CellPress

Review

Breaking TADs: How Alterations of Chromatin Domains Result in Disease

Darío G. Lupiáñez,[1,2,3,@] Malte Spielmann,[1,2,3] and Stefan Mundlos[1,2,3,*]

Spatial organization is an inherent property of the vertebrate genome to accommodate the roughly 2 m of DNA in the nucleus of a cell. In this nonrandom organization, topologically associating domains (TADs) emerge as a fundamental structural unit that is thought to guide regulatory elements to their cognate promoters. In this review we summarize the most recent findings about TADs and the boundary regions separating them. We discuss how the disruption of these structures by genomic rearrangements can result in gene misexpression and disease.

Trends

Chromosome conformation capture techniques have enabled scientists to study the 3D organization of the genome, identifying topologically associating domains (TADs) as fundamental regulatory units of the genome.

TADs and their boundary regions are critical for correct gene expression and their disruption can cause disease.

TAD maps of the human genome allow better prediction of the effects of structural variations.

CRISPR/Cas technology can be used to model human structural variants and understand their pathomechanisms.

Spatial Genome Organization

Major achievements have been made in functionally annotating the roughly 20 000 genes in our genome and to associate mutations/variants with specific diseases and pathomechanisms. Much less is known about the functional importance of the noncoding genome and its possible role in disease etiology. Recent studies have shown that the great majority of the genome is, in one way or another, involved in gene regulation [1]. Many noncoding elements are highly conserved in evolution and have been shown to drive specific patterns of expression. These sequences, generally called enhancers, are short (50–1500 bp) regions of DNA that can be bound by transcription factors. If placed next to a basic promoter and a reporter gene (usually LacZ), they are able to drive expression in specific spatial and temporal patterns [2–4]. Mammalian genomes are thought to contain more than 100 000 such regulatory sequences. With a comparable number of genes similar to those in lower organisms, this tremendous number of regulatory elements is thought to account for the pleiotropy of genes observed in higher organisms [5], thereby conferring tissue-specific patterns of expression [1,6]. This complexity raises the question of how regulation is achieved in general and how regulatory elements are able to find their correct target genes to ensure precise gene expression.

One important aspect of gene regulation appears to be related to the spatial organization of the genome in the nucleus. On a larger scale, chromosomes display a nonrandom nuclear organization highly influenced by their gene density and transcriptional status [7–10]. On a subchromosomal scale, the 3D organization of chromatin brings pairs of genomic sites that lie far apart along the linear genome into proximity. This 'looping' is part of the overall chromosomal folding process but also involves protein-mediated contacts between regulatory sequences (enhancers) and gene promoters. Such contact can result in the assembly of RNA polymerase II at the core promoter and the consecutive cell type-specific activation of transcription. The recent development of chromosome conformation capture (3C) and its derivatives (e.g., 4C, 5C, ChiaPet, Hi-C) have made it possible to explore in more detail the 3D architecture of the genome by quantifying chromatin looping via a proximity ligation assay [11,12] (Figure 1). Data obtained

[1]Max Planck Institute for Molecular Genetics, RG Development and Disease, 14195 Berlin, Germany
[2]Institute for Medical and Human Genetics, Charité – Universitätsmedizin Berlin, 13353 Berlin, Germany
[3]Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité – Universitätsmedizin Berlin, 13353 Berlin, Germany

