nature genetics

Composition and dosage of a multipartite enhancer cluster control developmental expression of *Ihh* (Indian hedgehog)

Anja J Will[1,2], Giulia Cova[1,2], Marco Osterwalder[3], Wing-Lee Chan[1,2,4], Lars Wittler[5], Norbert Brieske[1], Verena Heinrich[6], Jean-Pierre de Villartay[7], Martin Vingron[6], Eva Klopocki[8], Axel Visel[3,9,10], Darío G Lupiáñez[1,2,4,11] & Stefan Mundlos[1,2,4,11]

Copy number variations (CNVs) often include noncoding sequences and putative enhancers, but how these rearrangements induce disease is poorly understood. Here we investigate CNVs involving the regulatory landscape of *IHH* (encoding Indian hedgehog), which cause multiple, highly localized phenotypes including craniosynostosis and synpolydactyly[1,2]. We show through transgenic reporter and genome-editing studies in mice that *Ihh* is regulated by a constellation of at least nine enhancers with individual tissue specificities in the digit anlagen, growth plates, skull sutures and fingertips. Consecutive deletions, resulting in growth defects of the skull and long bones, showed that these enhancers function in an additive manner. Duplications, in contrast, caused not only dose-dependent upregulation but also misexpression of *Ihh*, leading to abnormal phalanges, fusion of sutures and syndactyly. Thus, precise spatiotemporal control of developmental gene expression is achieved by complex multipartite enhancer ensembles. Alterations in the composition of such clusters can result in gene misexpression and disease.

Work by the Encyclopedia of DNA Elements (ENCODE) Consortium and others has helped to characterize a wide catalog of regulatory elements, also referred to as enhancers, that control developmental gene expression in many species[3–5]. One of the most intriguing characteristics of these elements is their tendency to arrange in clusters, displaying redundancy in reporter assays and similarities in transcription factor occupancy[6,7]. Previous studies in *Drosophila melanogaster* showed that the observed redundancy may provide the system with robustness and spatiotemporal precision[8–10]. However, how the complex patterns of gene expression during development are achieved and why this involves elements with apparently redundant or overlapping functions remain elusive. CNVs generally include noncoding regions of the genome and can thus interfere with the composition and dosage of regulatory elements, but the effects of such alterations are poorly understood.

We investigated the effects of deletions and duplications upstream of *IHH*, a master gene of skeletal development involved in chondrocyte differentiation, joint formation and osteoblast differentiation. Accordingly, *Ihh* inactivation in mice results in extreme shortening of bones, joint fusions and almost absent ossification, ultimately causing early lethality[11]. Interestingly, patients carrying duplications at this locus display completely different phenotypes, including craniosynostosis, syndactyly and polydactyly[1,2], indicating alternative pathomechanisms. To define the regulatory landscape of *Ihh*, we performed circular chromosome conformation capture and sequencing (4C–seq) in embryonic day (E) 14.5 developing limbs and compared our findings to published data sets[12]. Our data show that the *Ihh* promoter interacts preferentially with the third intron of the upstream neighboring gene *Nhej1* (**Fig. 1** and **Supplementary Fig. 1**), in a genomic region affected in all reported disease-associated duplications. The region contains multiple sites positive for H3K4me1 and H3K27ac (indicative of active enhancers) and binding sites for CTCF, an architectural protein involved in facilitating enhancer–promoter contact by looping. The convergent CTCF motif orientation observed across the locus might facilitate the interactions measured in the 4C–seq experiments (**Fig. 1** and **Supplementary Fig. 2**)[13–16].

In mice in which a *lacZ* reporter cassette (*Sleeping Beauty*)[17] was inserted to capture the regulatory capacity of the region, a pattern consistent with *Ihh* expression was observed, that is, activity in condensing digits, growth plates, fingertips and skull sutures. Using a combination of H3K27ac and H3K4me1 ChIP–seq signal in E14.5 limbs[18], evolutionary conservation[19] and our 4C–seq interaction profiles, we

[1]Max Planck Institute for Molecular Genetics, RG Development and Disease, Berlin, Germany. [2]Institute for Medical and Human Genetics, Charité–Universitätsmedizin Berlin, Berlin, Germany. [3]MS 84-171, Lawrence Berkeley National Laboratory, Berkeley, California, USA. [4]Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité–Universitätsmedizin Berlin, Berlin, Germany. [5]Department of Developmental Genetics, Max Planck Institute for Molecular Genetics, Berlin, Germany. [6]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany. [7]Genome Dynamics in the Immune System Laboratory, INSERM, UMR 1163, Institut Imagine, Université Paris Descartes, Sorbonne Paris Cité, Paris, France. [8]Institute of Human Genetics, Biocentre, Julius Maximilians University Würzburg, Würzburg, Germany. [9]US Department of Energy Joint Genome Institute, Walnut Creek, California, USA. [10]School of Natural Sciences, University of California, Merced, California, USA. [11]These authors jointly directed this work. Correspondence should be addressed to S.M. (mundlos@molgen.mpg.de) or D.G.L. (lupianez@molgen.mpg.de).

Received 21 February; accepted 28 July; published online 28 August 2017; doi:10.1038/ng.3939

defined nine regions with enhancer potential and validated them in mouse transgenic enhancer activity assays[20] (**Fig. 1**). Embryos were analyzed at two time points, E14.5 and E17.5, to capture *Ihh* expression domains during digit development (fingertips and cartilage anlagen) and bone growth (skull sutures and growth plates), respectively. Five of the tested elements showed activity at both stages (**Fig. 1**), whereas additional elements were active only at E17.5 (**Supplementary Fig. 3**). We scored the activity of each element in the previously identified

