act not only on their immediate neighborhood or on their target gene(s) but more generally across large domains. Our high-resolution characterization of the *Shh* regulatory domain highlights that the potential to respond to a given enhancer shows peaks and troughs throughout an otherwise largely permissive interval. This potential can be different, depending on the promoter; for example, insertions immediately adjacent to *Rnf32*, which does not respond to the ZRS, showed expression in the ZPA; inversely, some insertions next to *Shh* were inactive (which could be also due to competition). But we also uncovered substantial variation in expression between insertions of the same reporter, even when separated by only a few kilobases. This variation indicates that other factors than promoter sequence modulate responsiveness. We found a good correlation between the physical proximity to the ZRS, as measured by 4C, and the propensity to respond to its enhancer activity. This suggests that the *Shh* region folds in a pattern that acts as a mold for enhancer action (Figure 7). This framework is flexible, as it comprises only a few regions that are completely unresponsive (Figure 7A). The re-activation of unresponsive positions after internal rearrangements (DUP/DEL(5–8)) indicates that these positions are not necessarily locally repressed but simply excluded from contacting enhancers. Interestingly, some responsive positions showed contact frequencies that were as low as unresponsive regions, revealing either the influence of other factors or the limits of 4C to measure some interaction parameters (e.g., duration of contacts in the context of an ensemble of dynamic conformations) (Fudenberg and Mirny, 2012; Giorgetti et al., 2014).



**Figure 6. Distance-Dependent Rescue of ZRS Activity on *Shh***

(A–D) Schematic representation of the series of inversions generated from C2. Red, blue, and white circles indicate putative enhancer elements that are progressively moved away from *Shh* by the inversions. A green arrow identifies the end of the *Shh* TAD.

(E–H) Skeletons of E18 embryos, including close-up views of the hindlimb (I–L). Scale bar, 2 mm.

(E and I) Control embryo (*Shh*<sup>del/+</sup>).

(F and J) *INV(6-C2)/Shh*<sup>del</sup>.

(G and K) *INV(4-C2)/INV(4-C2)*.

(H and L) *INV(2-C2)/INV(2-C2)*.

md, mandibule; fl, forelimb; hl, hindlimb; ph, phalanges; mt, metatarsal bones; ts, tarsal bones; ti, tibia; fi, fibula; cdv, caudal vertebrae. Arrowheads and asterisks point to deformed structures (cyclopia (1-e), vtf, vestigial partially fused tibia-fibula; pb, proboscis replacing anterior head structures). Photo in (E) was assembled from two images of the same embryo using Adobe Photoshop Photomerge script.

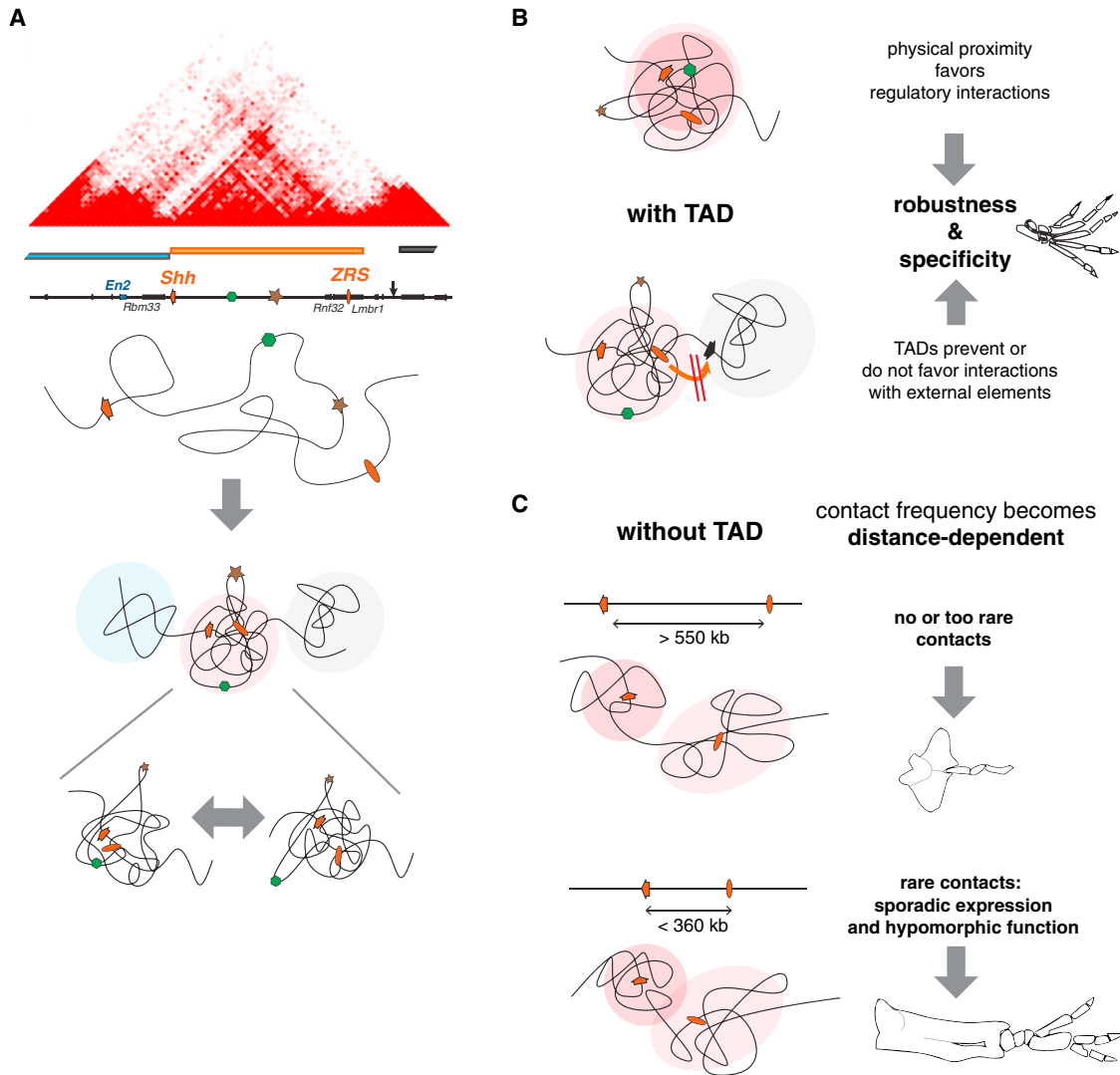
See also Figure S6.

### Tethers and Insulators?

Previous studies have suggested that sequences close to or within the ZRS may target it to *Shh* (Amano et al., 2009; Lettice et al., 2014), through the formation of a large loop (Williamson et al., 2011). Our ability to detect the action of the ZRS throughout the TAD with a reporter gene argues against the need for a specific promoter to respond to the ZRS. The inability of the ZRS to contact *Shh* in most inversions further shows that the ZRS cannot find *Shh* if it is not located in the same TAD, demonstrating the absence of a TAD-independent system targeting the ZRS to *Shh*.

Recent studies have substantiated models proposing that enhancers act within a space delineated by insulators (Downen et al., 2014). The existence of insulators is widely supported by experimental evidence that identified short regions that can

block enhancer-gene interactions (Hark et al., 2000; Lupiáñez et al., 2015; Tsujimura et al., 2015), and our data do not challenge the general existence and role of insulators. Yet, previous studies have also suggested that, even in the absence of specific insulators, certain loci show restricted enhancer-promoter interactions (Kokubu et al., 2009), questioning the universal necessity for insulators. Supporting this alternative view, many TAD boundaries appear not to be strict boundaries but correspond to a gradual effect, in terms of contact frequencies or blocking enhancer activities. At the *Shh* locus, the centromeric boundary between *Rbm33* and *Shh* appeared much more marked than the telomeric one between the ZRS and *Lmbr1*, both from a structural (based on 4C and Hi-C data) and a regulatory (changes in enhancer responsiveness) viewpoint, suggesting the telomeric boundary may be less robust (or organized). Modeling of



### Figure 7. TADs Organize Robustness and Specificity of Long-Distance Interactions

(A) TADs show an internal structure that determines the propensity of a region to be contacted by another one, hence defining enhancer responsiveness. *Shh* gene, orange arrow; ZRS, orange oval; an enhancer cold-spot, brown star; a responsive spot, green hexagon. The region folds into three different TADs (blue, orange, and dark gray bars), each of which likely corresponds to a dynamic ensemble of 3D conformations (below) (Fudenberg and Mirny, 2012; Giorgetti et al., 2014). The light colored area represents the region effectively explored by an element (e.g., the ZRS), i.e., with sufficient contact frequency to elicit a transcriptional response. The cold spot is located outside this zone, whereas the responsive spot can come in proximity of the ZRS.

(B) TADs contribute to long-distance regulatory interactions by favoring proximity between otherwise distant regions (the two colored ovals represent the regions explored by *Shh* and the ZRS, respectively; the extent of overlap indicates frequent interactions). Elements located in distinct TADs do not influence genes located in the adjacent ones, not necessarily because of active insulation but simply because of the absence of a mechanism compensating for the buffering effect of genomic distances.

(C) Without TADs, contacts between distant regions are too rare to be functional or lead only to sporadic gene activation producing variable phenotypic outcomes.

See also Figure S7.

insulator action has indicated their effect is largely distance insensitive (Doyle et al., 2014). The restoration of a functional ZRS-*Shh* interaction in the INV(C2) allele, when the distance separating *Shh* and the ZRS is reduced, therefore argues against the presence of a strict, well-defined insulator element. Interestingly, whereas *Lmbr1* is not responsive to the ZRS in mice, its ortholog in the more compact chicken genome shows distinct expression in the ZPA (Maas and Fallon, 2004). Based on our

observations, we suggest that large genomic distances can act as a buffer for regulatory interactions, without the need to invoke the presence of specific insulators.

### Overcoming the Dampening Effect of Long Genomic Distances

According to simple polymer models, contact frequency should decline sharply with increasing distances. Yet, Hi-C data have

revealed that, below 700 kb (approximately corresponding to the size of TADs), interactions occur more frequently than predicted, suggesting that loops and long-lived crosslinks may facilitate interactions at shorter scales (Doyle et al., 2014; Mateos-Langerak et al., 2009). A recent study modeling the *Igh* locus emphasized the importance of spatial confinement to establish interactions (Lucas et al., 2014) and proposed that this is the main determinant for enhancer-promoter communication. Our data also demonstrate that the interactions weaving the *Shh* TAD are necessary for efficient long-distance enhancer-promoter interactions; in the context of the *Shh* TAD, genomic distance has a minimal effect on enhancer-promoter interactions, whereas distance becomes a critical factor when this TAD is disrupted. TADs increase interaction frequency between elements and reduce the otherwise limiting effect of genomic distances. TADs can therefore actively extend the functional reach of enhancers to distantly located target genes. It will be important to see to what extent genes are dependent on this functionality of TADs or if other, independent mechanisms have also evolved to ensure proper long-range regulation.

### The Nature and Function of TADs: Loops and Compaction

The principles that lead to TAD formation are still debated (Barbieri et al., 2013; Dekker and Mirny, 2016), although mounting evidence suggests that loops between CTCF sites, possibly mediated by cohesin complexes, are involved (Merkenschlager and Odom, 2013; Zuin et al., 2014). The presence and relative orientation of CTCF sites at both ends of the *Shh* TAD (Figure S7) partially fit with recent CTCF-based models (de Wit et al., 2015; Guo et al., 2015). Yet, our observations also show noticeable deviations from what could be predicted from such models.

Firstly, some of our sensors integrated beyond the CTCF site separating *Rbm33* and *Shh* showed expression in the ZPA, implying that the ZRS is not blocked by this CTCF site or limited to a strictly defined CTCF loop. With respect to the predicted CTCF loops, the ZRS would be just outside the CTCF loop containing *Shh*, and the WT and INV configurations would be similar, whereas their functional outcome is strikingly distinct. In contrast, one would expect a more important effect in DUP(C1-Z), as the ZRS is now moved away from the potential CTCF-mediated loop containing *Shh*. Our functional data therefore underline that binding and orientation of CTCF are not sufficient to predict regulatory outcomes.

