Functional dissection of the *Sox9–Kcnj2* locus identifies nonessential and instructive roles of TAD architecture

Alexandra Despang[1,2,3], Robert Schöpflin[1,2,3], Martin Franke[1,7], Salaheddine Ali[1,2,3], Ivana Jerković [1,2,8], Christina Paliou [1,2], Wing-Lee Chan[2], Bernd Timmermann[4], Lars Wittler[5], Martin Vingron [6], Stefan Mundlos [1,2,3]* and Daniel M. Ibrahim [1,2,3]*

The genome is organized in three-dimensional units called topologically associating domains (TADs), through a process dependent on the cooperative action of cohesin and the DNA-binding factor CTCF. Genomic rearrangements of TADs have been shown to cause gene misexpression and disease, but genome-wide depletion of CTCF has no drastic effects on transcription. Here, we investigate TAD function in vivo in mouse limb buds at the *Sox9–Kcnj2* locus. We show that the removal of all major CTCF sites at the boundary and within the TAD resulted in a fusion of neighboring TADs, without major effects on gene expression. Gene misexpression and disease phenotypes, however, were achieved by redirecting regulatory activity through inversions and/or the repositioning of boundaries. Thus, TAD structures provide robustness and precision but are not essential for developmental gene regulation. Aberrant disease-related gene activation is not induced by a mere loss of insulation but requires CTCF-dependent redirection of enhancer–promoter contacts.

Chromosomes are organized in a specific three-dimensional (3D) structure in the nuclear space, a phenomenon that is directly linked to gene regulation[1–3]. On a gene locus level, this organization is characterized by regions of high interaction called TADs that are separated from each other by so-called boundaries[4,5]. TADs bring distant *cis*-regulatory elements such as promoters and enhancers into proximity, whereas boundaries are thought to act as insulators to preclude inappropriate enhancer–promoter interactions with neighboring genes or regulatory elements. This concept provides a basic framework for long-range gene regulation[6] and also has important implications for the interpretation of genomic rearrangements (structural variations)[7].

One key component for TAD and boundary formation is the zinc-finger transcription factor CTCF, which acts in concert with the multisubunit protein complex Cohesin[8]. In the currently prevailing model, TAD formation is the result of a loop extrusion process in which cohesin molecules extrude a chromatin loop, and thereby bring distant DNA fragments into spatial proximity[9,10]. In this model, CTCF-binding sites act as a barrier for the extrusion machinery in an orientation-dependent manner. This view is supported by the finding that a large fraction of TAD boundaries harbor clusters of CTCF-binding sites that are characteristically positioned in divergent orientation[9,11].

The importance of the CTCF–cohesin machinery for higher-order chromatin architecture has further been corroborated by experimental approaches that allow for the temporary genome-wide depletion of CTCF or various subunits of the cohesin complex,

circumventing their absolute requirement for cell survival[8,12–15]. Cells in which CTCF or cohesin is depleted lose most of their TAD structures. In spite of this dramatic loss in 3D genome organization, however, only modest effects on gene expression were observed[8,12,15]. Less than half of the regulated genes exhibit elevated expression, suggesting only spurious gains in enhancer–promoter interactions in the absence of TADs and boundaries. These results are seemingly in contrast with previous findings in which the rearrangement of TADs and their boundaries were shown to have dramatic effects on gene regulation, resulting in congenital disease or cancer[16–18]. The basis for this apparent discrepancy remains unclear, raising the question about the functional importance of TADs for gene regulation and the proposed molecular pathology of structural variants.

Here, we dissect the role of CTCF and TAD architecture for gene regulation in a developmental in vivo setting in mice. We created a series of genome-engineered mice with targeted mutations at the *Sox9–Kcnj2* locus and analyzed their effect on 3D chromatin architecture, gene regulation and phenotype. The *Kcnj2* and *Sox9* TADs are separated by a strong boundary, but a fusion of the TADs, as indicated by Hi-C, was achieved only after removal of all major CTCF sites at the boundary and within the TAD. TAD fusion, however, was not accompanied by major gene-regulatory effects, suggesting that long-range gene regulation does not exclusively rely on intact TAD structures. In contrast, inversions and the insertion of boundary elements were able to redirect regulatory activity inducing enhancer–promoter rewiring, gene misexpression and developmental phenotypes. Thus, TADs and, in particular, CTCF sites are
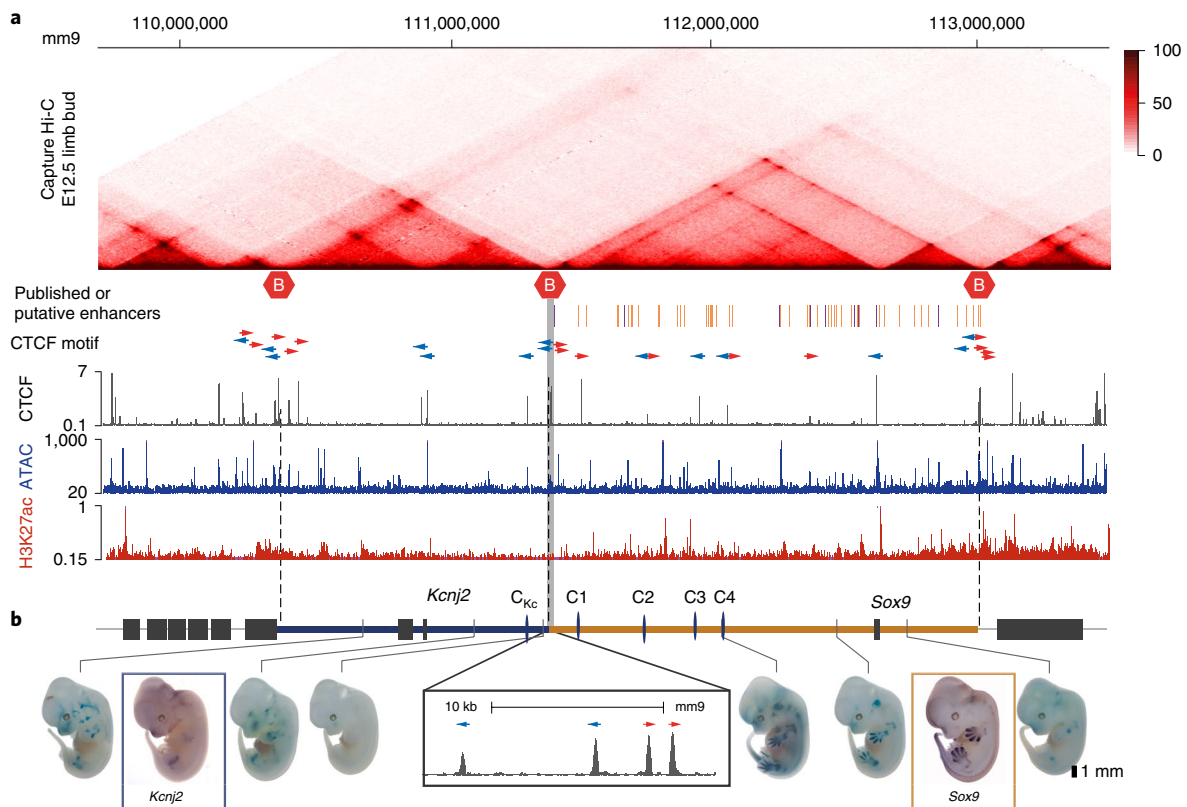
Fig. 1 | TAD configuration and regulatory activity at the *Sox9* locus. a, cHi-C from E12.5 mouse limb buds with boundaries are indicated by red hexagons. CTCF ChIP-seq, ATAC-seq and H3K27ac are shown below (*n* = 1). Bars represent putative (orange) or published (purple) enhancers identified by ATAC-seq and H3K27ac ChIP-seq signals. **b**, Schematic of the locus. Genes are indicated by black bars; TADs of *Kcnj2* (blue) and *Sox9* (orange) are shown. Boundary region between *Sox9* and *Kcnj2* TADs is highlighted in gray; magnification below shows the cluster of four divergent CTCF sites within a 15-kb region. Other major CTCF-binding sites are indicated and labeled as C1, C2, C3, C4 and C$_{Kc}$. Lower panel shows the activity of regulatory sensors in E12.5 embryos inserted at indicated positions. The expression pattern (*WISH*) of *Kcnj2* and *Sox9* is shown for comparison.

not essential for correct developmental gene expression, but they can induce gene misexpression when redirected.

Results

Two TADs define the regulatory landscape at the *Sox9–Kcnj2* locus. *Sox9* and *Kcnj2* are two adjacent genes with distinct expression patterns in the developing limb bud that are separated by a 1.7-megabase (Mb) gene desert (Fig. 1). In embryonic day (E) 12.5 limb buds, *Sox9* is expressed in the cartilage anlagen of the developing limbs, whereas *Kcnj2* is only weakly expressed in the distal zeugopod (Fig. 1b). Capture Hi-C (cHi-C) from mouse limb buds shows that the locus is divided in two TADs, one harboring *Sox9* and the other *Kcnj2* and *Kcnj16* (ref. [16]). The TAD boundary is characterized by two pairs of CTCF-binding sites with divergent orientation, showing strong loop formation with their neighboring boundaries (Fig. 1a). Within the *Sox9* TAD, a nested substructure with various loops is linked to at least four additional CTCF-binding sites (C1–C4). To profile the regulatory landscape of the locus in more detail, we performed transposase accessible chromatin sequencing (ATAC-seq) and H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) from E12.5 limb buds. Within the *Sox9* TAD, we identified 52 putative enhancers positive for H3K27ac and ATAC-seq, 9 of which were previously shown to drive limb expression in transgenic reporter assays (Fig. 1a and Supplementary Table 2). To capture the *cis*-regulatory activity of the locus, we used regulatory sensors (*β-globin* minimal promoter with *LacZ* reporter gene) that we integrated at various positions within the *Kcnj2* and *Sox9* TADs (Fig. 1b). LacZ staining of E12.5 embryos revealed that

all sensors within the *Sox9* TAD recapitulated the *Sox9* expression pattern, whereas sensors integrated within the *Kcnj2* TAD reflected the endogenous expression pattern of *Kcnj2*. This shows that the 3D genome organization in the two TADs corresponds with the regulatory domains of the *Kcnj2* and *Sox9* genes.