*Correspondence: mundlos@molgen.mpg.de (S. Mundlos).
[@]Twitter: @dariloops

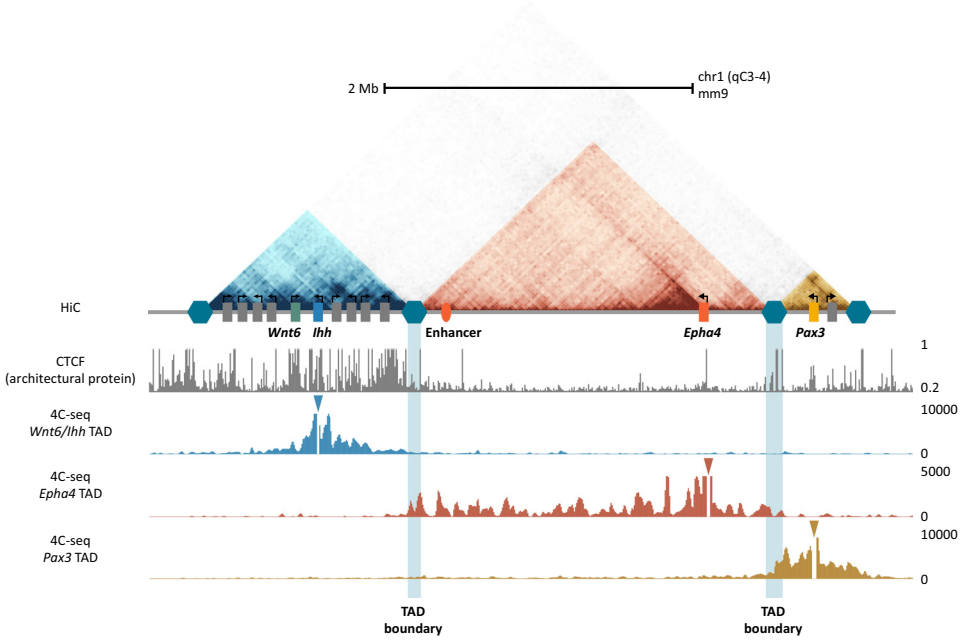

Figure 1. Higher-Order Chromatin Folding and Topologically Associating Domain (TAD) Structure at the *Eph receptor A4* (*Epha4*) Locus. Hi-C interactions are shown in a heat map in which each dot reflects two interaction pairs. The resulting interaction profile shows the formation of triangles (schematically enhanced in color) that represent individual TADs. There is a high degree of interaction within each TAD but little contact between TADs. Abrupt changes in the directionality of contacts demarcate boundary regions (blue hexagon). Of note is the very large TAD containing one (*Epha4*) gene, whereas the flanking TADs are much smaller (right) or contain many genes (left). Below, the binding profile of the CCCTC-binding factor (CTCF) transcription factor is shown. Note the scarcity of binding sites in the *Epha4* TAD and the enrichment at the boundaries. CTCF is also associated with gene promoters. The 4C-seq profiles of the viewpoints *Indian hedgehog* (*Ihh*), *Epha4*, and *Paired box3* (*Pax3*) are depicted below. Note that the interaction profiles are restricted to the respective TADs. Data from [23,30].

with these technologies have greatly expanded our knowledge of genome folding and have helped to elucidate how disruptions of genomic architecture, as they can be induced by structural variations, result in diseases such as congenital malformations and cancer. In this review we discuss the recent discoveries regarding the 3D architecture of the genome and their implications for understanding human disease.

Our Genome in 3D

Spatial folding of chromatin is required to establish long-range contacts between enhancers and their targets [13]. By using FISH and 3C techniques, some of these loops have been molecularly dissected, revealing some of the properties and dynamics of the folding process and its regulatory elements. First, enhancers can act over large genomic distances. In the case of the *Sonic hedgehog* (*SHH*) gene, for example, a prominent limb enhancer (ZRS) is located more than 1 Mb away from the transcription start site of the gene [14,15]. Many other developmentally active genes are surrounded by a similarly large gene desert. Some have been studied in more detail, showing that they contain a large number of regulatory sites that in their entirety appear to regulate the target gene's expression [16,17]. Second, enhancers can ignore nearby genes in favor of other, more distally located promoters, indicating that distance is not the determining factor that drives enhancer–promoter contacts. In line with these observations are recent genome-wide studies that show that only 7% of distal elements establish contact with the nearest promoter [18]. Third, looping, although necessary to initiate transcription, is not always

sufficient to drive gene expression. For example, FISH experiments at the *Shh* locus revealed that the ZRS enhancer is able to contact the *Shh* promoter in expressing and non-expressing cells [19]. Similar results were observed using 4C; the vast majority of DNA contacts are established long before gene expression, with only a small percentage showing transcription-dependent specificity [20]. Fourth, chromatin looping can persist in the absence of regulatory elements. Deletion of the ZRS enhancer, for example, does not impede the contact of the surrounding region with the *Shh* promoter [19]. The exact rules governing these principles, and whether they are universally applicable, are not fully deciphered but are crucial to understanding how gene regulation is achieved.

Hi-C is an expansion of the 3C technologies that uses purification of ligation products followed by massively parallel sequencing [8]. In contrast to 4C, which measures the contacts from one position (viewpoint) with the rest of the genome (one with all), Hi-C allows the identification of large numbers of long-range specific interactions across the entire genome in an unbiased manner (all with all) [11]. As they are derived from cell populations, Hi-C data reflect average configurations of regulatory landscapes, which largely agree with recent observations made by single Hi-C experiments [21]. Predictive modeling data, however, suggest that these conformations reflect a myriad of transient contacts differing from cell to cell, with some more favored than others [22]. Beyond the mere detection of enhancer–promoter interactions, these Hi-C data have revealed genome-wide maps of spatial proximity confirming the presence of chromosomal territories [8].

