Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions

Highlights

- Disruptions of TADs lead to de novo enhancer-promoter interactions and misexpression
- Misexpression occurs when CTCF-associated TAD boundary elements are disrupted
- Structural variations disrupting TAD structures can cause malformation syndromes
- Different phenotypes can result from one enhancer acting on different target genes

In Brief

Disease-associated structural variants, when affecting CTCF-associated boundary elements, cause pathogenicity by disrupting the structure of topologically associated chromatin domains leading to ectopic promoter interactions and altered gene expression.

Accession Numbers

GSE66383

Authors

Dario G. Lupiáñez, Katerina Kraft, ..., Axel Visel, Stefan Mundlos

Correspondence

mundlos@molgen.mpg.de
Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions

Dario G. Lupiñez,1,2 Katerina Kraft,1,2 Verena Heinrich,2 Peter Krawitz,1,2 Francesco Brancati,3 Eva Klopopki,4 Denise Horn,2 Hülya Kaysери́li,6 John M. Opitz,6 Renata Laxova,6 Fernando Santos-Simarro,7,8 Brigitte Gilbert-Dussardier,9 Lars Wittler,10 Marina Borschwer,1 Stefan A. Haas,11 Marco Osterwalder,12 Martin Franke,1,2 Bernd Timmermann,13 Jochen Hecht,1,14 Malte Spielmann,1,2,14 Axel Visel,12,15,16 and Stefan Mundlos1,2,14,*

1Max Planck Institute for Molecular Genetics, RG Development & Disease, 14195 Berlin, Germany 2Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin, 13353 Berlin, Germany 3Medical Genetics Unit, Policlinico Tor Vergata University Hospital, 00133 Rome, Italy 4Institute of Human Genetics Biozentrum, Julius Maximilian University of Würzburg, 97070 Würzburg, Germany 5Medical Genetics Department, Istanbul Medical Faculty, Istanbul University, 34093 Istanbul, Turkey 6Department of Pediatrics, School of Medicine, University of Utah, Salt Lake City, UT 84108, USA 7Instituto de Genética Médica y Molecular (INGEMM), IdiPAZ, Hospital Universitario La Paz, 28046 Madrid, Spain 8U753 Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, 28046 Madrid, Spain 9Service de Génétique, C.H.U. de Poitiers, 86021 Poitiers, France 10Department Developmental Genetics, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany 11Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany 12Genomics Division, MS 84–171, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA 13Max Planck Institute for Molecular Genetics, Sequencing Core Facility, 14195 Berlin, Germany 14Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité Universitätsmedizin Berlin, 13353 Berlin, Germany 15U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA 16School of Natural Sciences, University of California, Merced, CA 95343, USA
*Correspondence: mundlos@molgen.mpg.de
http://dx.doi.org/10.1016/j.cell.2015.04.004

SUMMARY

Mammalian genomes are organized into megabase-scale topologically associated domains (TADs). We demonstrate that disruption of TADs can rewire long-range regulatory architecture and result in pathogenic phenotypes. We show that distinct human limb malformations are caused by deletions, inversions, or duplications altering the structure of the TAD-spanning WNT6/IHH/EPHA4/PAX3 locus. Using CRISPR/Cas genome editing, we generated mice with corresponding rearrangements. Both in mouse limb tissue and patient-derived fibroblasts, disease-relevant structural changes cause ectopic interactions between promoters and non-coding DNA, and a cluster of limb enhancers normally associated with Epha4 is misplaced relative to TAD boundaries and drives ectopic limb expression of another gene in the locus. This rewiring occurred only if the variant disrupted a CTCF-associated boundary domain. Our results demonstrate the functional importance of TADs for orchestrating gene expression via genome architecture and indicate criteria for predicting the pathogenicity of human structural variants, particularly in non-coding regions of the human genome.

INTRODUCTION

Approximately 5% of the human genome is structurally variable in the normal population, which includes deletions and duplications (collectively referred to as copy number variants, CNVs), as well as inversions, and translocations. Structural variations have received considerable attention as a major cause for genetic disease, promoting the search for CNVs as a standard diagnostic procedure in conditions such as intellectual disability and congenital malformations (Stankiewicz and Lupski, 2010; Swaminathan et al., 2012). The pathogenicity of many CNVs can be explained by their effect on gene dosage. In contrast, it is difficult to predict the consequences of balanced rearrangements, such as inversions, or the functional impact of CNVs that are limited to non-coding DNA. Such variants have the potential to disrupt the integrity of the genome, causing changes in the regulatory architecture that lead to pathogenic alterations of gene expression levels and patterns (Haraksingh and Snyder, 2013; Spielmann and Mundlos, 2013). However, the lack of a comprehensive understanding of the large-scale functional organization of the regulatory genome is a major limitation in predicting their potential pathogenicity.

New methods for enhancer identification and analysis of chromosome conformation have enabled substantial progress toward elucidating genome-wide regulatory interactions. Chromatin immunoprecipitation sequencing (ChIP-seq) performed directly on ex vivo tissues can reveal the location of distant-acting tissue-specific enhancer sequences at genomic scale.
In parallel, sequencing-based studies of DNA-DNA interactions have provided insight into the general conformation of the genome in living cells, as well as interactions between promoters and distant-acting transcriptional enhancers in specific cell types (Lieberman-Aiden et al., 2009). These data also show that enhancers can control multiple genes, frequently over hundreds of kilobases away from their target. Only a fraction of enhancers contact the nearest promoter, whereas most skip one or more genes (de Laat and Duboule, 2013). How the selective interaction of enhancers with their respective target genes is achieved remains largely unknown, but the organization of the genome in domains of interaction that are shielded from each other by boundaries appears to be critical. Genome-wide interaction studies by chromosome conformation capture-based approaches such as Hi-C (a high-throughput variant of the chromosome conformation capture technique [3C]) and 5C (chromosome conformation capture carbon copy) show that the genome is partitioned into megabase-scale topologically associated domains (TADs) (Dixon et al., 2012; Nora et al., 2012). These domains have been proposed to represent regulatory units within which enhancers and promoters can interact. They are separated by boundary regions that often contain CTCF binding sites or housekeeping genes representing de facto insulators that block interactions across adjacent TADs (Dixon et al., 2012). The importance of TAD structures is further supported by the finding that TAD boundaries appear to be largely static across different species and cell types. This suggests the existence of a preformed and stable topology that organizes the physical proximity between enhancers and their target genes. However, the observation that TADs exist regardless of transcriptional status has also raised questions regarding their role in cell- and tissue-specific regulatory processes (de Laat and Duboule, 2013). Furthermore, it has remained unclear if alterations in TAD structure, as they may occur in genomic rearrangements, can contribute to disease etiology.

The present study, we analyze the potential value of annotated TAD boundaries for understanding how structural variation in the human genome elicits pathogenic phenotypes. Focusing on families with rare limb malformations, we identified several rearrangements in the extended WNT6/IHH/EPHA4/PAX3 region and re-engineered them in mice. Through a series of circular chromosome conformation capture sequencing (4C-seq) experiments and expression studies in mouse limb tissue and human patient-derived cells, we show that the rearrangements disrupt the normal topology of protein-coding genes and their enhancers relative to TAD boundaries, resulting in inappropriate interactions and misexpression. Our results highlight the utility of considering the three-dimensional architecture of the genome for predicting the consequences of structural variation and reinforce that this approach may be useful for the analysis of structural variants in a wide spectrum of human disease phenotypes.

RESULTS

Disruptions of TAD Structure at the EPHA4 Locus Are Associated with Limb Phenotypes

The EPHA4 gene resides within a large gene desert flanked by a gene-dense region on the centromeric side and the PAX3 gene on the telomeric side. Hi-C data show that the region is organized into three adjacent TADs, the largest encompassing EPHA4 (Figure 1A) (Dixon et al., 2012). Studying the genetic causes of rare limb malformations, we identified a series of structural variants at the EPHA4 locus that potentially interfere with the integrity of this region. In mice, Epha4 is expressed during limb development and required for normal innervation of the limb, but inactivation of Epha4 does not cause changes in the limb skeleton (Helmbacher et al., 2000).

First, we investigated a dominantly inherited novel type of brachydactyly in three unrelated families, characterized by short digits predominantly on the preaxial (radial) side resulting in stub thumbs, short index fingers and a cutaneous web between the first and second fingers (Figure 1B; Figure S1). High-resolution array comparative genome hybridization (CGH) revealed heterozygous deletions of 1.75–1.9 Mb on chromosome 2q35-36 in all three affected families. All three deletions include the EPHA4 gene along with a large portion of its surrounding TAD and extend into the non-coding part of the adjacent PAX3 TAD, thereby removing the predicted boundary between the EPHA4 and PAX3 TADs.

