Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis

Bjørt K. Kragesteen [1,2,3,4,14], Malte Spielmann [1,2,14], Christina Paliou [1,2,3], Verena Heinrich[5], Robert Schöpflin[1,2,5], Andrea Esposito [6,7], Carlo Annunziatella[6], Simona Bianco [6], Andrea M. Chiariello[6], Ivana Jerković [1,4], Izabela Harabula[1], Philine Guckelberger [1], Michael Pechstein[1], Lars Wittler[8], Wing-Lee Chan[2], Martin Franke [1,2,3], Darío G. Lupiáñez [1,2,3,9], Katerina Kraft [1,2,3], Bernd Timmermann[10], Martin Vingron [5], Axel Visel [11,12,13], Mario Nicodemi [6,7*], Stefan Mundlos [1,2,3*] and Guillaume Andrey [1,2,3*]

The regulatory specificity of enhancers and their interaction with gene promoters is thought to be controlled by their sequence and the binding of transcription factors. By studying Pitx1, a regulator of hindlimb development, we show that dynamic changes in chromatin conformation can restrict the activity of enhancers. Inconsistent with its hindlimb-restricted expression, Pitx1 is controlled by an enhancer (Pen) that shows activity in forelimbs and hindlimbs. By Capture Hi-C and three-dimensional modeling of the locus, we demonstrate that forelimbs and hindlimbs have fundamentally different chromatin configurations, whereby Pen and Pitx1 interact in hindlimbs and are physically separated in forelimbs. Structural variants can convert the inactive into the active conformation, thereby inducing Pitx1 misexpression in forelimbs, causing partial arm-to-leg transformation in mice and humans. Thus, tissue-specific three-dimensional chromatin conformation can contribute to enhancer activity and specificity in vivo and its disturbance can result in gene misexpression and disease.

Embryonic development requires highly accurate and robust regulation of gene expression. Much of this information is encoded in *cis*-regulatory elements called enhancers, which are short stretches of DNA that drive gene expression over long distances, independently of their location or orientation, by physically contacting their target gene promoter[1]. Characteristically, enhancers are active in different tissues and/or at distinct time points, with their individual activities collectively adding up to the complex developmental expression pattern of their shared target genes. These features allow enhancers to regulate genes in a modular manner resulting in a vast combinatorial flexibility. Most current models suggest that the regulatory activity of enhancers is encoded within their sequence and translated into regulation via the binding of tissue-specific transcription factors.

Since enhancers can act over long distances, it is critical that their genomic activity is tightly restricted. The required regional specificity is at least in part achieved by the formation of chromatin domains called topologically associating domains (TADs) that insulate regions of high interaction from each other, thereby restricting the promoters an enhancer can potentially interact with. Structural variants such as deletions, duplications, or inversions can disrupt TADs and their insulating borders allowing enhancers to act on

neighboring domains, thus, causing ectopic enhancer–promoter interactions, gene misexpression, and consecutive disease[2]. This mutational mechanism, termed enhancer adoption or enhancer hijacking, appears to be common in both cancer and congenital disease[3,4]. While the overall configuration of TADs is rather stable between tissues and cells, interactions at the sub-TAD level are more dynamic reflecting the instructive, cell-type-specific interactions between cognate enhancers and promoters[5–7]. However, it remains unclear whether the assembly and disassembly of these interactions is causally involved in modulating gene transcription in vivo and if genetic alterations perturbing these interactions can cause gene misregulation and disease.

In the present study, we dissect the basic mechanisms of enhancer–promoter interaction and their role in gene regulation at the *Pitx1* locus, a gene that is critically required for establishing the identity and differentiation of hindlimbs[8]. During limb development, the *Pitx1* gene is only expressed in hindlimbs; *Pitx1* loss-of-function (LOF) causes a selective reduction of hindlimbs, but not forelimbs, as well as a partial loss of hindlimb characteristics[9–11]. The instructive character of *Pitx1* for hindlimb identity is further shown by ectopic expression of *Pitx1* in developing forelimbs, which is sufficient to cause partial forelimb-to-hindlimb conversion in mice and humans[8,12,13].

[1]RG Development & Disease, Max Planck Institute for Molecular Genetics, Berlin, Germany. [2]Institute for Medical and Human Genetics, Charité-Universitätsmedizin Berlin, Berlin, Germany. [3]Berlin-Brandenburg Center for Regenerative Therapies, Charité-Universitätsmedizin Berlin, Berlin, Germany. [4]Berlin-Brandenburg School for Regenerative Therapies, Charité-Universitätsmedizin Berlin, Berlin, Germany. [5]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany. [6]Dipartimento di Fisica, Università di Napoli Federico II, and Istituto Nazionale di Fisica Nucleare, Napoli, Complesso Universitario di Monte Sant'Angelo, Naples, Italy. [7]Berlin Institute of Health, MDC-Berlin, Berlin, Germany. [8]Department of Developmental Genetics, Max Planck Institute for Molecular Genetics, Berlin, Germany. [9]Epigenetics and Sex Development Group, Berlin Institute for Medical Systems Biology, Max-Delbrück Center for Molecular Medicine, Berlin-Buch, Germany. [10]Max Planck Institute for Molecular Genetics, Sequencing Core Facility, Berlin, Germany. [11]Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [12]U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA, USA. [13]School of Natural Sciences, University of California, Merced, CA, USA. [14]These authors contributed equally: Bjørt K. Kragesteen, Malte Spielmann. *e-mail: nicodem@na.infn.it; mundlos@molgen.mpg.de; andrey@molgen.mpg.de
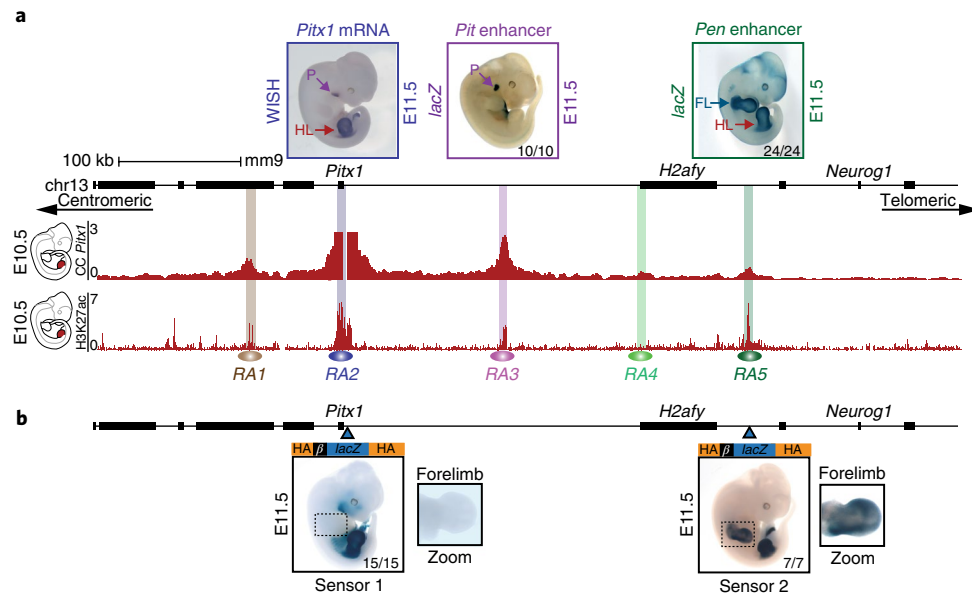
Fig. 1 | *Pitx1* regulatory landscape includes a pan-limb region. a, *Pitx1* regulatory landscape in hindlimbs. Top left: WISH of *Pitx1* at E11.5 showing hindlimb-specific and pituitary-specific (HL, and P, respectively) expression. Top middle: *lacZ* reporter activity at E11.5 of *RA3* (*Pit* enhancer) showing activity in the pituitary in 10 out of 10 embryos. Top right: *lacZ* reporter activity at E11.5 of *RA5* (*Pen* enhancer) demonstrating activity in forelimbs and hindlimbs in 24 out of 24 embryos (FL and HL). Upper red track: Capture-C in E10.5 hindlimbs using the *Pitx1* promoter as the viewpoint demonstrating chromatin interactions with regulatory anchors (*RA1*–*RA5*). Lower red track: H3K27ac ChIP-seq enrichment profile in E10.5 hindlimbs. **b**, *lacZ* sensors 1 and 2 reflect endogenous regulatory activity at E11.5. The blue arrowheads indicate the positions of the integrations. Sensor 1 recapitulates *Pitx1* expression pattern in 15 out of 15 embryos (compare with *Pitx1* WISH in **a**). Note the absence of forelimb staining. Sensor 2 shows a strong forelimb and hindlimb activity in 7 out of 7 embryos, similar to the *Pen* enhancer (compare with *Pen lacZ* activity in **a**).

Here, we demonstrate that *Pitx1* is regulated by an enhancer (*Pen*) that displays activity in both forelimbs and hindlimbs. The restriction of this activity to the hindlimb is associated with tissue-specific differences in three-dimensional (3D) chromatin structure enabling the *Pen* enhancer to control *Pitx1* transcription in hindlimbs only. The forelimb-specific 3D chromatin conformation that normally prevents *Pen* from physically contacting *Pitx1* can be reverted to the active hindlimb state by structural variants, which results in *Pitx1* misexpression in forelimbs. These alterations of the locus, whether engineered in mice or associated with disease in humans, result in the appearance of anatomical hindlimb features in the forelimb. Our results demonstrate that tissue-specific 3D chromatin architecture can refine enhancer activity into a precise enhancer–promoter interaction required for normal morphogenesis and that disruption of this process can result in gene misexpression and disease in both mice and humans.

Results

The regulatory landscape of *Pitx1* involves forelimb and hindlimb enhancer activity. *Pitx1* is expressed in the developing hindlimb buds, the first branchial arch, the pituitary gland, and parts of the gut (Fig. 1a)[14]. The gene resides at the centromeric edge of a 240-kb gene desert flanked by the housekeeping gene *H2afy* on the telomeric side (Fig. 1a). To determine the position of regulatory elements controlling *Pitx1* in hindlimb buds, we examined the Capture-C interaction profiles and chromatin immunoprecipitation sequencing (ChIP-seq) for the enhancer mark H3K27ac in hindlimb buds at mouse embryonic day (E) 10.5 (Fig. 1a)[15]. The chromatin interaction profile shows that the *Pitx1* regulatory landscape extends over 400 kb and forms several chromatin loops corresponding to H3K27ac peaks, termed regulatory anchors 1–5 (*RA1–RA5*, with the *Pitx1* promoter representing *RA2*, Fig. 1a). We next examined the ability of several of these elements to drive a *lacZ* reporter by single-copy insertion in transgenic mice (Fig. 1a,

Supplementary Fig. 1a)[16]. *RA3* displayed activity in the pituitary gland (*Pit* enhancer), recapitulating endogenous *Pitx1* expression in this tissue. Unexpectedly, *RA5*, and to a lower extent *RA4*, showed enhancer activity in both forelimb and hindlimb buds, which contrasts the hindlimb-restricted expression of *Pitx1*. Due to its strong activity in forelimbs and hindlimbs, we refer to *RA5* as a pan-limb enhancer (*Pen*)[12,17] (Fig. 1a).

Considering that these initial transgenic assays involved integration of the *lacZ* reporter gene constructs outside of their genomic context at the *Col1A1* locus, we also examined the endogenous regulatory activities within the native genomic locus by integrating two *lacZ* reporter cassettes, termed sensor 1 and 2, at the *Pitx1* promoter and next to *Pen*, respectively (Fig. 1b and Supplementary Fig. 1b). Sensor 1 recapitulated the entire *Pitx1* expression pattern and showed no activity in forelimbs. In contrast, sensor 2 showed strong staining in the forelimbs (Fig. 1b), further corroborating the presence of pan-limb regulatory elements within the *Pitx1* regulatory landscape.

The pan-limb enhancer *Pen* is a bona fide component of the *Pitx1* regulatory landscape. To functionally dissect the *Pitx1* regulatory landscape, we generated a series of mouse deletion mutants (Fig. 2a) using an adapted CRISPR/Cas9 protocol[18]. A homozygous 330-kb deletion (*Pitx1*[del1]), created by removing the entire telomeric gene desert including *RA3*, *RA4*, and the *Pen* enhancer (*RA5*), resulted in a 20-fold decrease in *Pitx1* expression in the hindlimb (Fig. 2a and Supplementary Fig. 2a). *Pitx1*[del1/del1] animals exhibited a size reduction of the femur and fibula, a loss of the tibia, and an absent patella, which is a hindlimb-specific structure, thereby phenocopying the *Pitx1* LOF phenotype (Fig. 2b)[9,10,19]. Thus, the critical region ensuring robust *Pitx1* hindlimb expression is embedded within these 330 kb.