Further deep sequencing of Hi-C libraries revealed suborders of chromosome organization at the megabase scale, designated TADs [23–25] (Figure 1). These regions interact more frequently with themselves than with the rest of the genome and comprise the majority of characterized enhancer–promoter pairs [23,25–27]. Most interactions are confined to one TAD with little contact with neighboring regions. Gene coregulation is also often observed inside TAD domains [24,28], mostly resulting from large regulatory domains over which specific enhancers exert their effects [20,27,29,30]. Other genomic-specific features correlate well with TADs, such as chromatin marks [23,25], DNA replication [23,31], LADs [23], and chromocenter association [32]. It was proposed that TADs represent a fundamental structural unit of the genome that is thought to direct regulatory elements to their cognate promoters [9,10,33].

The biological relevance of TADs is highlighted by the observation that they are conserved among species, cell types, and tissues [23,24]. However, it is not entirely clear whether they exist in bacteria [34], yeast [35], or plants [36,37], suggesting that other types of chromatin organization are possible. The analysis of syntenic regions between mice and humans revealed an overwhelming degree of TAD conservation [23]. This indicates strong negative selection for maintaining entire TADs throughout evolution. Another important aspect about TADs is their potential to serve as a breeding ground for gene pleiotropy as suggested by studies performed at the *Homeobox D* (*HoxD)* locus. Here, the same TADs have evolved to control expression in multiple tissues of vertebrates, including digits and genitalia, although with intra-TAD differences in enhancer–promoter interactions [38]. A similar organization is observed at the *Homeobox A* (*HoxA)* locus, suggesting that the ancestral gene cluster might have displayed the same organization before its duplication during the emergence of vertebrates. Interestingly, the same TAD organization is also present in teleost fishes for both loci, although they do not develop digits or genitalia [39]. Altogether, these data indicate the importance of TAD structures in evolution [40].

It is tempting to think of TAD organization as a preformed scaffold for the genome that guides enhancers to find the appropriate promoters while at the same time isolating them from other regions of the genome. However, the observation of TADs by C methods cannot exclude the

possibility that TADs might be a mere reflection of the contacts that occur within them. In that sense, the analysis of what separates TADs might hold the answer to this question.

Strong Boundaries – Good Neighbors

TADs appear to promote contacts within a domain and at the same time prevent contacts between neighboring domains. Also important for the separation of neighboring activities are so-called boundary regions, initially inferred from Hi-C data sets by measuring abrupt changes in the directionality of contacts [23]. These regions, ranging from a few to hundreds of kilobases, still cannot be precisely mapped at single-nucleotide resolution, either because the resolution of Hi-C data is insufficient or because they actually comprise large genomic regions. Boundaries are often associated with housekeeping genes, tRNAs, and short interspersed element (SINE) retrotransposons [23,25,41,42]. In addition, most boundaries are enriched in architectural proteins and boundary strength correlates with the number of architectural proteins clustered in these regions [43]. CCCTC-binding factor (CTCF) and cohesin are two of the most studied architectural proteins. However, their presence is not exclusive to boundaries [23], which is consistent with their reported function in intra-TAD loop formation [44–49]. Nonetheless, CTCF seems to be crucial for boundary function as its ablation affects TAD organization by decreasing intradomain and increasing interdomain contacts [50]. CTCF tends to interact and functionally associate with the cohesin complex [51–54]. In contrast to CTCF, depletion of cohesin has a prominent effect on short-range interactions while keeping TAD position and boundaries almost unaffected [50,55,56]. Based on these findings it has been proposed that cohesin might be required for intra-TAD organization although it is also involved in boundary formation, being recruited via CTCF to chromatin.

It is interesting to note that a striking correlation was found between CTCF motif orientation and looping, occurring in more than 90% of cases in a convergent manner [26,57]. The fact that CTCF can form dimers *in vivo* [58] suggested a model where pairs of CTCF/cohesin binding sites might interact together in an orientation-dependent fashion to shape the extension of TAD domains. A recent study confirmed this notion, showing that the inversion of the orientation of CTCF motifs can redirect DNA contact, altering enhancer–promoter interactions and reshaping TAD domains [59]. Moreover, this specific disposition on CTCF orientation at TAD boundaries appears to be conserved in mammals and deuterostomes [60].