Boundaries and internal CTCF sites act cooperatively to form TADs. To investigate the role of CTCF in maintaining the TAD structure at this locus, we generated mice with a series of alleles in which the four CTCF sites at the TAD boundary were deleted, followed by a consecutive deletion of five further sites within the TADs. cHi-C was performed from E12.5 limb buds to visualize the effects on TAD architecture and to quantify the contacts within the TADs (intra-TAD) and between the TADs (inter-TAD). We also produced virtual 4C interaction profiles from the cHi-C data to assess contact changes of *Sox9* and *Kcnj2* in these alleles (see Methods).

Deletion of the four CTCF sites that constitute the *Sox9–Kcnj2* TAD boundary (*ΔBor*) resulted in a moderate increase of contacts between TADs, but the two TADs remained largely separate (Fig. 2a,b)[16]. To test whether the intra-TAD CTCF sites contribute to TAD formation, we sequentially deleted, in addition to the boundary, one (*ΔBorC1*), two (*ΔBorC1–2*) and all four (*ΔBorC1–4*) of the major CTCF sites within the *Sox9–Kcnj2* TAD. Deletion of the C1 CTCF site together with the boundary deletion (*ΔBorC1*) led to a marked increase in contacts between the *Sox9* and *Kcnj2* TADs. Deletion of the C2 CTCF site (*ΔBorC1–2*) (Supplementary Fig. 1) and deletion of all four CTCF sites (*ΔBorC1–4*) led to a further increase in inter-TAD contacts and a near-complete fusion
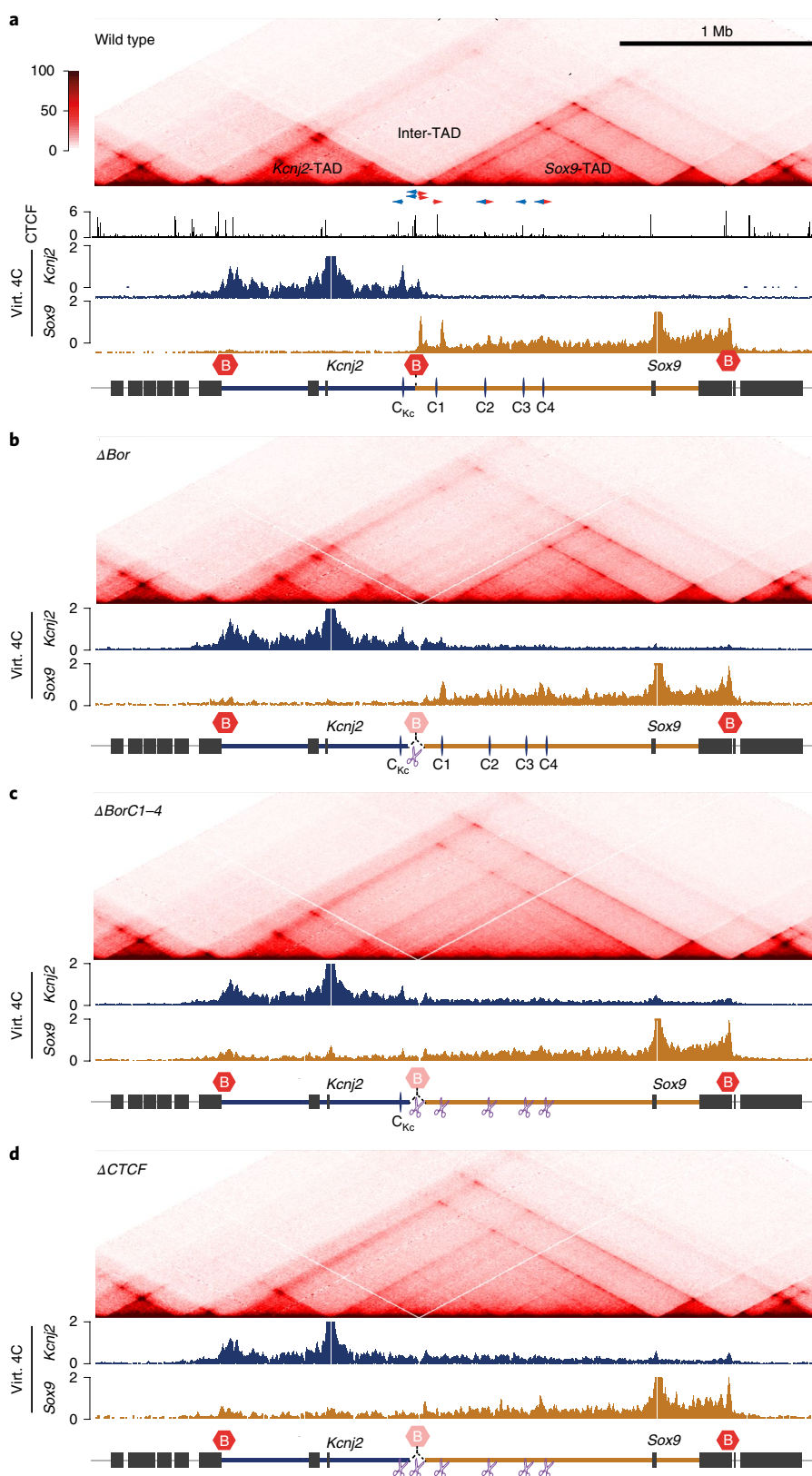
Fig. 2 | Progressive fusion of the *Kcnj2* and *Sox9* TADs on deletion of the TAD boundary and intra-TAD CTCF sites. cHi-C from E12.5 mouse limb buds. Virtual 4C with viewpoints at the *Kcnj2* (blue) or the *Sox9* promoter (orange) below. CTCF ChIP-seq with binding site orientation (red–blue) at the boundary and intra-TAD are highlighted. Two-headed arrow indicates two oppositely oriented sites (FIMO $P < 10^{-4}$) underlying the ChIP-seq peak. **a**, Wild-type cHi-C ($n = 1$). Dashed lines indicate *Kcnj2* and *Sox9* TADs and area of inter-TAD contacts. **b**, Eighteen-kilobase deletion (ΔBor) of the TAD boundary leaves the TAD configuration largely unchanged ($n = 1$). **c**, Deletion of the TAD boundary and targeted deletion of all four major intra-TAD CTCF-binding sites ($\Delta BorC1$–4, $n = 1$) cause TAD fusion. **d**, $\Delta CTCF$ shows further TAD fusion on deletion of the TAD boundary and all CTCF-binding sites between the *Kcnj2* and *Sox9* promoters (C1–C4 and C$_{Kc}$; $n = 1$). Virt., virtual.

between the *Sox9* and *Kcnj2* TADs (Fig. 3a and Supplementary Fig. 1). The strongest increase, however, was induced by the C1 site deletion. Whereas the internal TAD structure and loops disappeared, new interactions between the *Sox9* and *Kcnj2* promoters, as well as between the outer TAD boundaries, emerged. Further deletion of the remaining single major CTCF site in the *Kcnj2* TAD (*ΔCTCF*) abolished all major CTCF-binding sites between the *Sox9* and *Kcnj2* promoters. cHi-C from these animals showed a further increase of inter-TAD contacts between the former *Sox9* and *Kcnj2* TADs (Figs. 2d and 3a). Taken together, our data show that the TAD boundary deletion alone does not result in TAD fusion. Instead, formation and integrity of the *Sox9* and *Kcnj2* TADs are established by the TAD boundary in combination with the CTCF-mediated TAD substructure.

Loss of TAD insulation has minor effects on developmental gene regulation. We next investigated how the gradual fusion of TADs might affect the expression of *Kcnj2* and *Sox9* and the phenotype. We previously demonstrated that misexpression of *Kcnj2* in a *Sox9*-like pattern results in a malformation of the terminal phalanges (in humans, it is called Cooks syndrome, MIM 106995)[16]. Heterozygous loss of *Sox9*, in contrast, leads to a lethal skeletal phenotype characterized by bowing of long bones, cleft palate and rib abnormalities in heterozygous animals (in humans, it is called campomelic dysplasia, MIM 114290), and homozygous mutants do not form cartilage at all[19]. In addition, *Sox9* is an essential factor for testis development downstream of SRY, and its inactivation results in male-to-female sex reversal[20]. To monitor gene expression changes in embryos, we used quantitative PCR (qPCR) and whole-mount in situ hybridization (WISH). Phenotypes associated with loss of *Sox9* and/or gain of *Kcnj2* were assessed in mice by visual inspection (palate, claws), skeletal preparations, microtomography (μCT, digits) and testing fertility through breeding.

Despite the observed fusion of TADs, we did not observe any dramatic changes in gene expression. Even though the changes in *Sox9* gene expression were not substantial in the boundary deletion or the *ΔBorC1–4* mutants, we detected an ~10–15% reduction of *Sox9* expression in the *ΔCTCF* animals (Fig. 3b and Supplementary Fig. 2a). *Kcnj2*, which is only marginally expressed at E13.5, was not misexpressed in *ΔBor* animals but increased slightly to ~2-fold in all alleles. To detect whether there were changes in the patterns of expression in developing limbs, we performed WISH in E12.5 embryos. In all lines, *Sox9* expression was indistinguishable from wild-type embryos. Also, *Kcnj2* stayed unchanged, and no *Sox9*-like misexpression was detected in *ΔBor* or *ΔBorC1–2* animals (Fig. 3a and Supplementary Fig. 2b). However, on deletion of four or more CTCF sites in addition to the TAD boundary, faint *Kcnj2* expression in the digit anlagen was detected (*ΔBorC1–4* and *ΔCTCF*; Fig. 3c). Importantly, despite the fusion of TAD structures, all mutant animals were viable, bred to Mendelian ratios and had no detectable phenotype. In particular, no abnormality of the digits or claws was observed (Supplementary Fig. 2c).