Second, we studied the molecular cause of F-syndrome, a limb malformation syndrome characterized by severe and complex syndactyly, often involving the first and second fingers, and polydactyly of the feet (Figure 1C) (Grosse et al., 1999). F-syndrome had previously been mapped to this chromosomal region (2q36), but its genetic cause remained unknown (Camera et al., 1995; Thiele et al., 2004). We used whole-exome sequencing to detect mutations in genes located in the linkage interval but were not able to identify any potentially pathogenic changes. To search for non-coding mutations and structural variations, we used whole-genome sequencing. We detected a ~1.1-Mb heterozygous inversion in family F1 and a ~1.4-Mb heterozygous duplication, arranged in direct tandem orientation, in family F2. The telomeric breakpoints were located 1.4 Mb away from the EPHA4 gene within the gene desert in the case of the inversion and 1.2 Mb in the case of the duplication. The centromeric breakpoints were located centromeric and telomeric of WNT6 in the duplication and inversion, respectively (Figure 1C). Of note, both rearrangements bring the centromeric portion of the EPHA4-containing TAD into close proximity of the WNT6 gene.

Third, we studied a family that carries a heterozygous ~900-kb duplication in chromosomal region 2q35 that results in severe polysyndactyly and craniofacial abnormalities (Figure 1D) (Yuksel-Apak et al., 2012). The phenotype is reminiscent of the doublefoot (Dbf) mouse mutant, which also features massive polysyndactyly and was shown to be caused by a ~600-kb deletion affecting the same region (Babbs et al., 2008). Of note, both the human and the mouse alleles bring the IHH/Ihhi gene in proximity to the centromeric portion of the EPHA4-containing TAD.

Chromatin Interaction Landscape of the Extended WNT6/IHH/EPHA4/PAX3 Region

To elucidate the genetic basis of these birth defects, we sought to examine the regulatory landscape at this locus in more detail. In addition to EPHA4, we focused on the IHH, WNT6, and PAX3 genes due to their location near breakpoints in patients and their...
involvement in other developmental processes (Geetha-Loganathan et al., 2010; Goulding et al., 1994; St-Jacques et al., 1999). In each of the human disease alleles, at least one of the predicted TAD boundaries would be disrupted or its position changed relative to the four genes highlighted above (Figure 1). As illustrated in Figures 1 and 2, Hi-C data from human and mouse cells show a very similar TAD structure (Dixon et al., 2012), indicating that mice could serve as a model system for these diseases. To test if the TAD structure observed in embryonic stem cells (ESCs) by Hi-C is consistent with the chromatin conformation in the developing limb, we performed 4C-seq experiments in E11.5 mouse limb buds using the promoters of Epha4, as well as Ihh, Wnt6, and Pax3 as baits. For each promoter we observed interaction domains that were compatible with the Hi-C predicted TADs (Figure 2). These results confirm that the general TAD organization identified by Hi-C around the Epha4 locus is consistent with the interaction landscape during limb development obtained with 4C-seq.

**CRISPR-Mediated Re-engineering of Human Disease Alleles in Mice**

To study the effects of the structural variants observed in human patients in an experimentally accessible in vivo system, we generated mice with genome rearrangements recapitulating the human disease alleles. Using a protocol adapted for the introduction of large structural variants (Kraft et al., 2015), we co-transfected pairs of single guide RNAs (sgRNAs) into mouse ESCs to induce double strand breaks at desired positions.
To generate a mouse model for the human brachydactyly cases, mouse line DelB was created from ESC clones carrying a corresponding CRISPR-induced deletion (Figures 3A and 3B). Heterozygous DelB/+ mice showed shortening of the second and third digits due to hypoplastic phalanges, which was most pronounced in the middle phalanx of the second digit (Figure S2). This phenotype increased in severity when mice were bred to homozygosity, leading to severe shortening of the second and third digits due to very short middle phalanges, syndactyly between the second and third digits, and a deviation of these digits toward the radial side (Figure 3E, DelB/DelB). The phenotype is very similar to the human malformation that is also characterized by short thumb and index finger due to short or missing middle phalanges and partial syndactyly (Figure 1B). Thus, mutant mice with a deletion corresponding to the human disease alleles recapitulated the phenotype observed in patients. In addition to these digit malformations, homozygous DelB/DelB mice also showed the known phenotype resulting from Epha4 inactivation, i.e., a resultant hopping gait due to gross motor dysfunction (Helmbacher et al., 2000).

To create a mouse model of structural variants found in human F-syndrome patients, we reproduced the inversion observed in family F1 in mice. We obtained heterozygous and homozygous ESC clones carrying a 1.06-Mb CRISPR-induced inversion with breakpoints at comparable positions in the mouse locus (Figure 3C). Heterozygous as well as homozygous newborns generated via tetraploid aggregation died shortly after birth of unknown cause and did not show overt limb phenotypes or other morphological defects (data not shown).

Finally, we re-examined the previously described doublefoot (Dbf) mutant mouse strain due to the parallels in phenotype and genomic rearrangement with the human polydactyly patients (Figure 3D) (Yuksel-Apak et al., 2012). Dbf/+ mice have six to nine digits per limb in a mirror image position with loss of

Figure 2. TAD Organization around the Wnt6/Ihh/Epha4/Pax3 Locus
Hi-C profile derived from mouse ESCs (Dixon et al., 2012). Predicted TAD boundaries are indicated by vertical dashed lines. 4C-seq profiles from E11.5 mouse distal limbs with viewpoints on Wnt6, Ihh, Epha4, and Pax3 are shown. The y axis indicates number of normalized reads. 4C-seq interactions sharply drop at the Hi-C-predicted boundaries. The percentage of contacts per megabase within the TADs is shown on right. Orange bars represent contacts within corresponding TAD. Grey bars represent contacts located in the respective next two centromeric TADs and two telomeric TADs.
Figure 3. Ectopic Interaction of Wnt6, Ihh, and Pax3 with the Epha4 TAD and Misexpression in Mouse Models

(A) Schematic of the locus. Dashed lines indicate Epha4 TAD and boundaries. In situ hybridization (right) shows Epha4 expression in distal limbs of an E11.5 wild-type embryo (white arrow).

(B) Brachydactyly-like deletion generated by CRISPR/Cas (pink scissors). 4C-seq analysis with Pax3 as a viewpoint in E11.5 wild-type and heterozygous mutant distal limbs shows ectopic interaction with the centromeric part of the Epha4 TAD in the mutant (red box). Pax3 shows ectopic distal limb expression in mutants (white arrow).

(C) F-syndrome inversion generated by CRISPR/Cas (pink scissors). 4C-seq analysis with Wnt6 as a viewpoint shows ectopic interaction with the centromeric part of the Epha4 TAD in the mutant (red box). Wnt6 shows ectopic distal limb expression in mutants (white arrow).

(D) Dbf deletion. 4C-seq analysis with Ihh as a viewpoint shows ectopic interaction with most of the Epha4 TAD in mutants (red box). Ihh shows ectopic distal limb expression in mutants (white arrow).

(E and F) Autopod and corresponding skeletal preparations of wild-type and mutant mice at postnatal day 3. Insets in (E) show digit II. Note reduction of digit II in the homozygous brachydactyly-like deletion and massive polydactyly in Dbf/+.

See also Figures S2, S3, and S4.
anterior-posterior differences and no thumb-equivalent bipha-
langeal digit I (Figure 3F). Patients with polydactyly resulting
from duplication P1 have a very similar limb phenotype consist-
ing of severe polydactyly, complete fusion of digits (syndactyly),
and a mirror configuration of the digits (Figure 1D). In two of three
cases the mouse phenotypes closely resemble the human
congenital limb defects, further supporting the utility of these
mouse models.