Genome editing experiments, most based on the recent CRISPR/Cas technology (Box 1), have also been performed to elucidate the role of boundaries in TAD organization. For example, the deletion of a single CTCF boundary element at the *X inactive specific transcript* (*Xist*) locus resulted in partial fusion of two adjacent TADs and ectopic gene activation [24]. Similarly, the

Box 1. CRISPR/Cas9 Genome Editing Tool

The recent development of the CRISPR/Cas9 technology has led to wider use of genome editing and opens new possibilities to create mutations and structural variations in model systems [86]. CRISPR/Cas9 has revolutionized genome editing, which used to be very labor intensive and time consuming. Using combinations of clustered regularly interspaced short palindromic repeats (CRISPRs) and DNA donors it is now possible to obtain full and conditional gene knockouts, knock ins, and large chromosomal aberrations such as deletions, inversions, and duplications [87,88]. The currently used CRISPR/Cas9 system was originally identified as part of the adaptive immunity against viral infection in prokaryotes and is based on the hybridization of a guide RNA to a corresponding target DNA sequence. The guide RNA contains the hybridizing part, which is variable, and the Cas9-interacting regions, which allow the recruitment of the Cas9 endonuclease at the hybridization site [89]. Cas9 is a DNA endonuclease that generates double-stranded breaks (DSBs) in DNA target sequences that are then repaired by either nonhomologous end joining (NHEJ) or homology directed repair (HDR) [90]. NHEJ is a very error-prone mechanism and creates deletions or indels at the cutting site. A single guide RNA can therefore be used for simple gene knockouts or in combination with a donor sequence to induce specific mutations and knock ins. Furthermore, the use of two guide RNAs allows more complex chromosomal rearrangements to be created such as deletions, inversions, and duplications [88].

deletion of a CTCF element associated with a boundary at the *HoxA* locus in *Drosophila* resulted in a shift of boundary position and active chromatin spreading [61]. In both cases, the newly formed TAD was defined not by the next annotated boundary element but by another intra-TAD element. By contrast, larger boundary deletions including multiple CTCF sites at the *Eph receptor A4* (*Epha4*) locus result in an apparent complete fusion of adjacent TADs causing ectopic enhancer–promoter interaction, gene expression, and pathogenic phenotypes [30] (Figure 2). Similar results were observed at the human beta globin locus [59], suggesting that boundaries might comprise multiple elements, like CTCF, that need to be completely disrupted to abolish boundary function. Such an organization might confer robustness on TAD structures and allow buffering of the effects of mutations of binding sites. In agreement with this, a multispecies study was unable to detect any TAD fusion or splitting related to single CTCF binding site divergence at boundaries, postulating that genomic rearrangements of large boundary regions might be the major driving force in evolution for new TAD formation [57].

Besides deletions, large inversions can also induce TAD reorganization and rewiring of enhancer–promoter interactions. A 1-Mb inversion including the centromeric boundary of the *Epha4* locus, with divergently oriented CTCF sites, resulted in adjacent TAD reorganization without any apparent loss of boundary activity. Similar results were reported at the *Transcription factor AP-2, gamma/Bone morphogenetic protein 7* (*Tfap2c/Bmp7*) locus [62], suggesting that the inversion of clusters of CTCF elements with divergent orientation might not affect their boundary function, in contrast to single CTCF element inversions [59]. These data demonstrate