Thus, in spite of the observed TAD fusion, *Sox9* and *Kcnj2* expression remained largely unchanged with no phenotypic effect. These results indicated that enhancers were able to efficiently contact their target gene even without CTCF-mediated loops. Removal of all CTCF sites, however, led to a spillover of activity from the *Sox9* TAD to the *Kcnj2* TAD, indicating that TAD genome organization provides a certain degree of robustness and precision to gene expression at this locus.

Reorganization of 3D contacts by TAD boundaries and TAD substructure orientation. TAD fusions induced by large structural variants have been reported to cause gene misexpression and disease[17], yet the TAD fusion induced by CTCF site deletion occurred without major gene misexpression. To investigate this discrepancy

further, we produced mice with four different types of inversions and/or insertions: (1) an inversion of the *Sox9* regulatory domain including the TAD boundary (*InvC*); (2) an inversion of the *Sox9* regulatory domain without the boundary (*Inv-Intra*); (3) an insertion of the boundary alone without inverting the regulatory domain (*Bor-KnockIn*); and finally, (4) an inversion of the *Sox9* regulatory domain with the boundary removed (*InvCΔBor*) (summarized in Supplementary Fig. 3). This combination of alleles allowed us to dissect the role of TAD boundaries and substructure orientation in TAD formation and for structural variants.

cHi-C from the inversion of the centromeric 1.1 Mb of the *Sox9* TAD including the TAD boundary (*InvC*) (E12.5 limb buds) showed a fusion of the inverted part of the *Sox9* TAD with the *Kcnj2* TAD and a separation of the *Sox9* gene and its remaining TAD from the inverted part (Fig. 4a,b). cHi-C and virtual 4C showed that *Kcnj2* was now able to contact the C1–C4 CTCF sites in a similar fashion as *Sox9* in the wild-type situation (Fig. 4a,b).

To investigate the effect of the inverted TAD substructure versus the repositioned boundary, we produced a slightly smaller inversion not including the boundary (*Inv-Intra*). cHi-C of this intra-TAD inversion showed that the overall extent of interactions did not change. However, the redirection of CTCF sites resulted in an altered pattern of loop formation (Fig. 4c). The contacts between intra-TAD CTCF sites and the TAD boundary became stronger, whereas those with the *Sox9* promoter became weaker, as if the entire region had shifted its interaction toward the TAD boundary (Supplementary Fig. 4).

To test the effect of the TAD boundary alone, we used the boundary deletion background to insert a 6.3-kb construct carrying the four boundary CTCF sites 125 kb upstream of *Sox9* (*Bor-KnockIn*). cHi-C of this allele showed that the repositioned TAD boundary split the *Sox9* TAD into two domains. Similar to the TAD-spanning *InvC* inversion, the telomeric region containing the *Sox9* gene was now separated from the centromeric region (Fig. 4d). However, in contrast with the *InvC* inversion, the centromeric *Sox9* TAD did not fuse with the *Kcnj2* TAD but remained an isolated domain extending from the C1 CTCF site to the inserted boundary. Thus, the boundary was fully functional even at a different position, separating the *Sox9* TAD into two domains.

Finally, we wanted to test the effect of TAD substructure orientation without the influence of a nearby boundary. We therefore deleted the TAD boundary in the *InvC* inversion background (*InvCΔBor*). This resulted in a loss of insulation and reestablished *Sox9* contacts with its centromeric TAD. At the same time, the contacts between the inverted part of the *Sox9* TAD and *Kcnj2* were still present (Fig. 4e). Thus, this mutant resulted in a fusion of the entire *Sox9* and *Kcnj2* TADs, and both *Sox9* and *Kcnj2* were contacting the inverted *Sox9* regulatory region.

To compare how TAD orientation and boundary position affect the contact frequency, we quantified the contacts of the *Sox9* and *Kcnj2* promoters with all putative enhancers or with the entire inverted region (centromeric part of the *Sox9* TAD) (Fig. 4f, Supplementary Fig. 5 and Supplementary Table 2). Analysis of the TAD-spanning inversion (*InvC*) and the boundary knock-in (*Bor-KnockIn*) showed that the repositioned boundary caused a strong reduction of contacts of *Sox9* with its putative enhancers in the inverted region, a phenomenon that was much weaker when only the TAD substructure was inverted (*Inv-Intra* and *InvCΔBor*) or not observed when the CTCF sites were deleted (*ΔCTCF*) (Supplementary Fig. 5).

For *Kcnj2*, there was no increase in *Sox9* enhancer contacts for the *Inv-Intra* and *Bor-KnockIn* alleles. In contrast, *Kcnj2* gained contacts with *Sox9* enhancers in the *InvC* and *InvCΔBor* alleles and to a similar degree in the *ΔCTCF* allele. However, in contrast with the *InvC* or *InvCΔBor* alleles, *Sox9* continued to contact the region in the *ΔCTCF* allele (Supplementary Fig. 5).
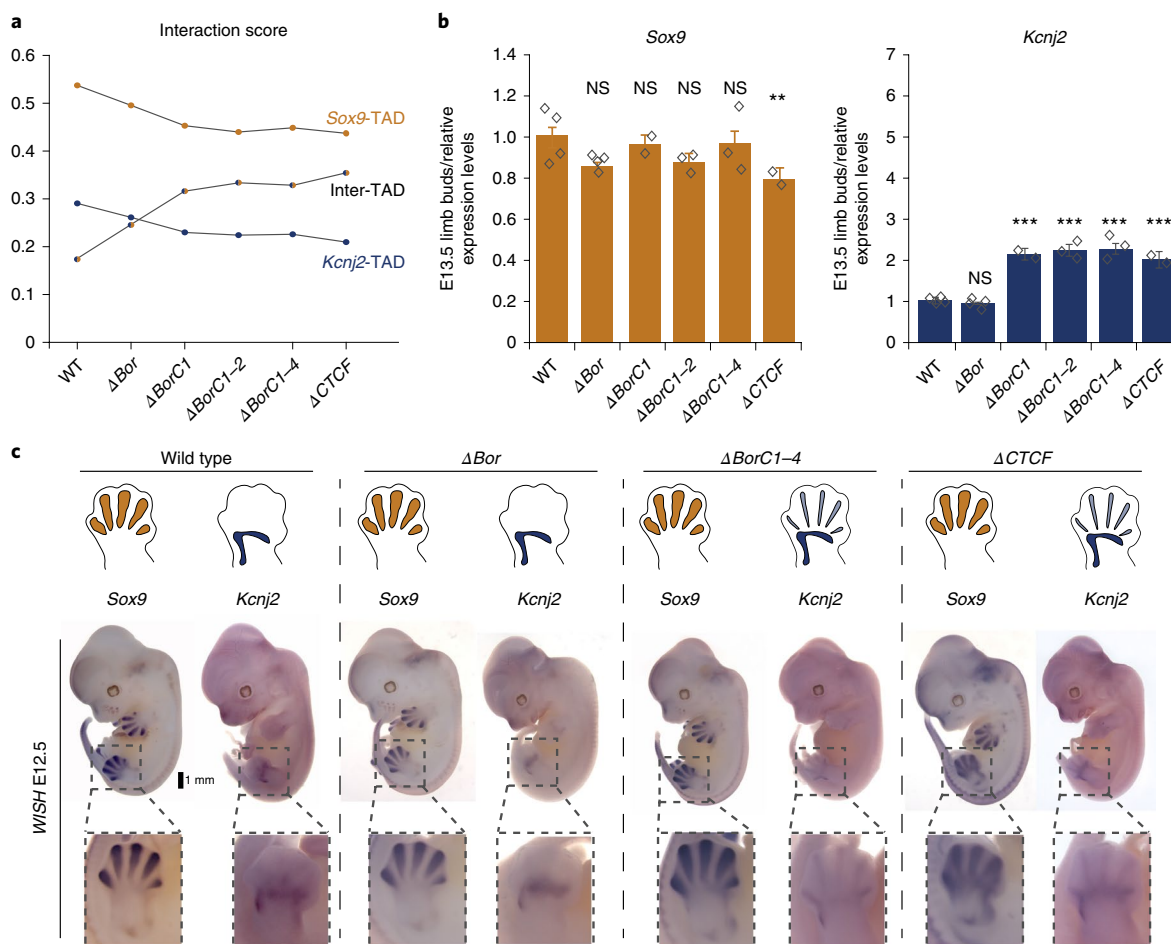
Fig. 3 | Effect of TAD fusion and deletion of CTCF-binding sites on gene expression and phenotype. a, Change of interaction score induced by consecutive removal of CTCF sites. Fraction of contacts in the *Kcnj2* TAD, *Sox9* TAD and between TADs (inter-TAD). The more CTCF-binding sites are deleted, the higher the contact frequency between the *Kcnj2* and *Sox9* TADs. **b**, Relative gene expression of *Kcnj2* and *Sox9* in E13.5 limb buds measured by qPCR. Values are normalized to *Gapdh* expression (wild type = 1). Bars represent the mean, error bars represent the standard deviation and diamonds indicate individual replicates. Significance in comparison with wild-type levels tested with one-sided, unpaired *t* test ($n = 2$–4 as indicated; **$P < 0.01$, ***$P < 0.001$). **c**, Expression pattern of *Sox9* and *Kcnj2* in E12.5 limbs (*WISH*, $n \geq 3$). Schematic is on top, WISH is below and detailed view of hindlimbs is at the bottom. *Sox9* is strongly expressed in the digit anlagen, whereas *Kcnj2* is expressed weakly in the distal stylopod. Note no change in *Sox9* expression and low degree of digit expression of *Kcnj2* in only *ΔBorC1–4* and *ΔCTCF*. NS, not significant.