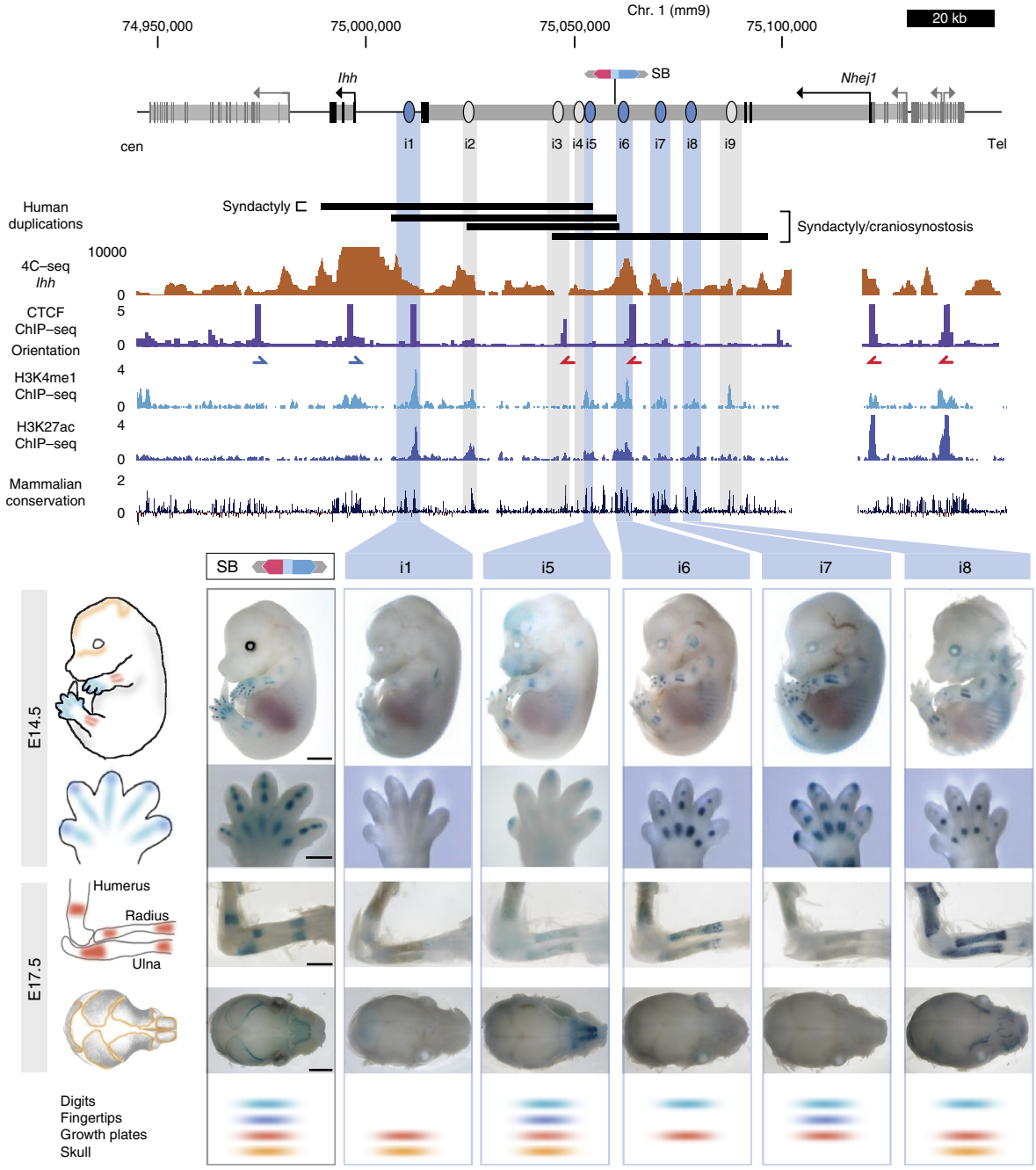

Figure 1 A cluster of enhancers interacts with the *Ihh* promoter during mouse development. Top, close-up view of the *Ihh* genomic region. Genes and their transcription start sites are indicated: black boxes, exons; gray boxes, introns. The position of the *lacZ* reporter insertion is shown (SB). Black bars indicate the size and position of previously described human duplications[1,2] converted to mouse genome coordinates. Findings from 4C–seq performed in E14.5 limbs using the *Ihh* promoter as the viewpoint are shown below. Note increased interactions with intron 3 of the adjacent *Nhej1* gene (see also **Supplementary Fig. 1**). The results of CTCF ChIP–seq performed in E14.5 limbs are shown (ENCODE)[3], where blue and red arrows indicate motif orientation. Additional tracks below show H3K4me1 and H3K27ac, as well as sequence conservation. This information was used to predict enhancers i1–i9, indicated by light blue and gray bars. Bottom, transgenic reporter assay (LacZ) of elements positive at E14.5 and E17.5 (marked in light blue; for each enhancer, an embryo and handplate at E14.5 and a dorsal view of a forelimb and a top view of the skull at E17.5 are shown). The regulatory activity of the region, as indicated by the activity of the inserted *lacZ* reporter (SB; black outline), is shown on the left. The lower panel shows scoring of each element for tissue specificity. Elements negative at E14.5 but with positive staining at E17.5 are marked in gray and shown in **Supplementary Figure 3**. Scale bars: 2,000 μm (embryos and skulls), 500 μm (handplates) and 1,000 μm (forelimbs).

regions (**Fig. 1**, **Supplementary Fig. 3** and **Supplementary Table 1**). This analysis highlighted the inherent complexity of the cluster, where almost every individual element displayed a unique pattern of activity. All elements gave a positive signal in growth plates, whereas other domains, like fingertips, were covered only by a small subset of enhancers (i5 and i7). This suggests that the enhancers in this cluster act in a modular fashion and that the degree of overlapping activity varies between tissues and developmental time points.

To evaluate the functionality of these elements, we deleted intron 3 of *Nhej1* (**Fig. 2**), which contains eight of the nine enhancers identified, using CRISVar[21]. *Nhej1* encodes a DNA repair protein essential for the non-homologous end-joining pathway, required for double-strand break repair. In humans, homozygous mutations in *NHEJ1* result in severe combined immunodeficiency (SCID) with microcephaly, growth retardation and sensitivity to ionizing radiation, reflecting a deficiency in DNA repair (MIM 611291)[22]. In contrast, *Nhej1*-knockout mice are viable and do not display any morphological phenotype[23,24]. µCT scans of *Nhej1*$^{-/-}$ skulls did not identify any abnormalities, indicating that *Nhej1* does not have a major role in skull and suture development (**Supplementary Fig. 4**). Mice homozygous for the *Nhej1* intronic deletion (Del(2–9)) displayed very short limbs, absent cortical bone and fused joints, as well as reduced skull ossification, very similar to the phenotypes observed upon *Ihh* inactivation[11]. Whereas *Nhej1* transcription levels remained basically unchanged (**Supplementary Fig. 5**), we observed a drastic reduction in *Ihh* mRNA levels in E13.5 limbs and E17.5 skulls (98% and 99% reduction, respectively), consistent with the observed phenotypes. Therefore, this genomic region contains most of the regulatory elements required for *Ihh* skeletal expression.