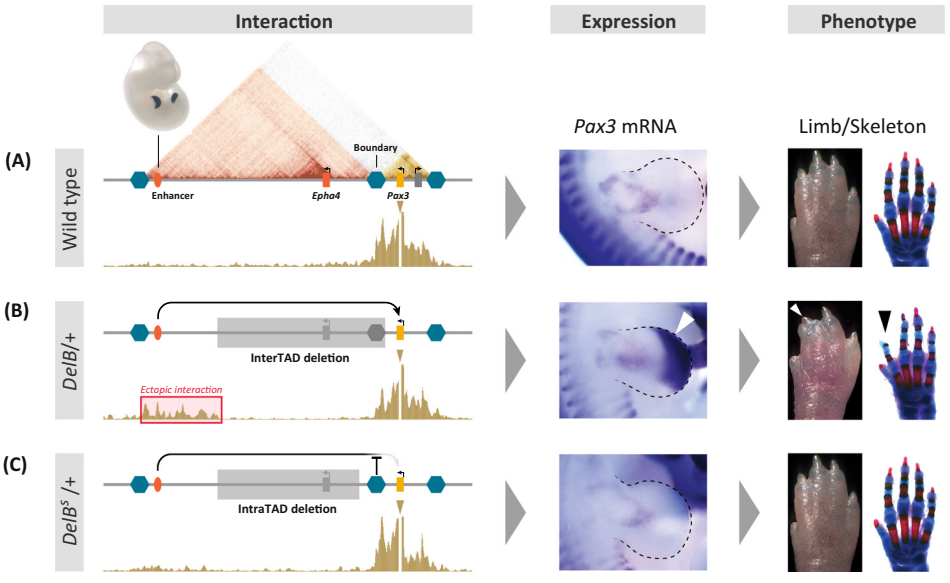

Figure 2. The Effect of Deletions on Topologically Associating Domain (TAD) Structure and Gene Expression at the *Eph receptor A4* (*Epha4*) Locus. (A) *Epha4* locus with TAD structure and corresponding 4C profile using the promoter of *Paired box3* (*Pax3*) as viewpoint. An 11.5-dpc embryo shows the LacZ reporter activity of the enhancer in the distal limb bud. Expression of *Pax3* is visualized by whole-mount *in situ* hybridization in limb buds of E11.5 embryos. The limb bud is indicated by a dashed line. Note the lack of expression in the distal limb. Some positive cells are seen in the proximal limb that correspond to migrating muscle cells. Wild-type forelimb and corresponding skeletal preparations stained with Alcian blue and Alizarin red are shown on the right. (B) A deletion removing *Epha4*, parts of the *Epha4* TAD, and the boundary region results in ectopic interaction of the *Pax3* promoter with enhancers that originally belonged to *Epha4*. This ectopic interaction results in ectopic expression of *Pax3* in the distal limb bud, which results in shortening of the first and second digits (arrow). (C) A 100-kb shift of the deletion toward *Epha4* removes similar parts of the region but leaves the boundary intact. This results in a normal 4C interaction profile similar to the wild type, with no ectopic expression and consequently resulting in a normal limb. Data from [4,23,30].

CellPress

Key Figure

Disruption of Topologically Associating Domain (TAD) Structure Causes Congenital Disease