Redirection of the TAD structure results in distinct regulatory effects and phenotypes. In contrast with the CTCF site deletions, the structural variants had pronounced regulatory and phenotypic effects. Animals carrying the *InvC* allele, in which the inverted centromeric *Sox9* TAD fused with the *Kcnj2* TAD, showed a clear loss of function of *Sox9* and a gain of function of *Kcnj2*. *Sox9* expression decreased by 20% in heterozygous and 50% in homozygous animals (Fig. 5b and Supplementary Fig. 6a). As a consequence, heterozygous animals displayed delayed ossification of the skeleton, and homozygous animals showed perinatal lethality with all hallmarks of a *Sox9* loss-of-function phenotype (that is, bowing of long bones, delayed ossification, cleft palate) (Supplementary Fig. 6c). *Kcnj2* expression increased up to 5-fold in homozygous embryos (Fig. 5b and Supplementary Fig. 6a). WISH in E12.5 embryos revealed a strong misexpression of *Kcnj2* in a *Sox9*-like pattern in the digit anlagen (Fig. 5c). Heterozygous *InvC* animals showed malformed terminal phalanges with high penetrance, the phenotype associated with *Kcnj2* misexpression in a *Sox9*-like pattern. Homozygous animals had severely dysplastic digits, preventing the development of this phenotype. Thus, the TAD-spanning inversion *InvC* reorganized the TADs at the locus and redirected regulatory activity from *Sox9* to *Kcnj2*.

In contrast, animals with the slightly smaller *Inv-Intra* inversion had no abnormal phenotype. Mice were viable and fertile, suggesting no major misregulation of *Sox9* or *Kcnj2*, which was confirmed by qPCR and WISH (Fig. 5b,c). These results indicate that the orientation of the internal TAD structure has no major effect as long as the boundary is intact.

Consistent with the cHi-C pattern, the knock-in of the border (*Bor-KnockIn*) showed a *Sox9* loss of function but no *Kcnj2* gain of function. *Sox9* levels were reduced by ~40% in homozygous E13.5 limb buds (Fig. 5b), and *Kcnj2* was upregulated ~2-fold, similar to the CTCF deletion alleles, and WISH showed no *Sox9*-like *Kcnj2* misexpression in E12.5 limb buds (Fig. 5a,c). Phenotypically, homozygous *Bor-KnockIn* animals died perinatally because of *Sox9*-related defects including cleft palate, short snout, shortened long bones and delayed ossification (Supplementary Fig. 6d). However, the skeletal phenotypes were less severe than those seen in the *InvC* animals. Importantly, in accordance with the regulatory effects of the boundary knock-in, the animals had normal phalanges (Supplementary Fig. 6d).

Finally, we tested gene expression and phenotypes in the *InvCΔBor* allele, in which the entire *Kcnj2* and *Sox9* TADs are fused, and the centromeric *Sox9* TAD contacted both *Sox9* and *Kcnj2*.
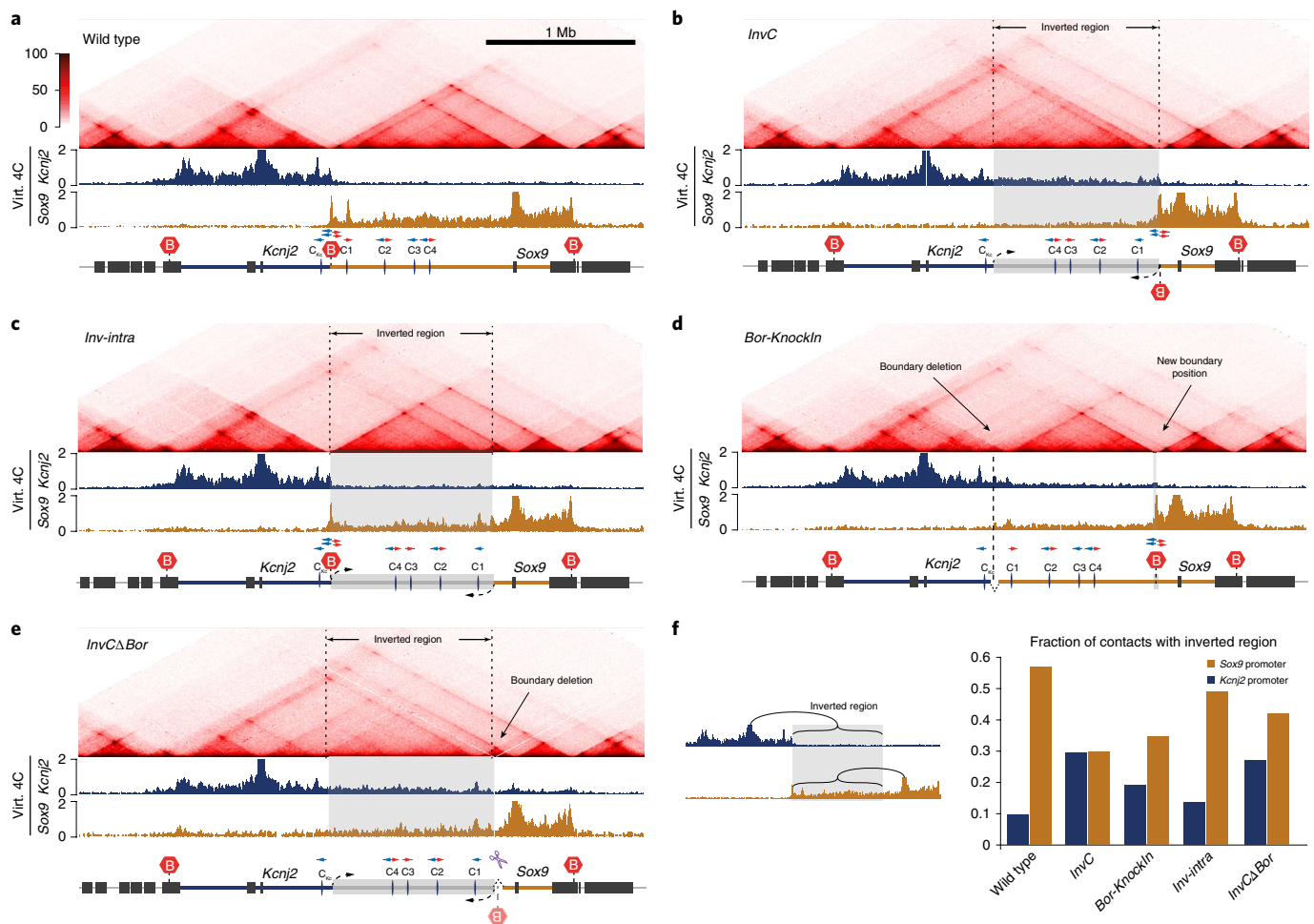
Fig. 4 | Boundaries and orientation of regulatory landscapes define TAD organization. cHi-C from E12.5 mouse limb buds, each mapped to a custom genome (**b–e**, $n = 1$), is shown. Virtual 4C with viewpoint at the *Kcnj2* (blue) or *Sox9* promoter (orange) below; tracks refer to the same genome as the cHi-C map above. Gray box indicates the extent of inverted region. **a**, Wild type. **b**, Inversion including boundary (*InvC*) leads to fusion of the inverted region with the *Kcnj2* TAD and separation of *Sox9* from its regulatory domain. **c**, Inversion excluding the boundary (*Inv-Intra*) has no major changes in TAD configuration. **d**, Repositioning of the boundary to a new position near *Sox9* (*Bor-KnockIn*) isolates *Sox9* from its TAD but does not cause the fusion with the *Kcnj2* TAD. **e**, InvC inversion with a deleted boundary (*InvCΔBor*) causes fusion of both TADs and reestablishes *Sox9* contacts with its TAD. **f**, Fraction of contacts with the inverted region for each allele. Bar diagram shows contacts of *Kcnj2* (blue) or the *Sox9* (orange) with the inverted region as a fraction of all contacts in the *Kcnj2* and *Sox9* TADs measured by 4C.

Consistent with the reestablished cHi-C interactions, *Sox9* expression was less severely reduced than in the *InvC* inversion (Fig. 5b and Supplementary Fig. 6a), and the *Sox9* expression pattern in E12.5 WISH was indistinguishable from wild-type *Sox9* expression (Fig. 5c). On a phenotypic level, the *InvCΔBor* allele rescued the *Sox9* loss-of-function effects of the *InvC* inversion. Homozygous *InvCΔBor* animals had no cleft palate and were viable. The only obvious *Sox9*-related phenotype of this allele was that homozygous males were infertile (Fig. 5c). Nonetheless, the *Kcnj2* gain-of-function effects were still present. E13.5 *Kcnj2* expression levels were 2.5-fold higher than wild type but lower than in the *InvC* allele (Fig. 5b and Supplementary Fig. 6a). WISH, however, showed a clear *Sox9*-like expression pattern in the limb buds. Most importantly, *InvCΔBor* animals showed the abnormal terminal phalanx phenotype, which is characteristic of *Sox9*-like misexpression of *Kcnj2* (Supplementary Fig. 6b).

Thus, changes in directionality of the TAD substructure (including their CTCF sites) and/or the repositioning of boundary elements at the *Sox9–Kcnj2* locus caused phenotype-inducing alterations in gene expression. Such redirecting of regulatory activity can result in either a gain of expression when *Sox9* enhancers are

forced to act on *Kcnj2*, or a loss when *Sox9* is disconnected from its regulatory domain.

Discussion

Here we functionally dissect the role of TAD boundaries, intra-TAD CTCF sites and directionality of TAD substructures for TAD formation and gene regulation in a developmental in vivo model at the *Sox9–Kcnj2* locus. With its two large TADs and the distinct expression patterns of its corresponding genes, as well as associated phenotypes, this region is ideally suited for this approach.