Structural Changes Cause Misexpression of Developmental Genes Resembling the Endogenous Epha4 Pattern

The general nature of the structural variations and the resem-
blance of phenotypes resulting from inversion and duplication
(F-syndrome) or duplication and deletion (polydacbylty/Dbf)
raise the possibility that these phenotypes are caused by
convergent alterations in gene regulation. In the case of Dbf
mice, ectopic expression of lhh in the embryonic limb was previ-
ously described (Babbs et al., 2008). To examine the new
CRISPR-engineered lines for aberrant expression, we per-
formed RNA-seq experiments in E11.5 limbs of wild-type,
DelB/+(brachydactyly-like deletion), InvF/InvF (F-syndrome-
like inversion), and Dbf/+ (polydacbylty) embryos. We analyzed
the chromosomal region around the Wnt6/Ihh/Epha4/Pax3 locus
(chromosome 1: 73000000–79000000, mm9), for altered levels
of gene expression related to the corresponding structural vari-
atation. We detected a significant upregulation of Pax3 in DelB/+ limbs, of Wnt6 in InvF/InvF limbs, and of lhh in the Delb/+ limbs,
whereas all other surrounding genes were unaltered or showed
only marginal increases in expression levels (Figure S3). As ex-
pected, Epha4, which is contained in the brachydactyly (DelB)
deletion, and all the genes located within the Dbf deletion were
downregulated. On the basis of these results, we analyzed the
expression patterns of Pax3, Wnt6, and lhh in the respective
mouse mutants by in situ hybridization at E11.5 and compared
them with the wild-type Epha4 expression pattern.

Epha4 is expressed in a distinct pattern in the developing limb,
mainly in the distal mesoderm with predominance to the anterior
side (Figure 3A, right). At the same developmental stage, Pax3 is
also expressed in the limb bud, but restricted to migrating mus-
cle cells, evident as faint staining in the proximal limb, and absent
from the developing hand plate (Figure 3B, top right). DelB/+
(brachydactyly-like deletion) mice showed strong misexpression
of Pax3 in the distal anterior part of the autopod, in a pattern
resembling endogenous Epha4 expression (Figure 3B, bottom
right). Wnt6 is normally expressed in the limb bud ectoderm,
but not the distal mesoderm where Epha4 is expressed (Fig-
ure 3C, top right). In InvF/InvF (F-syndrome-like inversion) mice,
we observed strong misexpression of Wnt6 in the distal
limb autopod mesenchyme, in a similar pattern to Epha4 in
wild-type controls and Pax3 in DelB/+ mice (Figure 3C, bottom
right). The same pattern of misexpression was observed for het-
erozygous InvF/+ mice (not shown). Finally, lhh is not expressed
at all in the limb bud autopod at E11.5 (Figure 3D, top right). Mis-
expression of lhh in the distal limb bud of Dbf mutants was pre-
viously demonstrated (Babbs et al., 2008), and comparison at
E11.5 revealed a striking resemblance of the expression pattern
with Epha4 in wild-type embryos (Figure 3D, bottom right). Taken
together, these results indicate that genes near the chromo-
somal breakpoints are misexpressed in all three mouse lines.
In all cases, the acquired expression domain closely resembles
the endogenous expression pattern of Epha4, suggesting that
regulatory sequences normally controlling Epha4 may play a
role in the pathogenesis of the human limb phenotypes.

4C-Seq Reveals Ectopic Interaction of Misexpressed Genes with the Epha4 TAD

To examine whether the structural variants result in aberrant chroma-
atin interactions that may explain the ectopic expression do-
 mains observed in mutant mice, we performed 4C-seq on distal
E11.5 limbs. We analyzed the different mutants and stage-
matched wild-type embryos, using the promoters of Pax3,
Wnt6, and lhh as baits (Figures 3B–3D). For the deletions (DelB
and Dbf), heterozygous animals were examined in order to mini-
mize the possibility of deleterious effects that otherwise may
result from homozygous deletion of genes. For the copy num-
ber-balanced inversion (InvF), we examined homozygous animals
because no genes are deleted and the absence of the wild-type
allele simplifies the interpretation of the 4C-seq data. In wild-
type distal limbs, 4C-seq experiments showed minimal interac-
tion of Pax3, Wnt6, and lhh with non-coding sequences in the
Epha4 TAD. In contrast, all three genes showed substantial inter-
action with the Epha4 TAD in the mutants. In the DelB/+ (brachy-
dactyly-like deletion) mice, Pax3 showed a novel interaction
domain of ~800 kb, spanning the remaining part of the Epha4
TAD flanking the centromeric breakpoint (Figure 3B). In InvF/
InvF (F-syndrome-like inversion), we detected strong interaction
of Wnt6 with a ~300-kb region that corresponds to the cen-
 tromeric part of the Epha4 TAD, brought to the vicinity of Wnt6
through the inversion (Figure 3C). For better visualization, we
mapped the 4C-seq data to a reference sequence that includes
the inversion as present in the mice (Figure S4). Comparison
with 4C-seq data from wild-type embryos showed that the inter-
action levels substantially exceed interactions with equidistant
sequences in wild-type limb buds. Finally, the Dbf/+ deletion
showed extensive interactions of lhh with sequences throughout
the entire Epha4 TAD (Figure 3D). Thus, all rearrangements re-
sulted in novel interactions within the Epha4 TAD and a fusion of
adjacent TADs. Remarkably, all novel interactions in the fused
TADs respected the adjacent boundaries, i.e., lhh and Wnt6 did
not show interaction with the Pax3 domain and vice versa.

Disrupted TADs Result in Ectopic Interaction in Patient Cells

Previous studies showed that TADs are highly stable across spe-
cies and cell lines (Dixon et al., 2012), raising the possibility that
patient-derived samples can provide direct insight into regulato-
ry aberrations that affect early embryonic development. To test
this paradigm, we applied 4C-seq to human adult fibroblasts
(HAFs) and compared the results to data from the mutant mouse
strains. We processed HAFs from a brachydactyly patient of
family B1, a patient with F-syndrome from family F2 (cells from
F1 were not available), and a polydactyly patient (P1) with
the duplication. We compared patient samples with HAFs
from age-, sex-, and passage-matched healthy control donors
(Figure 4). All controls showed a 4C-seq profile that was highly
similar to that observed in mouse limbs (Figure 2) and human Hi-C data (Figure 1, top). Similar to DelB mice, the human brachydactyly-associated deletion resulted in aberrant contact of the PAX3 promoter region with the centromeric part of the EPHA4 TAD (Figure 4A). Likewise, the F-syndrome-associated duplication showed an ectopic interaction domain in the centromeric regions of the EPHA4 TAD, closely resembling the interaction domain gained in InvF mice (Figure 4B). Finally, the polydactyly-associated human duplication resulted in an overlapping, smaller interaction domain in the most centromeric region of the EPHA4 TAD (Figure 4C). While individual reads could not be unambiguously traced to one of the two copies present in the duplication allele, one plausible explanation for this observation are ectopic interactions between the telomeric copies of Wnt6/Ihh and the centromeric copy of the duplicated regions of the EPHA4 TAD.

Figure 4. 4C-Seq Analysis in Patient-Derived Human Adult Fibroblasts

(A–C) Schematic representation of the locus and the structural variations are shown on top. 4C-seq profiles from human adult fibroblasts from healthy controls and patients with heterozygous structural variations are shown below. Red boxes indicate ectopic interaction. The EPHA4 TAD is indicated by dashed lines.

(A) Patient with brachydactyly-associated deletion. 4C-seq analysis with PAX3 as a viewpoint shows extensive ectopic interaction with the EPHA4 TAD (red box).

(B) Patient with the F-syndrome-associated duplication. The 4C-seq profile with WNT6 as a viewpoint shows ectopic interaction with the EPHA4 TAD (red box).

(C) Patient with the polydactyly-associated duplication. The 4C-seq profile with IHH as a viewpoint shows ectopic interaction with the centromeric region of the EPHA4 TAD (red box).
De Novo Interaction between Distal Limb Enhancers and Ectopic Target Genes upon TAD Reorganization

Comparisons between the 4C-seq profiles obtained from the different mutant mouse tissues and patient cells revealed a minimal common region of ~150 kb within the EPHA4 TAD (Figure 5A). 4C-seq analysis of distal wild-type mouse limbs at E11.5 showed that this region frequently interacts with the EPHA4 promoter during normal development across 1.66 Mb.
of the intervening gene desert (Figure 5A). To identify enhancers with regulatory activity during limb development, we screened public ChIP-seq data for enhancer-associated chromatin marks, DNase hypersensitive sites (HS) data from equivalent developmental stages (E11.5) for open chromatin, and sequence conservation in this region (Figure 5B). To examine the strongest candidate sequences identified through this screen in more detail, we studied the in vivo activity pattern of five candidate enhancers. Two of these regions were already studied and documented in the VISTA enhancer browser database (Visel et al., 2007). The remaining three were tested for enhancer activity using a transgenic mouse LacZ enhancer reporter system.