To further dissect the locus, we created three additional deletion alleles, one removing the intergenic space between *Pitx1* and
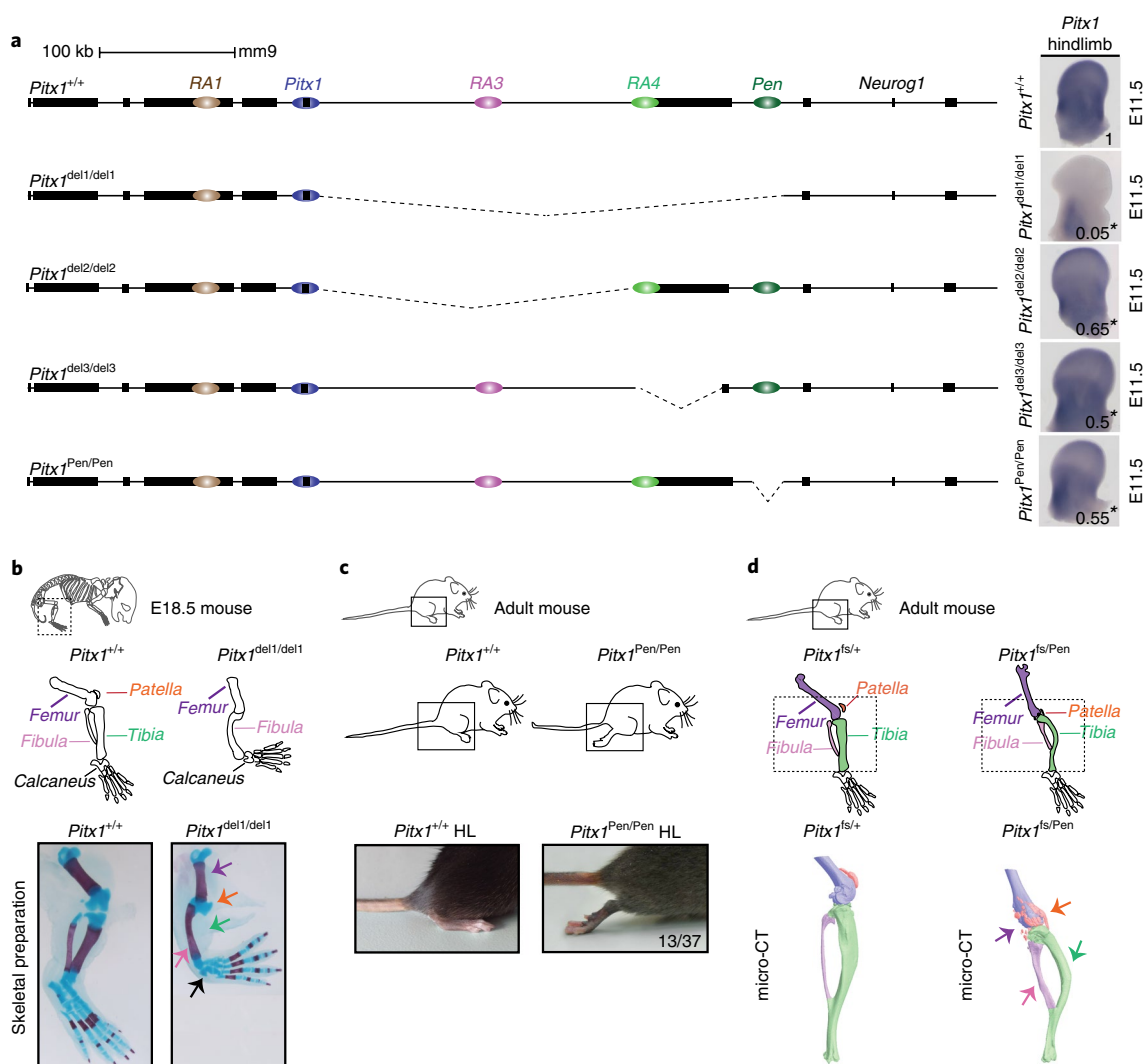
Fig. 2 | Genetic dissection of *Pitx1* regulation highlights the contribution of a pan-limb enhancer (*Pen*). a, CRISPR/Cas9 genetic dissection of the *Pitx1* regulatory landscape with mRNA quantification and visualization of E11.5 hindlimbs using qRT–PCR and WISH. The numbers indicating the average of the qRT–PCR fold change compared to WT are displayed on the pictures (see Supplementary Fig. 2 for quantification). We used a one-tailed *t*-test to evaluate the significance of *Pitx1* reduction; *$P < 0.05$. Strong reduction of *Pitx1* expression in *Pitx1*del1/del1 (picture) indicates that its hindlimb regulation is confined to the 330-kb region upstream of the gene ($P = 0.0002$, $n = 4$). *Pitx1*del2/del2 mutants with a 234-kb deletion of the gene desert upstream of *Pitx1* show 35% loss of *Pitx1* expression ($P = 0.025$, $n = 4$). *Pitx1*del3/del3 mutants carry a 58-kb deletion comprising the *H2afy* gene and show 50% loss of *Pitx1* expression ($P = 0.019$, $n = 3$). *Pitx1*Pen/Pen mice carry a 1.2-kb deletion of the *Pen* enhancer region and show a 45% loss of *Pitx1* expression ($P = 0.012$, $n = 4$). **b**, *Pitx1*del1/del1 alcian blue and alizarin red skeletal staining at E18.5 demonstrate a *Pitx1* LOF phenotype with reduction of the femur, loss of patella and tibia, and small calcaneus (5 skeletal stained samples out of 5). **c**, *Pitx1*Pen/Pen adult mice show a partially penetrant (13 animals out of 37) hindlimb club foot phenotype. **d**, *Pitx1*fs/Pen mice display fully penetrant hindlimb malformation (20 animals). Micro-computed tomography (micro-CT) of a *Pitx1*fs/Pen (right) compared to *Pitx1*fs/+ (left) adult mice showed a reduced femur (purple), deformed tibia (green) and fibula (pink), as well as a fragmented patella (orange).

H2afy (*Pitx1*del2), a second one removing the *H2afy* gene body as well as *RA4* (*Pitx1*del3), and one removing the *Pen* enhancer itself (*Pitx1*Pen). In all three cases, mice with a homozygous deletion exhibited a 35–50% reduction in hindlimb expression of *Pitx1* at E11.5 (Fig. 2a and Supplementary Fig. 2b–d), establishing that all three subregions are required for normal *Pitx1* expression during hindlimb development. Notably, the deletion of *Pen* was sufficient to induce a partially penetrant clubfoot (13/37), which is reminiscent of the phenotype observed in mice and humans with *Pitx1* haploinsufficiency (Fig. 2c)[20]. To confirm the genetic link between *Pen* and *Pitx1*, we created a *Pitx1* LOF allele (*Pitx1*fs) by introducing a frameshift mutation in exon 1, upstream of the region encoding the DNA-binding domain. Compound heterozygous *Pitx1*fs/Pen mice exhibited a fully penetrant clubfoot phenotype, as well as a size

reduction of the femur, a deformation of the tibia and fibula, and a partially fragmented patella (Fig. 2d).

Taken together, this allelic series establishes that normal hindlimb-specific expression of *Pitx1* requires *Pen*, an enhancer with strong activity both in forelimbs and hindlimbs. Thus, our observations challenge the assumption that the ultimate in vivo activity pattern of an enhancer is solely specified by the presence of binding sites for tissue-specific transcription factors and suggest that *Pen* activity is further modulated by additional effects.

Active and repressed regions show distinct tissue-specific interactions. In search of a mechanism modulating the endogenous activity pattern of the *Pen* enhancer, we examined the local chromatin architecture of the *Pitx1* locus in developing forelimbs and hindlimbs.
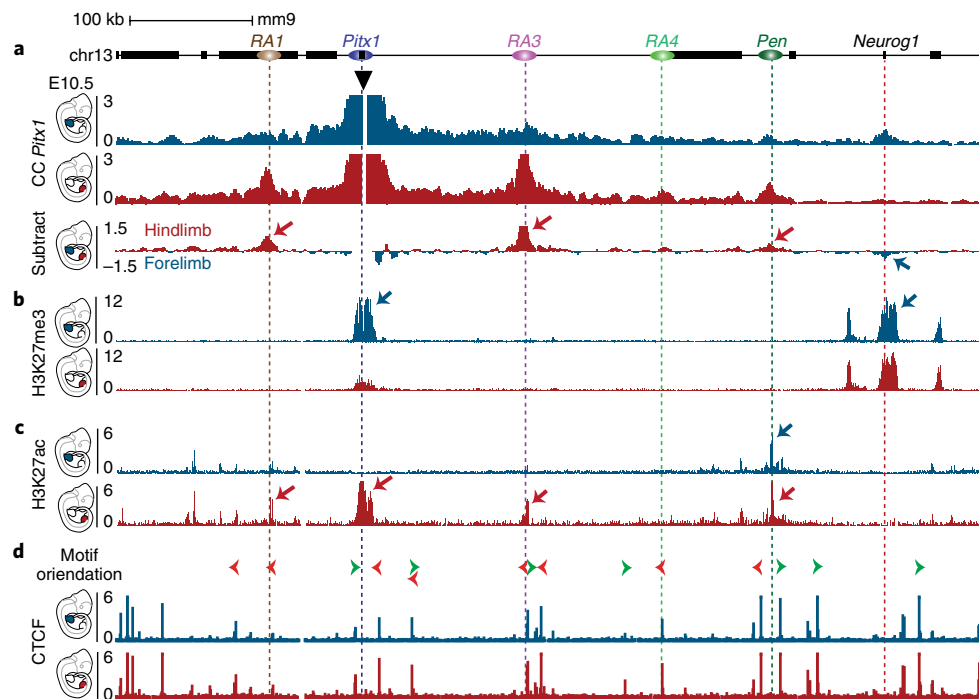
Fig. 3 | Chromatin architecture in the forelimb and hindlimb demonstrates tissue-specific interaction with active and repressed regions. a, Capture-C of *Pitx1* promoter in E10.5 forelimb (blue) or hindlimb (red) tissues. The subtraction of the two tracks shows the differences in chromatin interactions, whereby hindlimb-specific interactions are indicated by red arrows and a forelimb-specific interaction with a blue arrow. **b**, ChIP-seq tracks of H3K27me3 in E10.5 forelimb (blue) or hindlimb (red) tissues show the forelimb-specific repressive interaction between the *Pitx1* and *Neurog1* regions. **c**, ChIP-seq tracks of H3K27ac in E10.5 forelimb (blue) or hindlimb (red) tissues show the deposition of hindlimb-specific acetylation at *RA1*, *Pitx1*, and *RA3*, while the *Pen* region shows acetylation in both forelimb and hindlimb tissues. **d**, ChIP-seq tracks of the transcription factor CTCF in E10.5 forelimb (blue) or hindlimb (red) tissues show little changes in binding. Orientation of the CTCF motif is indicated by the green and light red arrowheads above the tracks.

First, we compared Capture-C interaction tracks from the *Pitx1* promoter in both tissues at E10.5 (Fig. 3a). In forelimbs, we observed minimal interactions of *Pitx1* across the *RA3–RA5* region; instead, we found that the *Neurog1* gene was the strongest distal site of interaction with *Pitx1*. In contrast, in hindlimbs *Pitx1* interacts strongly with *RA1*, *RA3*, and *Pen*, but not with *Neurog1*. These results indicate that the *Pitx1* promoter is embedded into fundamentally different chromatin architectures in forelimbs and hindlimbs.

To further determine the underlying chromatin features associated with these two configurations, we assessed the coverage of H3K27me3 and H3K27ac histone modifications, as well as the binding profile of the architectural transcription factor CTCF in both limb tissues (Fig. 3b–d)[15]. The overall distribution of both chromatin modifications is nearly identical between both tissues, with the exception of the *Pitx1* region. In forelimb tissue, *Pitx1* shows a strong repressive H3K27me3 signal, whereas in hindlimb tissue the *Pitx1* gene body is depleted for H3K27me3 and enriched for the activating H3K27ac mark. Intriguingly, this differential pattern shows a correlation between the dominant histone marks at *Pitx1* and those of the main interaction sites identified by Capture-C (Fig. 3a). In forelimbs, *Pitx1* strongly interacts with *Neurog1*, which is marked by H3K27me3, whereas in hindlimbs *Pitx1* strongly interacts with *RA3* and *Pen*, which are marked by H3K27ac. No major differences in CTCF binding at any of the regulatory anchors were evident, suggesting that the tissue-specific changes observed at the locus are not due to differential binding of CTCF (Fig. 3d). These results indicate that two divergent states of chromatin architecture maintain *Pitx1* either in an inactive state, by establishing an interaction with the H3K27me3-repressed *Neurog1* region, or in an active state, through interactions with active H3K27ac-enriched regions.