Next, we generated a series of specific deletions to assess the functional redundancy within this enhancer cluster (**Fig. 2**). Homozygous deletion of the enhancers located in the telomeric part of the intron (Del(4–9)) resulted in a lethal growth defect almost as severe as that observed with deletion of the entire intron, confirming that the most relevant enhancers are located in this telomeric region. Deletion of only the three central enhancers (Del(4–6)) reduced *Ihh* expression by approximately 70% in all tissues tested, whereas deletion of the three more telomeric enhancers (Del(7–9)) resulted in a 60% reduction in expression (**Fig. 2**). Both mutants were viable and phenotypically normal, but they showed a delay in skull ossification (**Fig. 2**) and a 10% reduction in bone length (**Supplementary Fig. 6**). All deletions except Del(7–9) resulted in loss of *Ihh* fingertip expression, indicating

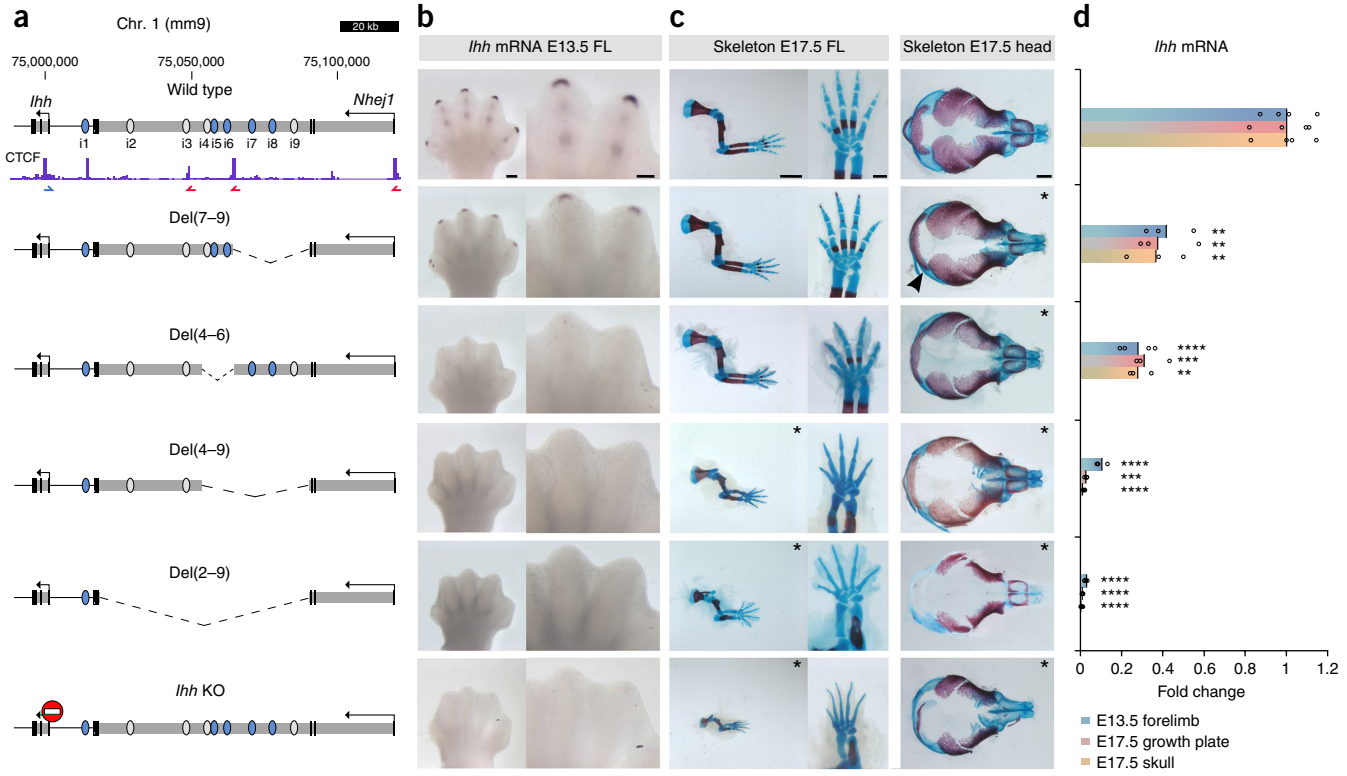

Figure 2 Deletions of regulatory elements highlight additive control of *Ihh* expression. (**a**) Deletions generated by CRISVar[21] at the *Ihh* locus. *Ihh* knockout (KO) is shown for comparison (stop sign). Findings from CTCF ChIP–seq performed in E14.5 limbs are shown (ENCODE)[3], where blue and red arrows indicate motif orientation. Each deleted chromosomal region is represented as a dashed line. Note that Del(4–9) and Del(7–9) delete only one intronic CTCF-binding site, maintaining another intact. (**b**) *In situ* hybridization shows *Ihh* expression in handplates (E13.5). Note expression in digit tips and condensing digits in wild-type embryo and loss of expression in all deletions encompassing enhancer i5. FL, forelimb. Scale bars, 200 µm (handplates). (**c**) Skeletal staining of forelimbs, autopods and skulls (E17.5). Mutants displaying abnormal phenotypes are indicated by asterisks. Both Del(2–9) and Del(4–9) result in massive reduction of limb size and reduced ossification similar to *Ihh* knockout, whereas Del(4–6) and Del(7–9) mice did not show noticeable limb abnormalities. All mutants studied displayed skull defects (delayed ossification), an effect that was less prominent in Del(7–9) mutants (arrowhead). Scale bars: 2,000 µm (forelimbs), 500 µm (autopods) and 1,000 µm (skulls). (**d**) *Ihh* qPCR analysis in E13.5 forelimb, E17.5 growth plate (elbow) and skull. Deletion of intron 3 of *Nhej1* encompassing enhancers i2–i9 results in almost complete loss of *Ihh* expression in all tissues, whereas smaller deletions partially reduce expression. Bars represent the mean of $n \geq 3$ different individuals (circles). Two-sided Student's *t* test, $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