At E11.5, four of the five regions showed reproducible LacZ reporter activity in the developing limb. Three of these enhancers, clustered in a 30-kb region, showed a high degree of spatial overlap with the endogenous expression pattern of EphA4, as well as the ectopic expression domains of Pax3, Ihh, and Wnt6 gained in DelB, Dbf, and InvF mice, respectively (Figures 5C and 5D; Figure S5).

To confirm that the interactions in this region previously observed in mouse wild-type and mutant limbs involve these enhancers, we performed 4C-seq using the enhancer cluster as bait. The interaction profile revealed that this region interacts frequently with EphA4 during development in wild-type distal limbs (Figure S6, top). Next we tested if these enhancers had contact with the promoters of Pax3, Wnt6, or Ihh in the distal limbs of mutants compared with wild-type controls (Figure S6, bottom). This analysis confirmed that in all three cases, as a consequence of the different structural variations, new interactions are established between limb enhancers located inside the EphA4 TAD and genes outside the domain.

**Boundaries of the EphA4 TAD Determine the Pathogenicity of Structural Variants**

All structural variants examined do not only change the arrangement of genes, enhancers, and predicted TAD boundaries relative to each other but also result in changes in the distance between enhancers and their possible target genes. To investigate if the observed ectopic interactions are caused by disruption of boundary elements or merely by distance effects, we examined the role of the putative boundaries at either side of the EphA4 TAD in additional mouse mutants. Regions with boundary-like properties were suggested by the Hi-C data (Dixon et al., 2012). The analysis of CTCF ChIP-seq data in human and mouse cell lines and tissues (ENCODr Project Consortium, 2004) showed an absence of CTCF binding sites within the EphA4 TAD and the presence of several CTCF peaks at each boundary region (Figure 6A). To investigate the possible role of the border regions flanking the EphA4 TAD, we generated DelB<sup>S</sup> and Dbf<sup>S</sup> mutant mice carrying deletions similar to the DelB (brachydactyly-like) and Dbf (polydactyly-like) mutants, except that the region containing the predicted boundary element was left intact (Figures 6B and 6C). Animals carrying these deletions had normal limbs and did not show any other abnormalities. Moreover, in situ hybridization for Ihh and Pax3 showed that these genes were not misexpressed and had retained their endogenous pattern of expression (Figures 6B and 6C, right).

To compare the interaction profiles in the absence or presence of boundary elements, we performed 4C-seq experiments using Ihh or Pax3 as baits in E11.5 distal limbs. This analysis revealed that the ectopic interactions of the corresponding gene with the EphA4 TAD observed in DelB and Dbf mice is reduced in DelB<sup>S</sup> and Dbf<sup>S</sup> mice (Figures 6B and 6C, left). Vice versa, 4C-seq using the enhancer cluster as bait also showed reduced interaction with the Ihh or Pax3 promoter (Figure S7). Thus, the presence of the boundary elements was sufficient to prevent inappropriate cross-TAD chromatin interactions, ectopic expression of non-target genes, and the morphological phenotypes resulting from this misexpression.

**DISCUSSION**

Structural variations are common in the human genome, but often difficult to interpret. Their inherent complexity necessitates model systems in which the human situation can be faithfully recapitulated and studied. Here we used an adapted CRISPR/Cas genome editing protocol to reproduce human rearrangements in mice (Kraft et al., 2015). Using this approach, we re-engineered three human malformation-associated rearrangements and investigated their effect on chromatin higher-order structures and gene function.

**Disruption of TADs Results in Aberrant DNA Domain Topology and Gene Misexpression**

TADs are stable units of genomic architecture that have been proposed to partition the genome into large regulatory units (Dixon et al., 2012). To investigate the effect of structural variants on TAD integrity, we examined limbs from the different mutant mouse strains using 4C-seq, which provides higher resolution than Hi-C. In all cases, we observed ectopic interaction of Pax3, Wnt6, or Ihh with the EphA4 TAD. The extent of interaction varied across the mutants, but included in all cases a minimal overlapping region of 150 kb (chromosome 1:75694480–75684058, mm9). Based on the hypothesis that this region might contain regulatory elements that drive the misexpression of Pax3, Wnt6, and Ihh, we screened the region for enhancers and identified a cluster of regulatory elements driving limb expression. The pattern driven by these enhancers overlaps with the endogenous limb expression of EphA4 and is very similar to the misexpression domains of Pax3, Wnt6, and Ihh observed in the mutant strains. Taken together, these data suggest that the target genes near the breakpoints were adopted by EphA4 enhancers, which in turn results in their misexpression (Figure 7). In DelB<sup>+</sup> mice we did not detect any other regulated genes. In both the Inv/InvF and the Dbf/+ mice, only one gene, Cyp27a and Fev, respectively, showed significant upregulation at overall marginal expression levels. Given the function of these genes in cholesterol metabolism (Cyp27a) and the central serotonin system (Fev), a contribution to the phenotype seems unlikely. Ectopic expression of Wnt6 can cause limb malformations in the chick via its anti-chondrogenic effect (Geetha-Loganathan et al., 2010), and misexpression of hedgehog proteins can induce polydactyly via the disruption of the anterior-posterior GLI3 gradient (Lettice et al., 2002). While the mechanisms by which ectopic expression of Pax3 may affect
skeletal morphology remain to be established, the observed misexpression domains in combination with the morphogenetic potential of Wnt6 and hedgehog proteins offer a plausible molecular explanation for at least two of the human phenotypes observed.

Our 4C-seq data using the Eph4 enhancers as a viewpoint (Figure S6) also show that the regions of ectopic interaction cover many other genes besides the identified targets Pax3, Ihh, and Wnt6. Nevertheless, expression analysis by RNA-seq showed no substantial upregulation of these genes, indicating that either enhancer-promoter distance or other unknown factors contribute to the receptiveness of a promotor to respond to the enhancer. In a Drosophila in vitro system, housekeeping and developmental promoters can respond to different classes of enhancers (Zabidi et al., 2015). It is possible that similar intrinsic specificities help to guide enhancer-promoter in vertebrate genomes. Here, the activated genes are all developmental genes expressed during limb development, indicating that there

Figure 6. Boundary Elements at Both Sides of the Eph4 TAD Prevent Ectopic Expression of Neighboring Genes
(A) CTCF ChIP-seq track in E14.5 mouse limbs (ENCODE/LICR). Red boxes and octagons mark clusters of CTCF peaks located at the boundary of the Eph4 TAD. The grey box indicates Eph4 TAD. 4C-seq profiles were generated from distal limb buds at E11.5. All data were obtained from heterozygous animals. Aberrant interactions are indicated by red boxes. Pink scissors indicate CRISPR/Cas-induced breakpoints in each deletion.
(B) A deletion (DelB$^S$) excluding the boundary region and CTCF cluster at the telomeric side of the Eph4 TAD (red octagon) was generated and compared with the brachydactyly-like deletion (DelB, including the CTCF cluster). The log2 ratio of the 4C-seq signal of DelB/DelB$^S$ shows increased interaction with the Eph4 TAD in the DelB deletion when compared with DelB$^S$ deletion (red box). Pax3 (right) shows normal expression of DelB$^S$/+ deletion mice, in contrast to Pax3 misexpression in DelB/+ mice (white arrow).
(C) A deletion (Dbf$^S$) excluding the boundary region and CTCF cluster at the centromeric side of the Eph4 TAD (red octagon) was generated and compared with the doublefoot deletion (Dbf, including the CTCF cluster). The log2 ratio of the 4C-seq signal of Dbf/Dbf$^S$ shows increased interaction with the Eph4 TAD in the Dbf deletion when compared with Dbf$^S$ deletion (red box). Ihh (right) shows an absence of limb expression in Dbf$^S$/+ deletion mice, in contrast to Ihh misexpression in Dbf/+ deletion mice (white arrow).
See also Figure S7.
may be a preference toward genes that are poised to get activated in this tissue.

**Conservation of TAD Structure across Species, Tissues, and Rearrangements**

Comparison of TADs across different mouse and human cell types suggests that their boundaries are largely conserved (Dixon et al., 2012). We hypothesized that this conservation allows for the analysis of disrupted regulatory interactions that occur in vivo during early embryonic development, using patient samples collected long after limb morphogenesis has ended. To test this approach, we performed 4C-seq in human adult fibroblasts from patients with the brachydactyly-associated deletion, the F-syndrome-associated duplication, and the polydactyly-associated duplication (Yuksel-Apak et al., 2012). Fibroblasts from healthy control individuals showed interaction domains highly similar to wild-type developmental mouse limb buds, whereas patient fibroblasts recapitulated the aberrant interactions observed in the respective mutant mouse strains. The observed chromatin interactions appear to be independent from gene expression levels, since EPHA4 and PAX3 are expressed at robust levels, WNT6 is expressed at very low levels, and IHH transcript is not detectable in HAFs (data not shown). While the ectopic interaction as well as the overall configuration was similar between human fibroblasts and mouse limb tissue, the distribution of peaks within the TADs was different. This is likely due to the different transcriptional activity and differentiation status of the analyzed samples (Nora et al., 2012). Regardless of such variation, our data indicate overall strong congruence between the developing mouse tissue and human-derived fibroblasts in the overall configuration of TADs. These results demonstrate the potential of patient-derived chromatin interaction data for gaining insight into the pathology of transient processes that occur during embryonic development.