The deletion of the *Sox9–Kcnj2* TAD boundary did not have a major effect on the overall TAD configuration. Intra-TAD CTCF sites needed to be deleted in addition to the TAD boundary to achieve effective TAD fusion. Interestingly, not all intra-TAD CTCF sites contributed equally to insulation, suggesting different insulator potential of individual binding sites. Our results, however, do not show a direct relation of CTCF binding strength or motif orientation to the impact on cHi-C insulation. Our data demonstrate a redundancy in spatial separation of TADs that originates from the combinatorial action of CTCF sites at the TAD boundary and within the TAD. A similar resilience of TAD structures was recently
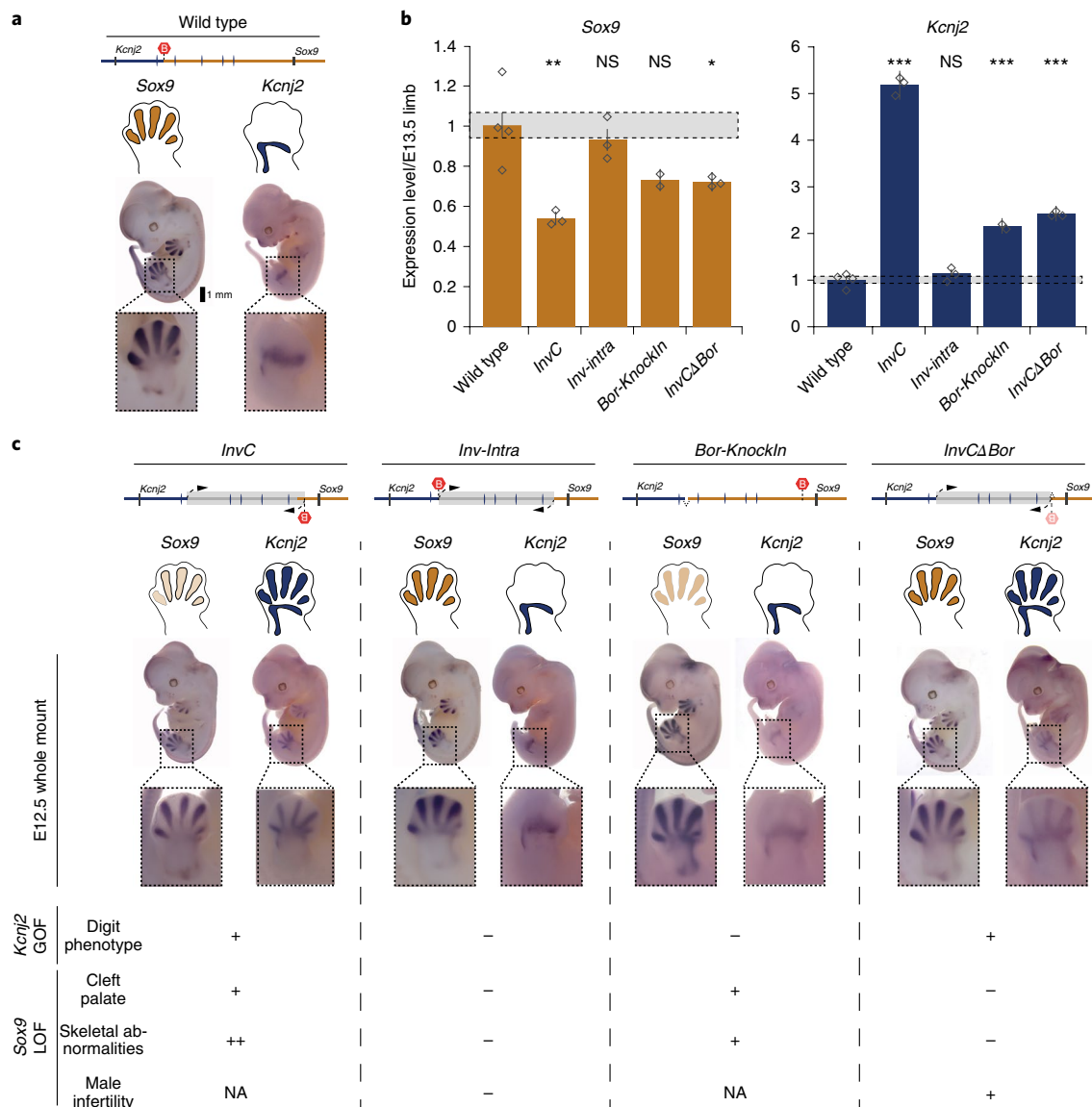
Fig. 5 | Redirection of regulatory activity results in *Kcnj2* misexpression and loss of *Sox9* expression. a, Wild-type gene expression pattern of *Sox9* and *Kcnj2* (*n* = 6). **b**, Relative gene expression of *Kcnj2* and *Sox9* in E13.5 limb buds measured via qPCR. Values are normalized to *Gapdh* expression (wild type = 1). Bars represent the mean, error bars represent the standard deviation and diamonds indicate individual replicates. Significance in comparison with wild-type levels tested with one-sided, unpaired *t*-test (*n* = 2–4 as indicated; *P < 0.05, **P < 0.01, ***P < 0.001). **c**, Schematic of structural variant. Below shows WISH of *Sox9* and *Kcnj2* in E12.5 limbs (*n* ≥ 3). Schematic is on top, WISH is below and detailed view of hindlimbs is at the bottom. *Kcnj2* gains *Sox9*-like expression pattern in digits in *InvC* and *InvCΔBor*. *Sox9* expression is visibly reduced in *InvC* embryos. Summary of phenotypes in animals induced by gain of *Kcnj2* expression (GOF) and/or loss of *Sox9* expression (LOF). NA, not available; NS, not significant.

reported for TADs at the HoxD gene cluster, which itself acts as a strong TAD boundary. Here, only a 400-kb deletion encompassing the entire HoxD cluster and two flanking genes leads to fusion of the centromeric and telomeric HoxD TADs[21]. Previous work on boundaries, which deleted specific CTCF sites at the HoxA cluster[22] or removed 58 kb surrounding a boundary CTCF site at the *Xist* TAD[4], suggested that other factors in addition to boundaries must contribute to TAD formation. With the deletion series presented here, we demonstrate the importance of intra-TAD CTCF sites for TAD formation through buffered and redundant mechanisms, and highlight the function of boundaries as strong insulators at the edges of TADs. At the *EphA4* locus the presence of a TAD boundary in pathogenic deletions determines whether two TADs fuse, demonstrating that it functions as a potent insulator. The resulting TAD fusion at the *EphA4* locus leads to misexpression of *Pax3*. This, however, is the

result of a deletion that removes not only the boundary but, in addition, the majority of the *Epha4* TAD and alters the entire 3D structure at the locus.

In contrast, the serial deletions of CTCF sites here leave the overall configuration of the locus intact but modify the barrier function between TADs. We thereby provide direct evidence that TADs are formed by a seemingly redundant system of CTCF sites at the TAD boundaries and within the TADs.

We addressed the role of CTCF at an individual locus in an in vivo developmental setting and thereby avoid the genome-wide effects associated with a loss of CTCF. Surprisingly, the gradual fusion of the two neighboring TADs was accompanied by only mild changes in gene regulation, indicating that the enhancers were able to find, contact and regulate their cognate promoters. The large number of putative enhancers in the region indicates that *Sox9* is

regulated by a complex interplay of many enhancers distributed throughout the TAD, and that no single CTCF site is essential for enhancer function.

The loss of TAD structures at the *Sox9* locus also had no phenotypic consequences, indicating that there were no substantial effects on *Sox9* or *Kcnj2* regulation throughout development. These results are in agreement with the weak gene expression changes observed on CTCF depletion[8,15], suggesting that enhancer–promoter communication can function independently of TADs and CTCF-mediated genome architecture. Such a mechanism might be mediated by homotypic interaction of TFs that bind at distal enhancers and their cognate promoters, and could promote transcriptional condensates as recently proposed[23]. In such a scenario, TADs would act as 3D scaffolds that optimize such interaction hubs, without being essential to establish them.

Another reason for the mild effects on gene regulation is that the deletion of CTCF sites does not affect cohesin recruitment to the chromatin, which is independent of CTCF[24]. Thus, also in the absence of CTCF sites, cohesin complexes at the *Sox9* locus can still facilitate enhancer–promoter interaction. What does change, however, are the limits for the extrusion complexes that are normally set by CTCF. The consecutive removal of the boundary and intra-TAD CTCF sites leads to increasing contacts of *Kcnj2* with the active *Sox9* regulatory landscape and no major change of *Sox9* contacts with its enhancers. However, this results in only an incremental increase in *Kcnj2* expression indicating that spurious contacts do not directly result in gene regulation. Interestingly, the contact frequency of *Kcnj2* with *Sox9* enhancers was similar for the inversion and the *ΔCTCF* alleles, in spite of the fact that they have a very different regulatory and phenotypic outcome. This effect might be caused by competition of *Sox9* and *Kcnj2* promoters for the same enhancers in the *ΔCTCF* allele, whereas the interaction of *Sox9* with its enhancers is prevented in the inversion by the repositioned boundary. However, because of lethality of *Sox9* promoter deletions, this cannot easily be assessed in an in vivo setting.

Overall, the loss of CTCF-mediated insulation was accompanied by a loss of *Sox9* and a gain of *Kcnj2* expression induced by a spread of regulatory activity from the *Sox9* to its neighboring TAD. Thus, at this locus, TADs and their boundaries are not essential for developmental gene expression, but they confer precision and robustness. For *Sox9* and *Kcnj2*, the relatively mild expression changes had no phenotypic effect. However, in other cases where precision and insulation are essential, such leaky expression can result in disease phenotypes[22].

Our findings support the idea that the spatial separation into TADs and enhancer–promoter interaction represent two independent layers of long-range gene regulation. These layers stabilize each other but are not inherently linked. Furthermore, they contradict the generally accepted idea that enhancers are promiscuous[6]; that is, they can and will activate any promoter in their vicinity. This is supported by reports that show that various mechanisms including promoter type[25], histone modifications, proximity to target promoter and openness of chromatin[26] influence the transcriptional response of promoters to enhancers.

Our in-depth analysis between 3D chromatin structure and gene regulation focuses on one locus, but the link between structure and function might differ in other TADs. However, a similar analysis would not be feasible on a genome-wide scale, highlighting the importance of a comprehensive approach for understanding gene regulation at any given locus.