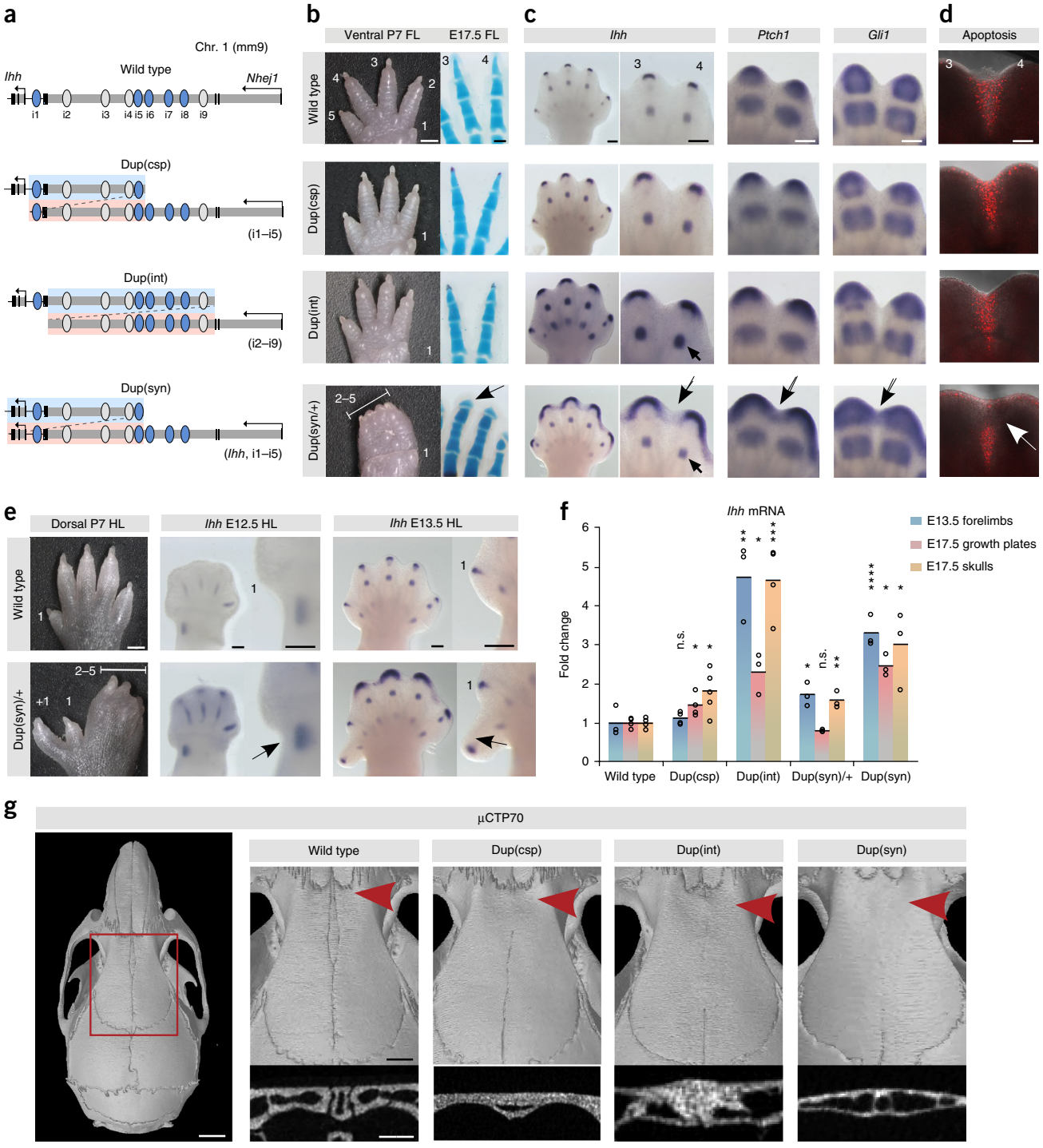

Figure 3 Duplications of enhancer elements result in *Ihh* over- and misexpression. (**a**) Duplications generated by CRISVar[21]. The duplicated fragments are shown in blue and pink. (**b**) Left, forelimb morphology (postnatal day (P) 7). Dup(int) and Dup(csp) mice (homozygous) are normal, but Dup(syn)/+ mice display 2/5 syndactyly. Right, skeletal staining at E17.5. Staining shows short and broad terminal phalanges in Dup(syn)/+ mice (arrow). Scale bars: 1,000 µm (P7) and 200 µm (E17.5 handplates). (**c**) *In situ* hybridization shows increased and broadened expression of *Ihh* and its downstream effectors *Ptch1* and *Gli1* in the mutants (Dup(csp) < Dup(int) < Dup(syn)/+). Expression domains in Dup(syn)/+ mice extend into distal interdigital space (arrows). Note the increased *Ihh* expression in digit condensations in Dup(int) as compared to Dup(syn)/+ mice (small arrow), also observed across the entire handplate. Scale bars, 200 µm. (**d**) Apoptosis in interdigital space (red signal). Note the lack of signal in the distal region in the Dup(syn)/+ mutant (arrow). Scale bar, 200 µm. (**e**) Hindlimb morphology of Dup(syn)/+ mice. Note the preaxial polydactyly and syndactyly 2/5. *In situ* hybridization shows increased *Ihh* expression (arrows) in the preaxial region (insets). HL, hindlimb. Scale bars: 1,000 µm (P7) and 200 µm (E12.5 and E13.5). (**f**) *Ihh* qPCR analysis. Duplications increase *Ihh* expression. High levels in Dup(csp) forelimbs (no phenotype) result from digit condensations, while moderate upregulation in Dup(syn)/+ forelimbs (syndactyly) derives from fingertips. Bars represent the mean of $n \geq 3$ different individuals (circles). Two-sided Student's *t* test, $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$; ns, not significant. (**g**) µCT skull analysis (P70). The red square indicates enlargement of the metopic suture region (right). Below, cross-sections of metopic sutures (red arrowheads). All mutants display complete suture fusion (maximum effect in Dup(int)). Scale bars: 2 mm (skull), 1 mm (enlargement), 0.5 µm (cross-section).

that element i5 acts as a major regulator for this region. These results demonstrate that *Ihh* expression is controlled by a cluster of redundant enhancers, which appear to act in an additive manner.

To understand the mechanisms underlying pathogenic duplications in the *IHH* locus, we duplicated the entire *Nhej1* intron (Dup(int)), equivalent to the sequence deleted in Del(2–9). In addition, we reengineered two of the previously described human duplications: Dup(csp), encompassing the region between enhancers i1 and i5 (reengineered human duplication causing craniosynostosis Philadelphia type[1,2]), and Dup(syn), which includes *Ihh* and the upstream region up to enhancer i5 (reengineered human duplication causing syndactyly Lueken type[2]) (**Fig. 3a**). Dup(int) and Dup(csp) mutants did not show gross morphological alterations in the heterozygous or homozygous state. In contrast, Dup(syn)/+ mice showed complete cutaneous syndactyly of digits 2/5 in fore- and hindlimbs (**Fig. 3b**), thus recapitulating the human phenotype.