Capture Hi-C (CHi-C) highlights a switch in 3D chromatin architecture in forelimbs versus hindlimbs. In contrast to other loci, public Hi-C data did not show the presence of tissue- or species-invariant TADs or contact domains encompassing the entire *Pitx1* regulatory landscape[21,22]. To characterize 3D chromatin folding at this locus in developing tissues using higher definition, we produced CHi-C interaction maps encompassing a 3-Mb region around *Pitx1* from mouse forelimbs and hindlimbs at E11.5 (for a map of the entire region, see Supplementary Fig. 3a). We found that the locus is divided into subdomains separated by the previously characterized regulatory anchors and that multiple interactions occur between the various regulatory anchors as well as with *Neurog1* (Fig. 4a,b). Within the whole assessed region, only the *Pitx1* regulatory landscape displayed extensive differences between forelimb and hindlimb tissues, as shown by the subtraction of forelimb and hindlimb CHi-C maps (Fig. 4c and Supplementary Fig. 3a). Moreover, we could recapitulate the forelimb-specific interaction between *Pitx1* and the functionally unrelated *Neurog1* gene, as well as the hindlimb-specific interactions between *Pitx1* and *RA1*, *RA3*, and *Pen* (Figs. 3a and 4c). We also observed that in hindlimbs, interaction frequency increased within subdomains and decreased between them. This insulation effect was further confirmed by computing the insulation score in forelimb and hindlimb tissues, which showed a pronounced insulation between subdomains in hindlimbs compared to forelimbs (Fig. 4c)[23]. As a control, we produced CHi-C maps from E10.5 midbrain tissues, where *Pitx1* is transcriptionally repressed. We did not observe any 3D changes at the *Pitx1* locus in comparison to forelimb tissue, indicating that the forelimb configuration represents the inactive state of chromatin architecture at the locus (Supplementary Fig. 3b).
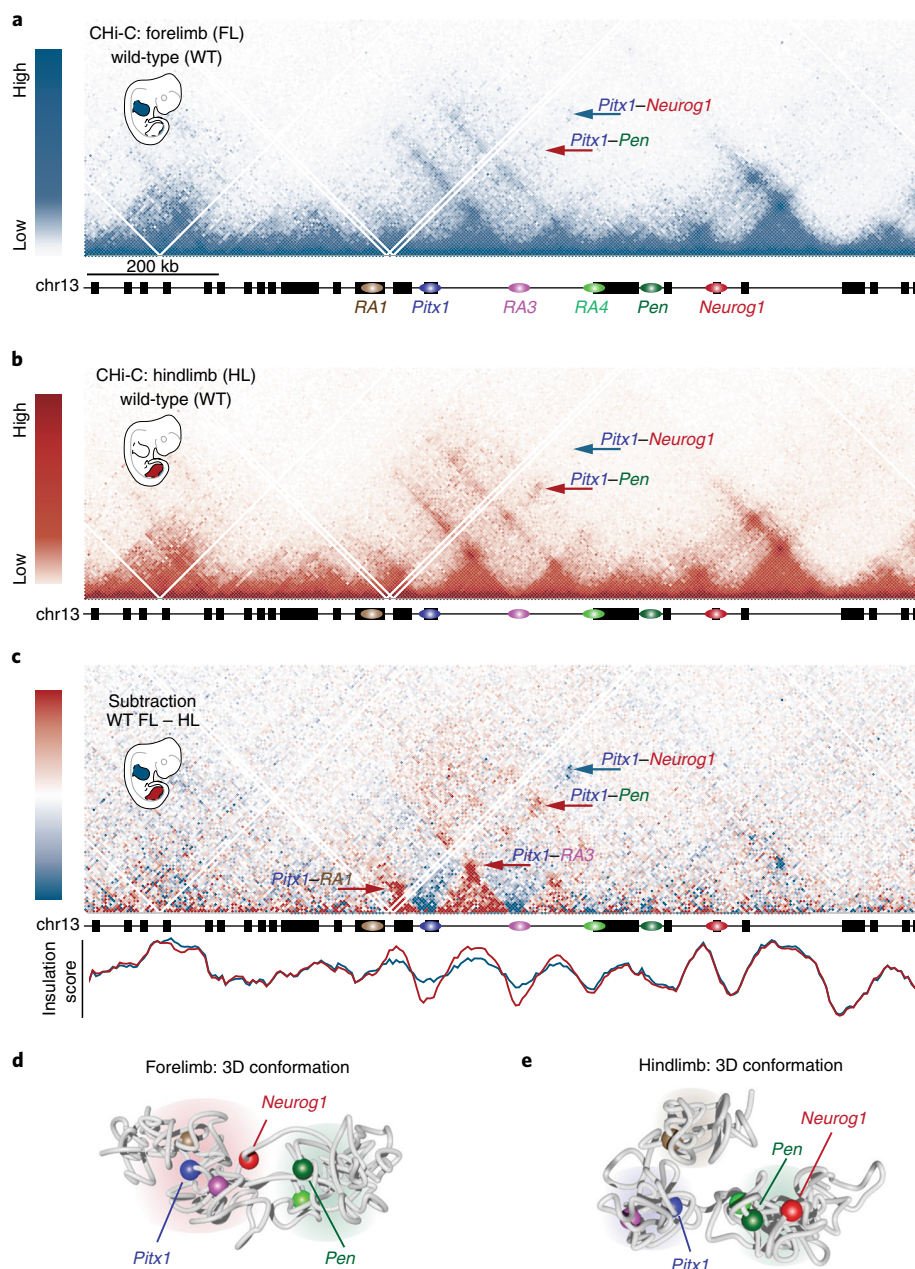
Fig. 4 | Tissue-specific 3D chromatin architecture constrains *Pen* and *Pitx1* interaction. a, CHi-C map of the forelimb at E11.5 shows the presence of chromatin subdomains separated by the regulatory anchors. *Pitx1* shows moderate contacts with *RA3* and *RA4*, weak contacts with *Pen* (see red arrow), and a distal interaction with the *Neurog1* region (blue arrow). **b**, CHi-C map of the hindlimb at E11.5 displays more condensed subdomains separated by the regulatory anchors. *Pitx1* interacts strongly with *RA3*, *RA4*, and *Pen* (red arrow), and shows little interaction with the *Neurog1* region (blue arrow). **c**, Top: CHi-C subtraction between forelimb and hindlimb tissue at E11.5. Chromatin interactions more prevalent in forelimb or hindlimb tissues are shown in blue or red, respectively. The interaction between *Pitx1* and *Neurog1* that is more prevalent in forelimbs is indicated with a blue arrow. Chromatin interactions between *Pitx1* and *RA1*, *Pitx1* and *RA3*, and *Pitx1* and *Pen* that are more prevalent in hindlimbs are indicated with red arrows. Below: Insulation score measuring insulation in hindlimbs (red) and forelimbs (blue). Note the increased insulation at *RA1* and *RA3* in hindlimbs. **d**, 3D model of forelimb CHi-C data show segregation of the *Pitx1* locus into two chromatin hubs containing (1) *RA1*, *Pitx1*, *RA3*, and *Neurog1*, and (2) *RA4* and *Pen*. Note that *Pen* and *Pitx1* are separated from each other. **e**, 3D model of hindlimb CHi-C data show a hindlimb-specific configuration of the *Pitx1* locus into three chromatin hubs containing (1) *RA1*, (2) *Pitx1* and *RA3*, and (3) *RA4*, *Pen*, and *Neurog1* (see also Supplementary Videos 1 and 2). Note the proximity between *Pen* and *Pitx1*.

To facilitate the interpretation of the CHi-C data beyond two-dimensional interaction maps, we computed 3D models of the *Pitx1* locus conformation in forelimbs and hindlimbs from CHi-C data using a polymer physics-based conformation prediction approach (Supplementary Fig. 4)[24]. High correlation between the modeled and experimental data as well as a subtraction map between forelimb and hindlimb models, showed that the model captured the

essential of the chromatin structure (Supplementary Figs. 4 and 5). In forelimbs, the derived thermodynamic ensemble of 3D structures of the locus shows a separation into two major chromatin hubs, one containing *Pitx1*, *RA3*, and *Neurog1*, and another containing *Pen* and *RA4* (Fig. 4d, Supplementary Video 1). *Pitx1* and *Pen* are positioned on distal faces of their respective hubs, leaving *Pitx1* in closer proximity to the repressed *Neurog1* gene. In contrast, the hindlimb
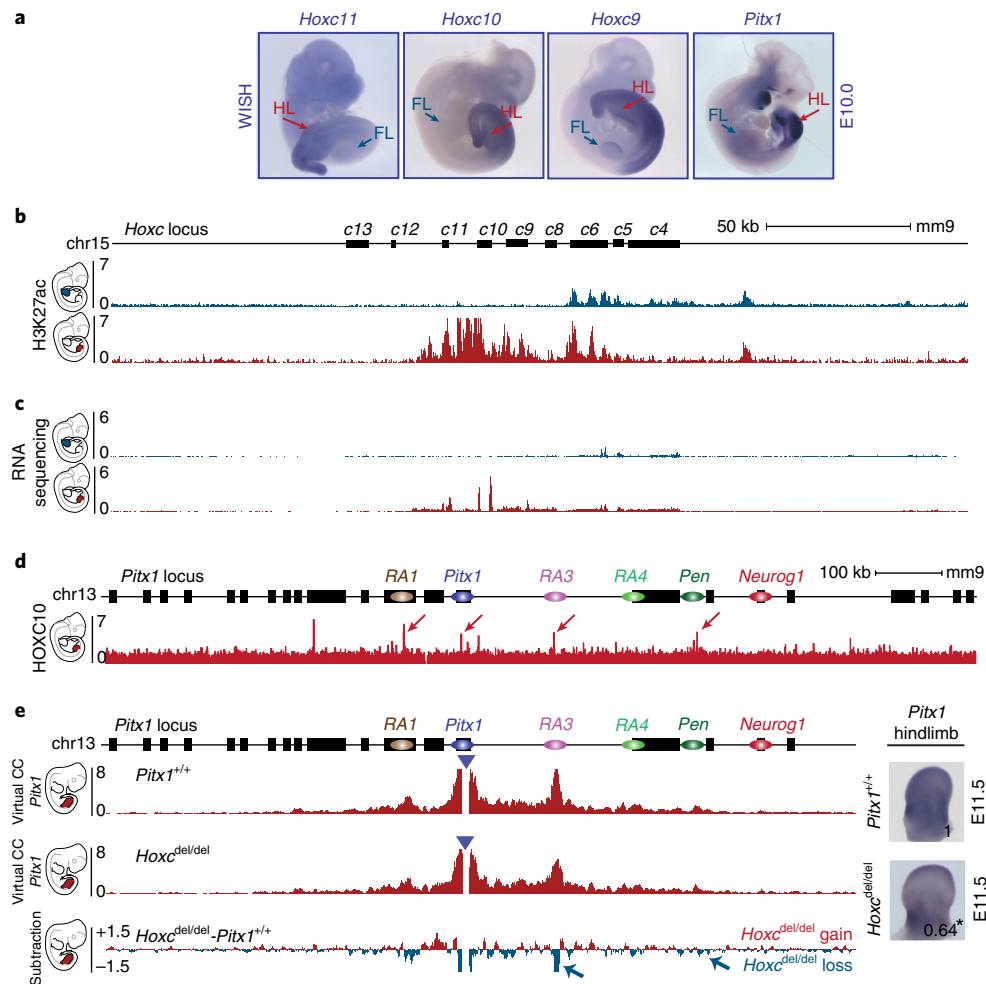
Fig. 5 | *Hoxc* genes participate in *Pitx1* hindlimb regulation. a, WISH of *Hoxc11*, *Hoxc10*, *Hoxc9*, and *Pitx1* in E10.5 embryos. (WISH was carried out at least two times per probe.) Note the overlapping patterns of expression in lateral plate mesoderm and hindlimbs (HL) and absence of transcript in forelimbs (FL). **b**, ChIP-seq tracks for H3K27ac in E10.5 forelimbs (blue) and hindlimbs (red) show the active state of posterior *Hoxc* genes in hindlimbs and anterior *Hoxc* genes in forelimbs. **c**, RNA sequencing profiles of E10.5 forelimb (blue) and hindlimb buds (red) showing expression of posterior *Hoxc* genes in hindlimbs and anterior *Hoxc* genes in forelimbs. **d**, ChIP-seq track of HOXC10 showing binding at *RA1*, *Pitx1*, *RA3*, and *Pen* (see red arrows)[25]. **e**, CHi-C in *Hoxc*del/del hindlimb tissue at E11.5. Derived virtual Capture-C profiles using the *Pitx1* viewpoint (blue triangle) show a significant loss of chromatin interactions between *Pitx1* and *RA3* as well as *Pitx1* and *Pen* (blue arrow, false discovery rate = 0.05). Right: WISH and qRT–PCR of *Pitx1* in hindlimb buds show reduced expression and significant 36% loss of mRNA expression. We used a one-tailed *t*-test to evaluate the significance of *Pitx1* reduction and found a *$P = 0.02$ ($n = 4$ biologically independent WT and mutant hindlimb pairs). The number on the picture indicates the average of qRT–PCR fold changes compared to WT (see Supplementary Fig. 7b).

3D model consists of three chromatin hubs containing: (1) *RA1*; (2) *Pitx1* with *RA3*; and (3) *RA4* together with *Pen* and *Neurog1* (Fig. 4e). Distinct from the forelimb configuration, *Pitx1* and *Pen* are physically closer and face each other on the surface of their respective chromatin hubs, whereas the repressed *Neurog1* gene is comparatively farther from *Pitx1* and *Pen* through its positioning on the hub's opposing face (Fig. 4e, Supplementary Figs. 5 and 6, Supplementary Video 2).

Taken together, these results show fundamentally different positioning of *Pitx1* relative to the pan-limb enhancer *Pen* and *Neurog1* in forelimbs and hindlimbs. In the forelimb, *Pitx1* is physically disconnected from the enhancer and associated with the repressed *Neurog1* gene, consistent with lack of *Pitx1* expression in the forelimb. In the hindlimb, *Pitx1* is in close vicinity of *Pen*, but segregated from the repressed *Neurog1* gene. Thus, our findings identify distinct, tissue-specific 3D arrangements of the *Pitx1* landscape effectively silencing *Pen* during forelimb development.