Beyond the underlying mechanism(s), the decisive factor governing enhancer-promoter functional interactions is the frequency of physical interactions between these elements. In this respect, the relative degree of insulation (which essentially is how TADs are identified) is far less important than the 3D volume of a TAD and its internal dynamics. Addition or removal of sequences normally present in the *Shh* domain have a small but noticeable impact on enhancer action, whereas interspersing external sequences into the *Shh* TAD, like with INV(6-C2), leads to a loss of compaction associated with the TAD and reduced long-range interactions. This shows that interaction frequency within a TAD may depend on its internal sequence or chromatin organization and not only on loops determined at its extremities.

Controlling regulatory interactions is an essential function of genomes, and current models have put a lot of emphasis on insulation (Downen et al., 2014; Lupiáñez et al., 2015). In animals with more compact genomes, insulators may be critical to avoid unwanted interactions between close neighbors, which could explain why *Drosophila* evolved multiple types of insulators (Yang and Corces, 2012). But in animals with large genomes and large intergenic distances, genomic distance per se can often suffice to limit functional interactions. In these conditions, promoting long-range interactions becomes crucial to ensure robustness of a system that would otherwise depend on rare, stochastic collisions. Absence of such a mechanism would lead to phenotypic variability as illustrated by INV(2-C2) animals (Figure 7). In this view, the formation of compact genomic domains like TADs and the diverse mechanisms that ensure both robust and specific long-range regulatory interactions may have been essential to expand the genomic toolbox of evolution.

## EXPERIMENTAL PROCEDURES

### Transgenic Mice

The founder ShhSB-C1, ShhSB-C2, and Z2D mice were generated by homologous recombination in E14 embryonic stem cells (ESCs). We inserted a Sleeping Beauty transposon that carries a *LacZ* reporter gene and a *loxP* site at chr5:29,413,901 for ShhSB-C1 and at position chr5:29,854,582 for ShhSB-C2. The ShhSB-C2 insert also contained a second *loxP* site outside the transposon. For Z2D, the ZRS enhancer (chr5:29,641,240–29,642,424) was substituted with a *Dach1* limb enhancer (chr14:97,485,490–97,486,674) (Visel et al., 2007). Remobilization of the SB transposon and mapping of new insertions was performed as described (Ruf et al., 2011). Targeted rearrangements were produced by in vivo recombineering (Hérault et al., 1998; Spitz et al., 2005). *Shh<sup>del</sup>* mice carry a deletion of the second and third exon of *Shh*, produced by *Cre*-mediated recombination of the *Shh-nLZ* transgene (Gonzalez-Reyes et al., 2012). *Shh-nLZ* mice were kindly provided by Andreas Kottmann (Columbia University, New York) and are referred to as *Shh::LacZ* mice in this paper. All lines were maintained by breeding with C57BL/6J mice. Genomic positions are given for using the mm9/NCBI37 assembly. Mouse experiments were conducted in accordance with the guidelines in place at the European Molecular Biology Laboratory, as defined and overseen by its Institutional Animal Care and Use Committee, in accordance with the European Directive 2010/63/EU.

### LacZ Staining, Whole-Mount In Situ Hybridization, and Skeletal Preparation

LacZ staining, whole-mount in situ hybridization, and skeletal preparation were performed according to standard protocols. Full details are in the Supplemental Experimental Procedures.

### Quantitative RT-PCR

Total RNA was isolated from microdissected tissue embryos using a PureLink RNA Mini Kit (Invitrogen) with on-column DNase I treatment; 200 ng to 1  $\mu$ g of isolated RNA was reverse transcribed with a ProtoScript M-MuLV First Strand cDNA Synthesis Kit (New England Biolabs) using oligo-dT as primer. qPCR was performed on an ABI7500 system with SYBR Green (Applied Biosystems), and analyzed using the  $\Delta\Delta C_T$  method. For data normalization, *TBP*, *GusB*, or *Hif1* was used as the reference gene, and each condition was normalized to stage-matched littermate controls. Primers are listed in Table S3.

### 4C-Seq

4C libraries were generated from microdissected embryonic limb tissues following published protocols (Simonis et al., 2007; van de Werken et al., 2012b) (see Supplemental Experimental Procedures for details). 4C libraries were generated by PCR amplification with primers containing barcodes and Solexa adapters (see Table S3). Viewpoints were analyzed in duplicate and approximately 40 libraries were pooled per sequencing lane. All samples were subjected to 50 bp single-read sequencing using Illumina HiSeq 2000.

For the analysis of 4C libraries, FASTQ files were de-multiplexed based on barcode and viewpoint primer sequences, allowing no mismatch (first eight bases were used). Primer sequences were trimmed keeping the first restriction site, and de-multiplexed libraries were aligned to the mm9 reference genome using Bowtie version 1.0.0 (Langmead et al., 2009). Aligned reads were then mapped to an in silico *NlaIII*-digested mm9 genome in order to filter out non-informative reads. Only reads mapping to fragment ends in the correct orientation were kept and assigned to the corresponding fragment end. Fragment read counts correspond to the sum of the counts obtained for each of their extremities. We assessed the quality of the libraries by determining the percentage of mapped reads and the percentage of reads mapping in *cis* (intra-chromosomal reads relative to the viewpoint) for cross-linking and digestion efficiency (van de Werken et al., 2012a). All samples showed similar library quality based on these parameters (see Table S4). 4C-seq reads were filtered as described in Klein et al. (2015) and down-sampled to match the number of the library with the lowest read count. Interaction values with the viewpoint were calculated using two measures: normalized raw read counts smoothed across 11 fragments and a hit percent rate (Denholtz et al., 2013), including a minimal threshold. For the latter approach, we transformed the 4C signal to a binary value (0 or 1) for each fragment, depending on whether the normalized read count was below or above a certain threshold (e.g., over 1, 10, or 100 counts). Fragments that fulfill the threshold criteria are termed hits. We calculated the hit percentage in a given window (e.g., 25, 51, or 101 fragments) as an estimate of the contact frequency and reliability of a given fragment. We compared the effect of different thresholds and window size on the reproducibility of the signals obtained with biological replicates (see log<sub>2</sub> ratios plots in Figure S2). For the experiments displayed here, parameters with a 51-fragment binning size (~10–20 kb length) with a read count threshold of 10 showed robust and reproducible contact patterns across the region of interest and were therefore used.

For the analysis of 4C data from samples carrying a genomic inversion, we inverted the reference genome in silico between the breakpoint coordinates and removed the fragments containing the breakpoints. To estimate the asymmetry of the interaction profiles, we calculated cumulative count distributions on each side of the viewpoint by using the counts of the sub-sampled libraries. In this analysis, we disregarded the fragments located at a distance less than 10 kb from the viewpoint to reduce the strong influence of the most proximal fragments. Data have been deposited on ArrayExpress (E-MTAB-4980).

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.devcel.2016.10.015>.

## AUTHOR CONTRIBUTIONS

F.S. conceived the project and designed it with O.S. O.S., L.P., T.A., and S.R. performed the experiments. F.K. and W.H. performed computational analysis of 4C data. F.S., O.S., L.P., S.R., and T.A. analyzed the data. O.S. and F.S. wrote the manuscript with input and comments from the other authors.

## ACKNOWLEDGMENTS

We thank the members of the EMBL Laboratory Animal Resources Facility, particularly Silke Feller, for animal welfare and husbandry and the EMBL Genomics Core Facility for help with sequencing. We are also grateful for the kind support of Wouter de Laat and Patra Klous in establishing the 4C protocol. O.S. was supported by a PhD fellowship from the Louis-Jeantet Foundation. S.R. was supported by an EMBL Interdisciplinary Postdoc (EIPOD) Fellowship under Marie Curie Actions COFUND. This work was supported by the European Molecular Biology Laboratory and by the Deutsche Forschungsgesellschaft (DFG grant: SP 1331/3-1, to F.S.).

Received: May 19, 2016

Revised: August 1, 2016

Accepted: October 18, 2016

Published: November 17, 2016

## REFERENCES

- Amano, T., Sagai, T., Tanabe, H., Mizushima, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* 16, 47–57.
- Anderson, E., Devenney, P.S., Hill, R.E., and Lettice, L.A. (2014). Mapping the Shh long-range regulatory domain. *Development* 141, 3934–3943.
- Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299–308.
- Barbieri, M., Fraser, J., Lavitas, L.-M., Chotalia, M., Dostie, J., Pombo, A., and Nicodemi, M. (2013). A polymer model explains the complexity of large-scale chromatin folding. *Nucleus* 4, 267–273.
- Bauer, D.E., Kamran, S.C., Lessard, S., Xu, J., Fujiwara, Y., Lin, C., Shao, Z., Canver, M.C., Smith, E.C., Pinello, L., et al. (2013). An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* 342, 253–257.
- Benko, S., Fantes, J.A., Amiel, J., Kleinjan, D.-J., Thomas, S., Ramsay, J., Jamshidi, N., Essafi, A., Heaney, S., Gordon, C.T., et al. (2009). Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nat. Genet.* 41, 359–364.
- Bhatia, S., Bengani, H., Fish, M., Brown, A., Divizia, M.T., de Marco, R., Damante, G., Grainger, R., van Heyningen, V., and Kleinjan, D.A. (2013). Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am. J. Hum. Genet.* 93, 1126–1134.
- Butler, J.E., and Kadonaga, J.T. (2001). Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev.* 15, 2515–2519.
- Calhoun, V.C., and Levine, M. (2003). Long-range enhancer-promoter interactions in the *Scr*-*Antp* interval of the *Drosophila* Antennapedia complex. *Proc. Natl. Acad. Sci. USA* 100, 9878–9883.
- Calhoun, V.C., Stathopoulos, A., and Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila* Antennapedia complex. *Proc. Natl. Acad. Sci. USA* 99, 9243–9247.
- Chetverina, D., Aoki, T., Erokhin, M., Georgiev, P., and Schedl, P. (2013). Making connections: insulators organize eukaryotic chromosomes into independent cis-regulatory networks. *Bioessays* 36, 163–172.
- Chiang, C., Litingtung, Y., Lee, E., Young, K.E., Corden, J.L., Westphal, H., and Beachy, P.A. (1996). Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* 383, 407–413.
- D'haene, B., Attanasio, C., Beysen, D., Dostie, J., Lemire, E., Bouchard, P., Field, M., Jones, K., Lorenz, B., Menten, B., et al. (2009). Disease-causing 7.4 kb cis-regulatory deletion disrupting conserved non-coding sequences and their interaction with the FOXL2 promoter: implications for mutation screening. *PLoS Genet.* 5, e1000522.
- de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 502, 499–506.
- de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Versteegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., and de Laat, W. (2015). CTCF binding polarity determines chromatin looping. *Mol. Cell* 60, 676–684.
- Dekker, J., and Mirny, L. (2016). The 3D genome as moderator of chromosomal communication. *Cell* 164, 1110–1121.
- Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149, 1233–1244.
- Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M., and Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell* 13, 602–616.
- Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P., and Grosveld, F. (1997). The effect of distance on long-range chromatin interactions. *Mol. Cell* 1, 131–139.
- 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.

- Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336.
- Downen, 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.
- Doyle, B., Fudenberg, G., Imakaev, M., and Mirny, L.A. (2014). Chromatin loops as allosteric modulators of enhancer-promoter interactions. *PLoS Comput. Biol.* **10**, e1003867.
- ENCODE Project Consortium, Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.
- Flavahan, W.A., Drier, Y., Liao, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suvà, M.L., and Bernstein, B.E. (2015). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114.
- Fudenberg, G., and Mirny, L.A. (2012). Higher-order chromatin structure: bridging physics and biology. *Curr. Opin. Genet. Dev.* **22**, 115–124.
- Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E.M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100.
- Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. *Mol. Cell* **49**, 773–782.
- Giorgetti, L., Galupa, R., Nora, E.P., Piolot, T., Lam, F., Dekker, J., Tiana, G., and Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157**, 950–963.
- Gonzalez-Reyes, L.E., Verbitsky, M., Blesa, J., Jackson-Lewis, V., Paredes, D., Tillack, K., Phani, S., Kramer, E.R., Przedborski, S., and Kottmann, A.H. (2012). Sonic hedgehog maintains cellular and neurochemical homeostasis in the adult nigrostriatal circuit. *Neuron* **75**, 306–319.
- 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.
- Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**, 486–489.
- Hérault, Y., Rassoulzadegan, M., Cuzin, F., and Duboule, D. (1998). Engineering chromosomes in mice through targeted meiotic recombination (TAMERE). *Nat. Genet.* **20**, 381–384.
- Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., et al. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458.
- Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulitou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* **46**, 205–212.
- Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M., and Epstein, D.J. (2006). A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. *Development* **133**, 761–772.
- Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294.
- Kieffer-Kwon, K.-R., Tang, Z., Mathe, E., Qian, J., Sung, M.-H., Li, G., Resch, W., Baek, S., Pruetz, N., Grøntved, L., et al. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* **155**, 1507–1520.
- Klein, F.A., Pakozdi, T., Anders, S., Ghavi-Helm, Y., Furlong, E.E.M., and Huber, W. (2015). FourCSeq: analysis of 4C sequencing data. *Bioinformatics* **31**, 3085–3091.
- Kmita, M., Fraudeau, N., Hérault, Y., and Duboule, D. (2002). Serial deletions and duplications suggest a mechanism for the collinearity of Hoxd genes in limbs. *Nature* **420**, 145–150.
- Kokubu, C., Horie, K., Abe, K., Ikeda, R., Mizuno, S., Uno, Y., Ogiwara, S., Ohtsuka, M., Isotani, A., Okabe, M., et al. (2009). A transposon-based chromosomal engineering method to survey a large cis-regulatory landscape in mice. *Nat. Genet.* **41**, 946–952.
- Kwon, D., Mucci, D., Langlais, K.K., Americo, J.L., DeVido, S.K., Cheng, Y., and Kassis, J.A. (2009). Enhancer-promoter communication at the *Drosophila* engrailed locus. *Development* **136**, 3067–3075.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.
- Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). 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.
- Lettice, L.A., Williamson, I., Devenney, P.S., Kilanowski, F., Dorin, J., and Hill, R.E. (2014). Development of five digits is controlled by a bipartite long-range cis-regulator. *Development* **141**, 1715–1725.
- Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98.
- Lucas, J.S., Zhang, Y., Dudko, O.K., and Murre, C. (2014). 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. *Cell* **158**, 339–352.
- 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.
- Maas, S.A., and Fallon, J.F. (2004). Isolation of the chicken *Lmbr1* coding sequence and characterization of its role during chick limb development. *Dev. Dyn.* **229**, 520–528.
- Mateos-Langerak, J., Bohn, M., de Leeuw, W., Giromus, O., Manders, E., Verschure, P., Indemans, M., Gierman, H., Heermann, D., van Driel, R., et al. (2009). Spatially confined folding of chromatin in the interphase nucleus. *Proc. Natl. Acad. Sci. USA* **106**, 3812–3817.
- Merkenschlager, M., and Odom, D.T. (2013). CTCF and cohesin: linking gene regulatory elements with their targets. *Cell* **152**, 1285–1297.
- Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606.
- Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., and Duboule, D. (2011). A regulatory archipelago controls hox genes transcription in digits. *Cell* **147**, 1132–1145.
- Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). Transcription. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017–1021.
- Niedermaier, M., Schwabe, G.C., Fees, S., Helmrich, A., Brieske, N., Seemann, P., Hecht, J., Seitz, V., Stricker, S., Leschik, G., et al. (2005). An inversion involving the mouse *Shh* locus results in brachydactyly through dysregulation of *Shh* expression. *J. Clin. Invest.* **115**, 900–909.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385.
- Ohtsuki, S., Levine, M., and Cai, H.N. (1998). Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. *Genes Dev.* **12**, 547–556.
- Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013).

- Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295.
- Prescott, S.L., Srinivasan, R., Marchetto, M.C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F.H., Swigut, T., and Wysocka, J. (2015). Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83.
- Prud'homme, B., Gompel, N., and Carroll, S.B. (2007). Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci. USA* **104** (Suppl 1), 8605–8612.
- Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680.
- Ruf, S., Symmons, O., Uslu, V.V., Dolle, D., Hot, C., Ettwiller, L., and Spitz, F. (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat. Genet.* **43**, 379–386.
- Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., and Shiroishi, T. (2005). 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.
- Sagai, T., Amano, T., Tamura, M., Mizushina, Y., Sumiyama, K., and Shiroishi, T. (2009). A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings. *Development* **136**, 1665–1674.
- Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116–120.
- Simonis, M., Kooren, J., and de Laat, W. (2007). An evaluation of 3C-based methods to capture DNA interactions. *Nat. Methods* **4**, 895–901.
- Smemo, S., Tena, J.J., Kim, K.-H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371–375.
- Spitz, F., Gonzalez, F., and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* **113**, 405–417.
- Spitz, F., Herkenne, C., Morris, M.A., and Duboule, D. (2005). Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes. *Nat. Genet.* **37**, 889–893.
- Sur, I.K., Hallikas, O., Vähärautio, A., Yan, J., Turunen, M., Enge, M., Taipale, M., Karhu, A., Aaltonen, L.A., and Taipale, J. (2012). Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* **338**, 1360–1363.
- Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400.
- Tsujimura, T., Klein, F.A., Langenfeld, K., Glaser, J., Huber, W., and Spitz, F. (2015). A discrete transition zone organizes the topological and regulatory autonomy of the adjacent tfap2c and bmp7 genes. *PLoS Genet.* **11**, e1004897.
- Tsukiji, N., Amano, T., and Shiroishi, T. (2014). A novel regulatory element for Shh expression in the lung and gut of mouse embryos. *Mech. Dev.* **131**, 127–136.
- Uslu, V.V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N.A., Marioni, J.C., and Spitz, F. (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nat. Genet.* **46**, 753–758.
- van de Werken, H.J.G., de Vree, P.J.P., Splinter, E., Holwerda, S.J.B., Klous, P., de Wit, E., and de Laat, W. (2012a). 4C technology: protocols and data analysis. *Methods Enzymol.* **513**, 89–112.
- van de Werken, H.J.G., Landan, G., Holwerda, S.J.B., Hoichman, M., Klous, P., Chachik, R., Splinter, E., Valdes-Quezada, C., Oz, Y., Bouwman, B.A.M., et al. (2012b). Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **9**, 969–972.
- Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C Reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10**, 1297–1309.
- Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92.
- Visel, A., Rubin, E.M., and Pennacchio, L.A. (2009). Genomic views of distant-acting enhancers. *Nature* **461**, 199–205.
- Wasserman, N.F., Aneas, I., and Nobrega, M.A. (2010). An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. *Genome Res.* **20**, 1191–1197.
- Williamson, I., Hill, R.E., and Bickmore, W.A. (2011). Enhancers: from developmental genetics to the genetics of common human disease. *Dev. Cell* **21**, 17–19.
- Wunderle, V.M., Critcher, R., Hastie, N., Goodfellow, P.N., and Schedl, A. (1998). Deletion of long-range regulatory elements upstream of SOX9 causes campomelic dysplasia. *Proc. Natl. Acad. Sci. USA* **95**, 10649–10654.
- Yang, J., and Corces, V.G. (2012). Insulators, long-range interactions, and genome function. *Curr. Opin. Genet. Dev.* **22**, 86–92.
- Zabidi, M.A., Arnold, C.D., Scherhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556–559.
- Zuin, J., Dixon, J.R., van der Reijden, M.I.J.A., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T.A., van Ijcken, W.F.J., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. USA* **111**, 996–1001.
- Zuniga, A., Michos, O., Spitz, F., Haramis, A.-P.G., Panman, L., Galli, A., Vintersten, K., Klasen, C., Mansfield, W., Kuc, S., et al. (2004). Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. *Genes Dev.* **18**, 1553–1564.



**Developmental Cell, Volume 39**

**Supplemental Information**

**The *Shh* Topological Domain Facilitates the Action  
of Remote Enhancers by Reducing  
the Effects of Genomic Distances**

**Orsolya Symmons, Leslie Pan, Silvia Remeseiro, Tugce Aktas, Felix Klein, Wolfgang Huber, and François Spitz**

## SUPPLEMENTARY MATERIALS

### **Inventory of Supplementary Materials.**

Figure S1, related to Figure 1. Regulatory potential is non-uniformly distributed throughout the whole *Shh* locus

Figure S2, related to Figure 2. Topological organisation at the *Shh* locus in different limb compartments

Figure S3, related to Figure 3. Effects of changing distances within the *Shh* TAD.

Figure S4, related to Figure 4. Consequences of TAD disrupting-alleles on gene expression and phenotypes

Figure S5, related to Figure 5. 4C profiles in INV(6-C2) alleles.

Figure S6, related to Figure 6. Phenotypic consequences of a TAD-breaking inversion series.

Figure S7, related to Figure 7. Chromatin and CTCF sites organisation around the *Shh*-ZRS locus.

Table S1, related to Figure 1. List of the different insertions of the regulatory sensor.

Table S2, related to Figure 2. Responsiveness to the ZRS and genomic features of the different insertions.

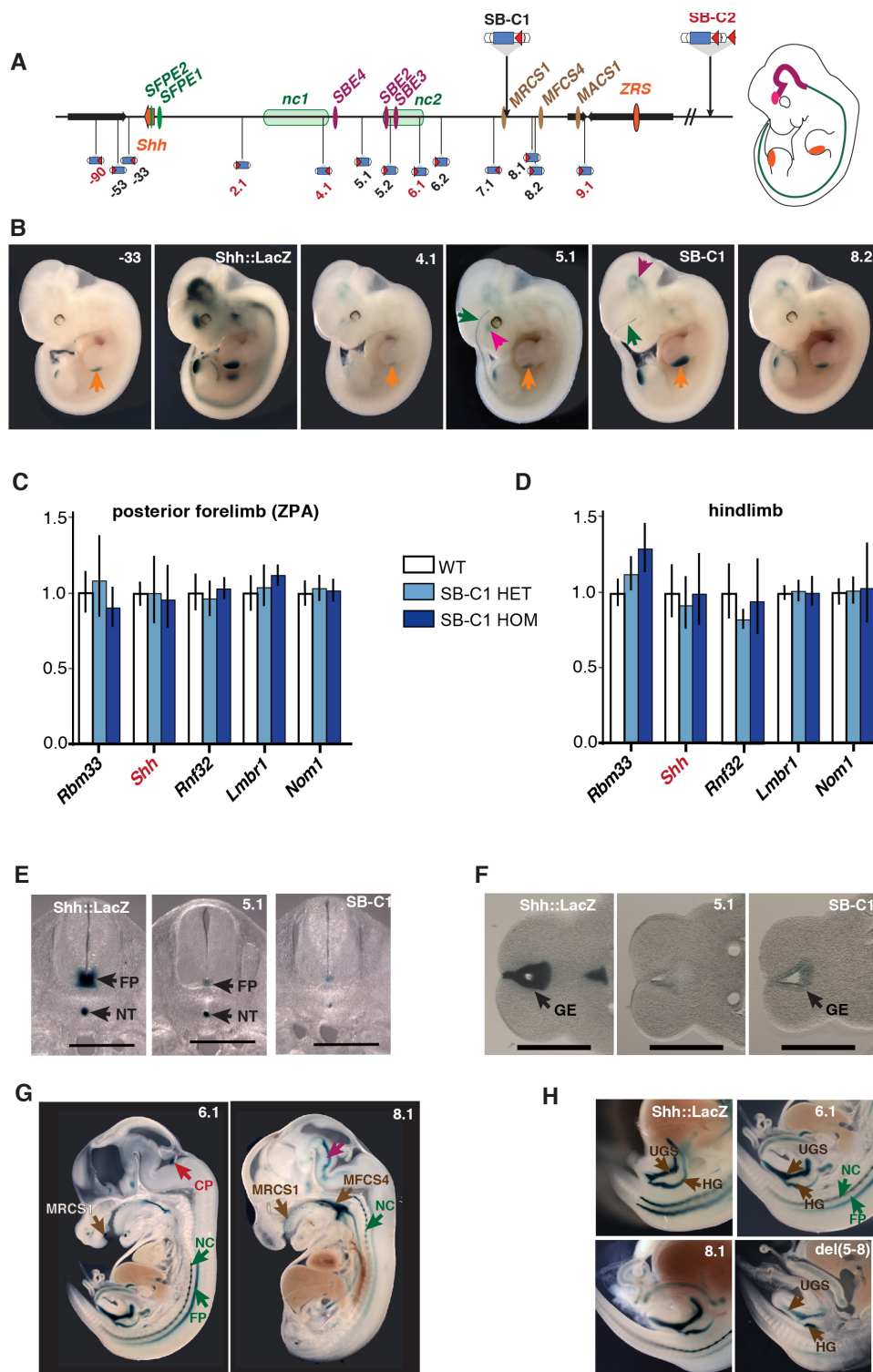
Table S3, related to Experimental Procedures. List of primer sequences.