**Perturbation of TAD Structure Results in the Formation of New TADs**

All structural variants examined in this study result in aberrant interactions of a regulatory unit with ectopic-target genes that normally do not take place because enhancers interact only with their respective target promoters. TADs have been proposed to play an important role in establishing appropriate enhancer-promoter interactions by providing a structural scaffold that limits the distance and direction over which enhancers operate. However, it remains unclear if the observed partitioning of the genome is a cause or consequence of pervasive enhancer-promoter interactions and what functional role boundary regions play in this process. In vitro experiments at the Xist locus, for example, demonstrated reorganization of a TAD and spill-over of activity upon deletion of a 58-kb element corresponding to the boundary region (Nora et al., 2012). In the present study, variants that delete TAD boundaries result in interaction across the domains with no apparent new boundary formation and thus represent an apparently seamless fusion of neighboring TADs (Figure 7). The interactions within these apparent new TADs as well as their exact boundaries and their impact on the overall three-dimensional architecture of the locus will have to be resolved by more quantitative analysis methods. However, the 4C-seq data obtained in this study suggest that the new TADs are defined by the next adjacent boundaries and that their functionality is not impaired. In the duplications the boundaries are not removed but the duplicated copy is flanked by an additional boundary, resulting in ectopic interaction within the newly formed TAD. Accordingly, our 4C-seq results show increased frequency of interaction of IHH (polydactyly family) and WNT6 (F-syndrome family) with centromeric parts of the EPHA4 domain (Figure 7, Duplication). In the case of the inversion, the EPHA4 enhancer cluster and the adjacent boundary is moved next to WNT6. This results in the ectopic interaction with WNT6 and the formation of a new TAD that is now confined by the former EPHA4 centromeric boundary (Figure 7, Inversion). Similar results were obtained at the Ttap2/Bmp2 locus, showing that inversions can result in reorganization of TAD structure by shifting boundaries (Tsujimura et al., 2015). Thus, a boundary element can be inverted or moved to a different chromosomal region without losing its functionality. However, the minimal region to determine boundary functionality is still unknown, and therefore...
the consequences of deleting only the boundary elements remain to be tested.

**Boundary Structures Are Important for TAD Integrity**

We experimentally challenged the assumption that TAD boundary elements are functional and relevant for disease pathogenesis by creating deletions that leave the proposed boundary regions on either side of the EphA4 TAD intact. Both regions contained a cluster of binding sites for CTCF, a factor involved in boundary formation (Dixon et al., 2012; Van Bortle et al., 2014). We engineered additional variants of the Dbf as well as the brachydactyly-associated rearrangements, this time leaving the predicted boundary regions undelated. No phenotypes and no misexpression of either Ihh or Pax3 were observed. Furthermore, 4C-seq experiments in these mice showed decreased frequency of interaction of the target genes Pax3 and Ihh with the Eph4 domain. Thus, leaving the proposed boundary regions intact diminishes all molecular phenotypes and averts morphological aberrations by preventing ectopic interactions.

Distance between regulatory elements and their target genes may be another determining factor. At the HoxD locus, for example, duplications within the TAD that result in an increase in the distance between promoter and enhancers were shown to result in an impairment of activation (Montavon et al., 2012). While we cannot rule out that distance effects contribute to the attenuation of molecular phenotypes, the difference between the deletion sizes is 100 kb for the Dbf/Dbf alleles and 200 kb for the DelB/DelB alleles, corresponding to 17% and 12% of the total deletion size, respectively. It appears unlikely that these minor differences in distance alone are sufficient to explain the reversion of molecular phenotypes to near-wild-type levels, given that similar differences in deletion size of ~200 kb are present across the brachydactyly families (B1, B2, and B3) without apparent effect on the phenotype.

**A Framework for Interpreting Human Structural Variation**

Depending on their size and position, structural variants may disrupt higher-order genomic organization. In this study we present a conceptual framework for the interpretation of such variants using genome-wide chromatin interaction datasets. Our results reinforce the notion that the pathogenicity of a substantial proportion of human disease-associated deletions results from ectopic enhancer-promoter interaction causing gene misexpression due to the elimination of annotated boundaries (Ibn-Salem et al., 2014). The aberrant chromatin interactions observed at the EPHA4 locus in the present study exemplify how the disruption of TAD structure by eliminating or interfering with boundary elements can lead to the functional rewiring of gene-enhancer interactions (schematically shown in Figure 7).

This model also illustrates how different types of large-scale structural changes can converge to give rise to the same phenotype. We showed that duplication and an inversion, as observed in the F-syndrome cases, or duplication and a deletion, as in the polydactyly family/mouse mutant, result in nearly identical molecular changes and morphological defects. In both cases the target gene interacts with the same non-coding genomic region and exhibits similar ectopic expression domains, despite the fundamentally different nature of the underlying structural mutation. In the deletions, a boundary is removed from the genome, permitting contact of the enhancer with genes outside of the TAD. In contrast, in the inversion and the duplications the enhancer is placed next to the new target gene. In the latter rearrangements the boundaries are left intact, but their new position no longer restricts contact of the enhancer with the target gene (Figure 7). The effect of such rearrangements is a gain of function via misexpression of one or several target genes. On the other hand, rearrangements of comparable size that do not interfere with TAD boundaries can be without consequence, as shown for the DelB and Dbf mutants. Thus, considering overall TAD structure and in particular the integrity of the boundaries and their position relative to genes and enhancers appears critical when predicting ectopic and potentially pathogenic enhancer-promoter interactions. As this situation depends on the presence of several factors including available enhancers and receptive genes that can give rise to phenotypes when misexpressed, other pathogenic mechanisms such as loss or gain of gene function have to be considered when interpreting structural variations. Notably, such predictions can be experimentally tested even if the affected tissue is not available. As shown in this study, normal and abnormal TAD structure and enhancer-promoter interactions may be preserved across developmental tissues and adult human cells such as fibroblasts. Hence, 4C-seq in patient cells can yield diagnostically valuable information to predict effects of structural variants on gene regulation and thus disease etiology. While the present study focused on one locus and one set of related morphological phenotypes, TAD data for the entire human and mouse genome are becoming available at increasing resolution (Jin et al., 2013; Ruo et al., 2014). These data should also help to interpret rearrangements in regions with higher gene density. Thus, the general principles uncovered in this study and the resulting approaches for the interpretation of structural variation in human phenotypes are expected to be applicable to other genomic loci and a wide spectrum of genetic conditions caused by structural variants.

**EXPERIMENTAL PROCEDURES**

**Human Material**

Venous blood samples and skin biopsies and were obtained from the patients and controls by standard procedures. Written informed consent was obtained from all individuals studied to participate in this study. This study was approved by the Charité Universitätsmedizin Berlin ethics committee.

**Identification of Human Structural Variations**

All experiments were done with genomic DNA extracted from blood. Brachydactyly deletions were identified using array CGH.

F-syndrome-associated duplication and inversion were identified using next generation sequencing (NGS, see Supplemental Information). Breakpoints for each structural variation were identified by breakpoint spanning PCR and Sanger sequencing.

**Generation of Transgenic Animals Using CRISPR/Cas**

Mouse ESCs carrying structural variations were created using a CRISPR/Cas-based protocol (Kraft et al., 2015) (see Supplemental Information). The size and position of the human structural variations (hg19) were converted to the mouse genome (mm9) using the USCS IROver tool. CRISPR guides were designed using the CRISPR design tool based on the algorithm described in Hsu et al. (2013) to place guide sequences within close proximity of the predicted
breakpoints (see Table S1). To minimize off target effects, guide sequences were chosen to have a quality score above 85%.

Embryos and live animals from ESCs were generated by tetraploid complementation (Artus and Hadjantonakis, 2011). For each structural variation, at least two independent clones were aggregated. Genotyping was performed by PCR analysis. Guide primers, genotyping primers, and breakpoint coordinates are summarized in Table S1.