Homeobox C (*Hoxc*) cluster genes control *Pitx1* chromatin folding and transcription. We next identified *trans*-acting factors that might contribute to the hindlimb-specific and thus active 3D chromatin configuration of the *Pitx1* locus. *Pitx1* itself displays an extremely restricted expression in early hindlimb buds and could participate in maintaining its own expression by establishing and sustaining the active 3D chromatin configuration through a positive feedback mechanism. However, mice with homozygous *Pitx1* frameshift LOF mutations (*Pitx1*fs/fs) showed only a slight change in the interaction between *Pitx1* and *RA3* in mutant hindlimbs (Supplementary Fig. 7a). The marginal contribution of *Pitx1* to the folding of its own locus in *trans* suggested the contribution of other putative upstream transcription factors. Posterior *Hoxc* genes (*Hoxc9* to *Hoxc11*) are candidate regulators of *Pitx1* transcription, as they are specifically expressed in the posterior lateral plate mesoderm at E9.5, from where hindlimb buds take their origin, and subsequently in hindlimb buds (Fig. 5a,b,c). Moreover, ChIP-seq data
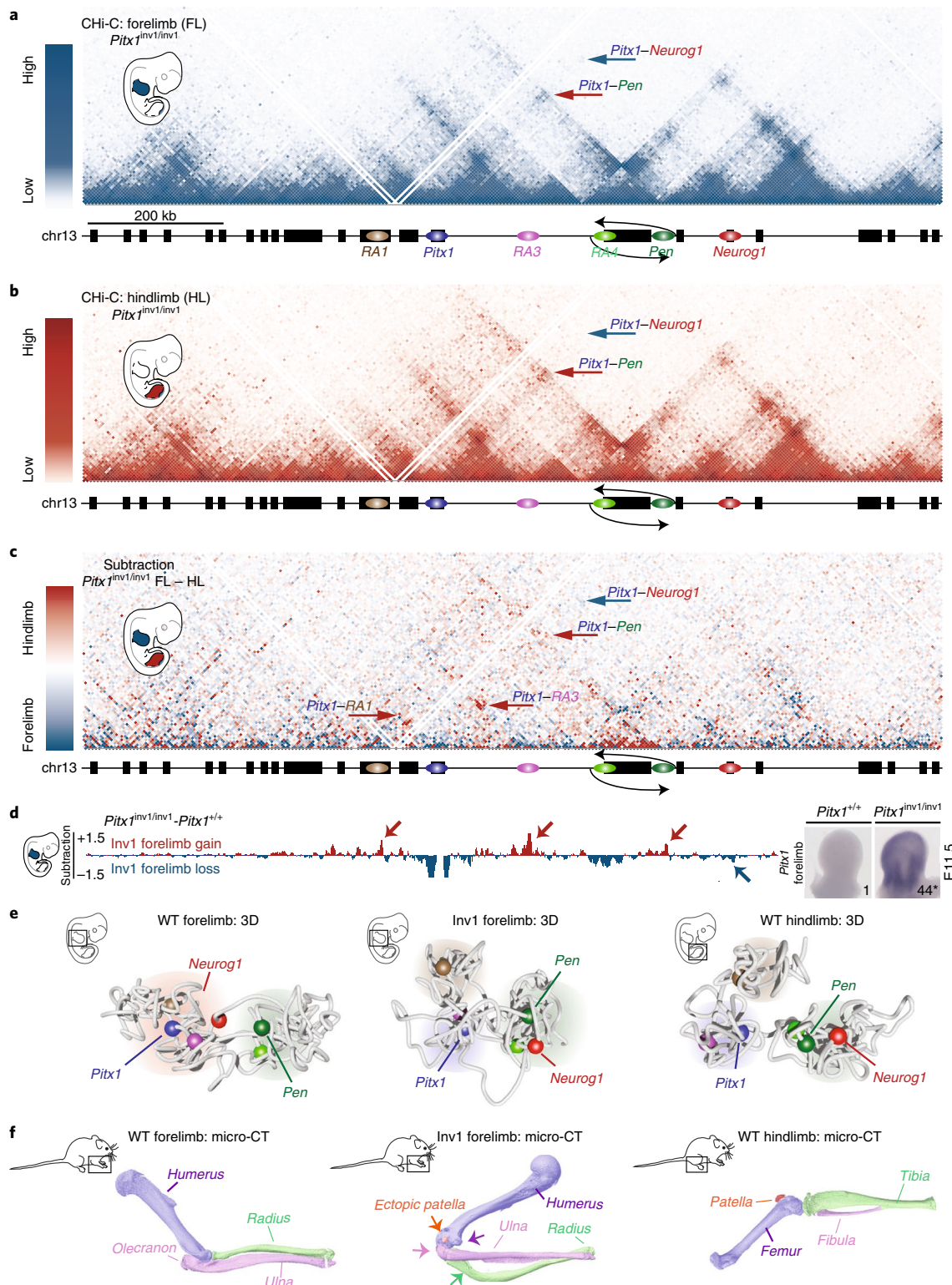
Fig. 6 | Ectopic *Pitx1–Pen* interaction, transcriptional endo-activation, and limb malformation induced by chromatin misfolding. a, CHi-C profiles of the *Pitx1* locus in the *Pitx1*[inv1/inv1] forelimb mapped on the WT genome. Note the absence of *Pitx1-Neurog1* interaction (blue arrow) and the presence of a *Pitx1–Pen* interaction (red arrow). **b**, CHi-C profiles of the *Pitx1* locus in *Pitx1*[inv1/inv1] hindlimbs. **c**, Subtraction maps of *Pitx1*[inv1/inv1] forelimbs and hindlimbs. Note the high similarity of 3D chromatin structure between both tissues in comparison to WT animals (see Fig. 4c). **d**, Left: subtraction of forelimb WT and *Pitx1*[inv1/inv1] chromatin interactions show gain of interaction between *Pitx1* and *RA1*, *RA3*, *RA4*, and *Pen* (red arrows) and loss of interaction with *Neurog1* (blue arrow) in mutants. Right: WISH and qRT–PCR *Pitx1* quantification of forelimb tissue demonstrate a strong ectopic expression and 44-fold increase of mRNA expression. We used a one-tailed *t*-test to evaluate the significance of *Pitx1* increase and found a *$P = 1.8 \times 10^{-7}$ ($n = 4$ biologically independent WT and mutant forelimb pairs); see Supplementary Fig. 10a. **e**, 3D model of the *Pitx1* locus conformation in WT forelimb, *Pitx1*[inv1/inv1] forelimb, and WT hindlimb. Note that *Pitx1*[inv1/inv1] conformation is hindlimb-like. **f**, Twenty *Pitx1*[inv1/inv1] adult mice displayed malformed forelimbs and micro-computed tomography (micro-CT) showed a distorted rotation of the elbow joint (purple arrow), a reduced olecranon (pink arrow), a deformed ulna, a deformed radius (green arrow), and an ectopic patella (orange arrow).
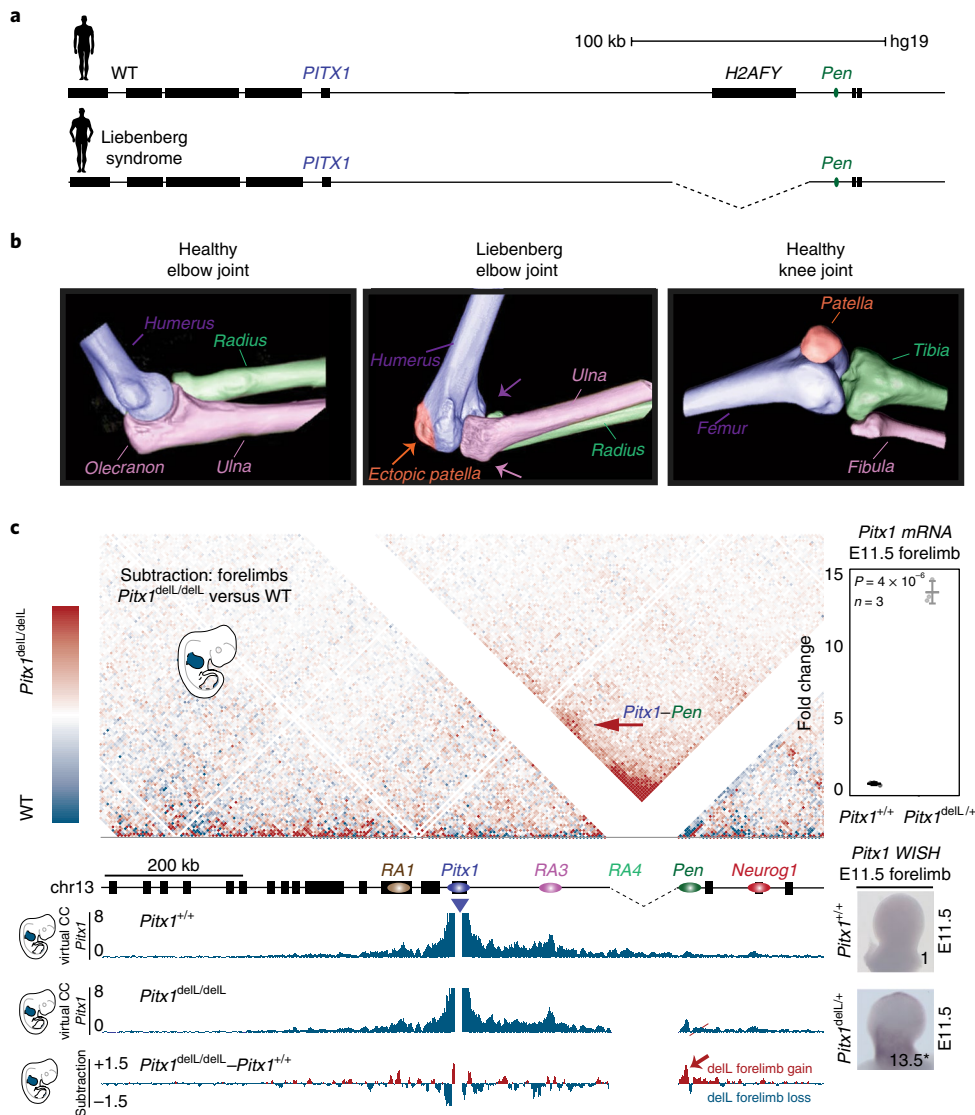
Fig. 7 | Liebenberg syndrome is caused by ectopic *Pen*–*Pitx1* interactions. a, Liebenberg syndrome is caused by deletions at the *PITX1* locus in human patients. **b**, 3D computed tomography scan of healthy human elbow joint, human Liebenberg syndrome elbow joint, and healthy human knee joint. Note that the Liebenberg elbow joint adopts morphologies of the knee joint. **c**, The subtraction of WT and *Pitx1*$^{delL/delL}$ CHi-C from forelimbs shows gain of interaction between *Pitx1* and *Pen* (red arrow). Virtual Capture-C profiles using *Pitx1* as a viewpoint derived from CHi-C in E11.5 WT and *Pitx1*$^{delL/delL}$ forelimb tissues and the subtraction of both tracks further show the gain of chromatin interaction with *Pen* (red arrow). Right: *Pitx1* WISH and qRT–PCR of *Pitx1*$^{delL/+}$ forelimb tissue showing a 13.5-fold ectopic expression. We used a one-tailed *t*-test to evaluate the significance of *Pitx1* increase and found a *$P = 4 \times 10^{-6}$; the s.d. is represented by error bars; the measure of center is the average of the data points; $n = 3$ biologically independent WT and mutant forelimb pairs.

of HOXC10 binding display enrichment at *Pitx1* itself as well as at the *RA1*, *RA3*, and *Pen* regions (Fig. 5d)[25]. We thus hypothesized that HOXC transcription factors might participate in the regulation/folding of the *Pitx1* locus. To investigate this possibility, we engineered a deletion encompassing the entire *Hoxc* cluster (*Hoxc*$^{del/del}$). Consistent with previous reports, these mice had normal limbs[26], but showed a 36% reduction in *Pitx1* expression in hindlimbs at E11.5 (Fig. 5e and Supplementary Fig. 7b). Interaction maps of mutant versus wild-type (WT) hindlimbs exhibited a significant loss of interaction between *Pitx1* and *RA3* as well as between *Pitx1* and *Pen* (Fig. 5e and Supplementary Fig. 7b). These results suggest a multifactorial control of *Pitx1* regulation and chromatin architecture involving HOXC proteins and likely other factors.