Table S4, related to Experimental Procedures. List and characteristics of 4C libraries

Supplementary Experimental Procedures

Supplementary References

**Figure S1 (related to Figure 1).**

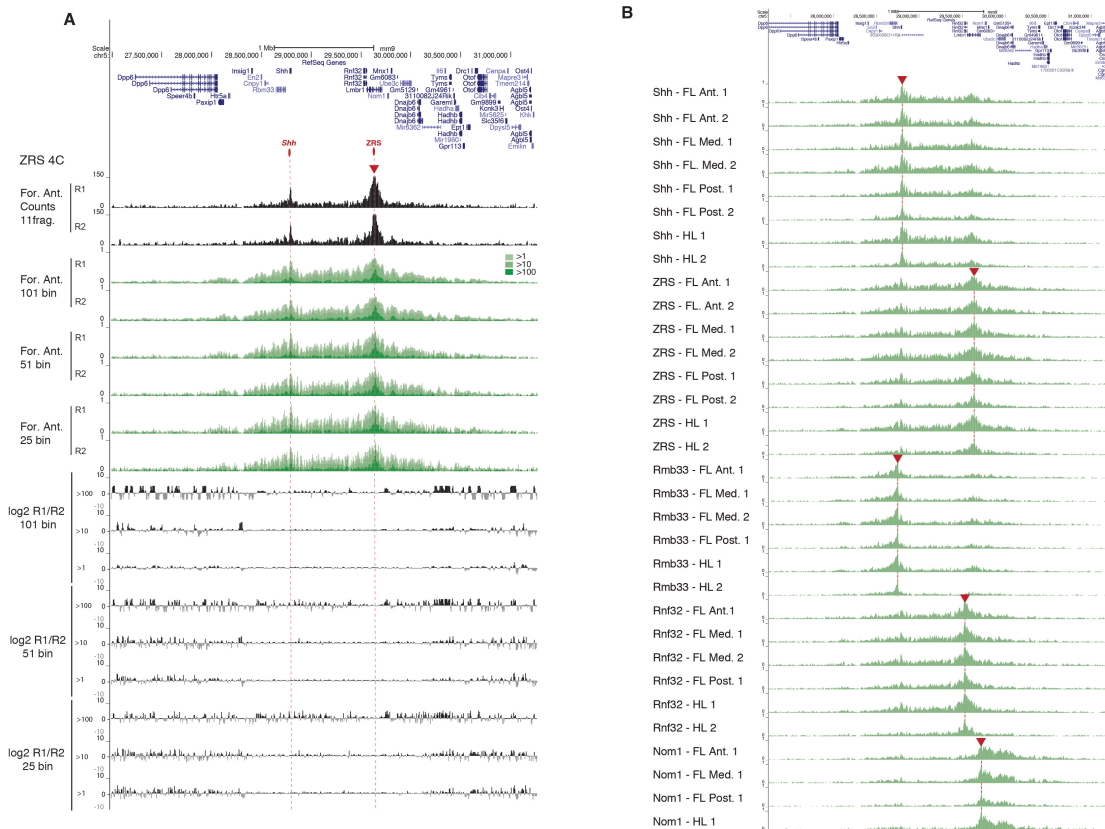


**Figure S1, related to Figure 1. Regulatory potential is non-uniformly distributed throughout the whole *Shh* locus.**

(A) *Shh* is regulated by an array of tissue-specific long-distance enhancers, which are distributed over a 900kb region. Their location along the locus and a schematic representation of their expression domains is shown, with overlapping/complementary enhancer activities indicated by matching colours. To capture the regulatory potential of the enhancers we inserted a transposable regulatory sensor at two different positions in the locus (indicated as SB-C1 and SB-C2). We subsequently remobilized the transposons and obtained multiple re-

integration events in the proximity of *Shh*. The positions of a subset of obtained insertions are shown below the locus. Black and red labels denote whether insertions were derived from SB-C1 or SB-C2, respectively. **(B)** Insertions at the locus capture the activity of different subsets of *Shh* enhancers, as revealed by lacZ staining of whole embryos. Most expression domains are observed throughout the locus, with most insertions. As examples, few domains are highlighted, with the colours of the arrows corresponding to the likely associated (orange: posterior limb buds/ZRS; green: notochord&floor plate/nc1/nc2/SFPE1/2; purple/pink: midbrain domains and ZLI. The lacZ pattern from a *Shh:lacZ* reporter line is shown as reference for the *Shh* expression pattern. **(C, D)** Gene expression in posterior forelimb (C) or hindlimb (D) of embryos heterozygous or homozygous for the SB-C1 insertion, as shown by quantitative RT-PCR. **(E-F)** Similar to the observations made on whole embryos, sections of lacZ-stained embryos also reveal non-uniform capture of regulatory activity **(G)** Differential capture of enhancer activity seen in E13.5 embryos. **(H)** Close-up view of the uro-genital region. SFPE: *Shh* floor plate enhancer, SBE: *Shh* brain enhancer, nc: notochord, MRCS: Mammal-reptile conserved sequence, MACS: Mammal-amphibian conserved sequence, MFCS: Mammal-fish conserved sequence, ZRS: ZPA-regulatory sequence, FP: floorplate, GE: genital epithelium, UGS: urogenital sinus, HG: hindgut, CP: choroid plexus, NC: notochord.

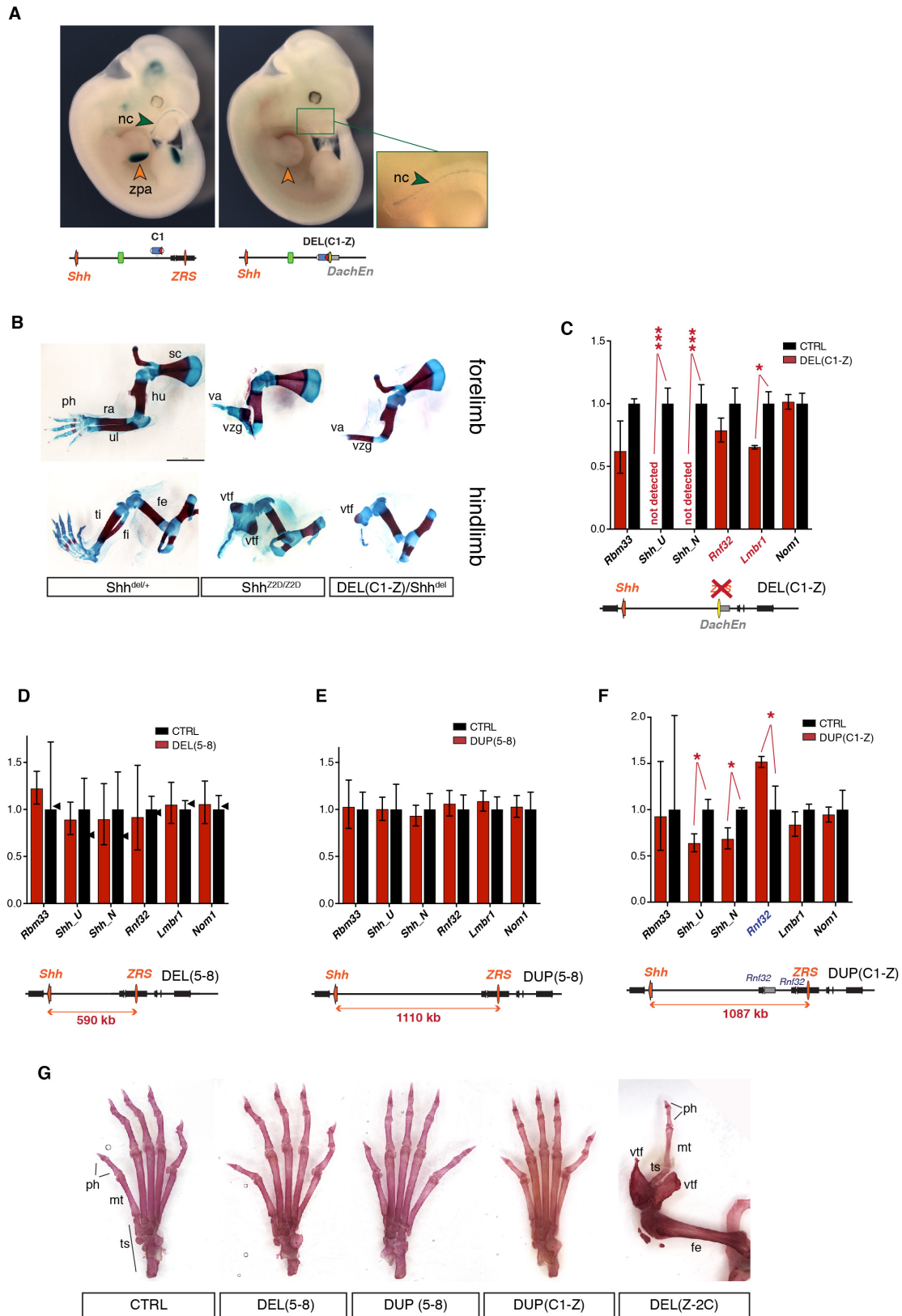
**Figure S2 (related to Figure 2).**



**Figure S2, related to Figure 2. Topological organisation at the *Shh* locus in different limb compartments.**

**(A)** Biological replicates show very similar 4C profiles, regardless of threshold and bin size. Top panels show interaction profiles in the anterior forelimb obtained using different bin sizes (as labelled on y axis) and using different thresholds (Hit percent rate, intensity of green shows increasingly higher cut-off values), lower panel shows log2 ratios between replicates. **(B)** Comparison of 4C signals in different limb compartments and different viewpoints. The position of the viewpoints is shown by red arrows. FL: forelimb, HL: hindlimb, Ant: anterior, Post: posterior, Med: middle, 1 and 2 indicate replicates.