**RNA-Seq**

E11.5 distal limbs were microdissected from wild-type or mutant embryos. RNA was isolated from tissue samples using the RNasy Mini Kit (QIAGEN). Samples were sequenced using Illumina HiSeq technology according to standard protocols.

**In Vivo Enhancer Validation**

Putative enhancer regions were selected on the basis of public tracks for H3K27Ac ChIP-seq (Cotney et al., 2012). DNase HS, p300 ChIP-seq data from 11.5 limbs (Visel et al., 2009b), and conservation available at the UCSC genome browser (http://genome.ucsc.edu). Selected regions were amplified by PCR from mouse genomic DNA and cloned into a Hsp68-promoter-LacZ reporter vector as previously described (Visel et al., 2009b) (see Table S2). Transgenic embryos were generated and tested for LacZ reporter activity at E11.5. All animal work performed at Lawrence Berkeley National Laboratory was reviewed and approved by the institutional Animal Welfare and Research Committee (AWRC).

**In Situ Hybridization and Skeletal Preparations**

In situ hybridization was performed according to standard protocols. Probes from Pax3, Ihh, Wnt6, and EphA4 were generated by PCR amplification using mouse limb cDNA. For skeletal preparation, specimens were stained according to standard Alcian blue/Alizarin red protocols.

**4C-Seq**

4C-seq libraries were generated from microdissected tissues or cells as described previously (van de Werken et al., 2012). BgIII or HindIII (6-bp cutters) were used as primary restriction enzymes. Csp6I or DpnII were used as secondary restriction enzymes. For each viewpoint, a total of 1.6 mg of each library was amplified by PCR (primer sequence in Table S3). Samples were sequenced with Illumina Hi-Seq technology according to standard protocols.

**ACCESSION NUMBERS**

All data have been deposited at GEO: GSE66383.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.04.004.

**ACKNOWLEDGMENTS**

D.G.L. is supported by the Fundación Alfonso Martín Escudero. This research was supported by grants from the Deutsche Forschungsgemeinschaft, from the Berlin Institute for Health, and the Max Planck Foundation to S.M. M.O. was supported by a Swiss National Science Foundation (SNSF) fellowship. A.V. was supported by NIH grants R01HG003988, U54HG006997, and U01DE024427. Research conducted at the E.O. Lawrence Berkeley National Laboratory was performed under Department of Energy Contract DE-AC02-05CH11231, University of California. We thank Daniel Ibrahim and Guillaume Andrey for comments on the manuscript. We also thank Nicole Rössner, Aisata Stege, Karol Macura, Nadine Lehmann, Anne Heß, and Christin Franke for technical support.

Received: January 5, 2015
Revised: February 12, 2015
Accepted: March 24, 2015
Published: May 7, 2015

**REFERENCES**


Figure S1. Phenotype of Preaxial Brachydactyly, PAX3 Type, Related to Figure 1
Representative members of families B1, B2, and B3. Top: Affected child of family B1. Note partial cutaneous syndactyly, short thumb and index finger as well as broad terminal phalanges. Corresponding radiograph on right showing short middle phalanges and broad, partially duplicated distal phalanx of index finger. Feet of same individual are shown below. Note partial syndactyly of hallux with second toe, and medial deviation. Corresponding radiograph on right shows short metatarsal I and medially deviated hallux and second toe. Middle: Affected individual at different ages of family B2. Left side shows hands of adult person. Note short, stub thumb and short index fingers with broad terminal phalanges. Palmar view shows abnormal flexion creases of index finger. Radiograph of hands on right top side at age 2 years. Note short abnormal metacarpal I and abnormally shaped distal phalanges of first and second finger. Below, radiograph of feet at age 1 year. Note short metatarsals I and broad, deviated hallux. Bottom: Affected child and mother of family B3. Hands of child are shown on left. Note similarity to families B1 and B2. Note broad thumb and additional flexion creases of index finger (4 instead of 3) on palmar view. Feet from same individual are shown on right top. Note medial deviation of toes and broad hallux. Mother of child is shown on right bottom. Note short thumbs, index and middle fingers.
Figure S2. Phenotype of Heterozygous DelB/+ Mice, Related to Figure 3
Phenotype of heterozygous DelB/+ mice at age p3 compared to wild-type litter mates. Note short second and third digits. In skeletal alizarin red stainings, a short middle phalanx of the second digit is visible.
**Brachydactyly-like deletion (DelB+/+)**

<table>
<thead>
<tr>
<th>gene</th>
<th>chromosome</th>
<th>start-end</th>
<th>FPKM control</th>
<th>FPKM mutant</th>
<th>log2(fold_change)</th>
<th>q_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ephod*</td>
<td>chr1</td>
<td>77367184-77515088</td>
<td>61.6997</td>
<td>32.0769</td>
<td>-0.943727</td>
<td>0.00428574</td>
</tr>
<tr>
<td>Pax3</td>
<td>chr1</td>
<td>78101266-78197134</td>
<td>0.285077</td>
<td>21.6602</td>
<td>6.24755</td>
<td>0.00428574</td>
</tr>
</tbody>
</table>

**F-syndrome-like inversion (InvF/InvF)**

<table>
<thead>
<tr>
<th>gene</th>
<th>chromosome</th>
<th>start-end</th>
<th>FPKM control</th>
<th>FPKM mutant</th>
<th>log2(fold_change)</th>
<th>q_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyp27a1</td>
<td>chr1</td>
<td>74713573-7473890</td>
<td>3.38099</td>
<td>8.32742</td>
<td>1.30043</td>
<td>0.00428574</td>
</tr>
<tr>
<td>Wnt6</td>
<td>chr1</td>
<td>74771891-74785319</td>
<td>12.8715</td>
<td>70.8773</td>
<td>2.46115</td>
<td>0.00428574</td>
</tr>
</tbody>
</table>

**Polydactyly-like deletion (Dbf+/+)**

<table>
<thead>
<tr>
<th>gene</th>
<th>chromosome</th>
<th>start-end</th>
<th>FPKM control</th>
<th>FPKM mutant</th>
<th>log2(fold_change)</th>
<th>q_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fov</td>
<td>chr1</td>
<td>74881508-74885419</td>
<td>0</td>
<td>0.362812</td>
<td>inf</td>
<td>0.00428574</td>
</tr>
<tr>
<td>Ihh</td>
<td>chr1</td>
<td>74945318-74951651</td>
<td>0.152798</td>
<td>5.29448</td>
<td>5.11479</td>
<td>0.00428574</td>
</tr>
<tr>
<td>Cnpdp1</td>
<td>chr1</td>
<td>75134553-75142368</td>
<td>19.6089</td>
<td>8.78013</td>
<td>-1.1592</td>
<td>0.00428574</td>
</tr>
<tr>
<td>Fam134a</td>
<td>chr1</td>
<td>75142785-75147909</td>
<td>22.976</td>
<td>11.9152</td>
<td>-0.947325</td>
<td>0.00774012</td>
</tr>
<tr>
<td>Zfand2b</td>
<td>chr1</td>
<td>75168645-75171629</td>
<td>16.1509</td>
<td>6.46836</td>
<td>-1.32014</td>
<td>0.0108332</td>
</tr>
<tr>
<td>Abcb6*</td>
<td>chr1</td>
<td>75171716-75180392</td>
<td>28.3344</td>
<td>8.88020</td>
<td>-1.67389</td>
<td>0.069227</td>
</tr>
<tr>
<td>Atg9b*</td>
<td>chr1</td>
<td>75180965-75191923</td>
<td>13.6562</td>
<td>3.42275</td>
<td>-1.33246</td>
<td>0.00428574</td>
</tr>
<tr>
<td>Ankef1*</td>
<td>chr1</td>
<td>75192150-75210778</td>
<td>20.4849</td>
<td>8.37645</td>
<td>-1.29015</td>
<td>0.0207677</td>
</tr>
<tr>
<td>Dnpep*</td>
<td>chr1</td>
<td>75307895-75317637</td>
<td>38.2119</td>
<td>20.7968</td>
<td>-0.877663</td>
<td>0.00774012</td>
</tr>
<tr>
<td>Gmpno*</td>
<td>chr1</td>
<td>75435929-75453179</td>
<td>29.6456</td>
<td>11.9868</td>
<td>-1.31724</td>
<td>0.0184657</td>
</tr>
<tr>
<td>Chof,Obs1*</td>
<td>chr1</td>
<td>75477456-75506452</td>
<td>54.5827</td>
<td>27.7491</td>
<td>-0.976005</td>
<td>0.00428574</td>
</tr>
<tr>
<td>Stk11p*</td>
<td>chr1</td>
<td>75521528-75537335</td>
<td>22.0375</td>
<td>10.5425</td>
<td>-1.06375</td>
<td>0.0249175</td>
</tr>
</tbody>
</table>