Altered chromatin folding induces *Pitx1* endogenous activation and limb malformations.. To investigate if the inactive forelimb

configuration can be converted to the active hindlimb state and induce *Pitx1* transcription, we perturbed the 3D chromatin conformation of the locus. One of the major differences between forelimbs and hindlimbs is the interaction of *Pitx1* with the repressed gene *Neurog1*. To investigate the role of this interaction we first deleted the entire (45 kb) H3K27me3-enriched region surrounding *Neurog1* (*Neurog1*$^{del/del}$). This did not result in new chromatin interactions along the chromosome, nor did it result in upregulation of *Pitx1* in the forelimb (Supplementary Fig. 8a,b). Next, we deleted the other fragment involved in this repressive interaction: a 16-kb region comprising the *Pitx1* Polycomb-repressed domain itself (*Pitx1*$^{del/del}$). To detect any regulatory changes due to this deletion, we integrated a *lacZ* regulatory sensor in the *RA3* region. In WT and mutant animals, *lacZ* reporter activity did not show activity in forelimbs, but only in hindlimbs (Supplementary Fig. 8c). Therefore, the loss of *Pitx1* itself is not sufficient to confer forelimb activity to the regulatory
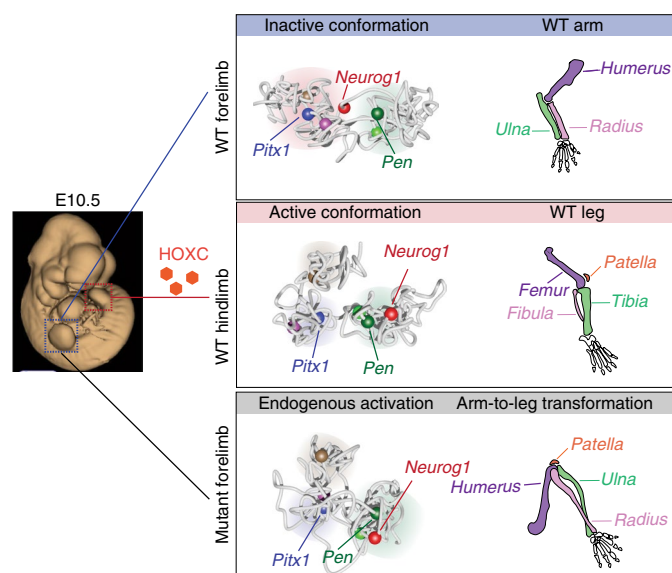
Fig. 8 | Change in chromatin configuration restricts the activity of *Pen* to the hindlimbs by separating the enhancer from its cognate promoter in forelimbs. The *Pitx1* locus can adopt several functional conformations. In E10.5 embryos (left), the forelimb and hindlimb buds appear morphologically similar. In forelimb tissues (upper panel), the *Pitx1* locus adopts an inactive conformation, whereby *Pitx1* interacts with the repressed *Neurog1* gene and is disconnected from its active enhancer regions, including the pan-limb enhancer *Pen*. The absence of *Pitx1* expression allows for normal arm morphogenesis. In hindlimb tissues (middle panel), the *Pitx1* locus adopts an active conformation, whereby *Pitx1* is disconnected from the repressed *Neurog1* gene and is in close proximity with its active enhancer *Pen*. In hindlimbs, upstream transcription factors such as HOXC (orange hexagons) are partially controlling the activity of the gene as well as locus conformation. The expression of *Pitx1* in hindlimb buds allows for normal leg morphogenesis. Finally, as a result of structural variants (lower panel), misfolding of the locus structure leads to *Pitx1* endog-activation, specifically disconnecting *Pitx1* from *Neurog1* and taking on an active conformation allowing ectopic interactions with its active enhancer *Pen*. This regulatory endo-activation of *Pitx1* in forelimbs results in a partially transformed morphogenesis in which the arms are partially transformed into legs.

landscape. Altogether, these results suggest that the repressive interaction between *Pitx1* and *Neurog1* in the forelimb is not necessary to maintain the hindlimb restriction of *Pen* activity.

Next, we hypothesized that *Pitx1* might be activated by inducing a hindlimb-like 3D configuration. This was achieved by inverting a 113-kb fragment containing *Pen* and *RA4* (*Pitx1*^inv1^). In contrast to WT tissues, the 3D chromatin organization of the *Pitx1* locus was nearly identical in the forelimbs and hindlimbs of *Pitx1*^inv1/inv1^ embryos (Fig. 6a–c). Subtraction between WT and *Pitx1*^inv1/inv1^ virtual Capture-C from the *Pitx1* viewpoint in forelimbs showed several hallmarks of WT hindlimb architecture in *Pitx1*^inv1/inv1^ forelimbs: a gain of interaction between *Pitx1*, *RA3*, and *Pen*, as well as a diminished interaction with *Neurog1* (Fig. 6d). Furthermore, modeling of the *Pitx1*^inv1^ locus from forelimb CHi-C data showed a 3D conformation strikingly similar to the conformation of WT hindlimbs (Fig. 6e, Supplementary Fig. 9, Supplementary Video 3).

In contrast to WT forelimbs, 3D models of the *Pitx1*^inv1/inv1^ locus architecture in the forelimb showed the formation of three chromatin hubs and a closer proximity between *Pen* and *RA4* to *Pitx1*, similar to the hindlimb conformation. Thus the chromatin structure of the *Pitx1* locus in mutant limb tissues appeared to have adopted a constitutive hindlimb-like configuration. *Pitx1*^inv1/inv1^ forelimb buds

displayed a 44-fold increase in *Pitx1* expression at E11.5 (Fig. 6d, Supplementary Fig. 10a). A similar forelimb expression change was also observed in *Pitx1*^del2/del2^ animals, in which the *RA4* element is present, but not in *Pitx1*^del3/del3^ animals, in which it is deleted, thereby making a potential insulator function of the *RA4* element unlikely (Supplementary Fig. 10b,c).

Pitx1^inv1/inv1^ adult mice displayed a partial homeotic arm-to-leg transformation, characterized by a reduction of the olecranon, loss of the deltoid crest, bowing of the radius, and presence of an ectopic patella in adult mice (Fig. 6f). As a control, we inverted a slightly smaller genomic region (99 kb), which leaves *Pen* at its original location (*Pitx1*^inv2^, Supplementary Fig. 10d). Adult *Pitx1*^inv2/inv2^ mice had a normal skeleton (not shown) and did not show ectopic expression of *Pitx1* in the forelimbs at E11.5 (Supplementary Fig. 10d). This finding confirms the direct effect of the *Pen* element and its position on the misexpression of *Pitx1* in *Pitx1*^inv1/inv1^ embryos. Together, these results suggest that tissue-specific changes of the *Pitx1* locus conformation modify the activity of the *Pen* enhancer, thus providing an additional layer of regulatory control.

Ectopic forelimb interaction between *Pen* and *Pitx1* induces Liebenberg syndrome. Interestingly, the skeletal alterations observed in *Pitx1*^inv1/inv1^ mice resemble those in humans affected by Liebenberg syndrome (Fig. 7a,b)[12,13]. To further assess the possibility that Liebenberg syndrome-associated structural mutations cause disruptions of the normal 3D architecture of the *PITX1* locus, we reengineered a human Liebenberg deletion in mice (*Pitx1*^delL^). This 99-kb deletion spans the entire *H2afy* gene body, placing *Pen* at the relative position of *RA4*, which is similar to the inversion in *Pitx1*^inv1^ mice. *Pitx1*^delL/+^ forelimb buds show a 14-fold upregulation of *Pitx1* at E11.5 (Fig. 7c). Moreover, CHi-C data of *Pitx1*^delL/delL^ forelimbs demonstrate a misfolding of the *Pitx1* regulatory landscape, which is characterized by the gain of interaction between *Pitx1* and *Pen*.

These results offer a mechanistic explanation for the phenotype observed in Liebenberg syndrome. The partial arm-to-leg transformation observed in this human malformation is caused by pathogenic misexpression of *PITX1* in the forelimb/arm due to an altered 3D conformation of the locus that permits ectopic interactions between *PITX1* and its endogenous *Pen* enhancer. We propose the term 'regulatory endo-activation' for this phenomenon in which the lack of modulation of an endogenous enhancer by appropriate chromatin conformation causes the ectopic activation of a gene. Our results highlight the potential importance of enhancer activity modulation by chromatin structure for human disease.

Discussion
In the present study, we show that the hindlimb-restricted *Pitx1* expression is partially controlled by a regulatory element (*Pen*) that is active in forelimbs and hindlimbs. The required tissue specificity of the enhancer is facilitated by modifying its position in the 3D chromatin space. This change in chromatin configuration restricts the activity of *Pen* to the hindlimbs by separating the enhancer from its cognate promoter in forelimbs (Fig. 8). The effect of 3D chromatin conformation on the locus becomes evident in our CHi-C maps and is clearly visible in our 3D simulations, which demonstrate fundamentally different configurations in forelimbs versus hindlimbs. The biological significance of this chromatin conformation is further exemplified by our finding that structural variations can convert the inactive 3D conformation of the forelimb into the active hindlimb-like conformation, and that it is sufficient to trigger *Pitx1* expression and change the basic morphogenetic program of forelimb versus hindlimb. However, it is important to note that in WT hindlimbs, the *Pen* enhancer only partially contributes to *Pitx1* expression. It is thus unclear if its interaction with the *Pitx1* promoter in hindlimbs occurs before *Pitx1* transcriptional onset, hence acting as the sole determinant of *Pitx1* activation, or if it is rather

supporting a preexisting, weaker activation of *Pitx1* by uncharacterized enhancers.

The spatiotemporal specificity and robustness of gene expression is achieved through the interaction of a gene's promoter with several enhancers driving complementary or overlapping regulatory patterns[5,27,28]. In this paradigm, multiple enhancer activities add up to produce the complex expression pattern of the target gene. However, the comparison of individual enhancer activities outside of their normal genetic context with their endogenous gene expression in vivo has shown that subsets of enhancer activities are suppressed endogenously through uncharacterized regulatory processes[29,30]. Furthermore, large-scale transgenic studies have shown that enhancer activity does not always match the expression patterns of the surrounding genes[17]. In all of these cases, conformation-driven modulation of enhancer activity offers a potential explanation for the observed transcriptional patterns. Another intriguing possibility raised by our findings is that genes requiring strong transcriptional dynamics might take advantage of similar unspecific but potent enhancers to ensure robust and rapid transcriptional onset. In this theory, chromatin dynamics allows for the formation of poised transcriptional pockets containing all cis and trans components needed for robust transcriptional initiation, which could be activated near-instantaneously through a change in chromatin conformation. Our results indicate that dynamic 3D chromatin architecture can play a determinant role in modulating enhancer activities, thereby ensuring transcriptional robustness.

To investigate possible mechanisms that control the active versus inactive chromatin state in the developing limb, we aimed at perturbing the interaction of *Pitx1* in the forelimb with its Polycomb-repressed neighbor gene, *Neurog1*. This type of interaction has previously been shown to be mediated by Polycomb group proteins[31,32]. Our results, however, demonstrate that the disruption of the interaction itself is insufficient to trigger gene activation. In contrast, a rearrangement of the locus, placing *Pen* at the position of *RA4* (*Pitx1*[Inv1]), resulted in *Pitx1* transcriptional activation in forelimb tissue and loss of interaction with *Neurog1*, suggesting that this Polycomb-mediated interaction is a consequence of the *Pitx1* transcriptional status.

Our results indicate that the 3D conformation directly impacts the specific regulation of gene expression, raising the question of how this status is initiated and maintained. Searching for upstream factors we found that genes from the *Hoxc* cluster partially control the hindlimb-specific 3D chromatin architecture of the *Pitx1* locus as well as its transcription. Similarly, the HOXA13 and HOXD13 transcription factors have been involved in setting tissue-specific chromatin interactions at the *Hoxd* locus during distal limb development[33]. *Hox* genes thereby arise as modulators of tissue-specific 3D chromatin architecture at downstream target loci. However, it remains unclear whether HOX proteins directly contribute to the 3D chromatin architecture or if they primarily control enhancer activities that in turn promote tissue-specific 3D chromatin folding.

In contrast to previous studies, in which structural variants caused disease by altering chromatin architecture and insulation between neighboring TADs[3,34], we describe a pathomechanism where structural variants trigger the formation of an active chromatin structure leading to ectopic transcription and disease. Compared to established cases in which enhancers activate genes that are not their physiological targets[3,35,36], we show that in the Liebenberg-associated *PITX1* alleles examined here, chromatin misfolding permits the erroneous interaction between a gene and its endogenous, normally spatially sequestered enhancer. To distinguish this process from previously described enhancer adoption scenarios, we propose the term 'regulatory endo-activation' for this peculiar type of pathogenic dysregulation.

Several studies have implicated *Pitx1* as a key factor in the morphological evolution of limbs in vertebrates. In all species investigated,

the relationship between *Pitx1* and *H2afy* is maintained. However, the linkage between *Pitx1* and *Neurog1* is lost in teleost fish, supporting the idea that this region was changed and/or gained function during the evolution of tetrapods (see Supplementary Fig. 11). The *Pen* enhancer is not found in fish but is present in all tetrapods (Supplementary Fig. 12). Interestingly, deletions encompassing this enhancer alter *Pitx1* expression in certain pigeon breeds, which display a partial leg-to-wing homeotic transformation, evident from the appearance of leg feathers[37]. In sticklebacks, the loss of pelvic structures has been attributed to regulatory mutations deleting a pelvic-specific enhancer (*Pel*) of *Pitx1*[38,39]. This enhancer, however, is distinct from *Pen* and is not present in other species (Supplementary Fig. 11). The stickleback *Pel* enhancer is thus unlikely to be related to *Pitx1* expression in tetrapods, suggesting that the regulation of *Pitx1* expression in hindlimbs arose via a different regulatory mechanism than that of pelvic fins. The presence of a pan-limb enhancer in tetrapods that is silenced via a specific type of 3D genome architecture may suggest that the ancestral type of *Pitx1* expression was not hindlimb-specific, but rather of the pan-limb type and that the development of morphologically distinct forelimbs and hindlimbs was a subsequent evolutionary event.