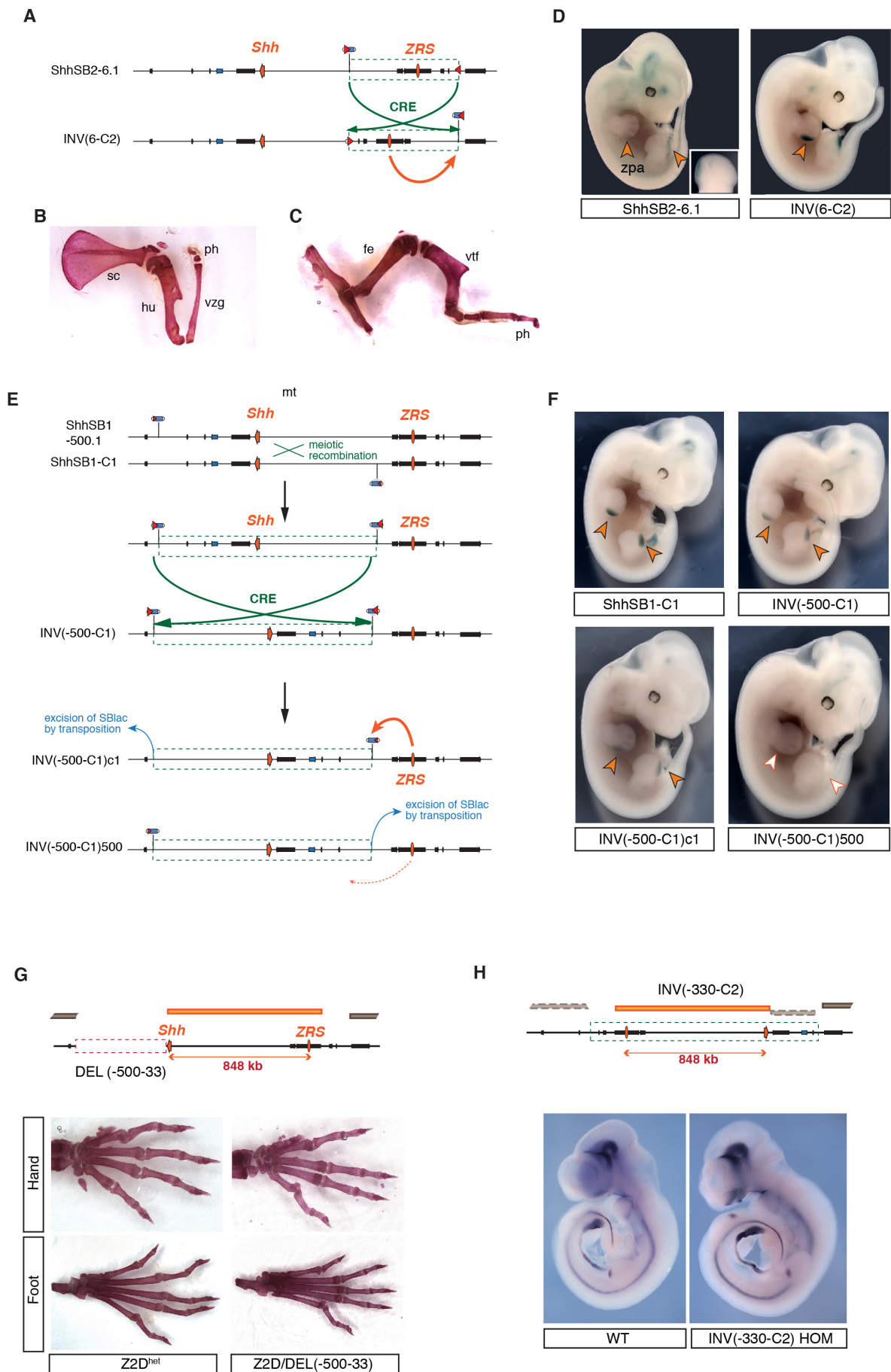
**Figure S3 (related to Figure 3).**



**Figure S3, related to Figure 3. Changing distances within the *Shh* TAD.**

**(A)** LacZ expression of the regulatory sensor at the SB-C1 position in the native locus (left) and in the context of DEL(C1-Z) allele (right). The orange arrowhead indicates expression (or lack thereof) in the ZPA, the green arrowhead expression in the notochord. A schematic outline of the corresponding locus is indicated below the pictures, with the ZRS shown as an orange oval, the region with notochord enhancers shown in green. In the DEL(C1-Z) allele the ZRS is replaced by the Dach enhancer (hs126 (Visel et al., 2007), yellow oval), which appeared to be essentially inactive when inserted at this position (Figure S3B and G). **(B)** Forelimb (top) and hindlimb (bottom) morphology of E18 embryos. Both homozygous *Shh*<sup>Z2D</sup> embryos and DEL(C1-D)/*Shh*<sup>del</sup> embryos have typical *Shh* limb loss-of-function phenotypes, but the phenotype is weaker in *Shh*<sup>Z2D</sup> embryos, especially in the hindlimb where a fused tibia-fibula element remains. sc: scapula; hu: humerus; fz: fused zeugopod; ra: radius, ul: ulna, ph: phalanges, fe: femur, ti: tibia, fi: fibula, vzg: fused zeugopod, vtf: vestigial partially fused tibia-fibula, va: vestigial autopod. **(C-F)** Gene expression, assessed by RT-qPCR in DEL(C1-Z) (C), DEL(5-8) (D), DUP(5-8) (E) and DUP(C1-Z) (F) E11 forelimb buds. Homozygous mutant samples are in red (n=3), stage-matched wild-type samples from the same litters (n=3) are used as control, except for (D), where wild-type samples include embryos from separate litters (the arrows indicate the expression level in wild-type littermates of the mutants). **(G)** Foot (hindlimb) skeletons of adult mice with different rearranged alleles. Alleles are *in trans* of either *Shh*<sup>del</sup> (for DUP(5-8), DUP(C1-Z), DEL(Z2C)) or of a ZRS replacement (Z2D allele, for CTRL and DEL(5-8)) ts: tarsus, mt: metatarsus, ph: phalanges, fe: femur, vtf: vestigial partially fused tibia-fibula.

**Figure S4 (related to Figure 4).**

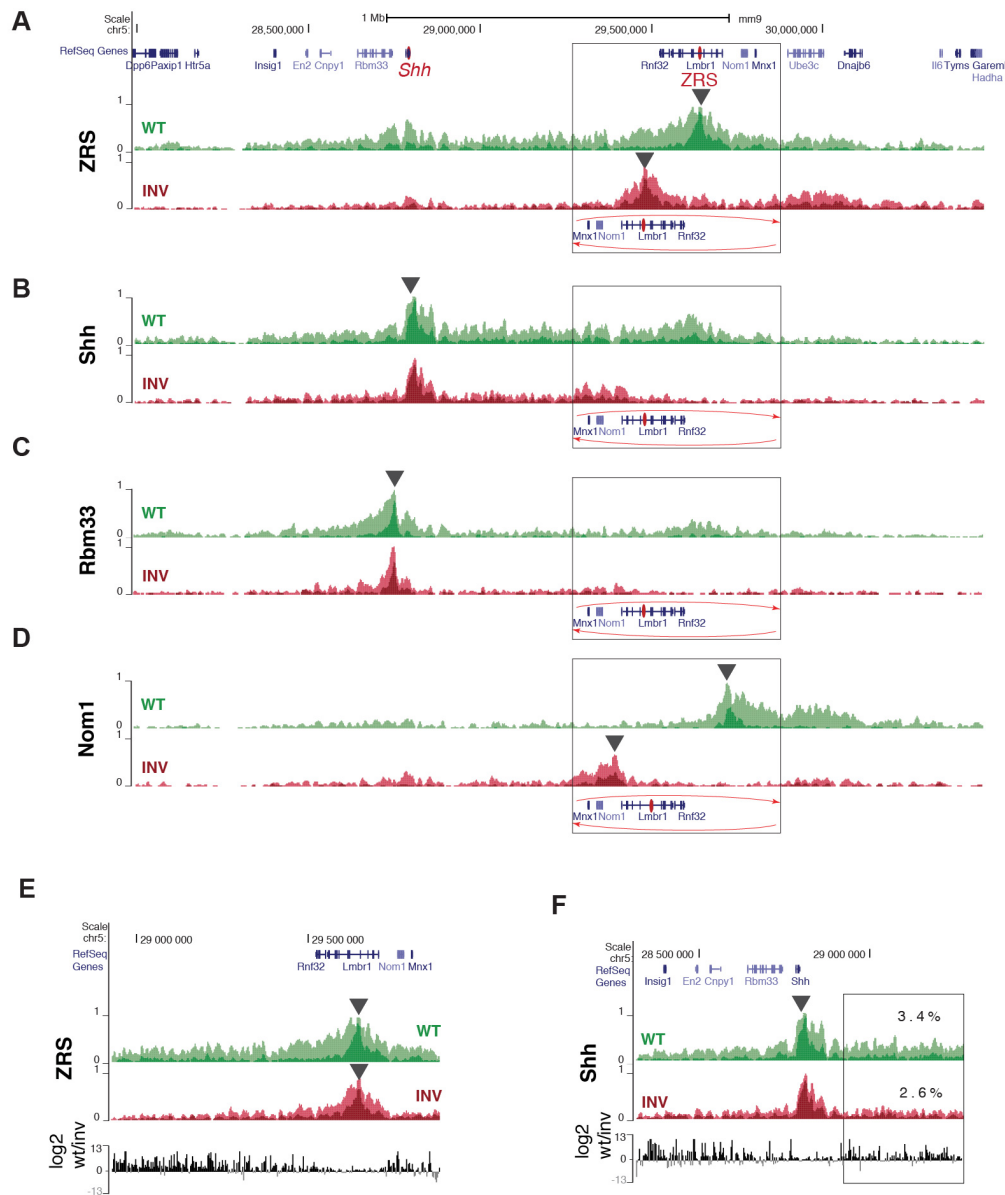


**Figure S4, related to Figure 4. Consequences of TAD disrupting-alleles.**

**(A)** Generation of the INV(6-C2) allele. The parental allele (ShhSB2-6.1) carried a *loxP* site at the SB-C2 location (telomeric of the ZRS) and a *loxP* site in opposite orientation in *cis* at the ShhSB2-6.1 position, allowing for *Cre*-mediated inversion of the DNA between the two *loxP* sites. **(B-C)** Adult fore- and hindlimb skeletons of INV(6-C2)/ *Shh*<sup>Z2D</sup> mice, showing typical monodactyl and fused zeugopod phenotypes, characteristic for *Shh* loss-of-function in the limb. sc: scapula, hu: humerus, vzg: fused zeugopod, ph: phalange, fe: femur, vtf: vestigial tibia-fibula. **(D)** Expression of the regulatory sensor before (ShhSB1-6.1) and after (INV(6-C2)) inversion. The orange arrowhead shows expression in the ZPA. **(E)** Generation of the INV(-500-C1) allele. The starting point for the inversion were two insertions at positions ShhSB1-500kb (centromeric of *Shh*) and SB-C1 (between *Shh* and the ZRS). The two insertions had *loxP* sites in opposite orientations in *trans*, which were brought in *cis* through a meiotic recombination event during breeding. This intermediate allele was then subjected to *Cre*-mediated inversion. The obtained inversion allele (INV(-500-C1)) also contained two copies of the regulatory sensor at both ends of the inversion, which could be removed independently via remobilization of the transposon, giving rise to the INV(-500-C1)C1 and the INV(-500-C1)500 alleles. **(F)** *LacZ* expression in the ZPA of embryos with the INV(-500-C1) inversion allele. The parental ShhSB1-C1 allele is displayed as control, the ShhSB1-500kb insertion had no expression (not shown). The original inverted allele with two copies of the regulatory sensor at both breakpoints (INV(-500-C1)), and the two alleles after remobilization of one copy of the regulatory sensor (INV(-500-C1)C1 and INV(-500-C1)500) are shown. The orange arrowhead highlights expression in the ZPA, the white-and-red arrowhead lack thereof. **(G)** Hand and foot morphology of adult DEL(-500-33)/*Shh*<sup>Z2D</sup> mice. The position of the deletion is indicated by the dashed red rectangle on the outline of the locus. The morphology of heterozygous *Shh*<sup>Z2D</sup> mice is shown as control. **(H)** *Shh* expression in INV(-330-C2) embryos. The position of the inversion is indicated by the dashed green rectangle on the outline of the locus, orange and brown bars represent TADs. *Shh* expression was detected by whole-mount in situ hybridization in wild-type and homozygous INV(-330-C2) embryos at E10.5.