With our series of inversions and knock-in alleles, we were able to dissect the relevance of TAD substructure and TAD boundaries for altering gene expression. In the TAD-spanning *InvC* inversion, the repositioned TAD boundary serves as a strong insulator separating *Sox9* from its regulators. At the same time, redirection of the TAD substructure creates new loops with the *Kcnj2* gene and

fuses the *Sox9* regulatory domain with the *Kcnj2* TAD. This combination is apparently sufficient to connect the *Sox9* enhancers to the *Kcnj2* promoter, thereby overcoming their inherent affinity for the *Sox9* promoter. As a consequence, we observe a *Sox9* loss-of-function phenotype and a *Kcnj2* gain-of-function. Thus, misexpression and disease can be induced by redirecting TAD substructures and enhancer activity, but not by removing them. In this context, boundary and TAD substructure function together, but the substructure with its CTCF sites cannot override a boundary. However, substructure inversion is needed to achieve pathogenic misexpression.

Our results help to explain the apparent discrepancy between the modest effects of CTCF–cohesin depletion on transcription and the drastic effects of TAD reorganization in pathogenic structural variants. Based on our findings, it is to be expected that, because of the high redundancy of CTCF sites in maintaining TAD structure, most smaller structural variants are likely to be tolerated with respect to TAD structure and enhancer–promoter interaction. To result in aberrant gene activation, the rearrangement will likely need to actively reorganize 3D chromatin contacts and thereby connect a regulatory region to the new target gene. Such effects can be achieved by deleting boundaries together with adjacent divergent CTCF sites or by redirecting regulatory activity through inversions or duplications. The repositioning of boundaries, in contrast, can result in loss of expression induced by disconnecting a gene from its regulatory domain. Thus, structural variant-induced misexpression is not caused by the simple removal of barriers or the effect of single enhancer–promoter rewiring. Rather, it is the result of connecting larger regulatory structures with novel target genes through CTCF-mediated loops.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0466-z.

References

1. Bonev, B. & Cavalli, G . Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).
2. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
3. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
4. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
5. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
6. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361**, 1341–1345 (2018).
7. Spielmann, M., Lupiáñez, D. G. & Mundlos, S. Structural variation in the 3D genome. *Nat. Rev. Genet.* **19**, 453–467 (2018).
8. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944.e22 (2017).
9. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
10. Sanborn, A.L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad Sci. USA* **112**, E6465–E6465 (2015).
11. Vietri Rudan, M. et al. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10**, 1297–1309 (2015).
12. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
13. Wutz, G. et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).

14. Busslinger, G. A. et al. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, 503–507 (2017).

15. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017).

16. Franke, M. et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).

17. Lupianez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).

18. Weischenfeldt, J. et al. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat. Genet.* **49**, 65–74 (2016).

19. Bi, W., Deng, J. M., Zhang, Z., Behringer, R. R. & de Crombrugghe, B. Sox9 is required for cartilage formation. *Nat. Genet.* **22**, 85–89 (1999).

20. Barrionuevo, F. et al. Homozygous inactivation of Sox9 causes complete XY sex reversal in mice. *Biol. Reprod.* **74**, 195–201 (2006).

21. Rodriguez-Carballo, E. et al. The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* **31**, 2264–2281 (2017).

22. Narendra, V., Bulajić, M., Dekker, J., Mazzoni, E. O. & Reinberg, D. CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev.* **30**, 2657–2662 (2016).

23. Boija, A. et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855.e16 (2018).

24. Wendt, K. S. et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).

25. Zabidi, M. A. et al. Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556–559 (2014).

26. Kraft, K. et al. Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. *Nat. Cell Biol.* **21**, 305–310 (2019).

Author contributions

D.M.I. and S.M. conceived the project. A.D., D.M.I. and M.F. performed cHi-C. R.S. and D.M.I. performed the computational analysis with input from M.V. D.M.I., A.D., S.A. and M.F. produced transgenics, carried out transgenic validation and performed expression–phenotypic analysis. C.P. performed ATAC-seq. A.D., I.J. and D.M.I. performed ChIP-seq. B.T. oversaw sequencing of the cHi-C, ChIP-seq and ATAC-seq. L.W. performed morula aggregation. W.-L.C. performed micro-CT analysis. D.M.I., A.D. and S.M. wrote the manuscript with input from the authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0466-z.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.M. or D.M.I.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Mouse embryonic stem cell targeting and transgenic strains. Embryonic stem cell (ESC) culture was performed as described previously[16]. A list of single guide RNAs (sgRNAs) used to generate the various deletions and inversions is given in Supplementary Table 1. For targeting and retargeting of CTCF sites, sequence-verified ESC lines were retargeted with either one or two pX459-sgRNAs. For each CTCF site deletion, structural variants were excluded, and modified CTCF sites for both alleles were verified through PCR amplification of the cut site, followed by subcloning and Sanger sequencing of several PCR products. The ESCs were used for further experiments only if the successful modification of both alleles could be verified. The results were later validated using the cHi-C sequencing data. Embryos and live animals from ESCs were generated by diploid or tetraploid complementation[27]. Genotyping was performed by PCR analysis.

CRISPR-guided knock-in in mouse ESCs. For targeting the *Sox9*–*Kcnj2* TAD boundary, a 6.3-kb construct containing the four CTCF sites (C1 site (mm9 chr11:111384818–111385832) followed by C2–C4 (chr11:111393908-111399229)) was cloned into a targeting vector with asymmetric homology arms (HA1: chr11:112511756–112514691, HA2: chr11:112514692-12519932) using standard cloning procedures.

For knock-in of targeting constructs without selection marker, the targeting construct was transfected in combination with a pX459-sgRNA vector. Importantly, the targeting construct did not contain either an intact PAM site or guide sequence. Puromycin selection and clonal ESC line generation were performed as described previously. Successfully targeted ESC lines were screened using PCR and validated for locus-specific integration after successful establishment of the ESC line. Validation of the homozygous TAD boundary knock-in was performed bioinformatically using the cHi-C data.

LacZ-Sensor mouse lines. The SB-*Kcnj2* and SB-*Sox9* alleles described in Franke et al.[16] were used for remobilization of the SB transgene, following the protocol in Ruf et al.[28], to generate new SB insertion sites (LacZ-Sensors) at the locus (*Kcnj2* TAD: SB20, SB16, SB24; *Sox9* TAD: SB23, SB18). An additional LacZ-Sensor in the *Sox9* TAD (mid-*Sox9*) was targeted directly using CRISPR-guided knock-in as described earlier. Asymmetric homology arms (0.8 and 1.5 kb) with mutated PAM sites and restriction sites for cloning of the LacZ transgene were obtained from IDT and cloned into a plasmid vector. The beta-Globin-LacZ transgene was then inserted into the Acc65I linearized targeting vector using Gibson assembly. The list of primers used for cloning and genotyping targeting construct and pX459-sgRNAs is provided in Supplementary Table 1.

Generation of mice. Mice from transgenic and genome-edited ESCs were generated by diploid or tetraploid aggregation[27], maintained by crossing them with C57Bl.6/J mice and genotyped by PCR. Primers for genotyping can be provided on request. All animal procedures were conducted as approved by the local authorities (LAGeSo Berlin) under the license numbers G0368/08 and G0247/13.

ChIP-seq. Limb buds from E12.5 embryos were microdissected and digested with Trypsin–EDTA 0.05% (Gibco) for 10 min at 37 °C, shaking at 900 r.p.m. The cells were mixed with 10% FCS–PBS, and a single-cell suspension was obtained using a 40-μm cell strainer (Falcon). After centrifugation, cells were fixed in 1% paraformaldehyde (PFA)–10% FCS–PBS for 10 min at room temperature. Cells were then lysed in 3C-Lysis buffer (50 mM Tris, pH 7.5; 150 mM NaCl; 5 mM EDTA; 0.5% NP-40; 1.15% Triton X-100; protease inhibitors (Roche)), and nuclei were pelleted by centrifugation. For sonication, nuclei were resuspended in sonication buffer (10 mM Tris–HCl, pH 8.0; 100 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 0,1% Na-deoxycholate; 0.5% *N*-lauroylsarcosine; protease inhibitors (Roche complete)). Chromatin was sheared using a Bioruptor until reaching a fragment size of 200–500 base pairs (bp). Lysates were clarified from sonicated nuclei, and protein–DNA complexes were immunoprecipitated overnight at 4 °C with the respective antibody. A total of 10–15 μg chromatin was used for each replicate ChIP. Antibodies: H3K27ac (C15410174; Diagenode) and CTCF (Active Motif: 613111). ChIP-seq was performed as in ref. [29]. Sequencing libraries were prepared using the Nextera adapters.

ATAC-seq. ATAC-seq (assay for transposase-accessible chromatin using sequencing) experiments were performed according to Buenrostro et al.[30] In brief, E12.5 limb buds were dissected and homogenized using the Ultra Turrax T8 disperser (IKA). Per replicate, 5×10^4 cells were washed in cold PBS and lysed in fresh lysis buffer (10 mM Tris–HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% (v/v) Igepal CA-630) for 10 min while being centrifuged. The transposition reaction was performed using the Nextera Tn5 Transposase (Nextera kit, no. FC-121-1030; Illumina) according to the manufacturer's instructions. After 30 min at 37 °C, DNA was purified using the MinElute PCR Purification kit (no. 28004; Qiagen), and the transposed DNA was eluted in 10 μl of elution buffer and stored at −20 °C. Barcoded adapters were added to the transposed fragments by PCR with the NEBNext High Fidelity 2× PCR Master Mix (no. M0541; NEB) with nine PCR cycles. Samples were then purified using the AMPure XP beads (no. A63881;

Agencourt) and eluted in 20 μl, and controlled before sequencing using Bioanalyzer 2100 (Agilent). ATAC-seq samples were paired-end sequenced (75-bp) using Illumina technology.

cHi-C. Distal limbs (hand–foot plate) from homozygous E12.5 embryos were microdissected in PBS trypsinized in 500 μl collagenase solution (0.1% Collagenase type 1a (Sigma no. C9891), 0.1% (w/v) Trypsin, 5% FCS in DMEM:HAM'S F-12 (1:1)) for approximately 15 min at 37 °C. Samples were transferred in a 50-ml falcon tube through a 40-μm cell strainer and complemented with 10% FCS/PBS. 37% formaldehyde (no. 252549; Sigma-Aldrich) was added to a final concentration of 2%, and cells were fixed for 10 min at room temperature. The reaction was quenched by adding glycine (final concentration: 125 mM). Fixed cells were washed twice with cold PBS and lysed using fresh lysis buffer (10 mM Tris, pH 7.5, 10 mM NaCl, 5 mM MgCl₂, 0.1 mM EGTA complemented with protease inhibitor) to isolate nuclei. After 10-min incubation in ice, cell lysis and nuclei suspension were assessed microscopically. Nuclei were centrifuged for 5 min at 480*g*, washed once with PBS and snap frozen in liquid N₂.