References

1. Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
2. Spielmann, M., Lupiáñez, D. G. & Mundlos, S. Structural variation in the 3D genome. *Nat. Rev. Genet.* **19**, 453–467 (2018).
3. Weischenfeldt, J. et al. Pan-cancer analysis of somatic copy-number alterations implicates *IRS4* and *IGF2* in enhancer hijacking. *Nat. Genet.* **49**, 65–74 (2017).
4. Redin, C. et al. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat. Genet.* **49**, 36–45 (2017).
5. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506 (2013).
6. Freire-Pritchett, P. et al. Global reorganisation of *cis*-regulatory units upon lineage commitment of human embryonic stem cells. *eLife* **6**, e21926 (2017).
7. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572.e24 (2017).
8. DeLaurier, A., Schweitzer, R. & Logan, M. *Pitx1* determines the morphology of muscle, tendon, and bones of the hindlimb. *Dev. Biol.* **299**, 22–34 (2006).
9. Marcil, A., Dumontier, E., Chamberland, M., Camper, S. A. & Drouin, J. *Pitx1* and *Pitx2* are required for development of hindlimb buds. *Development* **130**, 45–55 (2003).
10. Lanctôt, C., Moreau, A., Chamberland, M., Tremblay, M. L. & Drouin, J. Hindlimb patterning and mandible development require the *Ptx1* gene. *Development* **126**, 1805–1810 (1999).
11. Nemec, S. et al. *Pitx1* directly modulates the core limb development program to implement hindlimb identity. *Development* **144**, 3325–3335 (2017).
12. Spielmann, M. et al. Homeotic arm-to-leg transformation associated with genomic rearrangements at the *PITX1* locus. *Am. J. Hum. Genet.* **91**, 629–635 (2012).
13. Al-Qattan, M. M., Al-Thunayan, A., Alabdulkareem, I. & Al Balwi, M. Liebenberg syndrome is caused by a deletion upstream to the *PITX1* gene resulting in transformation of the upper limbs to reflect lower limb characteristics. *Gene* **524**, 65–71 (2013).
14. Szeto, D. P., Ryan, A. K., O'Connell, S. M. & Rosenfeld, M. G. P-OTX: a PIT-1-interacting homeodomain factor expressed during anterior pituitary gland development. *Proc. Natl Acad. Sci. USA* **93**, 7706–7710 (1996).

15. 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).

16. Beard, C., Hochedlinger, K., Plath, K., Wutz, A. & Jaenisch, R. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* **44**, 23–28 (2006).

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

18. Kraft, K. et al. Deletions, inversions, duplications: engineering of structural variants using CRISPR/Cas in mice. *Cell Rep.* **10**, 833–839 (2015).

19. Szeto, D. P. et al. Role of the Bicoid-related homeodomain factor Pitx1 in specifying hindlimb morphogenesis and pituitary development. *Genes Dev.* **13**, 484–494 (1999).

20. Alvarado, D. M. et al. Pitx1 haploinsufficiency causes clubfoot in humans and a clubfoot-like phenotype in mice. *Hum. Mol. Genet.* **20**, 3943–3952 (2011).

21. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

22. 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).

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

24. Bianco, S. et al. Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat. Genet.* **50**, 662–667 (2018).

25. Jain, D. et al. Regulatory integration of Hox factor activity with T-box factors in limb development. *Development* **145**, dev159830 (2018).

26. Suemori, H. & Noguchi, S. *Hox C* cluster genes are dispensable for overall body plan of mouse embryonic development. *Dev. Biol.* **220**, 333–342 (2000).

27. Will, A. J. et al. Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). *Nat. Genet.* **49**, 1539–1545 (2017).

28. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167**, 1170–1187 (2016).

29. Symmons, O. & Spitz, F. From remote enhancers to gene regulation: charting the genome's regulatory landscapes. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **368**, 20120358 (2013).

30. 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).

31. Schoenfelder, S. et al. Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat. Genet.* **47**, 1179–1186 (2015).

32. Joshi, O. et al. Dynamic reorganization of extremely long-range promoter-promoter interactions between two states of pluripotency. *Cell Stem Cell* **17**, 748–757 (2015).

33. Beccari, L. et al. A role for HOX13 proteins in the regulatory switch between TADs at the *HoxD* locus. *Genes Dev.* **30**, 1172–1186 (2016).

34. Lupiáñez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: how alterations of chromatin domains result in disease. *Trends Genet.* **32**, 225–237 (2016).

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

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

37. Domyan, E. T. et al. Molecular shifts in limb identity underlie development of feathered feet in two domestic avian species. *eLife* **5**, e12115 (2016).

38. Shapiro, M. D. et al. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).

39. Chan, Y. F. et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327**, 302–305 (2010).

Author contributions

G.A., S.M., B.K.K., and M.S. conceived the project. G.A., B.K.K., and M.F. performed the CHi-C. V.H., R.S., and M.V. performed the computational analysis. M.S., B.K.K., I.H., I.J., P.G., K.K., and D.G.L. produced the transgenic reporter and carried out transgenic validation. G.A., B.K.K., M.S., C.P., M.P., and P.G. performed the knockout and knockin studies. B.T. sequenced the CHi-C samples. L.W. performed morula aggregation. W.L.C. performed the micro-computed tomography analyses. M.N. conceived the polymer modeling study. A.E., C.A., S.B., and A.M.C. ran the related computer simulations and analyses. G.A., S.M., M.S., B.K.K., and A.V. wrote the manuscript with input from the remaining authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0221-x.

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

Correspondence and requests for materials should be addressed to M.N. or S.M. or G.A.

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

Methods

Cell culture and mice. *CRISPR/Cas9-engineered allelic series.* The *Pitx1*[fs/+] mouse line was generated from embryonic stem cells (ESCs) with a heterozygote 31-bp frameshift deletion in the first exon induced by the transfection of a single single-guide RNA (sgRNA) cloned in the pX459 vector[40]. Mouse mutants with larger deletions or inversions were created using the adapted CRISPR/Cas9 method—CRISPR/Cas-induced structural variants[18]. In brief, sgRNA targeting regions were designed using the Benchling and Zhang laboratory tools (see URLs) and were selected to bear a minimal off-target score (Supplementary Table 1). The sgRNA was cloned into the px459 vector (Addgene). Cell culturing was done according to standard procedure[40]. On day 1, CD1 mouse embryonic fibroblasts feeder cells were seeded onto a 6-well plate. On day 2, 400,000 G4 ESCs were seeded for each transfection. On day 3, 2 h before transfection, the ESC medium without penicillin-streptomycin was added. For transfection, a DNA mix consisting of 8 µg of each pX459-sgRNA vector was combined with 125 µl Gibco Opti-MEM reduced serum medium (Thermo Fisher Scientific); a transfection mix consisting of 25 µl FuGene HD (Promega) and 100 µl OptiMEM was incubated at room temperature for 15 min before being added drop by drop onto the cells. On day 4, three 6-cm dishes of DR4 puromycin-resistant feeders were seeded for each transfection. On day 5, targeted G4 ESCs were split onto three DR4 6-cm dishes and a 48-hour selection was initiated by adding puromycin to the ESC medium (final concentration 2 µg ml⁻¹). On day 7, selection was stopped and recovery initiated by using standard ESC medium. The recovery period was ~4 days. On day 11, individual clones (~300 per construct) were picked from the plate and transferred into 96-well plates with CD1 feeders. After 3 days, ESC cultures were split in triplicates, two for freezing and one for growth and DNA collection. Genotyping was performed using PCR and quantitative PCR (qPCR) analyses.

ESCs and feeder cells were tested for *Mycoplasma* contamination using the MycoAlert detection kit (Lonza) and MycoAlert Assay Control Set (Lonza). Genetically modified G4 ESCs were used to produce embryos through tetraploid aggregation, and genotyping confirmed the presence of the desired mutations in the cells and later in embryos. DR4 and CD1 feeder cell lines were directly produced from mouse embryos originating from DR4 and CD1 mice crosses, respectively.

CRISPR/Cas9 knockin of lacZ sensors. Transgenic mice carrying the *lacZ* sensor were created as described earlier with the following changes. The *lacZ* sensor targeting constructs were synthesized using VectorBuilder (see URLs), whereby the vector contained the *lacZ* reporter gene together with 50 bp of the β-globin promoter flanked by asymmetric homology arms of 750 bp and 2 kb in size, respectively[41]. The sgRNA was designed to hybridize with the genomic region corresponding to the first 50 base pairs of the homology arm (Supplementary Table 1). The corresponding region in the targeting vector was mutated to avoid undesired double-strand breaks. For transfection, 4 µg of the *lacZ* sensor targeting construct was cotransfected with 8 µg of px459 encoding the sgRNA.

lacZ enhancer testing. For enhancer testing, candidate regions were selected on the basis of public tracks for conservation, H3K27ac ChIP-seq, DNase hypersensitive site, and p300 ChIP-seq from 11.5 limbs[15,42]. The flippase (FLP)-flippase recognition target (FRT) system was used as the targeting strategy of the C2 ESC line, which has a phosphoglycerate kinase neomycin selection cassette flanked by FRT sites and an ATG-less, promoter-less hygromycin cassette targeted downstream of the *Col1A1* locus[16]. First, enhancer candidate regions of interest were amplified by PCR from mouse genomic DNA with NotI overhangs and subcloned into the pBluescript-*lacZ* vector linearized with NotI. Next, the enhancer sequence with the *lacZ* cassette was cloned into a phosphoglycerate kinase targeting vector containing FRT sites carrying the ATG needed for the puromycin resistance of the C2 cells. The transfection procedure was as follows. On day 1, CD1 feeders were seeded onto a 6-well plate. On day 2, 800,000 C2 cells were seeded onto the 6-well plate. On day 3, cells were washed twice with PBS and 1.75 ml ESC medium with leukemia inhibitory factor and without penicillin-streptomycin was added. The DNA mix consisting of 9 µg of targeting construct, 3 µg FLP-encoding vector, 1 µl Lipofectamine LTX Plus reagent (Thermo Fisher Scientific) tipped with OptiMEM up to 125 µl, and the transfection mix consisting of 20 µl Lipofectamine and 125 µl OptiMEM, were incubated separately at room temperature for 5 min, combined, and further incubated at room temperature for 30 min. The transfection-DNA mix was added drop by drop onto the C2 cells. DR4 puromycin-resistant feeder cells were seeded onto 6-cm dishes, three for each targeting construct. On day 4, each transfected C2 well was split onto three DR4 feeder dishes. On day 5, selection was started by using ESC medium containing hygromycin B (final concentration 150 µg ml⁻¹), and then continued until clones were picked and transferred into 96-well plates with CD1 feeders. After another 3 days, cultures were split in triplicates, two for freezing and one for growth and DNA collection. Positive clones were thawed and grown on CD1 feeders until they reached an average of 4 million cells. Three vials were frozen and DNA was collected from the rest of the cells for genotyping. An overview of tested enhancer candidate regions is shown in Supplementary Fig. 1a. Genetically modified C2 ESCs were used to produce embryos through tetraploid aggregation, and genotyping confirmed the presence of the desired mutations in the cells and

later in the embryos. DR4 and CD1 feeder cell lines where directly produced from mouse embryos originating from DR4 and CD1 mice crosses, respectively.

Aggregation of mouse ESCs. Embryos and live animals were generated from ESCs by diploid or tetraploid complementation, after thawing a frozen ESC vial seeded on CD1 feeders and grown for 2 days[43]. Female mice of the CD1 strain were used as foster mothers. Several mouse lines were maintained by crossing them with C57BL6/J mice.

Animal procedures. All animal procedures were carried out in accordance with institutional, state, and governmental regulations (LAGeSo Berlin, G0247/13).

lacZ reporter staining. For whole-mount *lacZ* reporter staining (with some adjustments[44]), E11.5 mouse embryos were dissected in cold PBS, fixed in 4% paraformaldehyde (PFA) in PBS on ice for 30 min, followed by 3× washing with the *lacZ* buffer (2 mM MgCl₂, 0.01% sodium deoxycholate, 0.02% Nonidet P-40 in PBS). The embryos were then incubated in staining solution (0.5 mg ml⁻¹ X-gal, 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide in *lacZ* buffer) at 37 °C for a few hours to overnight until the desired staining was obtained. After staining, embryos were washed in *lacZ* buffer and imaged using a ZEISS SteREO Discovery.V12 with cold light source CL9000 microscope and Leica DFC420 digital camera (Leica Microsystems). Embryos were stored at 4 °C in 4% PFA in PBS.

RNA isolation and quantitative reverse transcription PCR (qRT–PCR). To quantify mRNA levels in WT and mutant mice, E11.5 forelimb or hindlimb buds were microdissected in cold PBS, immediately snap-frozen, and stored at −80°C. To isolate RNA, 500 µl guanidinium thiocyanate was added to the tissue and homogenized using a pestle; 100 µl chloroform was then added, the samples were pulse-vortexed for 15 s, and centrifuged at 4 °C at 9,300g for 15 min. The supernatant was transferred to a new tube and mixed with 1 volume of 70% ethanol, and then loaded onto a RNeasy Mini Kit (QIAGEN) column. The rest of the RNA isolation was done according to the manufacturer's instructions. Complementary DNA was generated using the Superscript III First-Strand Synthesis System (Thermo Fisher Scientific) whereby 300 ng of RNA was reverse-transcribed using random hexamer primers. To quantify the relative abundance of transcripts, qRT–PCR analyses of 3–8 pairs of forelimb or hindlimb bud pairs in technical triplicates was done using the GoTaq qPCR Master Mix (Promega).