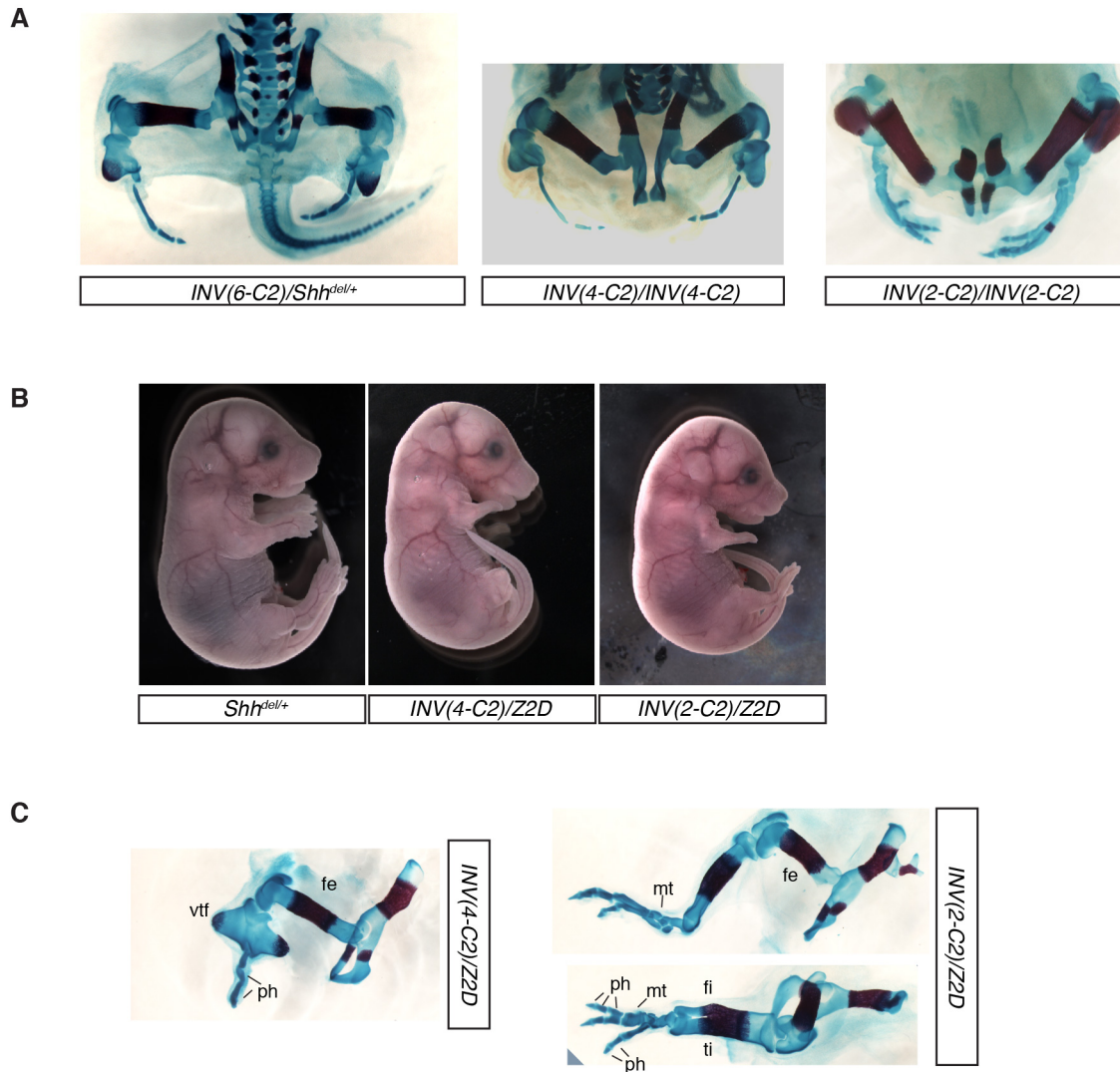


**Figure S5 (related to Figure 5).**



**Figure S5, related to Figure 5. 4C profiles in INV(6-C2) alleles. (A-D)** For each viewpoint, the normalized profiles obtained from WT (green) and INV (red) hindlimb samples are plotted on their respective genomic configurations (i.e. with an inversion of the [6-C2] genomic segment for INV). The position of the viewpoints is indicated with the black arrowhead. The inverted region is boxed, and the new position of the genes in the INV allele is depicted. **(E)** Comparison of the interaction profile of the ZRS between WT and INV in the inverted region (plotted with the same orientation) **(F)** Same comparison as in (E) for the interaction profile of *Shh* between WT and INV. The box delimitates the intra-TAD segment not affected by the inversion and the percentage of counts contained within it.

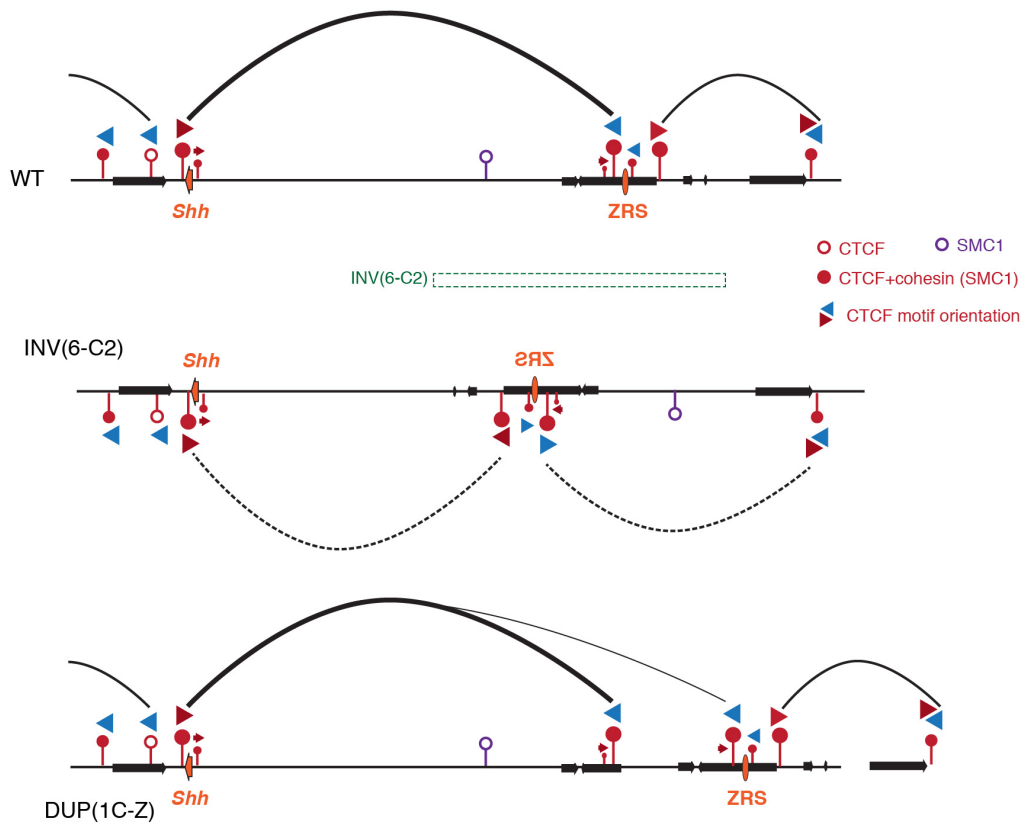
**Figure S6 (related to Figure 6).**



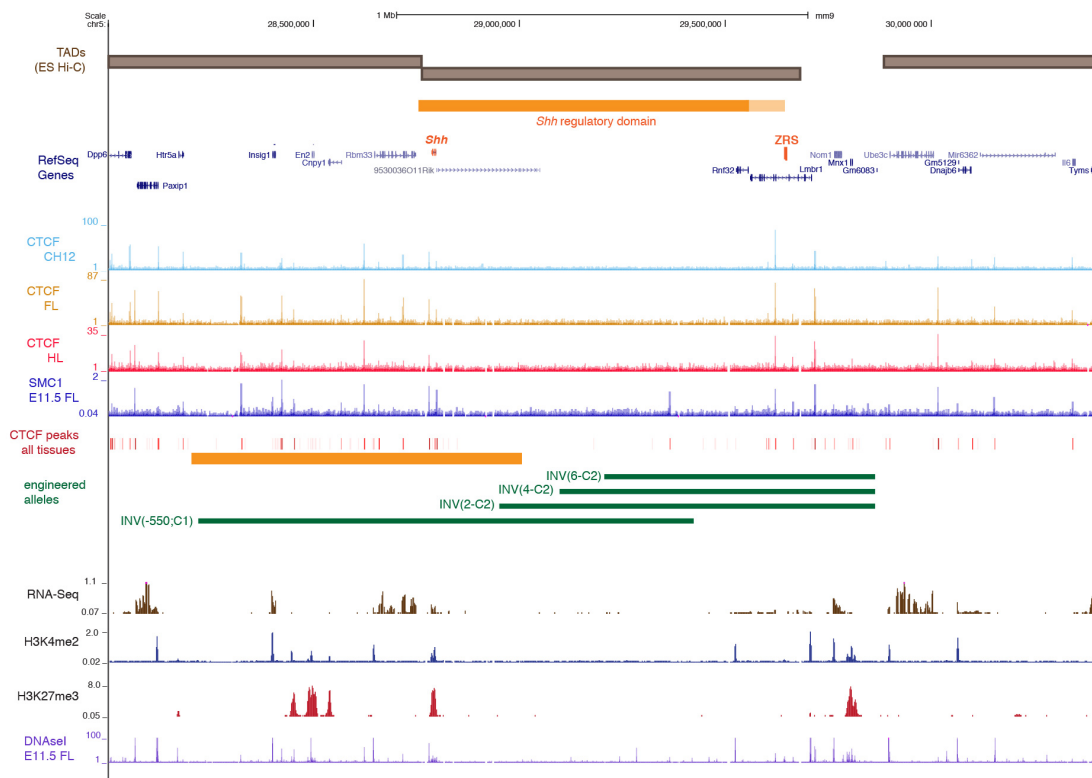
**Figure S6, related to Figure 6. Phenotypic consequences of a TAD-breaking inversion series. (A)** Hindlimb skeletons of E18 embryos carrying the *INV(6-C2)*, *INV(4-C2)* and *INV(2-C2)* alleles. **(B)** Morphology of E13 embryos that are compound heterozygous for an inversion and the *Z2D* allele. Most of the gross phenotypic abnormalities seen in homozygous embryos or in embryos that were compound heterozygous over a *Shh* deletion are restored, with the exception of the truncated limb. A heterozygous *Shh<sup>del</sup>* embryo at the same stage is also shown for comparison. **(C)** Hindlimb skeletons of E18 embryos that are compound heterozygous for *INV(4-C2)* and *Z2D* (a functional null allele of the *ZRS*) or for *INV(2-C2)* and *Z2D*. fe: femur, vtf: vestigial partially fused tibia-fibula, ph: phalange, mt: metatarsus, fi: fibula, ti: tibia.

**Figure S7 (related to Figure 7).**

**A**



**B**



**Figure S7, related to Figure 7. Chromatin organisation around the *Shh*-ZRS locus.** (A) Distribution and orientation of CTCF sites in the *Shh*-ZRS interval. Sites are shown by lollipops (size corresponding to ChIP-Seq peak height) and arrowheads indicate CTCF site relative orientation, based on the motif(s) identified below the peaks using JASPAR (Mathelier et al., 2016). Predicted loops by the CTCF-orientation model (de Wit et al., 2015; Guo et al., 2015; Rao et al., 2014) are shown by black lines. Our 4C data does not show evidence of the isolated neighborhood predicted for INV(6-C2). In DUP(1C-Z), the formation of the normal CTCF-CTCF loop with the ZRS-proximal site should position the ZRS far outside of this loop (B). Chromatin map of the locus, indicating the different TADs (in brown) (Dixon et al., 2012) and the *Shh* regulatory domain (in orange) identified (this work and (Symmons et al., 2014)), as well as the distribution of CTCF and cohesin binding sites. ChIP-Seq tracks from CH12 cells (Shen et al., 2012) as well as from embryonic limbs (forelimb – FL, hindlimb HL) indicate CTCF and the cohesin subunit SMC1 (Demare et al., 2013). CTCF peaks from 19 tissues (Shen et al., 2012) were clustered on one track, with red intensity indicating the proportion of tissues where they were detected. Gene expression is shown by RNA-Seq (E10.5 forelimb (Cotney et al., 2012)). H3K4me2 (Demare et al., 2013), and DNaseI hypersensitivity (Mouse ENCODE Consortium et al., 2012) highlight accessible promoters and other regions in the locus. H3K27me3 track (Cotney et al., 2012) shows that repressive marks are only focused on developmentally-regulated gene promoters.