Skeletal staining showed that the syndactyly of Dup(syn) mutants did not involve bony fusions. Digits and joints developed normally, but terminal phalanges were broad and short. *In situ* hybridization experiments in E13.5 limbs identified major changes in fingertips, where *Ihh* expression was not only increased but also broadened. These effects were weak in Dup(csp) mice, more pronounced in Dup(int) mice and most prominent in Dup(syn)/+ mice, in which *Ihh* expression extended into the distal interdigital space (**Fig. 3c**). Accordingly, the expression domains of the hedgehog downstream targets *Gli1* and *Ptch1* were broadened, and fusion of the normally separate domains was observed that was most pronounced in Dup(syn)/+ mutants. Except for *Bmp4* and *Nog*, we did not observe abnormalities in other genes or pathways involved in syndactyly and interdigital cell death (**Supplementary Fig. 7**), suggesting that hedgehog signaling alone is sufficient to induce this type of syndactyly. Next, we quantified interdigital apoptosis, which is required for digit separation[25]. Consistent with the observed phenotypes, we detected strong signal in the interdigital space in wild-type, Dup(csp) and Dup(int) embryos, but an absence of signal in the distal region in Dup(syn)/+ embryos (**Fig. 3d**). Thus, upregulation and misexpression of *Ihh* in fingertips beyond a certain threshold resulted in abnormalities of the distal phalanges, most likely by interfering with the phalanx-forming region[26], and syndactyly due to suppression of interdigital apoptosis.

In addition, Dup(syn) mutants displayed preaxial polydactyly on hindlimbs (50% penetrance; **Fig. 3e**). One major cause of polydactyly is ectopic activation of hedgehog signaling at the anterior developing limb bud[27,28]. Interestingly, Dup(syn)/+ embryos showed a prominent increase in *Ihh* expression in the distal zeugopod during hindlimb development starting at E12.5 (expression was absent at E10.5 and E11.5). As IHH is a potent diffusible morphogen, we hypothesize that the increased expression might interfere with the anterior–posterior hedgehog gradient. Thus, the observed phenotype seems to be the result of a loss of precision in spatiotemporal expression, indicating that, similar to the syndactyly, an increase in enhancer dosage can have site-specific effects.

Expression profiling by quantitative RT–PCR (qPCR) was used to quantify the effect of the duplications on gene expression (**Fig. 3f**). Whereas *Nhej1* and other nearby genes showed no alteration in expression (**Supplementary Fig. 5**), all mutants analyzed displayed increased *Ihh* expression in the skull and limbs, with the highest expression levels observed in Dup(int) mutants (up to fivefold upregulation). *In situ* hybridization of Dup(int) forelimb autopods (**Fig. 3c**) showed increased expression mainly in digits, whereas in Dup(syn) mutants the expression increase was most prominent in fingertips, consistent with the syndactyly observed. To investigate the effect of increased *Ihh* expression on skull development and suture formation, a detailed μCT analysis was performed (**Fig. 3g**). This analysis identified fusion of the metopic suture (craniosynostosis) in all mutants, but this phenotype was most pronounced in Dup(int) mice. The phenotypes observed in our mouse mutants (syndactyly, polydactyly and craniosynostosis) accurately recapitulate previous observations in human patients[1,2] (**Supplementary Fig. 8**). Thus, the induced changes in enhancer composition and dosage resulted in a disturbance of the levels and precision of gene expression, thereby causing abnormal development and disease. Interestingly, the observed phenotypes did not always correlate with the number of duplicated elements but appeared to be influenced by other factors such as the position of the duplication and the arrangement of individual elements relative to the cluster.

To investigate a possible effect of spatial configuration on the duplicated alleles, we performed 4C–seq experiments in E14.5 limbs (viewpoint at *Ihh*; **Fig. 4a**). In Dup(int)/+ mutants (with duplication of enhancers i2–i9), we observed increased interactions across the entire duplicated region. In contrast, Dup(syn)/+ mutants (with duplication of *Ihh* and enhancers i1–i5) only showed increased contact with the centromeric region of the enhancer cluster, suggesting that the centromeric *Ihh* copy created its own regulatory domain containing only the duplicated regulatory elements i1–i5 (**Fig. 4b**). The presence of two divergently oriented CTCF-binding sites near the promoter of the telomeric *Ihh* copy might explain this domain separation by limiting chromatin interaction beyond these elements. Moreover, the larger contact areas in Dup(int)/+ mutants correlate with the observed levels of *Ihh* upregulation as compared to Dup(syn)/+ mutants. As illustrated in **Figure 4c**, the syndactyly in Dup(syn)/+ mice is likely due to two types of interactions between the major fingertip enhancer i5 and the two copies of *Ihh*: one type involves long-range interactions and the other the presence of the i5 enhancer in direct proximity to *Ihh*. Together, these interactions result in localized upregulation of *Ihh* expression in the fingertips. Increased expression mediated by disconnection from a repressor element is unlikely, as none of the deletions studied resulted in any observable upregulation of *Ihh*. To further evaluate whether the observed limb phenotypes in the *Ihh*-containing duplication (Dup(syn)) merely corresponded to a gene-dosage effect, we crossed Dup(syn)/+ mice with $Ihh^{+/-}$ mice or with mice lacking the enhancer cluster (Del(2–9) mice). In both cases, double-heterozygous mice displayed the same syndactyly and polydactyly as was observed in Dup(syn)/+ mice (**Supplementary Fig. 9**), indicating that misexpression was due to the specific, partially duplicated regulatory landscape.