Whole-mount in situ hybridization. The *Pitx1* mRNA expression in E11.5 mouse embryos was assessed by whole-mount in situ hybridization (WISH) using a digoxigenin-labeled *Pitx1* antisense riboprobe transcribed from a cloned *Pitx1* probe (PCR DIG Probe Synthesis Kit, Roche). Whole embryos were fixed overnight in 4% PFA/PBS. The embryos were washed in PBS with Tween (PBST, 0.1% Tween), dehydrated stepwise in 25%, 50%, and 75% methanol/PBST, and finally stored at −20°C in 100% methanol. The WISH protocol was as follows. On day 1, embryos were rehydrated on ice in reverse methanol/PBST steps, washed in PBST, bleached in 6% H₂O₂/PBST for 1 hour and washed in PBST. Embryos were then treated in 10 µg ml⁻¹ proteinase K/PBST for 3 min, incubated in glycine/PBST, washed in PBST, and finally refixed for 20 min with 4% PFA/PBS, 0.2% glutaraldehyde, and 0.1% Tween 20. After further washing steps with PBST, embryos were incubated at 68 °C in L1 buffer (50% deionized formamide, 5× saline sodium citrate, 1% SDS, 0.1% Tween 20 in diethyl pyrocarbonate, pH 4.5) for 10 min. Next, embryos were incubated for 2 h at 68 °C in hybridization buffer 1 (L1 with 0.1% transfer RNA and 0.05% heparin). Afterwards, embryos were incubated overnight at 68 °C in hybridization buffer 2 (hybridization buffer 1 with 0.1% transfer RNA and 0.05% heparin and 1/500 digoxigenin-*Pitx1* probe). On day 2, removal of the unbound probe was done through a series of washing steps 3 × 30 min each at 68 °C: L1; L2 (50% deionized formamide, 2× saline sodium citrate pH 4.5, 0.1% Tween 20 in diethyl pyrocarbonate, pH 4.5); and L3 (2× saline sodium citrate pH 4.5, 0.1% Tween 20 in diethyl pyrocarbonate, pH 4.5). Subsequently, embryos were treated for 1 hour 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 Tris-buffered saline, 0.1% Tween 20 (TBST) 1 (140 mM NaCl, 2.7 mM KCl, 25 mM Tris-HCl, 1% Tween 20, pH 7.5). Next, embryos were blocked for 2 h at room temperature in blocking solution (TBST 1 with 2% fetal bovine serum and 0.2% bovine serum albumin), followed by incubation at 4 °C overnight in blocking solution containing 1/5,000 anti-digoxigenin-alkaline phosphatase. On day 3, removal of unbound antibody was done through a series of washing steps 8 × 30 min at room temperature with TBST 2 (TBST with 0.1% Tween 20 and 0.05% levamisole/tetramisole) and left overnight at 4 °C. On day 4, staining of the embryos was initiated by washing at room temperature with alkaline phosphatase 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). Limb buds from at least three embryos were analyzed from each mutant genotype. The stained limb buds were imaged using a ZEISS SteREO Discovery.V12 with cold light source CL9000 microscope and Leica DFC420 digital camera.

Skeletal preparation. E18.5 fetuses were processed and stained for bone and cartilage as described previously[45]. Fetuses were kept in H₂O for 1–2 h at room

temperature and heat-shocked at 65 °C for 1 minute. The skin was taken off and the abdominal and thoracic viscera were removed using forceps. The fetuses were then fixed in 100% ethanol overnight. The next day, the cartilage was stained using alcian blue staining solution (150 mg l[−1] alcian blue 8GX in 80% ethanol and 20% acetic acid). On the third day, fetuses were rinsed and postfixed in 100% ethanol overnight. On the fourth day, initial clearing was done by incubating the fetuses for 20 min in 1% potassium hydroxide in H_2O, followed by alizarin red (50 mg l[−1] alizarin red S in 0.2% potassium hydroxide) staining of bones overnight. On the fifth day and onward, rinsing and clearing was done using low concentrations of potassium hydroxide. The stained embryos were dissected in 80% glycerol and limbs were imaged using a ZEISS SteREO Discovery.V12 with cold light source CL9000 microscope and Leica DFC420 digital camera.

Micro-computed tomography. Limbs of control and mutant mice were wrapped in plastic film and scanned ex vivo using a SkyScan 1172 high-resolution micro-computed tomography system (Brucker microCT). The scanning was done with a 0.5-mm aluminum filter at 80 kV and 120 μA at a 10-μm resolution. 3D model reconstruction was done with the SkyScan image analysis software (computed tomography analyzer and computed tomography volume).

CHi-C. *SureSelect design.* The CHi-C SureSelect library was designed over the genomic interval (mm9, chr13: 54,000,001–57,300,000) using the SureDesign tool from Agilent.

Preparation of chromosome conformation capture (3C) libraries. 3C libraries were prepared from homozygous E11.5 forelimb and hindlimb buds as described previously[46]. CHi-C experiments were performed as quadruplets (WT) or duplicates (mutants). Per biological replicate, 5–6 pairs of limb buds (~3×10^6 cells) were microdissected in PBS at room temperature. A single-cell suspension was obtained by incubating the tissue in 500 μl Gibco trypsin-EDTA 0.05% (Thermo Fisher Scientific) at 37 °C for 10 min shaking at 900 r.p.m.. The cells were resuspended and homogenized using a 0.40-μm cell strainer (Falcon) and diluted in 10% fetal bovine serum/PBS. Cells were fixed by adding 650 μl 37% formaldehyde (Sigma-Aldrich) with a final concentration of 2% and mixed for 10 min at room temperature. Fixation was quenched using 1.425 M glycine (Merck) on ice and immediately centrifuged at 260*g* for 8 min. Supernatant was removed, the pellet resuspended in lysis buffer (final concentration of 10 mM Tris, pH 7.5, 10 mM NaCl, 5 mM $MgCl_2$, 0.1 M EGTA, 1× cOmplete protease inhibitors (Sigma-Aldrich)) and incubated on ice for 10 min. Removal of lysis buffer was done by centrifugation at 400 *g* for 5 min at 4 °C, followed by removal of supernatant, snap-freezing, and storage at −80 °C. On the next day, the pellet was resuspended in 520 μl 1× DpnII buffer (Thermo Fisher Scientific), and incubated with 7.4 μl 20% SDS shaking at 900 r.p.m. at 37 °C for 1 hour. Next, 75 μl 20% Triton X-100 was added and the pellet was left shaking at 900 r.p.m. at 37 °C for 1 hour. A 15-μl aliquot was taken as a control for undigested chromatin (stored at −20 °C). The chromatin was digested using 40 μl 10 U μl[−1] DpnII buffer shaking at 900 r.p.m. at 37 °C for 6 h; 40 μl of DpnII was added and samples were incubated overnight, shaking at 900 r.p.m. at 37 °C. On the third day, 20 μl DpnII buffer was added to the samples, shaking for 5 more hours at 900 r.p.m. at 37 °C. The DpnII restriction enzyme was inactivated at 65 °C for 25 min. A 50-μl aliquot was taken to test digestion efficiency (stored at −20 °C). Next, the digested chromatin was diluted and religated in 5.1 ml H_2O, 700 μl 10× ligation buffer (Thermo Fisher Scientific), 5 μl 30 U μl[−1] T4 DNA ligase (Thermo Fisher Scientific), incubated at 16 °C for 4 h, and shaken manually 3 times. The ligated samples were incubated for a further 30 min at room temperature. The chimeric chromatin products and test aliquots were de-cross-linked overnight by adding 30 μl and 5 μl proteinase K, respectively, and incubated at 65 °C overnight. On the fourth day, 30 μl or 5 μl of 10 mg ml[−1] RNase was added to the samples and aliquots, respectively, and incubated for 45 min at 37 °C. Next, chromatin was precipitated by adding 1 volume phenol-chloroform to the samples and aliquots, vigorously shaking them, followed by centrifugation at 4,000 r.p.m. at room temperature for 15 min. The upper phase containing the chromatin was transferred to a new tube. One hundred percent ethanol was added to the aliquots, and the samples were frozen for 30 min, centrifuged at 5,000 r.p.m. for 45 min at 4 °C, washed with 70% ethanol, and resuspended in 20 μl 10 mM Tris-HCl, pH 7.5; 7 ml H_2O, 1 ml 3M NaAc, pH 5.6, and 35 ml 100% ethanol were added to the samples. The samples were frozen at −20 °C for 3 h. The precipitated chromatin was isolated by centrifugation at 5,000 r.p.m. for 45 min at 4 °C. The chromatin pellet was washed with 70% ethanol and further centrifuged at 5,000 r.p.m. for 15 min at 4 °C. Finally, the 3C library chromatin pellet was dried at room temperature and resuspended in 10 mM Tris-HCl, pH 7.5. To check the 3C library, 600 ng were loaded on a 1% gel together with the undigested and digested aliquots. The 3C library was then sheared using a Covaris sonicator (duty cycle: 10%; intensity: 5; cycles per burst: 200; time: 6 cycles of 60 s each; set mode: frequency sweeping; temperature: 4–7 °C). Adaptors were added to the sheared DNA and amplified according to the manufacturer's instructions for Illumina sequencing (Agilent). The library was hybridized to the custom-designed SureSelect beads and indexed for sequencing (50–100 bp paired-end) following the manufacturer's instructions (Agilent).

Quantification and statistical analysis. *qRT–PCR analysis.* The $2^{-\Delta\Delta CT}$ method[47] was used to calculate the fold change between WT and mutant samples. In the forelimb, using a one-tailed Student's *t*-test, we tested the ability of mutations (*Pitx1*[del2/del2], *Pitx1*[del3/del3], *Pitx1*[inv1/inv2], *Pitx1*[delL/delL], *Neurog1*[del/del]) to result in increased gene expression. In contrast, in the hindlimb we tested the ability of mutations (*Pitx1*[del1/del1], *Pitx1*[del2/del2], *Pitx1*[del3/del3], *Pitx1*[Pen/Pen], *Hoxc*[del/del]) to result in decreased *Pitx1* gene expression. The variance in mutant and WT samples was assumed to be equal. Individual qRT–PCR data points are shown in Supplementary Table 2.

Hi-C analysis. Preprocessing and mapping of paired-end sequencing data, as well as filtering of mapped di-tags, was performed with the HiCUP pipeline version 0.5.8[48] (Nofill: 1, no size selection; Format: Sanger). The pipeline used Bowtie2 version 2.2.6[49] (see URLs) for mapping short reads to the reference genome (NCBI37/mm9). Replicates were combined after mapping and filtering. Filtered di-tags were further processed with Juicer Tools (see URLs) to bin di-tags (5 and 10 kb bins); the map was normalized using Knight–Ruiz matrix balancing[50–52]. For this, only reads with a mapping quality score ≥30 were considered. The DNA-capturing step enriches genomic region chr13:54,000,001–57,300,000 on mm9 leading to three different regimes in the CHi-C map: (1) enriched versus enriched; (2) enriched versus non-enriched; and (3) non-enriched versus non-enriched. For binning and normalization, only di-tags in regime (1) were considered. Therefore, di-tags were filtered for the enriched region and the mm9 coordinates were shifted by 54,000,000 bp. For Juicer Tools, a custom chromosome size file containing only the enriched region on chr13 (length 3,300,000 bp) was used. After binning and normalization, coordinates were shifted back to their original values. All maps were processed on the WT reference genome to work with the same genomic coordinates across all samples. To account for differences between maps in their distance-dependent signal decay, maps were scaled pairwise by the sum of their subdiagonals. Therefore, each subdiagonal vector in one matrix was divided by its sum and multiplied by the average of the sums of both matrices. To normalize for sequencing depth, each map was additionally converted to reads per million. To avoid copy number biases, a region spanning all tested deletions (chr13:55,730,001–56,500,000) was not considered for the computation of scaling factors, for diagonal normalization, and reads per million normalization. CHi-C maps of count values and subtraction maps were visualized as heat maps, truncating all values above the 99th percentile for visualization purposes. The percentile was determined from absolute values within the same region used for normalization.

Differential interactions. Differential interactions were determined from the pairwise subtraction of normalized maps. For diagonal and reads per million normalization, as well as for the computation of *P* values, individual regions spanning the corresponding deletions were excluded, that is, chr13:56,140,001–56,260,000 for *Pitx1*[delL] and chr13:56,320,001–56,370,000 for *Neurog1*[del]. To avoid artifacts from Knight–Ruiz normalization due to low coverage, maps were analyzed at a 10-kb resolution; a single row with very low coverage (chr1:54,450,001–54,460,000) was excluded. To account for the distance dependence of the magnitude of differences, each difference value was subtracted by the mean and divided by the s.d. of the corresponding subdiagonal. For the computation of the mean and s.d. per subdiagonal, values above the 99.5th percentile were not included. *P* values for the z-transformed difference values were computed using a standard normal distribution and further corrected for multiple testing with false discovery rate[53]. In histograms, the distribution of z-scores does not fit the normal distribution perfectly, but shows a bell-like shape distribution.