---

**Figure S3. RNA-Seq in Mutant Limbs, Related to Figure 3**

RNA-seq on distal limbs at E11.5 from the brachydactyly-like deletion (DelB+/, top) and F-syndrome-like inversion (InvF/InvF, middle) and polydactyly-like deletion (Dbf+/, bottom) mouse mutants compared to wild-type controls. Upper panel shows genes with significant differential expression in the genomic region chr1:73000000–79000000. Lower panel shows the chromosomal region including all genes with significant upregulation. Red boxes indicate upregulated genes with the highest FPKM in mutants compared to wild-type. FPKM, fragments per kilobase of exon per million fragments mapped. Asterisks indicate genes deleted in mutant alleles. q-value cut-off = 0.05.
**Figure S4. 4C-Seq Interaction Profile in F-Syndrome Inversion, Related to Figure 3**

Schematic of locus (top) and 4C-seq interaction profile with Wnt6 as viewpoint in wild-type and inversion mutant (InvF/InvF). The inverted part is in black. An artificially inverted chromosome with the corresponding 4C-seq is shown at the bottom. Note high degree of ectopic interaction and sharp drop at the boundary region. Boundary position before and after inversion is indicated by red line.
Figure S5. Limb Enhancers in the Interacting Region of the EphA4 TAD, Related to Figure 5
Top: Schematic of the locus. Cen, centromeric; tel, telomeric. Red box indicates region of minimal overlap. This region was analyzed for enhancer activity. Gray box indicates the location of a cluster of enhancers with reporter activity overlapping with the endogenous pattern of expression of EphA4 in the distal limb. Below: LacZ reporter expression of hs1635 shows reproducible mesenchymal limb expression (black arrows). Putative enhancer mm1044 showed no reproducible pattern of expression in independent transgenic assays.
Figure S6. 4C-Seq Interaction Profile with Enhancer Cluster as Viewpoint, Related to Figure 5

Top: Interaction of enhancer with *Epha4* gene and locus. Below: from top to bottom interaction profiles of enhancer in brachydactyly-like deletion, F-syndrome inversion and *Dbf* mutant showing ectopic interaction with *Pax3*, *Wnt6* and *Ihh*, respectively.
Figure S7. Boundary Elements of the EphA4 TAD Prevent Ectopic Expression of Neighboring Genes, Related to Figure 6

4C-seq was performed as shown in Figure 6, using enhancer mm1036 located in the limb enhancer cluster within the EphA4 domain as bait. Aberrant interactions are indicated by red boxes.

(A) Interaction profiles of E11.5 distal limbs with mm1036 as viewpoint generated from DelB^S and DelB^mutants. Log2 ratio of the 4C-seq signal of DelB/DelB^S is shown below. The interaction with the Pax3 promoter is reduced in the DelB^S deletion when compared to DelB deletion (red box).

(B) Interaction profiles of E11.5 distal limbs with mm1036 as viewpoint generated from Dbf^S and Dbf^mutants. Log2 ratio of the 4C-seq signal of Dbf/Dbf^S shows that interaction with the Ihh promoter is reduced in the Dbf^S deletion when compared to Dbf deletion (red box).
Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions

Dario G. Lupiáñez, Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopacki, Denise Horn, Hülya Kayserili, John M. Opitz, Renata Laxova, Fernando Santos-Simarro, Brigitte Gilbert-Dussardier, Lars Wittler, Marina Borschiwer, Stefan A. Haas, Marco Osterwalder, Martin Franke, Bernd Timmermann, Jochen Hecht, Malte Spielmann, Axel Visel, and Stefan Mundlos
SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Culture of Human Adult Fibroblasts (HAF)
Skin fibroblasts were cultured in DMEM (Lonza) supplemented with 10% fetal calf serum (Gibco), 1% ultraglutamine (Lonza) and 1% penicillin/streptomycin (Lonza). Written informed consent was obtained from all individuals studied. This study was approved by the Charité Universitätsmedizin Berlin ethics committee.

Microarray-Based Comparative Genomic Hybridization (Array CGH), Next Generation Sequencing (NGS), Quantitative Real-Time PCR (qPCR), and Breakpoint Analysis
Array comparative genome hybridization (array CGH) was performed using a whole-genome 1M oligonucleotide array (Agilent). The 1M arrays were analyzed by Feature Extraction v9.5.3.1 and Cytogenomics v2.7.8.0, respectively (Agilent). Analysis settings: aberration algorithm: ADM-2; threshold: 6.0; window size: 0.2 Mb; filter: 5 probes, log2 ratio=0.29. The genomic profile was visualized using the Agilent CytoGenomics software (Agilent CytoGenomics v2.7.8.0.).
Array data were submitted to the DECIPHER database (http://decipher.sanger.ac.uk); accession numbers: BER284939, BER284938, BER284937, BER285015 and BER285016.
qPCR using genomic DNA of patients and healthy controls was performed to detect breakpoint regions of the deletions studied. Primers were distributed over flanking sequences obtained on the previous array CGH experiments. Primer sequences can be provided upon request.
The whole genome of a patient from F-syndrome family 2 was sequenced using by a 100 bp paired-end run on an Illumina HiSeq and yielded 120 Gb of raw sequence data corresponding to a mean coverage of around 40 x. Sequence reads were mapped to the human reference GRCh37.3 with Novoalign in paired-read mode (Novoalign V2.07; www.novocraft.com). The alignment was filtered for inverted reads defined as paired reads that mapped to opposite strands. The majority of these inverted reads represent artifacts that originate from in vitro fragment ligation. The distribution of these reads over the genome is well described by a Poisson distribution. To identify potential breakpoints of inverted chromosomes, we filtered for a significant enrichment of such inverted reads in a window size of 500 bp, representing half of
the mean fragment size. Two regions with a highly significant enrichment of inverted paired-ends (15 and 18 reads) indicated the breakpoints of the inverted segment.

For the detection of the duplication, the patient's genome was sequenced using a 100 bp paired-end protocol on an Illumina HiSeq. The mean fragment size was ~500 bp (see figure below, fragment is indicated by a red line: a+b+c+d+e). The sequenced ends (a and c+d+e) of these fragments were aligned to the reference hg19 by Novoalign in paired-end mode. The breakpoint of the heterozygous duplication was identified by filtering for read pairs in which one end could be aligned (a) within the region of interest (mapped F-syndrome locus), but not the other (c+d+e). Based on the hypothesis that the duplication was in tandem (which has been the case for all duplications tested in our lab), some of these unmapped sequence reads should span the breakpoint.

![Diagram of duplication](image)

To identify the breakpoint spanning reads, the middle 40 bp of the 100 bp reads (d) was removed and the remaining 30 bp fragments (c and e) were again aligned to the wt reference. For those reads that span the breakpoint both fragments reads should map at the ends of the duplication:

![Diagram of reference sequence](image)

Using this method we identified reads with the exact breakpoint which was validated by Sanger sequencing.
Generation of Structural Variations Using CRISPR/Cas

A pair of oligos for every guide was annealed, phosphorylated and ligated into the px330 CRISPR/Cas vector (Addgene). G4 ES cells (129/Sv × C57BL/6 F1 hybrid background)(George et al., 2007) were cultured on MEF feeder layers under standard ES cell culture conditions. Cells were co-transfected for 12 hours with a pSuperior.puro plasmid (Addgene) and pX330 plasmids containing the corresponding Crispr guides using FUGENE HD reagent (Promega) under manufacturer conditions. After 12 hours, cells were replated on DR4 mouse embryonic fibroblasts (MEF) feeder layers. After 1 day, Puromycin was added during 2 days at a concentration of 2ug/ml. Then cells were allowed to recover for 4-6 days. Single clones were isolated, expanded and genotyped by PCR.

RNA-Seq Data Analysis

50 bp paired-end reads from Illumina sequencing were mapped to the mouse reference genome (mm10) using the STAR mapper (Dobin et al., 2013) (splice junctions based on RefSeq ;options: - -alignIntronMin 20 --alignIntronMax 500000 --outFilterMismatchNmax 10). More than 87% of read pairs mapped uniquely and these reads were subsequently used for expression analysis via the Cufflinks package (Trapnell et al., 2012) (version 2.2.1; default settings). In a first step, transcripts of each sample were assembled using Cufflinks provided with reference gene annotations from Ensembl. The resulting assemblies were then merged via Cuffmerge and finally differential gene expression was determined by applying Cuffdiff.