Our study shows that a multipartite enhancer ensemble regulates *Ihh* expression in fingertips, digit condensations, growth plates and skull sutures. The described functional redundancy appears to be a common phenomenon of these types of enhancers, as was recently shown for the α-globin and *Wap* super-enhancers[29,30]. At the *Ihh* locus, we observed a complex scenario, as not all enhancers displayed the same combination of expression domains, a phenomenon also described for the HoxD cluster and *Fgf8* (refs. 31,32). This modular nature and, in particular, correct dosage appear critical in conferring the required precision of gene expression. This is supported by our finding that an increase in enhancer number resulted in an increase in gene expression. However, this effect was site specific and dependent not only on enhancer number but also on enhancer position. CNVs, and in particular duplications, may affect this delicate balance, thereby causing over- and/or misexpression resulting in disease. The reported duplications do not interfere with topologically associating domain (TAD) boundaries, as reported at

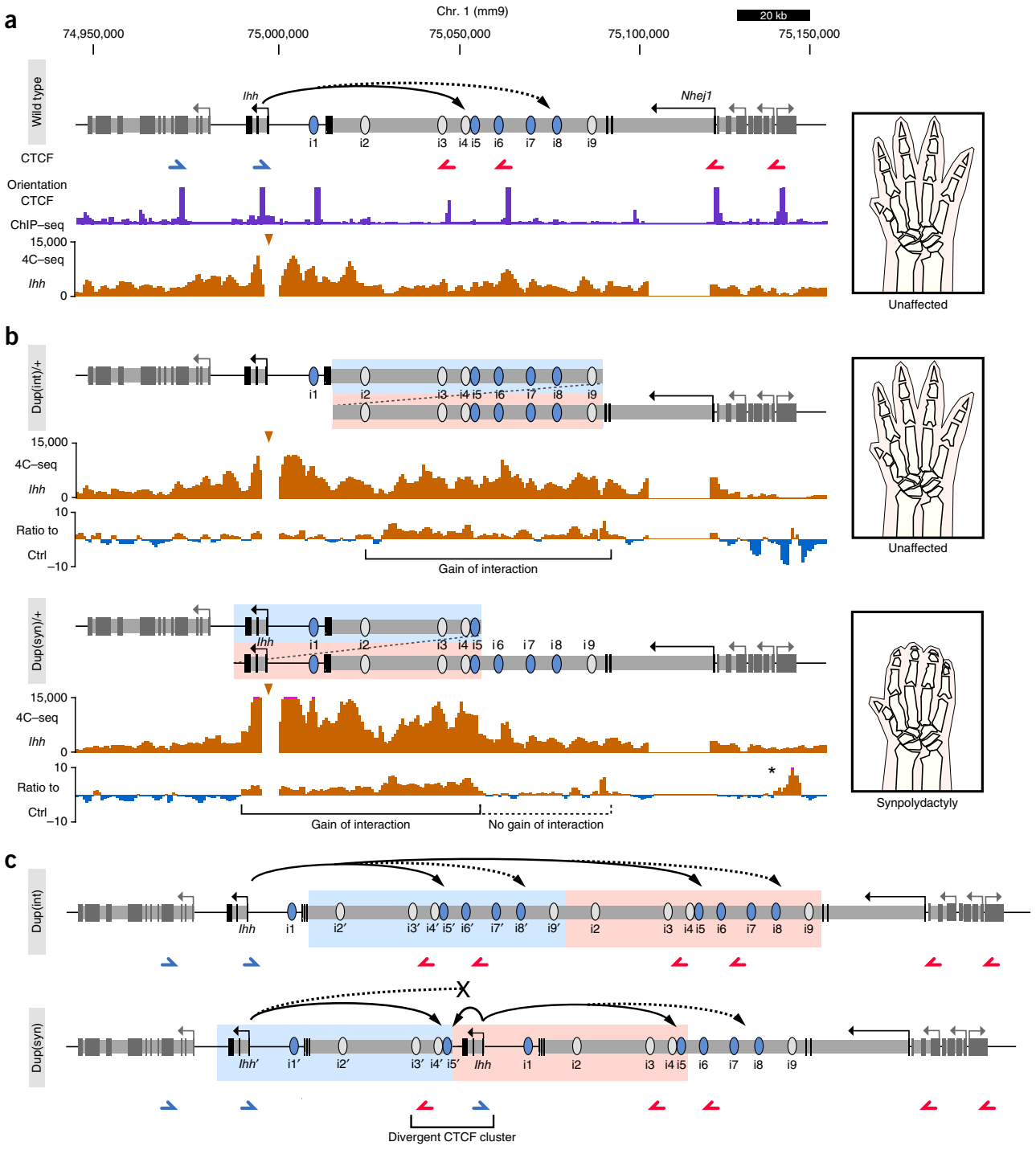

Figure 4 4C–seq identifies specific regulatory configurations in duplications. (**a**) Schematic of the wild-type *Ihh* locus. The continuous arrow indicates interaction between *Ihh* and enhancers i4–i6, while the discontinuous arrow indicates interaction of *Ihh* with i7–i9. CTCF ChIP–seq findings (E14.5 limbs; ENCODE)[3] are shown below, where blue and red arrows indicate motif orientation. 4C–seq results (viewpoint at the *Ihh* promoter) show the interaction profile in E14.5 handplates. A schematics of limb morphology is shown on the right. (**b**) Duplication of intron 3 of *Nhej1* (Dup(int); top) and a reengineered human duplication causing syndactyly (Dup(syn); bottom). Duplicated regions are shown in blue and pink. The 4C–seq profile (viewpoint at the *Ihh* promoter) and the ratio to wild-type E14.5 handplates control are shown for each duplication. Brackets indicate regions with gain of interaction. Note that there is no gain of interaction with the region containing enhancers i6–i9 (dashed bracket) in Dup(syn)/+ mice, indicating that the duplicated *Ihh* copy does not interact with this region. Observed phenotypes are schematically shown on the right. An asterisk indicates increased interactions with the *Cnppd1* and *Fam134a* genes, which do not have functional consequences (**Supplementary Fig. 5**). (**c**) Model of the regulatory interactions of the duplicated alleles. In Dup(int), *Ihh* can interact with the entire duplicated landscape, including two copies of the main digit enhancer (i5). In Dup(syn), both *Ihh* copies interact with a downstream copy of the i5 enhancer (long continuous arrows) but only the telomeric *Ihh* copy has access to i7–i9 (discontinuous arrows) because of the presence of the divergent CTCF cluster (bracket). Additionally, the duplicated i5 enhancer interacts with the telomeric *Ihh* copy because of its genomic proximity (short continuous arrow). Duplicated regions are shown in blue and pink.

the *Epha4* and *Sox9* loci[33,34], thus highlighting alternative mechanisms that should be considered when interpreting genomic duplications. Our study demonstrates the importance of analyzing regulatory elements in the complex setting of their native genomic environment, as reductionist approaches relying on reporter assays and deletions of individual enhancers insufficiently capture the multifaceted redundant and complementary functions of enhancer clusters.