3C libraries were prepared from at least three pairs of E12.5 forelimb and hindlimb buds per genotype as described previously[16]. In summary, nuclei pellets were thawed on ice and used for DpnII digestion, ligation and crosslink reversal. Religated products were sheared using a Covaris sonicator (duty cycle: 10%, intensity: 5, cycles per burst: 200, time: 2 cycles of 60 s each, set mode: frequency sweeping, temperature: 4–7 °C), and adapters were added to the sheared DNA and amplified according to Agilent instructions for Illumina sequencing. The library was hybridized to the custom-designed SureSelect library and indexed for sequencing (75 bp, paired end) following manufacturer's instructions. The custom-designed SureSelect library was described previously[16].

Expression analysis. RNA for qPCR was extracted from E13.5 mouse zeugopods. After dissection, samples were immediately frozen in liquid nitrogen, and individual embryos were genotyped. Tissue was lysed in RLT buffer and a .20-gauge syringe, and RNA extraction was performed using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. cDNA synthesis was performed with SuperScriptIII RT (Invitrogen) and polyT primers, and qPCR was performed on an ABI9700.

WISH. WISH using a digoxigenin (DIG)-labeled antisense riboprobe for *Sox9* and *Kcnj2* was transcribed from linearized gene-specific probes (PCR DIG Probe Synthesis Kit; Roche). Primers for P probe generation are listed in Supplementary Table 1.

For WISH, embryos were fixed overnight in 4% PFA in PBS. The embryos were washed in PBS with 0.1% Tween, dehydrated stepwise in 25%, 50% and 75% methanol/PBST and stored at −20 °C in 100% methanol. DNA extracted from individual embryo amnia were genotyped. For WISH, genotyped embryos were rehydrated on ice in reverse methanol/PBST steps, washed in PBS–Tween, bleached in 6% H₂O₂ in PBST for 1 h and washed in PBS–Tween. Embryos were Proteinase K treated (10 μg ml⁻¹ in PBS–Tween) for 3 min, and the reaction was quenched with glycine in PBS–Tween. Embryos were washed in PBS–Tween and finally refixed for 20 min with 4% PFA in PBS, 0.2% glutaraldehyde and 0.1% Tween. After further washing steps with PBS–Tween, embryos were incubated at 68 °C in L1 buffer (50% deionized formamide, 5× SSC, 1% SDS, 0.1% Tween 20 in diethyl pyrocarbonate (DEPC), pH 4.5) for 10 min. For prehybridization, embryos were incubated for 2 h at 68 °C in hybridization buffer 1 (L1 with 0.1% tRNA and 0.05% heparin). For subsequent probe hybridization, embryos were incubated overnight at 68 °C in hybridization buffer 2 (hybridization buffer 1 with 0.1% tRNA, 0.05% heparin and 1:500 DIG probe). The next day, unbound probe was removed through repeated washing steps; 3× 30 min at 68 °C with L1, L2 (50% deionized formamide, 2× SSC, pH 4.5, 0.1% Tween 20 in DEPC, pH 4.5) and L3 (2× SSC, pH 4.5, 0.1% Tween 20 in DEPC, pH 4.5).

For signal detection, embryos were treated for 1 h with RNase solution (0.1 M NaCl, 0.01 M Tris, pH 7.5, 0.2% Tween 20, 100 μg ml⁻¹ RNase A in H₂O), followed by washing in TBST 1 (140 mM NaCl, 2.7 mM KCl, 25 mM Tris–HCl, 1% Tween 20, pH 7.5) and blocking (2 h at room temperature in blocking solution (TBST 1 with 2% calf serum and 0.2% bovine serum albumin). Overnight incubation with Anti-Dig antibody conjugated to alkaline phosphatase (1:5,000) at 4 °C (no. 11093274910; Roche) was followed by 8× 30 min washing steps at room temperature with TBST 2 (TBST with 0.1% Tween 20 and 0.05% levamisole–tetramisole) and left overnight at 4 °C. Finally, embryos were stained after equilibration in AP buffer (0.02 M NaCl, 0.05 M MgCl₂, 0.1% Tween 20, 0.1 M Tris–HCl and 0.05% levamisole–tetramisole in H₂O) 3× 20 min, followed by staining with BM Purple AP Substrate (Roche). The stained embryos were imaged using a Zeiss Discovery V12 microscope and Leica DFC420 digital camera.

Microcomputer tomography. Autopods of 7-week-old control and mutant mice were scanned using a SkyScan 1172 X-ray microtomography system (Brucker microCT, Belgium) at 5-μm resolution. 3D model reconstruction was done with the SkyScan image analysis software CT-Analyser and CT-volume (Brucker microCT, Belgium).

ATAC-seq and ChIP-seq data processing. Paired-end ATAC-seq reads were trimmed to 30 bp to allow fragments with close-by transposition events (<50 bp) to map, that is, increase read coverage at nucleosome-free regions. Trimmed reads were mapped with Bowtie2 (ref. [31]), and duplicate fragments were removed using Picard RemoveDuplicates (https://broadinstitute.github.io/picard/). Coverage tracks (bigWig) were generated with deepTools2 (ref. [32]) for properly mapped read pairs (FLAG 0 × 2) with mapping quality ≥20.

For ChIP-seq, single-end reads were mapped with bowtie (v2.2.6) to mm9, filtered for mapping quality MAPQ ≥ 10, and duplicates were removed using samtools rmdup (v1.8). For generating coverage tracks, reads were extended to 300 bp (H3K27ac) or 200 bp (CTCF) and scaled to r.p.m. (10^6 per number of unique reads) using bedtools genomecov v2.27.1.

cHi-C data processing. Raw reads were preprocessed with cutadapt v1.15 to trim potential low-quality bases (-q 20 -m 25) and potentially remaining sequencing adapters (-a and -A option with Illumina TruSeq adapter sequences according to the cutadapt documentation) at the 3′ ends of the reads. Mapping, filtering and deduplication of the short reads were performed with the HiCUP pipeline v0.5.10 (ref. [33]) (no size selection, Nofill: 1, Format: Sanger). The pipeline was set up with Bowtie2 v2.2.6 (ref. [31]) for mapping short reads to reference genome mm9. For inversions and the Border-KI allele, reads were also mapped to a customized genome, derived from mm9 based on the genotyping of the mutant ESC lines. Juicer tools 0.7.5 (ref. [34]) was used to generate binned contact maps from valid and unique read pairs with MAPQ ≥ 30 and to normalize contact maps by Knight and Ruiz (KR) matrix balancing[3,34,35]. For the generation of contact maps, only reads pairs mapping to the enriched genomic region (chr11:109,010,001–114,878,000) were considered and shifted by the offset of the enriched genomic region (109,010,000 bp). For the import with Juicer tools, we used a custom chrom.sizes files containing only the size of the enriched part of the genome. Afterward, KR-normalized maps were exported at 10-kb resolution, and coordinates were shifted back to their original values.

Subtraction maps were generated from KR-normalized maps, which were normalized in a pairwise manner before subtraction. To account for differences between two maps in their distance-dependent signal decay, we scaled maps jointly across their subdiagonals. Therefore, the values of each subdiagonal of one map were divided by the sum of this subdiagonal and multiplied by the average of these sums from both maps. Afterward, the maps were scaled by 10^6/total sum. cHi-C maps and subtractions maps were displayed as heatmaps in which (absolute) values above the 97.5th percentile were truncated for visualization purposes. In the display of cHi-C maps, this threshold defined the value of 100 on a linear scale.

Interaction score for enhancer–promoter contacts. Count values of KR-normalized maps were scaled such that the 97.5th percentile equals 100. Then count values for the interaction between the bins covering the *Sox9–Kcnj2* TSS and the corresponding enhancer (Supplementary Table 2) were extracted from the map and plotted per genotype. When two enhancers were located in the same bin, they were shown in a single data point. Individual data points very close to the *Sox9* viewpoint were omitted.

Virtual 4C. To obtain individual interaction profiles for specific viewpoints with more fine-grained binning, we created virtual Capture-C-like interactions profiles from the same filtered bam files that were used for the cHi-C maps. Paired-end reads with MAPQ ≥ 30 were considered in a profile when one mate mapped to the viewpoint region, whereas the other one mapped outside of it. Contacts of the viewpoint region with the rest of the genome were counted per restriction fragment. Afterward, count data were binned to 1-kb bins. In case a restriction fragment overlapped with more than one bin, the counts were split proportionally. Afterward, each profile was smoothed by averaging within a sliding window of 5 kb

and scaled by 10^3/sum of its counts on the same chromosome. The viewpoint and a window ±5 kb around it were excluded from the computation of the scaling factor. The profiles were generated with custom Java code using htsjdk v2.12.0 (https://samtools.github.io/htsjdk/).