**Table S1, related to Figure 1. List of insertions of the reporter sensor within and around the *Shh* locus.**

The position of each insertion is given as well as its expression pattern, assessed by LacZ staining, at specific stages of embryonic development.

as .xls file

**Table S2, related to Figure 2. Comparison of responsiveness to the ZRS and genomic properties**

For each insertion, the expression in the ZPA is determined in a semi-quantitative manner (0=nul, 1=weak, 2=strong). At each position, different genomic parameters have been measured. Proximity to ZRS, as determined by 4C (with normalized read-counts or hitpercent), accessibility (with DNase I hypersensitivity, DHS, data from ENCODE-WashU), density of various histones marks (H3K27ac, H3K27me3 – data from Cotney et al 2012, Demare et al. 2013), as well as distance to different classes of repeats (LINE1, SINE) are given.

as .xls file

**Table S3, related to Experimental Procedures.** List of primer sequences.

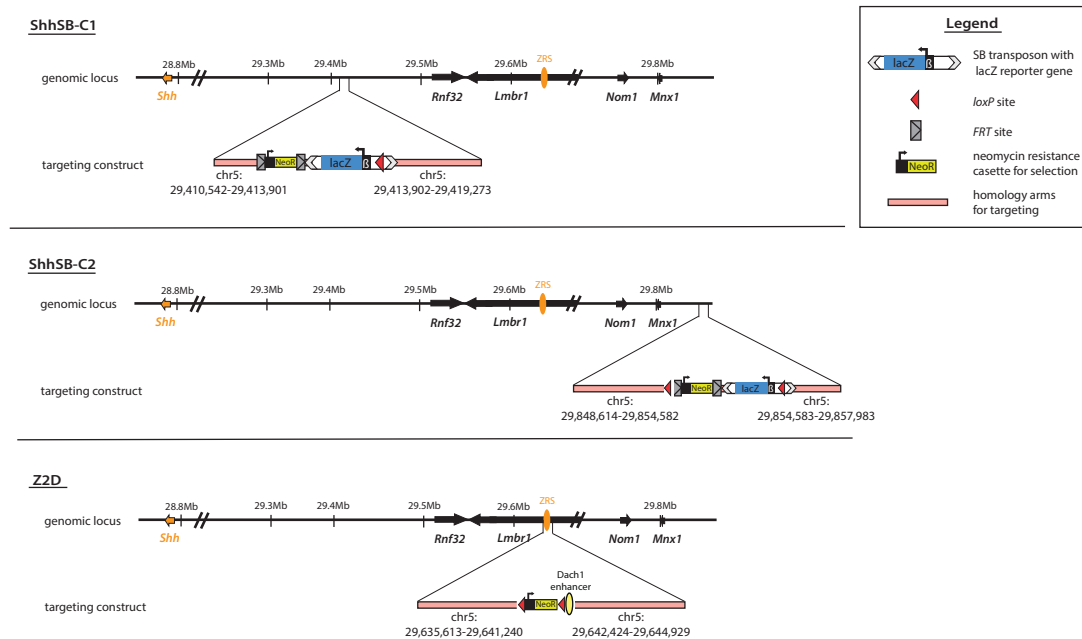
as .xls file

**Table S4, related to Experimental Procedures.** List and characteristics of 4C libraries

## Supplemental Experimental Procedures

### Transgenic mice

The founder **ShhSB1** and **ShhSB2** mice were generated by homologous recombination in E14 ES cells. For the targeting construct we inserted a neomycin resistance cassette (under the control of a PGK promoter), flanked by FRT sites, and with (ShhSB2) or without (ShhSB1) an adjacent *loxP* site into a pSB8 plasmid. The pSB8 (Chen et al., 2013) plasmid consists of a *LacZ* reporter gene (the regulatory sensor, driven by a human  $\beta$ -globin minimal promoter) and a *loxP* site, which are cloned between the terminal inverted/direct repeats of the Sleeping Beauty transposon. The resulting SB-FRT-neomycin-FRT(-*loxP*) construct was inserted between homology arms that had previously been PCR amplified and cloned into a pSK Bluescript plasmid. The homology arms corresponded to chr5:29,410,542-29,413,901 and chr5:29,413,902-29,419,273 (NCBI37/mm9) for ShhSB1 and chr5: 29,848,614-29,854,582 and chr5:29,854,583-29,857,983 (NCBI37/mm9) for ShhSB2. Targeting constructs were validated by restriction digest and sequencing, and released using NotI.



The **Z2D transgenic mouse line** was also generated by homologous recombination in E14 ES cells. In this line the ZRS enhancer (chr5:29,641,240-29,642,424) was substituted with a limb enhancer of the mouse *Dach1* gene (chr14: 97,485,490-97,486,674) and a single *loxP* site. To this end the *Dach1* enhancer and a *loxP* site were cloned between homology arms corresponding to chr5:29,635,613-29,641,241 and chr5:29,642,424-29,644,929 (NCBI37/mm9). The targeting construct was released using XbaI and XhoI. The *Dach1* enhancer was previously shown to drive expression of a reporter gene in the anterior part of the developing limb bud (element hs126 in the VISTA Enhancer Browser (Nobrega et al., 2003; Visel et al., 2007)).

Constructs were electroporated into E14 ES cells grown on inactivated DR3 feeders in ES cell medium: DMEM (Gibco, Cat.No.41965) with 15% FBS (PAN Biotech GmbH, Cat.No.2602), 1% L-glutamine (Gibco, Cat.No.25030-081), 1% penicillin/streptavidin (Gibco, Cat.No.15070-063), 1% Non-essential amino acids (Gibco, Cat.No.11140-050), 1% Sodium Pyruvate (Gibco, Cat.No.11360-070), 1% 2-mercaptoethanol (Sigma-Aldrich, Cat.No M7522) and 1000 U/ml Leukaemia inhibitory factor (LIF, Chemicon, ESG1107). When E14 cells reached 70-80% confluence they were trypsinised, counted and 13 million cells were incubated in electroporation buffer at room temperature with approximately 12 mg of DNA for 5 minutes and electroporated (240 V, 500  $\mu$ F, 4 mm cuvette).

Following electroporation, ES cells were then grown on feeder cells without selection for 3 days. On the third day gentamycin was added at a concentration of 200  $\mu$ g/ml for 2 days, and subsequently at a concentration 250  $\mu$ g/ml until the end of selection (~ 8-10 days).

After gentamycin selection 200-300 individual colonies per construct were picked and expanded. Integration of

the transgene was tested by long-range PCR and Southern blot. Positive ES clones were injected into mouse blastocysts, and chimeric males were bred to C57BL/6 for germ line transmission, which was confirmed by Southern blot and PCR. The neomycin selection cassette was subsequently removed by breeding to *hACTB::FLPe* transgenic animals. Following removal of the neomycin cassette the SB transposon was remobilized and new insertions were mapped as described before (Ruf et al., 2011).

Deletions and duplications were engineered by TAMERE (Hérault et al., 1998; Wu et al., 2007), using *loxP* sites in the same orientation in *cis* (for SB-C2 derived deletions) or in *trans*. Inversions were generated by *Cre*-mediated recombination using *loxP* sites in opposite orientation in *cis*. *loxP* sites in *cis* were obtained either using STRING (Spitz et al., 2005), i.e. breeding mice for meiotic recombination between 2 transposons in *trans*, or by using ShhSB2 remobs, where one *loxP* site remained fixed at the initial integration site (whereas the other was remobilized as transposon cargo). All rearrangements were verified by long-range PCR bridging the rearrangement breakpoint,

### Long-range PCR

For long range PCR, DNA was prepared from ES cells or tails by adding 200µl lysis buffer (100mM Tris pH=8.5, 5mM EDTA, 0.2% SDS, 200mM NaCl), complemented with 1:100 dilution of 10mg/ml stock of ProteinaseK. Cells were lysed overnight at 65°C. The next day, material was precipitated with 140µl isopropanol at -20°C for 2h, centrifuged for 15min, the supernatant removed, and the DNA pellet washed with 70% ethanol. After additional centrifugation and removal of supernatant, the pellet was dried, and resuspended in 60µl TE. Long-range PCR was performed using the Expand Long Range dNTPack (Roche, Cat. No. 4829042001) with 0.5-1µl DNA as template. To test ES cells for integration of the SB insert via homologous recombination we used one primer anchored within the transgenic construct and one primer in the genomic region flanking the homology arms of the targeting construct. To test for deletions, duplications and inversions in mice we used primers for the genomic regions flanking the region. All primers are listed in Table S4.

### Genotyping and genotyping strategies

For genotyping, tails or embryonic membranes were lysed overnight in lysis buffer (10 mM Tris-HCl pH 8, 50 mM KCl, 2 mM MgCl<sub>2</sub>, 0.1 mg/ml gelatin, 0.45% NP-40 and 0.45% Tween-20) supplemented with proteinase K (100 µg/ml) and heat inactivated. 1µl of lysate was then used as template in PCR, using an in-house preparation of *Thermus aquaticus* (Taq) DNA polymerase (produced from a construct provided by the IGBMC, Strasbourg) and 10x PCR buffer (500mM Tris-HCl pH=9.5, 150mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and 17.5mM MgCl<sub>2</sub>). Genotypes were determined by PCR using transgene-specific primers for the transposon (SB8), the transposase (Prm), *HprtCre*, *Shh<sup>del</sup>* and *Shh<sup>Z2D</sup>*. SBlac insertions were genotyped with one primer specific to the given insertion site and one primer specific for the transposon. All rearrangements were routinely genotyped through the presence of novel primer combinations. Primers are listed in Table S3.

### LacZ staining

E10.5, E11.5 or E12.5 mouse embryos were dissected in cold PBS, and fixed in PBS with 4% PFA on ice for 20, 30 or 40 min, respectively. Next, they were washed twice with ice-cold PBS and once at room temperature. Staining for β-galactosidase activity was performed overnight in a humid chamber at 37 °C. After staining, embryos were washed in PBS and stored at 4 °C.

### Whole-mount RNA In situ hybridization

For whole-mount RNA *in situ* hybridization embryos were collected at E10.5 and E11.5, placed in 4% PFA for overnight fixation at 4°C. Yolk sacs were collected and processed for genotyping. The next day embryos were washed 3 times for 5 minutes in PBS containing 0.1% Tween-20 (PBS-T) and gradually dehydrated in 30%, 50%, 70% and 100% methanol containing PBS-T. Dehydrated embryos were stored at -20°C.

The *Shh* ISH probe was *in vitro* transcribed from a plasmid from A. McMahon (linearized with HindIII, blunted with Klenow fragment and purified using QIAquick PCR Purification Kit (Qiagen, Cat. No. 28104)). 500ng DNA was used as template for DIG-labelled complementary probe synthesis with T3 polymerase using DIG RNA Labelling Kit (Roche). RNA probe was cleaned with GE Illustra ProbeQuant G50 kit (GE Healthcare Life Sciences, Cat.No. 28-9034-08), and eluted in 50µl RNase-free water. To remove any remaining DNA, the sample was treated with 1µl RNase-free DNase I.

Embryos were re-hydrated at 4°C in 70%, 50% and 30% methanol containing PBS-T and washed in PBS-T. After washing, embryos were bleached for 45 minutes to 1 hour in 6% H<sub>2</sub>O<sub>2</sub> (diluted in PBS-T) and washed again in PBS-T. Bleached embryos were permeabilised by ProteinaseK treatment at room temperature for 7 (E10.5 embryos) or 12 (E11.5 embryos) minutes. ProteinaseK activity was stopped by washing with 2mg/ml glycine solution for 5 minutes on ice. The embryos were post-fixed at room temperature in 4% PFA for 20

minutes, and washed in PBS-T. Next, embryos were washed for 10 minutes in W1 (5X SSC pH4.5, 50% ionized Formamide, 1% SDS, 0.1% Tween-20) at 65°C, and then pre-hybridised at 65°C in H2 (W1+ 5mg/ml Torula yeast RNA and 25µl of 100mg/ml heparin) for at least 2 hours. Embryos were incubated overnight at 65°C with the RNA probe in H2. The next day, embryos were washed for 3 times 30 minutes at 65°C in pre-heated W1, then also W2 (2X SSC pH=4.5, 50% ionized Formamide, 0.1% Tween-20), and finally once in W3 (2X SSC pH=4.5, 0.1% Tween-20). After equilibration at room temperature, embryos were washed in TBS-T (137mM NaCl, 20mM Tris, containing 1% Tween-20). For blocking, embryos were kept in blocking solution (TBS-T with 20µl fetal calf serum and 20µl of 100mg/ml BSA) for at least 2 hours, before adding DIG antibody (1:3000 dilution in blocking solution), and leaving them overnight at 4°C. On the third day, embryos were washed in TBS-T extensively, and kept in TBST overnight at 4°C.

The last day embryos were washed 3 times 10 minutes in NTMT (100mM Tris pH9.5, 100mM NaCl, 1% Tween-20), before staining with nitro-blue tetrazolium chloride and 5-bromo-4-chloro-3'-indolylphosphate p-toluidine salt (3.4µl/ml of 100mg/ml NBT and 3.5µL of 100mg/ml BCIP in NTMT) in a dark chamber. The staining reaction was stopped by washing in PBS after approx. 2-4 hours.

#### **Alizarin red staining of adult skeletons**

For alizarin red staining of bone adult mice (usually at 6 weeks or older) were euthanized by CO<sub>2</sub> inhalation. After visual inspection of limbs, particularly checking for syn- and polydactyly, skin and internal organs were removed, and bodies were then fixed in 95% ethanol for 7 days. After fixation soft tissue was digested and bone was stained by incubation in 2% KOH with 0.001% alizarin red (Fluka) for 2-3 days. Remaining tissue was cleared with glycerol.