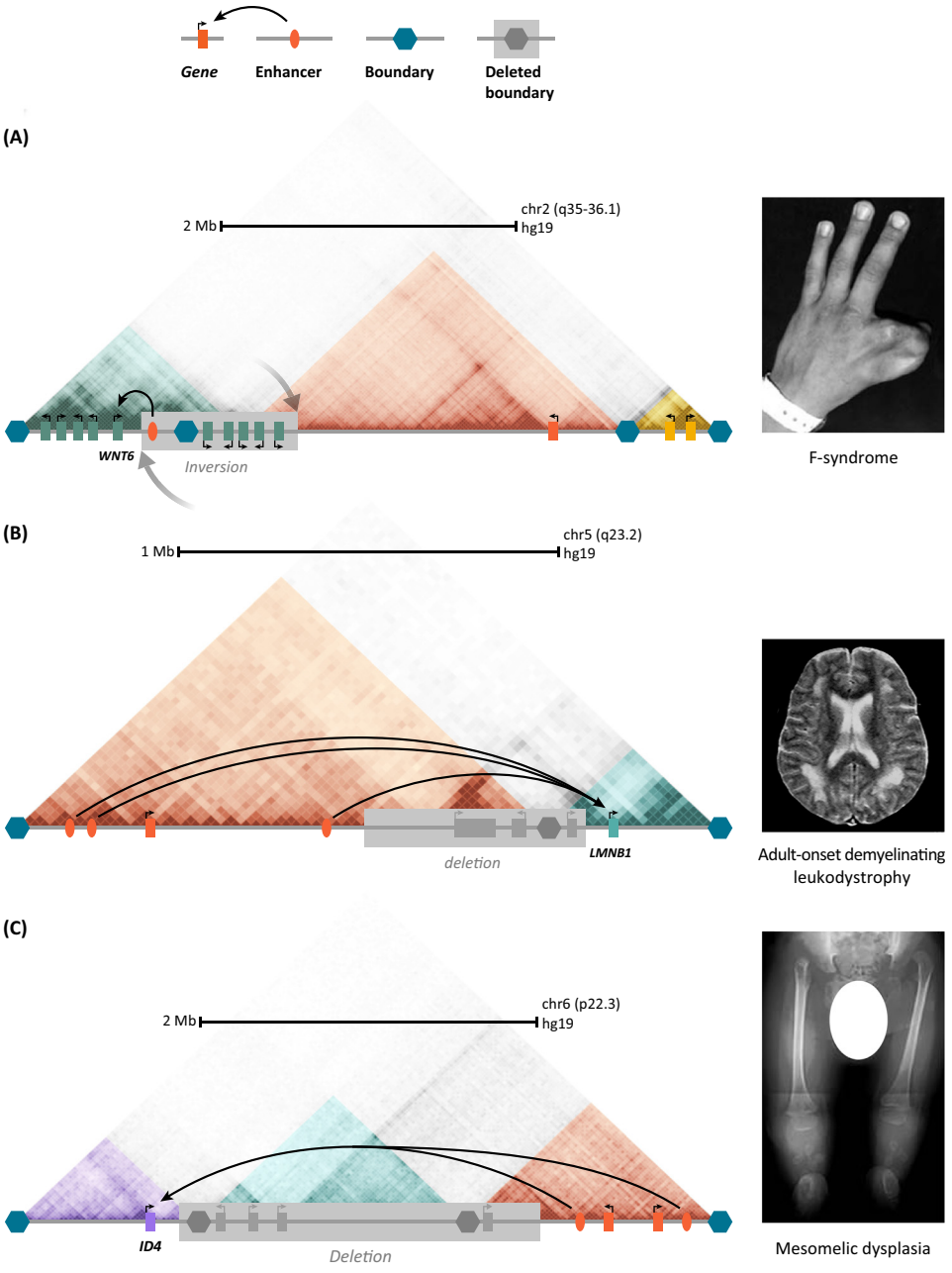

Figure 3. (A) F syndrome (also called acropectorovertebral dysgenesis) is a rare dominantly inherited skeletal disorder characterized by syndactyly of the first and second fingers. An inversion leaves the TAD boundary intact but places a cluster of limb enhancers from a neighboring TAD in front of WNT6 causing misexpression in digit 1 and 2. Data from [23,30]. (B) Autosomal-dominant adult-onset demyelinating leukodystrophy (ADLD) is a rare neurological disorder characterized by

that TADs and their boundaries play a key role in controlling gene expression. Structural variations such as deletions, inversions, or duplications have the potential to interfere with TAD structure by disrupting or repositioning boundaries. Recent studies highlight this as an important disease mechanism.

TADs and Disease

Testing for copy number variations (CNVs) (i.e., deletions and duplications) has become part of the diagnosis of genetic diseases such as congenital malformations and/or intellectual disability. High-resolution array CGH can reliably detect such variations and large studies show that a disease-causing variant can be identified in 10–20% of cases using this technology [63]. The results of such an investigation are generally interpreted by comparison with databases and by linking effects on gene dose with the phenotype. However, in many instances these explanations are unsatisfactory and the effect of the CNV remains unclear. With the detection of other structural variations such as inversions, insertions, or smaller CNVs by whole-genome sequencing, the demand for interpretation will increase. Some unexplained cases, especially in balanced translocations or inversions, have long been thought to be due to so-called 'position effects' [64–67]. With the discovery of TADs and our increased knowledge about enhancers and DNA folding, we can begin to understand these effects better. It is now clear that structural variations have the potential to alter the topological domain architecture of the genome by deleting, inverting, duplicating, or misplacing TAD boundaries, thereby allowing enhancers from neighboring domains to ectopically activate genes, causing misexpression and disease [68].

The relevance of TADs for genomic integrity and disease has recently been shown at the *EPHA4* locus [30]. *EPHA4* encodes ephrin receptor A4, a protein involved in the guidance of axon growth cones, the formation of tissue boundaries, cell migration, and segmentation. In accordance with its complex expression pattern, *Epha4* is surrounded by a large gene desert contained in one TAD of 2 Mb in size (Figure 1). Several large structural variations were identified at this locus that were all associated with limb malformations but of different types. A deletion encompassing *EPHA4* at the telomeric side resulted in brachydactyly (short digits) whereas an inversion and a duplication on the centromeric side involving part of the *EPHA4* TAD were shown to be associated with a complex form of syndactyly (fusion of fingers; Figure 3A, Key Figure). In addition, a family was investigated with a duplication whose breakpoints partially overlapped with a deletion present in a mouse mutant where 800 kb between the *Indian hedgehog* (*Ihh*) gene and the most centromeric part of the *Epha4* TAD are deleted. These two rearrangements, despite being of different natures, both resulted in massive polydactyly (seven or more fingers). The results indicated that different structural variations can result in similar but distinct malformation syndromes and that these were likely to be independent of the gene in this locus (i.e., *EPHA4*). The authors took advantage of the recent development of CRISPR/Cas9 genome editing (Box 1) and reengineered the human mutations (deletion and inversion) in mice. They showed that an enhancer cluster located in the *Epha4* TAD that normally regulates *Epha4* expression in the limb bud was now activating different genes depending on the breakpoint; that is, *Paired box3* (*Pax3*) in the brachydactyly, *Wingless-type MMTV integration site family, member*

progressive central nervous system demyelination due to overexpression of LMNB1. A 600-kb deletion including a TAD boundary was shown to result in pathological interactions between three strong forebrain enhancer elements and the *LMNB1* promoter resulting in cerebral lamin B1 overexpression and myelin degeneration [77,78]. Data from [23,78]. (C) Structural variations can alter the TAD architecture of the genome by deleting, duplicating, or inverting TADs and their boundaries, thereby allowing enhancers from neighboring domains to ectopically activate genes causing misexpression and disease [68]. Deletions on chromosome 6p22.3 have been shown to cause mesomelic dysplasia featuring hypoplastic tibiae and fibulae. The deletions span three TADs and remove two TAD boundaries. This brings several potential limb enhancers into close proximity with ID4, presumably resulting in misexpression in the developing limb bud. Data from [23,79].