URLs. FIMO, http://meme-suite.org/tools/fimo; JASPAR database, http://jaspar.binf.ku.dk/; Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo/.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the sequencing core, transgenic unit and animal facilities of the Max Planck Institute for Molecular Genetics for technical assistance. This study was supported by a grant from the Deutsche Forschungsgemeinschaft to S.M. and E.K. S.M. was supported by the European Community's Seventh Framework Programme, grant agreement 602300 (SYBIL). M.O. was supported by a Swiss National Science Foundation (SNSF) fellowship. A.V. was supported by National Institutes of Health grants R01HG003988, U54HG006997, U01DE024427 and U01DE024427.

AUTHOR CONTRIBUTIONS

A.J.W., D.G.L. and S.M. conceived the study and designed the experiments. M.O. and A.V. performed LacZ experiments and analysis of individual enhancers, and A.J.W. performed analysis of *Sleeping Beauty* insertion. A.J.W. and L.W. generated transgenic mouse models. W.-L.C., J.P.d.V. and G.C. contributed to histological analysis. A.J.W., N.B. and G.C. performed qPCR, *in situ* hybridization and phenotype analysis. G.C., N.B. and W.-L.C. provided technical support. E.K., M.O. and A.V. contributed to scientific discussion. A.J.W. performed 4C–seq experiments, and V.H., M.V. and D.G.L. performed bioinformatic analysis. A.J.W., D.G.L. and S.M. wrote the manuscript with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Barroso, E. *et al.* Identification of the fourth duplication of upstream *IHH* regulatory elements, in a family with craniosynostosis Philadelphia type, helps to define the phenotypic characterization of these regulatory elements. *Am. J. Med. Genet. A.* **167A**, 902–906 (2015).
2. Klopocki, E. *et al.* Copy-number variations involving the *IHH* locus are associated with syndactyly and craniosynostosis. *Am. J. Hum. Genet.* **88**, 70–75 (2011).
3. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
4. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
5. Bernstein, B.E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
6. Hong, J.W., Hendrix, D.A. & Levine, M.S. Shadow enhancers as a source of evolutionary novelty. *Science* **321**, 1314 (2008).
7. Perry, M.W., Boettiger, A.N., Bothma, J.P. & Levine, M. Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr. Biol.* **20**, 1562–1567 (2010).
8. Dunipace, L., Ozdemir, A. & Stathopoulos, A. Complex interactions between *cis*-regulatory modules in native conformation are critical for *Drosophilasnail* expression. *Development* **138**, 4075–4084 (2011).
9. Cannavò, E. *et al.* Shadow enhancers are pervasive features of developmental regulatory networks. *Curr. Biol.* **26**, 38–51 (2016).
10. Perry, M.W., Bothma, J.P., Luu, R.D. & Levine, M. Precision of hunchback expression in the *Drosophila* embryo. *Curr. Biol.* **22**, 2247–2252 (2012).
11. St-Jacques, B., Hammerschmidt, M. & McMahon, A.P. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes Dev.* **13**, 2072–2086 (1999).
12. 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).
13. Guo, Y. *et al.* CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015).
14. 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).
15. Sanborn, A.L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. USA* **112**, E6456–E6465 (2015).
16. Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Reports* **10**, 1297–1309 (2015).
17. 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).
18. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* **107**, 21931–21936 (2010).
19. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
20. 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).
21. Kraft, K. *et al.* Deletions, inversions, duplications: engineering of structural variants using CRISPR/Cas in mice. *Cell Rep.* http://dx.doi.org/10.1016/j.celrep.2015.01.016 (2015).
22. Buck, D. *et al.* Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* **124**, 287–299 (2006).
23. Li, G. *et al.* Lymphocyte-specific compensation for XLF/cernunnos end-joining functions in V(D)J recombination. *Mol. Cell* **31**, 631–640 (2008).
24. Vera, G. *et al.* Cernunnos deficiency reduces thymocyte life span and alters the T cell repertoire in mice and humans. *Mol. Cell. Biol.* **33**, 701–711 (2013).
25. Hernández-Martínez, R., Castro-Obregón, S. & Covarrubias, L. Progressive interdigital cell death: regulation by the antagonistic interaction between fibroblast growth factor 8 and retinoic acid. *Development* **136**, 3669–3678 (2009).
26. Witte, F., Chan, D., Economides, A.N., Mundlos, S. & Stricker, S. Receptor tyrosine kinase–like orphan receptor 2 (ROR2) and Indian hedgehog regulate digit outgrowth mediated by the phalanx-forming region. *Proc. Natl. Acad. Sci. USA* **107**, 14211–14216 (2010).
27. Anderson, E., Peluso, S., Lettice, L.A. & Hill, R.E. Human limb abnormalities caused by disruption of hedgehog signaling. *Trends Genet.* **28**, 364–373 (2012).
28. Hill, R.E. How to make a zone of polarizing activity: insights into limb development via the abnormality preaxial polydactyly. *Dev. Growth Differ.* **49**, 439–448 (2007).
29. Hay, D. *et al.* Genetic dissection of the α-globin super-enhancer *in vivo*. *Nat. Genet.* **48**, 895–903 (2016).
30. Shin, H.Y. *et al.* Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat. Genet.* **48**, 904–911 (2016).
31. Marinić, M., Aktas, T., Ruf, S. & Spitz, F. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev. Cell* **24**, 530–542 (2013).
32. Montavon, T. *et al.* A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
33. Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).
34. Lupiáñez, D.G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene–enhancer interactions. *Cell* **161**, 1012–1025 (2015).

ONLINE METHODS

Experimental design. No statistical methods were used to predetermine sample size. All experiments and analyses were performed using samples from at least three different animals and were repeated at least two times in the laboratory. Samples/animals were included or excluded according to genotype by PCR. Experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

ES cell targeting and transgenic mouse strains. Mouse embryonic stem (ES) cell culture was performed as described previously[21]. ES and feeder cells were tested for mycoplasma contamination using a Mycoalert detection kit (Lonza) and the Mycoalert assay control set (Lonza).

Duplications and deletions were generated in G4 ES cells (129/Sv × C57BL/6 F_1 hybrid) using CRISVar as described previously[21]. Target regions, sizes and guide sequences are listed in **Supplementary Table 2**. Embryos and live animals derived from ES cells were generated by diploid or tetraploid complementation[35]. Genotyping was performed by PCR analysis.