Interaction score between *Kcnj2* and *Sox9* TADs. The *Kcnj2* TAD was manually defined as genomic region chr11:110,340,001–111,400,000 and the *Sox9* TAD as chr11:111,400,001–113,030,000. Contact counts were summed within each TAD individually and within the region of the cHi-C map containing the contacts between the two TADs. To avoid a strong influence of the main diagonal, only contacts spanning more than 100 kb were considered in this analysis. The three sums of contact counts were normalized to represent fractions adding up to 1. Thus, the change of contact frequency between the two TADs was determined with respect to the intra-TAD contact frequency.

For the calculation of the *Sox9–Kcnj2* contacts with the centromeric *Sox9* TAD, the virtual 4C-seq interaction data were further processed. The contacts with the 'inverted region' represent the respective proportion of all *Sox9–Kcnj2* promoter contacts in the *Sox9* and *Kcnj2* TADs. The contacts <50 kb from the promoter were excluded from the calculation.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Datasets are available through the Gene Expression Omnibus (GEO) under accession numbers GSE78109 and GSE125294. Reagents, cell lines and other data supporting the findings of this study are available from the corresponding author upon request.

Code availability
Our pipeline for cHi-C data processing outlined above, as well as custom codes used in this study, will be provided upon request.

References
27. Artus, J. & Hadjantonakis, A.-K. Generation of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. *Methods Mol. Biol.* **693**, 37–56 (2011).
28. Ruf, S. et al. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat. Genet.* **43**, 379–386 (2011).
29. Andrey, G. et al. Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res.* **27**, 223–233 (2017).
30. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
31. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
32. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucl. Acids Res.* **44**, W160–W165 (2016).
33. Wingett, S. et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res.* **4**, 1310 (2015).
34. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
35. Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* **33**, 1029–1047 (2013).

nature research

Corresponding author(s):	Daniel M. Ibrahim Stefan Mundlos
Last updated by author(s):	Jun 1, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	no commercial, open source or custom software/code was used to collect the data in this study
Data analysis	cHi-C: Fastq files were processed with the HiCUP pipeline v.0.5.8 performing the mapping as well the filtering for valid and unique di-tags The pipeline was set up with Bowtie2 v2.2.6 Filtered di-tags were further processed with Juicer tools to bin di-tags (5 and 10 kb bins) and to normalize the map by Knight-Ruiz (KR) matrix balancing ChIP-seq and ATAC-seq: read mapping was performed using, bowtie, bam files were converted to begraph files using bedtools genomecoveragebed and normalized (per million reads). Bedgraph files were converted to bigwig files using UCSC's bedgraphtobigwig. Bigwig tracks for ATAC seq were created with deeptools2 Analysis and presentation of qRT-PCR experiments where performed using Microsoft Excel for Mac 2011 (Version 14.7.2) µCT analysis was performed using Skyscan image analysis software for the Skyscan 1172 X-ray microtomography system.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Datasets are available through the Gene Expression Omnibus (GEO) under accession number GSE78109 and GSE125294.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Reported on page 7: Sample size was at least 3 homozygous animals for qPCR and WISH. Phenotypic analysis for mutant mouse lines was carried out for at least 2 homozygous animals per analysis/developmental stage. The phenotype of the mutants of the study was fully penetrant, i.e. was either present or absent and no statistical test was applied in this case. For quantification of differential RNA expression we used heterozygous and homozygous littermates compared to their wildtype siblings, wherever possible, sample sizes of 2-6 individuals per datapoint are reported in the respective Figure legends. cHi-C was produced as singlicate from limbs buds of at least 3 homozygous embryos.
Data exclusions	Samples/ animals were included/ excluded according to genotype. Genotyping was established and verified in appropriate control experiments.
Replication	For phenotypic and WISH expression data at least 3 animals/embryos of the appropriate genotype were stained all embryos and animals gave reproducible staining/phenotypes. Individual datapoints for for qPCR expression replicates are presented in the figures.
Randomization	Several founder animals for each mouse line were used for establishing line stock with variable intercrosses between single founder and Bl6 wildtype animals. Random selection of animals for analysis and breeding. For the TAD-boundary knock-in two independent ESC-lines with homozygous knock-ins of the boundary were used with reproducible results
Blinding	No blinding was carried out, since mouse breeding and analysis required knowledge about the genotype at hand.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	polyclonal rabbit anti-CTCF antibody from Active Motif, Cat-No. 61311 polyclonal rabbit anti-H3K27ac antibody from Diagenode, Cat-No. c15410174 Anti-Digoxigenin conjugated to alkaline phosphatase from Roche, Cat-No. 11093274910)
Validation	The anitbodies are validated for use on mouse samples on the manufacturer's websites and give highly reporducible results with

published positive control data-sets in ChIP-seq. The specificity of the anti-Dig antibody was used by staining WISH embryos that were incubated without RNA probe and embryos that were incubated with sense RNA probes.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	G4-ESCs were obtained from Anders Nagy and subsequently used to generate custom genome-engineered cell lines for generation of mice used in this study.
Authentication	The pluripotent state of the ESCs used was authenticated by generation of highly chimeric, germ-line transmitting mice through di- and tetraploid complementation assays
Mycoplasma contamination	all cell lines were tested negative for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mouse lines described in this study were established in a Bl6/129Sv5 background and maintained through intercrosses with C57Bl.6/J mice. As this study concerns an autosomal trait, male and female mice of each genotype were used for analysis and breeding.
Wild animals	the study does not involve wild animals
Field-collected samples	the study does not involve field-collected samples
Ethics oversight	The study plan was approved by the Landesamt für Gesundheit und Soziales (LaGeSo), Berlin under licenses G0243/13 and G0346/13

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	Datasets are available through the Gene Expression Omnibus (GEO) under accession number GSE78109 and GSE125294.
Files in database submission	CTCF_E125_Rep1.bw CTCF_E125_Rep2.bw H3K27ac_E125_Rep1.bw H3K27ac_E125_Rep2.bw CTCF_E125_Rep1.fastq.gz CTCF_E125_Rep2.fastq.gz H3K27ac_E125_Rep1-2.bw H3K27ac_E125_Rep1.fastq.gz H3K27ac_E125_Rep2.fastq.gz ATAC_E125_Rep1.bw ATAC_E125_Rep1_R1.fastq.gz ATAC_E125_Rep1_R2.fastq.gz ATAC_E125_Rep2.bw ATAC_E125_Rep2_R1.fastq.gz ATAC_E125_Rep2_R2.fastq.gz cHi-C_BorKnockIn_E125_TR1_R1.fastq.gz cHi-C_InvCdelB_E125_TR1_R1.fastq.gz cHi-C_delBorC1-2_E125_TR1_R1.fastq.gz cHi-C_delBorC1_E125_TR1_R1.fastq.gz cHi-C_BorKnockIn_E125_TR1_R2.fastq.gz cHi-C_InvCdelB_E125_TR1_R2.fastq.gz cHi-C_delBorC1-2_E125_TR1_R2.fastq.gz cHi-C_delBorC1_E125_TR1_R2.fastq.gz cHi-C_InvC_E125_TR1_R1.fastq.gz cHi-C_InvIntra_E125_TR1_R1.fastq.gz cHi-C_delBorC1-2_E125_TR2_R1.fastq.gz cHi-C_delBorC1_E125_TR2_R1.fastq.gz

cHi-C_InvC_E125_TR1_R2.fastq.gz
cHi-C_InvIntra_E125_TR1_R2.fastq.gz
cHi-C_delBorC1-2_E125_TR2_R2.fastq.gz
cHi-C_delBorC1_E125_TR2_R2.fastq.gz
cHi-C_InvC_E125_TR2_R1.fastq.gz
cHi-C_InvIntra_E125_TR2_R1.fastq.gz
cHi-C_delBorC1-4_E125_TR1_R1.fastq.gz
cHi-C_delCTCF_E125_BR1_R1.fastq.gz
cHi-C_InvC_E125_TR2_R2.fastq.gz
cHi-C_InvIntra_E125_TR2_R2.fastq.gz
cHi-C_delBorC1-4_E125_TR1_R2.fastq.gz
cHi-C_delCTCF_E125_BR1_R2.fastq.gz
cHi-C_BorKnockIn_E125.KR_10kb.WashU.txt
cHi-C_InvCdelB_E125.KR_10kb.WashU.txt
cHi-C_delBorC1-2_E125.KR_10kb.WashU.txt
cHi-C_delBorC1_E125.KR_10kb.WashU.txt
cHi-C_BorKnockIn_E125.Raw_10kb.WashU.txt
cHi-C_InvCdelB_E125.Raw_10kb.WashU.txt
cHi-C_delBorC1-2_E125.Raw_10kb.WashU.txt
cHi-C_delBorC1_E125.Raw_10kb.WashU.txt
cHi-C_InvC_E125.KR_10kb.WashU.txt
cHi-C_InvIntra_E125.KR_10kb.WashU.txt
cHi-C_delBorC1-4_E125.KR_10kb.WashU.txt
cHi-C_delCTCF_E125_BR1.KR_10kb.WashU.txt
cHi-C_InvC_E125.Raw_10kb.WashU.txt
cHi-C_InvIntra_E125.Raw_10kb.WashU.txt
cHi-C_delBorC1-4_E125.Raw_10kb.WashU.txt
cHi-C_delCTCF_E125_BR1.Raw_10kb.WashU.txt

Genome browser session
(e.g. UCSC)

not applicable

Methodology

Replicates

each ChIP-seq experiment was performed in two biological replicates

Sequencing depth

CTCF ChIP-seq replicates were sequened 50bp with ca. 30 mio reads/replicate, H3K27ac replicates with 75bp single end with ca. 20 mio reads/replicate

Antibodies

polyclonal rabbit anti-CTCF antibody from Active Motif, Cat-No. 61311
polyclonal rabbit anti-H3K27ac antibody from Diagenode, Cat-No. c15410174

Peak calling parameters

no peak calling was performed

Data quality

enrichment was validated using spp cross-correlation plots

Software

mapping was performed using, bowtie, bam files were converted to begraph files using bedtools genomecoveragebed and normalized (per million reads) and then converted to bigwig files using UCSC's bedgraphtobigwig