6 (*Wnt6*) in the syndactyly, and *Ihh* in the polydactyly. All mutant phenotypes were caused by gene misexpression due to ectopic interaction of the enhancer with the target gene. However, this interaction was dependent on the disruption of one of the TAD boundaries (telomeric for *Pax3*, centromeric for *Ihh* and *Wnt6*), since shifting the deletions so that they did not include the boundary resulted in no interaction and normal mice.

As mentioned above, TADs were shown to be conserved among species and cell types [23,24]. The authors took advantage of this observation and performed 4C experiments using cells from patients. Remarkably, they were able to show the same ectopic interactions between the *EPHA4* enhancer and the respective target gene as observed in the mouse limb buds. Thus, TAD disruptions resulted in ectopic interactions and rewiring of enhancer–promoter interactions that were demonstrable even long after the event had occurred. This observation indicates that 4C technologies might be used in the future as diagnostic tools to investigate the interaction profile of rearranged genomes (Box 2).

Observations of diseases associated with structural variations before the discovery of the TADs can now be revisited. For example, the 11-Mb inversion causing the short digits (*Dsh*) phenotype in mice results in the misplacement of *Shh* in the domain of *Semaphorin 3c* (*Sema3c*), thereby inducing expression of *Shh* in the digits, corresponding to the *Sema3c* expression domain [69]. This interaction is possible because the inversion links *Shh* with the *Sema3c* TAD thereby exposing *Shh* to *Sema3c* regulatory sequences. Likewise, Lettice *et al.* mapped the breakpoints of an inversion at the *SHH* locus in a child with features of holoprosencephaly and severe limb malformation [70]. They were able to show that the inversion relocated the SHH transcription unit close to a highly conserved noncoding element that functions as a limb bud enhancer in mouse embryos. Moreover, the enhancer was able to drive ectopic expression of *Shh in vivo*, recapitulating the limb phenotype in the child. This study was one of the first to provide functional support for this novel type of long-range *cis*-regulatory mutation in which ectopic expression of a gene was driven by an enhancer that was not its own. The authors introduced the term 'enhancer adoption' for this mutational mechanism.

The deletion of a TAD boundary as a disease mechanism was proposed in Liebenberg syndrome, an autosomal-dominant condition in which the arms of the patient acquire morphological characteristics similar to those of the legs [71]. The authors identified disease-causing deletions that remove a gene [*H2A histone family member Y* (*H2AFY*)] 300 kb upstream of *Paired-like Homeodomain 1* (*PITX1*). In contrast to *H2AFY*, which shows no major phenotype when inactivated, *PITX1* was an interesting disease candidate. *Pitx1* determines hindlimb identity and misexpression of *Pitx1* in the forelimb of mice results in forelimb-to-hindlimb conversion. Like many housekeeping genes, *H2AFY* appears to function as a boundary element separating the *PITX1* TAD from the neighboring regulators. Without this barrier, an enhancer from a neighboring TAD is free to act on *PITX1* thereby inducing expression in the fore- and hindlimb. Recently, a tandem duplication that relocates the *PITX1* transcriptional unit in front of a strong forelimb enhancer was identified as the cause of Liebenberg syndrome [72].

So far the only study to investigate the role of deletions involving TAD boundaries in congenital disease on a larger scale used a bioinformatic approach [73]. Using the Human Phenotype Ontology database, the phenotypes of 922 deletion cases recorded in the DECIPHER database were related to monogenic diseases associated with genes in or adjacent to the deletions. The authors identified tissue-specific enhancers brought into the vicinity of developmental genes as a consequence of a deletion that included a TAD boundary. These enhancer–gene combinations were considered pathogenic when the phenotype was observed in the same tissue in which the enhancer was active. They compared this computationally with a gene-dose pathomechanism that attempted to explain the deletion phenotype based on haploinsufficiency of genes located

Box 2. C Techniques as Diagnostic Tools for Human Disease

Chromosome conformation capture (C) techniques are used to quantify DNA–DNA contacts in the nucleus by a proximity ligation assay. In this method, chromatin is first crosslinked and subsequently restriction-enzyme digested and ligated, generating a library of interacting DNA–DNA fragments bound by protein complexes. After crosslink reversion, interaction can be quantified using various methodologies (3C, 4C, 5C, Hi-C, and others). C techniques have made DNA folding measurable on a genome-wide scale, contributing enormously to our understanding of higher-order chromatin structure. In addition, some C techniques show potential to be used in the diagnosis of human diseases. 4C selects a specific locus or viewpoint to identify all of its genomic interacting partners (one with all). This method can be used to map breakpoints of structural variations at near-nucleotide resolution, providing confirmation and additional information about variants detected by other techniques such as array CGH [91]. Interestingly, this approach is also effective when breakpoints involve highly repetitive regions since it is able to capture the genomic regions surrounding them. Another variant called Hi-C takes advantage of an additional ligation enrichment step to identify every possible interaction in the genome (all with all), at the cost of reducing resolution to few kilobases. Both Hi-C and 4C data can be used to generate allele-specific interaction profiles and for phasing, thereby discriminating enhancer–promoter contacts between alleles [92–94].