4C-Seq Data Analysis

To achieve an overall high quality of the raw data, all short sequence reads were validated and reads that did not match the primary restriction enzyme sequence or the primer sequence were discarded, whereas one mismatch in the beginning or end of the primer sequence was allowed. After clipping the primer sequence from each read, all remaining reads were mapped by Novoalign (Novoalign V2.07; www.novocraft.com) to the reference sequences GRCh37/hg19 or NCBI37/mm9. As suggested by van de Werken et. al. (van de Werken et al., 2012) we used the number of mapped reads on the cis chromosome over the total number of mapped reads (cis/overall ratio) to analyze the quality of the experiment as this is often a reflection of the
cross-link efficiency. To visualize the data, we created files in bedGraph track format for the readcounts of each fragment or in a specified window of fragments. The viewpoint, adjacent undigested fragment and fragments 10 kb up and downstream were removed as they showed no biological significance. 4C-seq contacts were analyzed in the mouse region chr1:73000000-79000000 or human region chr2:217000000-225000000. A range of 5 fragments was used to normalize the data per million reads (RPM) over a sliding window and display the continuous-valued data in the figures. Log2 ratios were calculated by dividing fragment reads between different samples.
### SUPPLEMENTAL TABLES AND LEGENDS

<table>
<thead>
<tr>
<th>Mutant</th>
<th>sgRNAs</th>
<th>Human breakpoint (hg19)</th>
<th>Converted breakpoint (mm9)</th>
<th>CRISPR breakpoint (mm9)</th>
<th>Genotyping primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>DelB</td>
<td>ACGTCTGGTTAAGACCG</td>
<td>chr2:221,278,232</td>
<td>chr1:76,385,972</td>
<td>chr1:76,388,978</td>
<td>F: TCTGCCGACATAAAGAAAAG</td>
</tr>
<tr>
<td></td>
<td>AACCATCATTTGAATCGGCG</td>
<td>chr2:223,014,332</td>
<td>chr1:78,061,513</td>
<td>chr1:78,060,839</td>
<td>R: GACACCTGATGGGTAACATAGAG</td>
</tr>
<tr>
<td>InvF</td>
<td>ATGCTGACTCAATTCGCAGT</td>
<td>chr2:219,741,632</td>
<td>chr1:74,834,693</td>
<td>chr1:74,832,836</td>
<td>F1: CTTTCTGGTTCCGGATGAATC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F2: AGGGACTGAAAGGCAACAAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R2: CCCGAGGCTTGACATACTCT</td>
</tr>
<tr>
<td>DelB^S</td>
<td>ACGTCTGGTTAAGACCG</td>
<td></td>
<td></td>
<td></td>
<td>F: TCTGCCGACATAAAGAAAAG</td>
</tr>
<tr>
<td></td>
<td>AATAGTAGTACTTCGGGTAC</td>
<td></td>
<td></td>
<td></td>
<td>R: GTCATCTACAAGTATCAGGCTAGCA</td>
</tr>
<tr>
<td>Dbf^S</td>
<td>GAGGAGGTITTTGGCGCTAT</td>
<td></td>
<td></td>
<td></td>
<td>F: GACCCGCTGAAATAGTGG</td>
</tr>
<tr>
<td></td>
<td>AGACTAAGGATGCAATCCGG</td>
<td></td>
<td></td>
<td></td>
<td>R: TGCAGAAAATGGGTCTTCCCT</td>
</tr>
</tbody>
</table>

**Table S1. Guide RNAs and Genotyping Primers, Related to Experimental Procedures**

Converted breakpoints indicate the position of the human breakpoints in the mouse genome after conversion using the USCS liftOver tool.
<table>
<thead>
<tr>
<th>ID</th>
<th>Coordinates (mm9)</th>
<th>Coordinates (hg19)</th>
<th>Already reported in VISTA enhancer browser?</th>
</tr>
</thead>
<tbody>
<tr>
<td>hs1507</td>
<td>chr1:75,765,578-75,770,167</td>
<td>chr2:220,713,868-220,717,300</td>
<td>Yes</td>
</tr>
<tr>
<td>mm1036</td>
<td>chr1:75,780,943-75,784,020</td>
<td>chr2:220,727,652-220,730,756</td>
<td>No</td>
</tr>
<tr>
<td>mm1044</td>
<td>chr1:75,793,720-75,796,439</td>
<td>chr2:220,735,762-220,738,581</td>
<td>No</td>
</tr>
<tr>
<td>mm1042</td>
<td>chr1:75,797,576-75,801,905</td>
<td>chr2:220,739,202-220,745,868</td>
<td>No</td>
</tr>
<tr>
<td>hs1635</td>
<td>chr1:75,829,293-75,831,744</td>
<td>chr2:220,771,085-220,773,279</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Table S2. Elements Tested for Enhancer Activity, Related to Experimental Procedures**

Red text indicates species of DNA origin.
<table>
<thead>
<tr>
<th>ID</th>
<th>Read Primer Sequence 1 (5'-&gt;3')</th>
<th>Primer Sequence 2 (5'-&gt;3')</th>
<th>Genome</th>
<th>Fragment coordinates</th>
<th>Restriction enzymes (1st/2nd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wnt6 promoter</td>
<td>CTACACGACGCTCTCCGATCTCCACAGATGCCCTACATCAG</td>
<td>CAGACGTGTGCTCTCCGATCTCAGTTGGACAGCCTACTACC</td>
<td>mm9</td>
<td>chr1:74,817,076-74,820,240</td>
<td>BglII/Csp6I</td>
</tr>
<tr>
<td>Ihh promoter</td>
<td>CTACACGACGCTCTCCGATCTTTGCTCGGGATTCAAGAT</td>
<td>CAGACGTGTGCTCTCCGATCTGCATGTCTAGCATGGTGTA</td>
<td>mm9</td>
<td>chr1:74,991,196-74,999,700</td>
<td>BglII/Csp6I</td>
</tr>
<tr>
<td>mm1036</td>
<td>CTACACGACGCTCTCCGATCTACCTATCTTGGCCATTCAAGAT</td>
<td>CAGACGTGTGCTCTCCGATCTCCATAACCCAGACTGATG</td>
<td>mm9</td>
<td>chr1:75,779,362-75,783,890</td>
<td>BglII/Csp6I</td>
</tr>
<tr>
<td>Eph4 promoter</td>
<td>CTACACGACGCTCTCCGATCTTGCTTTTACCGAAATAAAA</td>
<td>CAGACGTGTGCTCTCCGATCTTCGAGGTCTAGGTTGGAAG</td>
<td>mm9</td>
<td>chr1:77,508,304-77,513,767</td>
<td>BglII/Csp6I</td>
</tr>
<tr>
<td>Pax3 promoter</td>
<td>CTACACGACGCTCTCCGATCTTTCTATCTCCCCATGACACC</td>
<td>CAGACGTGTGCTCTCCGATCTCCTCAAAACGTGCTCTG</td>
<td>mm9</td>
<td>chr1:78,190,864-78,201,005</td>
<td>BglII/Csp6I</td>
</tr>
<tr>
<td>WNT6 promoter</td>
<td>CTACACGACGCTCTCCGATCTGCAGCACTCAGCTCTGAAG</td>
<td>CAGACGTGTGCTCTCCGATCTGGTGAGAAGGCAAGAAG</td>
<td>hg19</td>
<td>chr2:219,722,14-219,725,114</td>
<td>BglII/Csp6I</td>
</tr>
<tr>
<td>IHH promoter</td>
<td>CTACACGACGCTCTCCGATCTAGAGTTGGGAAGGTTAAAG</td>
<td>CAGACGTGTGCTCTCCGATCTTTGCTGGTTGCTGGAAG</td>
<td>hg19</td>
<td>chr2:219,924,07-219,929,315</td>
<td>HindIII/DpnII</td>
</tr>
<tr>
<td>PAX3 promoter</td>
<td>CTACACGACGCTCTCCGATCTGTGAATGTTCAACAGAT</td>
<td>CAGACGTGTGCTCTCCGATCTCGTGAGAAGTCAGATG</td>
<td>hg19</td>
<td>chr2:223,155,16-223,168,649</td>
<td>BglII/Csp6I</td>
</tr>
</tbody>
</table>

**Table S3. 4C-Seq Primers, Related to Experimental Procedures**

Red text indicates 4C-seq primer.
SUPPLEMENTAL REFERENCES