A *Sleeping Beauty* (SB) cassette[17] was inserted in G4 ES cells at the center of the third intron of the *Nhej1* gene (chr. 1: 75,060,87; mm9), by homologous recombination using standard protocols[36]. The *Sleeping Beauty* transgene carries a single *lacZ* reporter gene with a minimal human β-globin promoter and a neomycin-resistance cassette, flanked by transposable elements. Coordinates and primer sequences for amplifying homology sequences are provided in **Supplementary Table 3**. Positive ES cell clones were injected into donor blastocysts to generate chimeras. The neomycin-resistance cassette was removed by crossing chimaeric animals with a Flpe-deleter line. Genotyping was performed by PCR analysis.

Mouse strains were maintained by crossing the strains with C57BL/6J mice. All animal procedures were conducted as approved by the local authorities (LAGeSo Berlin) under license numbers G0368/08 and G0247/13.

***In vivo* enhancer validation.** Putative enhancer regions were amplified by PCR from mouse genomic DNA and cloned into a *Hsp68* promoter–*lacZ* reporter vector as previously described[20] (**Supplementary Table 4**). Transgenic embryos were generated and tested for LacZ reporter activity at E14.5 and E17.5. All animal work performed at the Lawrence Berkeley National Laboratory was reviewed and approved by the institutional Animal Welfare and Research Committee (AWRC). Sample sizes were selected empirically on the basis of our previous experience of performing transgenic mouse assays for >2,000 total putative enhancers. A summary of all transgenic mice can be found in **Supplementary Table 1**. As all transgenic mice were treated with identical experimental conditions, and as there were no groups of animals directly compared in this section of the study, randomization and experimenter blinding were unnecessary and were not performed.

Quantitative real-time PCR. Handplates (E13.5), forelimb and hindlimb growth plates (E17.5) and cranium (E17.5) were dissected from wild-type and mutant embryos ($n \geq 3$) in ice-cold PBS/DEPC and immediately frozen in liquid nitrogen. RNA isolation was performed using the RNeasy kit (Qiagen), and cDNA was transcribed using the TaqMan Reverse Transcription kit (Roche) according to the specifications of the manufacturer. qPCR was performed using SYBR Green (Qiagen) on an ABI Prism HT 7900 Real-Time Cycler. *GAPDH* was used as an internal control, and fold changes were calculated by relative quantification ($2^{-\Delta\Delta Ct}$). Primers are summarized in **Supplementary Table 5**.

4C–seq. 4C–seq libraries were generated from microdissected E14.5 mouse forelimb tissue (digits 2–5) as described previously[37]. The starting material for all 4C–seq libraries was 5×10^6 to 1×10^7 cells. All 4C–seq experiments were carried out in heterozygous animals, and results were compared to those in wild-type controls. 4-bp cutters were used as primary (Csp6I) and secondary (BfaI) restriction enzymes. A total of 1 to 1.6 µg of DNA was amplified by PCR (primer sequences in **Supplementary Table 6**). All samples were sequenced with Illumina HiSeq technology according to standard protocols. 4C–seq experiments were carried out in two biological replicates in wild-type, Dup(int) and Dup(syn)/+ mutants. A representative result is shown in **Figure 4**.

For 4C–seq data analysis, reads were preprocessed and mapped to a corresponding reference (mm9) using BWA-MEM[38]; coverage was normalized as reported previously[34]. The viewpoint and adjacent fragments 1.5 kb up- and downstream were removed, and a window of two fragments was chosen to normalize the data per million mapped reads (RPM). To compare the interaction profiles of different samples, we obtained the $\log_2$-transformed fold change for each window of normalized reads. To obtain ratios, duplicated regions were excluded for calculation of the scaling parameter used in RPM normalization. Code is available upon request.

CTCF motif orientation analysis. Orientation of the motifs within conserved CTCF peaks was obtained using FIMO (see URLs) with standard parameters[39]. The CTCF motif[40] was obtained from the JASPAR database (see URLs).

Phenotypic analysis. Phenotypic analysis for mutant mouse lines was carried out for at least three animals per analysis and developmental stage (E17.5, P7 and P70), in homo- and heterozygous animals. The penetrance of phenotypes was determined by analyzing $n > 20$ animals, and a genotype was considered fully penetrant if all mutants were similarly affected.

Microcomputer tomography. Skulls and autopods of control and mutant mice ($n > 3$) were scanned using a Skyscan 1172 X-ray microtomography system (Brucker microCT, Belgium) at 10 µm resolution. 3D model reconstruction and length measurements were performed with the Skyscan image analysis software CT-Analyser and CT-volume (Brucker microCT, Belgium). Cross-sections were performed at 10 µm resolution. Relative length was determined relative to wild-type controls.

Whole-mount *in situ* hybridization and skeletal preparations. Whole-mount *in situ* hybridization was performed in wild-type and mutant E13.5 embryos ($n = 4$) according to standard procedures. All probes were generated by PCR amplification using mouse limb bud cDNA. For skeletal preparations, wild-type and mutant E17.5 embryos ($n = 4$) were stained with Alcian blue/Alizarin red according to standard protocols.

LacZ staining. E14.5 and E17.5 mouse embryos ($n > 5$) were dissected in cold PBS, fixed in 4% paraformaldehyde (PFA)/PBS on ice for 30 min, washed twice with ice-cold PBS, washed once at room temperature (19–24 °C) and then stained overnight for β-galactosidase activity in a humid chamber at 37 °C as previously described[17]. After staining, embryos were washed in PBS and stored at 4 °C in 4% PFA/PBS.

Statistical analyses. Results are presented as the mean ± s.d. of at least three independent biological replicates ($n \geq 3$). Statistical differences between the means were examined by two-sided Student's *t* test. $P < 0.05$ was considered statistically significant. A prespecified effect size was not defined.

Code availability. Custom computer codes used to generate results reported in the manuscript will be made available upon request.

Data availability. Sequencing data are available from the Gene Expression Omnibus under accession GSE95062.

35. 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).
36. Hooper, M., Hardy, K., Handyside, A., Hunter, S. & Monk, M. HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**, 292–295 (1987).
37. van de Werken, H.J. *et al.* 4C technology: protocols and data analysis. *Methods Enzymol.* **513**, 89–112 (2012).
38. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
39. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
40. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).

doi:10.1038/ng.3939