#### **Alizarin red/alcian blue staining of embryonic skeleton**

For skeletal preparation of late E17 embryos we used a modified protocol from (Wallin et al., 1994). Briefly, embryos were collected from pregnant females, and skin was removed by briefly incubating them in 60°C water. Internal organs were removed and embryos were fixed in 95% EtOH and 100% acetone for 1 week each. Next, they were transferred to alizarin red/alcian blue staining solution (1 volume 0.3% alcian blue (Fluka), 1 volume 0.1% alizarin red (Fluka), 1 volume acetic acid, 17 volumes 70% acetic acid) for 3 days. Finally, embryos were cleared by incubation in glycerol/1% KOH with an increasing ratio of glycerol.

#### **Generating 4C libraries**

For 4C libraries we collected the following embryonic mouse tissues:

- C57Bl/6 wild-type forelimbs (dissected into anterior, middle and posterior sections) and whole hindlimbs of E11.5 mouse embryos
- INV(6-C2)/del(-90kb-C2) whole forelimbs and hindlimbs. For this purpose heterozygous INV(6-C2) mice were crossed with heterozygous del(-90kb-ShhSB2) mice. Compound heterozygous embryos could phenotypically easily be distinguished from all other phenotypes due to their small size and truncated limbs.

All tissues were collected into PBS pH=7.4. Single cell suspensions were obtained by incubation in 1% trypsin at 37°C for 1 minute and rigorous pipetting with a blue P1000 tip. The cells were then immediately fixed with 2% formaldehyde in 10% FCS/PBS at room temperature for 10 minutes. Glycine was added to a final concentration of 0.1M and the samples were moved to ice to quench the cross-linking reaction. The cells were centrifuged at 8000 rpm for 8 minutes at 4°C, and the supernatant removed. Cells were lysed in 1ml cold lysis buffer (50mM HEPES pH=8, 150mM NaCl, 5mM EDTA, 0.5% NP-40 1.2% Triton X-100) for 10 minutes on ice. Successful lysis was verified by Methyl Green-Pyronin staining. Cell lysate was centrifuged at 8000 rpm at 4°C, the supernatant removed, and the pellet was frozen in liquid N<sub>2</sub> until sufficient material was available.

Once sufficient tissue had been collected, cells were resuspended in 62.5 µl 10X NlaIII restriction buffer, placed at 37°C and 15 µl 10% SDS was added. Cells were incubated for 1h at 37°C, while shaking at 900rpm to remove non-crosslinked proteins. Subsequently, SDS was sequestered, by adding 150µl 10% Triton X-100 and incubated at 37°C while shaking at 900 RPM. The sample was then digested by adding 400U (40 µl) NlaIII and incubation at 37°C. After 4h and 8h another 400U of enzyme was added (total: 1200U NlaIII), adjusting the volume and buffer concentration as necessary, and the restriction digest was continued overnight, during the two latter digests. The next day we verified digestion efficiency by de-crosslinking and proteinaseK-treating samples taken before and after digest and running on an 0.6% agarose gel. If digest was not sufficient further enzyme was added for another 4h.

Following sufficient digestion, enzyme was inactivated by incubating 20 minutes at 65°C and samples were transferred to a 50ml falcon tube. Samples were ligated in a 7ml volume using 700µl 10X ligation buffer (0.3M Tris-HCl pH=7.8, 0.1M MgCl<sub>2</sub>, 1.54% DTT, 0.5% ATP) and 50U T4 DNA ligase and incubated at 16°C overnight. We determined ligation efficiency by de-crosslinking and proteinaseK-treating 100µl of ligation reaction and running on an 0.6% agarose gel. If further ligation was necessary, we added fresh ATP and

continued ligation for 4-6h.

Once ligation was complete we de-crosslinked the sample overnight at 65°C in the presence of 30µl Proteinase K (10mg/ml). The next morning, we added 30µl RNase A (10mg/ml) and incubated the sample for 45 min at 37°C. We then extracted DNA by adding 7ml phenol-chloroform and precipitating DNA using 7ml water, 1.5 ml 2M NaAcetate pH 5.6, 7 µl 20mg/ml glycogen and 35 ml 100% ethanol. The resulting pellet was dissolved in 150µl Tris pH 7.5 at 37°C.

The obtained 3C library was then subjected to a second round of overnight digestion using 60U DpnII in a 500µl volume. The next day we verified digestion efficiency and subsequently inactivated the enzyme by incubating the sample at 65°C for 25 minutes. The sample was then subjected to phenol-chloroform extraction, resuspended in 100µl water and ligated overnight with 100U T4 DNA ligase in a final volume of 14.3ml. The next day the DNA was first phenol-chloroform extracted and then purified using the QIAquick PCR purification kit, using 3 columns per sample, eluting with 50µl 10mM Tris per column.

4C primers for the region of interest were selected from the Tanay lab 4C primer database (van de Werken et al., 2012), further tested for off-targets using BiSearch (Arányi et al., 2006) and only primers with <50 hits were kept. For quality assessment of primers we tested them on 25, 50, 100 and 200ng of 4C template and kept only those with reproducible banding pattern at all concentrations. The final list of 4C primers is included in the Table S4. Using these primers, we set up PCR reactions for sequencing with primers containing barcodes and the Solexa sequencing adapter at their 5' end.

For PCRs we used Expand Long Template system from Roche, using either 100 or 200ng of the 4C library as template and setting up 16 PCR reactions in parallel for each primer. PCR products were purified using Roche HighPure PCR product purification kit, adding an additional 1 min spin after the last wash step to remove all residual ethanol. The samples were eluted in 50µl 10 mM Tris pH=8.0, pooled and 1 µg each of approximately 40 PCR products were pooled to achieve a balanced base composition at the first 4 bases.



Supplementary Table S3 (related to Experimental Procedures) – Primer sequences  
as .xls file

Supplementary Table S4 (related to Experimental Procedure) – List of 4C libraries: summary information and quality controls.

Library	Raw Reads	Low Quality Reads	Mapped Reads	Mapping Ratio	Trans Counts	Cis Counts	Cis Ratio	Trans Counts	Cis Counts (-1Mb vwp)	Cis Ratio (-1Mb vwp)
Shh_28804_FA_1	1730807	0	1630977	0.94	693294	937683	0.57	693294	747864	0.21
Shh_28804_FA_2	1914379	0	1775919	0.93	965390	810529	0.46	965390	560664	0.21
Shh_28804_FM_1	2540591	0	2403406	0.95	973660	1429746	0.59	973660	1129771	0.24
Shh_28804_FM_2	1981742	0	1855936	0.94	656294	1199642	0.65	656294	881489	0.33
Shh_28804_FP_1	1939550	0	1754581	0.90	858547	896034	0.51	858547	680800	0.20
Shh_28804_FP_2	1380275	0	1305467	0.95	894833	410634	0.31	894833	260706	0.14
Shh_28804_H_1	1889315	0	1809985	0.96	730969	1079016	0.60	730969	867404	0.22
Shh_28804_H_2	1968711	0	1795179	0.91	1052237	742942	0.41	1052237	585991	0.13
Shh_28804_invF_1	2598357	0	2470169	0.95	1242232	1227937	0.50	1242232	990738	0.16
Shh_28804_invH_1	1745224	0	1607727	0.92	816800	790927	0.49	816800	559355	0.22
Shh_28806_FA_1	1395209	0	1330831	0.95	549917	780914	0.59	549917	582472	0.27
Shh_28806_FP_1	1576901	0	1498786	0.95	695601	803185	0.54	695601	579363	0.24
Shh_28806_H_1	1302506	0	1247398	0.96	493281	754117	0.60	493281	590136	0.25
Shh_28806_H_2	1531435	0	1457472	0.95	466535	990937	0.68	466535	782445	0.31
ZRS_FA_1	2113844	0	1991871	0.94	820855	1171016	0.59	820855	865876	0.27
ZRS_FA_2	2307615	0	2171679	0.94	978616	1193063	0.55	978616	803521	0.28
ZRS_FM_1	1179046	0	1083418	0.92	411309	672109	0.62	411309	474574	0.32
ZRS_FM_2	2056175	0	1920819	0.93	600006	1320813	0.69	600006	933723	0.39
ZRS_FP_1	2483997	0	2310022	0.93	1010325	1299697	0.56	1010325	965947	0.25
ZRS_FP_2	1923942	0	1773280	0.92	922651	850629	0.48	922651	608316	0.21
ZRS_H_1	2210918	0	2045358	0.93	843989	1201369	0.59	843989	906189	0.26
ZRS_H_2	2800697	0	2575402	0.92	1426290	1149112	0.45	1426290	848110	0.17
ZRS_invF_1	2688247	0	2433454	0.91	1477965	955489	0.39	1477965	585105	0.20
ZRS_invH_1	2366582	0	2201760	0.93	1145465	1056295	0.48	1145465	674751	0.25
Rbm33_FA_1	3663217	0	3521871	0.96	689794	2832077	0.80	689794	2630637	0.23
Rbm33_FM_1	3469042	0	3338762	0.96	627580	2711182	0.81	627580	2510876	0.24
Rbm33_FM_2	3620884	0	3480692	0.96	560906	2919786	0.84	560906	2638088	0.33
Rbm33_FP_1	3575961	0	3398238	0.95	727884	2670354	0.79	727884	2479788	0.21
Rbm33_H_1	3858835	0	3717274	0.96	635881	3081393	0.83	635881	2863760	0.25
Rbm33_H_2	3777146	0	3587211	0.95	923720	2663491	0.74	923720	2498848	0.15
Rbm33_invF_1	1354644	0	1309030	0.97	524548	784482	0.60	524548	720174	0.11
Rbm33_invH_1	1155364	0	1105181	0.96	377230	727951	0.66	377230	646859	0.18
Nom1_FA_1	782770	0	739755	0.95	371350	368405	0.50	371350	238591	0.26
Nom1_FM_1	585510	0	532208	0.91	249894	282314	0.53	249894	179185	0.29
Nom1_FP_1	408472	0	378472	0.93	203261	175211	0.46	203261	105908	0.25
Nom1_H_1	868031	0	818688	0.94	371599	447089	0.55	371599	316557	0.26
Nom1_invF_1	341516	0	322357	0.94	213545	108812	0.34	213545	58010	0.19
Nom1_invH_1	802372	0	745905	0.93	406257	339648	0.46	406257	188412	0.27
Rnf32_FA_1	1947975	0	1882565	0.97	650825	1231740	0.65	650825	985970	0.27
Rnf32_FM_1	2132872	0	2075854	0.97	681974	1393880	0.67	681974	1098263	0.30
Rnf32_FM_2	1370276	0	1295654	0.95	421023	874631	0.68	421023	637071	0.36
Rnf32_FP_1	1960298	0	1891582	0.96	714008	1177574	0.62	714008	930819	0.26
Rnf32_H_1	1880914	0	1833459	0.97	637401	1196058	0.65	637401	985745	0.25
Rnf32_H_2	2262878	0	2141327	0.95	892810	1248517	0.58	892810	1055835	0.18

## Supplemental References.

Arányi, T., Váradi, A., Simon, I., and Tusnády, G.E. (2006). The BiSearch web server. *BMC Bioinformatics* 7, 431.

Chen, C.-K., Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Smedley, D., and Spitz, F. (2013). TRACER: a resource to study the regulatory architecture of the mouse genome. *BMC Genomics* 14, 215.

Cotney, J., Leng, J., Oh, S., Demare, L.E., Reilly, S.K., Gerstein, M.B., and Noonan, J.P. (2012). Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Res* 22, 1069–1080.

Demare, L.E., Leng, J., Cotney, J., Reilly, S.K., Yin, J., Sarro, R., and Noonan, J.P. (2013). The genomic landscape of cohesin-associated chromatin interactions. *Genome Res* 23, 1224–1234.

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.-Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 44, D110–D115.

Mouse ENCODE Consortium, Stamatoyannopoulos, J.A., Snyder, M., Hardison, R., Ren, B., Gingeras, T., Gilbert, D.M., Groudine, M., Bender, M., Rajinder, K., et al. (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biology* 13, 418.

Nobrega, M.A., Ovcharenko, I., Afzal, V., and Rubin, E.M. (2003). Scanning human gene deserts for long-range enhancers. *Science* 302, 413.

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res* 35, D88–D92.

Wallin, J., Wilting, J., Koseki, H., Fritsch, R., Christ, B., and Balling, R. (1994). The role of Pax-1 in axial skeleton development. *Development* 120, 1109–1121.

Wu, S., Ying, G., Wu, Q., and Capecchi, M.R. (2007). Toward simpler and faster genome-wide mutagenesis in mice. *Nat Genet* 39, 922–930.