Virtual Capture-C profiles. To obtain more fine-grained interaction profiles, we generated virtual Capture-C-like profiles based on the filtered BAM files also used for the CHi-C maps, and defined several virtual 10-kb viewpoints. A read pair was considered in the profile when one mate mapped to the defined viewpoint region and the other one outside of it. Reads were counted per restriction fragment and binned further to a regular 1-kb grid. When a fragment spanned more than one bin, the count value was distributed proportionally to the overlaps. Afterwards, the profiles were smoothed by averaging over a sliding window of five bins. For coverage normalization, the profiles were divided by the sum of counts in the enriched region on chr13 and multiplied with 10^3. The region ± 5 kb around the viewpoint as well as the chr13:56,140,001–56,260,000 region for *Pitx1*[delL] and the chr13:56,320,001–56,370,000 region for *Neurog1*[del] were excluded from the computation of the scaling factor in the pairwise comparisons. The profiles were generated with a custom Java code using htsjdk version 1.139 (see URLs).

In the case of the *Neurog1*[del] virtual Capture-C to WT comparison, we created a chromosome-wide interaction profile instead of a profile for the enriched region only. Here we used a larger viewpoint region of 15 kb (chr13: 55,925,001–55,940,000), spanning the entire H3K27me3-enriched region to improve the coverage. Binning and scaling was performed as described previously.

3D polymer modeling. *The String & Binders Switch (SBS) polymer model.* To investigate the 3D structure of the *Pitx1* locus, we employed the SBS model[54], a polymer physics model previously shown to recapitulate fluorescence in situ hybridization, Hi-C, and generalized additive model data[24,55–57]. In the SBS model, a chromatin filament is represented as a self-avoiding chain of beads that include

binding sites for diffusing molecule binders; under the laws of physics, the polymer folds spontaneously because the binders can bridge and loop their cognate binding sites (see Bianco et al.[24]).

In each of the studied cases, forelimb and hindlimb tissues, and forelimb inv1 inversion, the specific SBS models for the *Pitx1* locus were established using a simulated annealing optimization procedure[24] that finds the minimal number of different types of binding sites (different colors in our notation) and their arrangement along the chain (Supplementary Figs. 4a, 4b, 9a), returning the best agreement between the corresponding CHi-C data (Supplementary Figs. 4c,d and 9B top panel) and the equilibrium pairwise contact map derived by the polymer model (Supplementary Figs. 4c,d and 9b bottom panel). Molecular dynamics (MD) computer simulations were used to derive an ensemble of the model equilibrium 3D conformations.

Simulations details. We modeled a broad genomic sequence encompassing the mouse *Pitx1* regulatory region to avoid boundary effects and focused on chr13:55,600,000–56,650,000 (mm9). Based on CHi-C interaction data (10-kb resolution), we used a polymer chain of $N=1,785$ beads. We ran MD simulations using the LAMMPS software[58] (see URLs). The model interaction potentials derive from classical studies of polymer physics[59]; in particular, the molar concentration of binders $= 135$ nmol l^{-1} and the scale of the bead-binder interaction energy is $E_{int} = 1.0$ kbT and $E_{int} = 8.1$ kbT, corresponding respectively to the coil and globule conformational state of the polymer[24]. The σ size of each bead of the polymer chain is approximately 17 nm[24]. The model and MD details are illustrated in Bianco et al.[24].

In our MD simulations, the polymer initial states were self-avoiding walk configurations, with the binders randomly distributed in the simulation volume, as in standard MD approaches[24,59]. To minimize finite-size effects, the simulation box was at least two times larger than the gyration radius of the self-avoiding walk polymer; periodic boundary conditions were used. The initial configurations evolved up to 5×10^8 time steps to approach stationarity, as measured by the plateauing of the gyration radius and mechanical energy, and confirmed by the polymer scaling exponents[24]. An ensemble of at least 10^2 different equilibrium configurations for each of the considered cases was derived using MD.

Contact matrices. To test our models against the experiments, we compared the CHi-C data against the average contact matrix computed from the ensemble of 3D polymer model conformations derived using MD. To compute the contact probability for all pairs of beads i and j, in each 3D conformation we counted if i and j were in contact, that is, if their physical distance r_{ij} was less than a threshold distance (see Bianco et al.[24]). To take into account the effects of cell population heterogeneity, that is, the possibility that the locus could be in different states (coil/globule) in different cells, we considered the contact matrix of the coil/globule mixture that maximized the Pearson correlation coefficient, r, with the CHi-C data[24]. An 80–20% mixture well describes all cases.

Correlation coefficients. To account for the effects of genomic proximity beyond Pearson's r between model-predicted and CHi-C contact matrices, we also computed the distance-corrected Pearson's r, that is, the correlation between the two matrices where the average contact frequency at each genomic distance has been previously subtracted. The MD model versus the CHi-C Pearson's r is 0.98 in WT forelimb, 0.98 in WT hindlimb, and 0.97 in the inv1 forelimb case (Supplementary Figs. 4c,d and 9b); the distance-corrected Pearson's r is 0.84 in WT forelimb, 0.82 in WT hindlimb, and 0.74 in the inv1 forelimb case. (Here, strong outliers above the 90th percentile are excluded.)

3D modeling results. Our modeling results of the *Pitx1* locus in hindlimbs, forelimbs, and the *Pitx1*[inv1/inv1] (inv1) forelimb case are summarized in Supplementary Figs. 4 and 9. Our polymer model of the locus involves 14 different types of binding sites, whose position and abundance along the genome in the three distinct cases is represented by the histograms shown in Supplementary Figs. 4a,b and 9a, where a different color is associated with each type of binding site. (Similar colors in different cell types are used just to simplify the presentation.) CHi-C and model-derived contacts maps are shown in Supplementary Figs. 4c,s and 9b. As illustrated by the comparatively high values of Pearson's r and distance-corrected Pearson's r, chromatin contacts in the different samples are well captured by our 3D modeling. To further evaluate the accuracy of the 3D models, subtractions between the CHi-C and SBS model-derived contact matrices were obtained and are shown in Supplementary Figs. 4e,f and 9c. Significant differences between experimental data and model are found in just a minor fraction of cases (<5%) and are sparsely located within the contact matrices. Significance is established with respect to the 95th and 5th percentile of a conservative standard random control model made of contact matrices having bootstrapped diagonals. Our polymer model can also capture the major architectural changes between forelimb and hindlimb, as displayed in Supplementary Fig. 5, where the subtraction between WT forelimb and hindlimb SBS models is shown. Finally, Supplementary Figs. 4g,h and 9d display, in each studied case, a representative 3D structure of the locus, selected from the ensemble of 'single-cell' model-derived conformations.

Polymer 3D representation. In Figs. 4d,e and 6e, a single representative configuration of the *Pitx1* locus in the globule state is shown for each different

cell type; to better visualize the relative positions of *Pitx1* and its regulatory anchors, a coarse-grained version of the simulated polymer is depicted. Instead, in Supplementary Figs. 4g,h and 9d, we show the entire polymer model. In all cases, the coordinates of each bead were interpolated with a smooth third-order polynomial splice curve. Figures were created using POV-Ray 3.7.0 (see URLs; Persistence of Vision Raytracer Pty. Ltd).

Relative physical distances. Our polymer model allows to measure the physical distances between the regions of interest. The changes in relative distance, shown in Supplementary Fig. 6, are the ratio $(d_{FL}-d_{HL})/d_{FL}$ of the distances in forelimbs (d_{FL}) and hindlimbs (d_{HL}) among *Pitx1* and its key regulatory regions averaged over the discussed state mixture.

Sequence alignment. The sequences and annotations of the extended *Pitx1* locus from human, chicken, frog, stickleback, spotted gar, and elephant shark were obtained from the Ensembl genome browser (see URLs) and aligned using the mVISTA LAGAN program[60] (see URLs). We used the following parameters: calc window, 100 bp; Min Cons Width, 100 bp; Cons Identity, 70%.

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

Online content. Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-018-0221-x.

Data availability

CHi-C datasets generated for this study are available in the Gene Expression Omnibus under accession GSE103676. Correspondence regarding 3D modeling should be addressed to M.N. (nicodem@na.infn.it).

References

40. Andrey, G. & Spielmann, M. CRISPR/Cas9 genome editing in embryonic stem cells. *Methods Mol. Biol.* **1468**, 221–234 (2017).
41. Byrne, S. M., Ortiz, L., Mali, P., Aach, J. & Church, G. M. Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic Acids Res.* **43**, e21 (2015).
42. Visel, A. et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
43. 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).
44. Lobe, C. G. et al. Z/AP, a double reporter for cre-mediated recombination. *Dev. Biol.* **208**, 281–292 (1999).
45. Mundlos, S. Skeletal morphogenesis. *Methods Mol. Biol.* **136**, 61–70 (2000).
46. Hagège, H. et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* **2**, 1722–1733 (2007).
47. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408 (2001).
48. Wingett, S. et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res.* **4**, 1310 (2015).
49. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
50. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
51. Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* **33**, 1029–1047 (2013).
52. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
53. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Methodol.* **57**, 289–300 (1995).
54. Nicodemi, M. & Prisco, A. Thermodynamic pathways to genome spatial organization in the cell nucleus. *Biophys. J.* **96**, 2168–2177 (2009).
55. Beagrie, R. A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).
56. Barbieri, M. et al. Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl Acad. Sci. USA* **109**, 16173–16178 (2012).
57. Barbieri, M. et al. Active and poised promoter states drive folding of the extended *HoxB* locus in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* **24**, 515–524 (2017).
58. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
59. Kremer, K. & Grest, G. S. Dynamics of entangled linear polymer melts: a molecular-dynamics simulation. *J. Chem. Phys.* **92**, 5057–5086 (1990).
60. Brudno, M. et al. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13**, 721–731 (2003).

natureresearch

Corresponding author(s): Stefan Mundlos and Guillaume Andrey

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

▶ Experimental design

1. Sample size

   Describe how sample size was determined.

   For expression analyses sample size contained at least 3 independent biological samples (pairs of limb buds), which is standardly accepted for this kind of experiment.
   For capture HiC, we used at least two independently obtained biological duplicates to prepare libraries in order to reach high coverage of every bin. In other similar studies, cHiC was done as singleton (See Franke et la., Nature, 2016 or Bianco et al., Nat. Gen. 2018).

2. Data exclusions

   Describe any data exclusions.

   There was no exclusion/inclusion of samples or animals in the analysis.

3. Replication

   Describe the measures taken to verify the reproducibility of the experimental findings.

   To quantify the relative abundance of transcripts, qRT-PCR analyses of 3-8 pairs of fore- or hindlimb buds from wildtype and mutant animals, in technical triplicates were done. The changes described in this work were consistant amongst replicates and tested using one sided t-test in all the cases.
   cHiC experiments were performed as quadruplets (wildtypes) or duplicates (mutants). cHiC maps were showing reproducibility in the pre-processing steps (HiCup pipeline) and were visually inspected for reproducibility and consistency. Finally, replicates where merged in order to obtain high coverage and high resolution.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   There was no randomization of experiments.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   Investigators were not blinded during experiments and outcome assessment. The genotyping of tissues had to occur before the preparation of RNA/cDNA of cHiC libraries and as a result investigators knew what samples they were handling.

   Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The <u>exact sample size</u> (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	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 any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒	A clear description of statistics including <u>central tendency</u> (e.g. median, mean) and <u>variation</u> (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars in <u>all</u> relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

▶ Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

> 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
>
> Picture and video of 3D models where produced using POV-RAY (Persistence of Vision Pty. Ltd. 2004)
>
> We used publicly available computer codes (LAMMPS) for our Molecular Dynamics simulations.
>
> The profiles were generated with custom Java code using htsjdk v1.139 (https://samtools.github.io/htsjdk/
>
> Analysis and display of qRT-PCR experiment where done on Excel and R, respectively

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

▶ Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

> No, there are no restriction

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

> We used Anti-Digoxigenin-AP antibody to perform WISH. The antibody is produced by Roche with the reference number 11093274910. We used the lot number 11266027.

10. Eukaryotic cell lines

    a. State the source of each eukaryotic cell line used.

    We used G4 Embryonic Stem cells (G4 ESCs). CD1 and DR4 feeder cell lines produced from CD1 and DR4 transgenic embryos were used to culture the G4 cells.

    b. Describe the method of cell line authentication used.

    Genetically modified ESCs were used to produced embryos through tetraploid aggregation and genotyping confirmed the presence of the desired mutations in the cells.
    DR4 and CD1 feeder cell lines where directly produced from mouse embryos originating from DR4 and CD1 mice crosses, respectively.

    c. Report whether the cell lines were tested for mycoplasma contamination.

    All cell lines tested negative for mycoplasma contamination.

    d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

    No commonly misidentified cell lines were used.

▶ Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

    Provide all relevant details on animals and/or animal-derived materials used in the study.

    Embryos and live animals were generated from wildtype or genetically engineered male G4 ESCs (129/Sv×C57BL/6 F1 Hybrid ES Cell) by diploid or tetraploid complementation. Several mouse lines were maintained by crossing them with C57BL6/J mice.
    For biochemical analyses (qRT-PCR or cHiC), embryos were collected at E11.5 and limbs where microdissected. Skeletal preparations where done from E18.5 embryos.
    Microtomography was performed on adult mice (3 months) legs and arms.
    All animal procedures were in accordance with institutional, state, and government regulations (Berlin:LAGeSo G0247/13).

Policy information about studies involving human research participants

12. Description of human research participants

    Describe the covariate-relevant population characteristics of the human research participants.

    We did not use human participants for this research