Since TADs are highly conserved between different cell types and species [23,24], C techniques can also be used to study human cells such as lymphoblasts or fibroblasts (Figure I). Based on this principle, the effect of structural variations on TAD organization can be determined even if the genes and the region under investigation are no longer active in these cells. The extent of enhancer–promoter contacts can be assessed and ectopic interactions can be identified [30,95]. This provides an invaluable tool to interpret the effects of structural variations, especially those that do not involve gene-dose effects. The presence of contacts and TADs in cells that do not express the gene appears to be somewhat paradoxical but is likely to reflect the basic function of TADs in configuring DNA folding. It is important to note, however, that the interactions within a TAD may not be directly related to gene regulation. Rather, they appear to reflect a potential for interaction if the appropriate environment is present (i.e., developmental state, cell differentiation). Therefore, certain studies might require a cell type that reflects more accurately the pathological conditions of a specific disease.

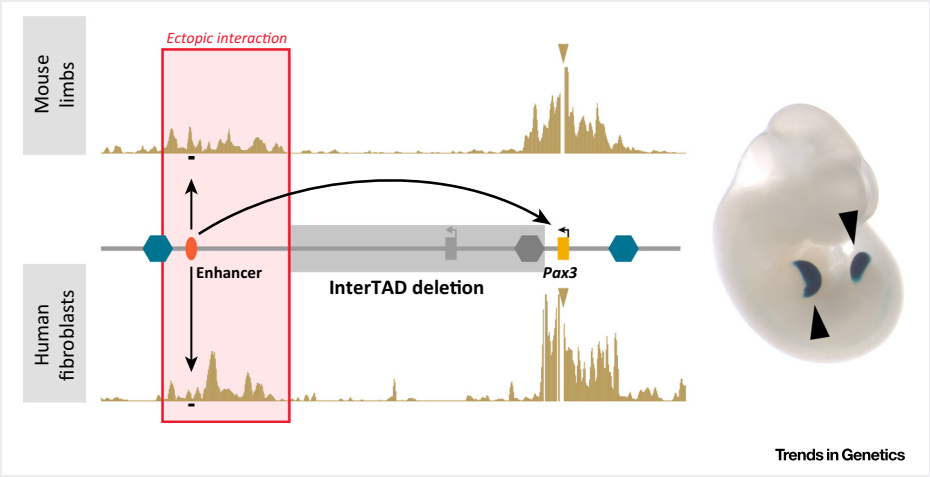

Figure I. 4C Detects Abnormal Interaction Profile in Mouse Limb Buds and Human Fibroblasts. Top panel shows 4C interaction profile from mouse limb buds with a deletion at the *Eph receptor A4* (*Epha4*) locus using the *Paired box3* (*Pax3*) promoter as viewpoint. Ectopic interaction of the *Pax3* promoter with the *Epha4* enhancer region is boxed. A major enhancer element driving expression in the limb bud was identified (region underlined) and tested in a LacZ transgenic mouse assay. As shown on the right, this enhancer drives expression in the distal limb buds and the tail region. Lower panel shows 4C from fibroblasts from a patient with limb malformation carrying the same deletion as the mouse shown above. Note the ectopic interaction of the *PAX3* promoter with *EPHA4* enhancers. The region spanning the interaction is very similar. The intensity of interactions, as indicated by the peaks of the 4C profile, are, however, different. Data from [4,30].

within the deletions. Up to 11% of the deletions could be best explained by enhancer adoption. Their results suggest that enhancer adoption caused by deletions of TAD boundaries may contribute to a substantial number of CNV phenotypes and should thus be taken into account in their medical interpretation. Although the data will need more experimental validation, one interesting example is the *Forkhead box G1* (*FOXG1*) locus. Mutations in *FOXG1* cause a congenital variant of Rett syndrome [74,75]. A recent study showed misregulation of *FOXG1* in cell lines derived from patients with a 60-kb telomeric deletion of *FOXG1* [76]. Taking the TAD

architecture of the locus into account, it seems likely that the deletion removes a TAD boundary and brings ectopic brain enhancers into the regulatory landscape of *FOXG1*. Several known brain enhancers are located close to the telomeric breakpoint and are free to act on *FOXG1* to cause misexpression in the brain of affected individuals leading to a Rett-like phenotype.

A similar pathomechanism was recently identified as a cause of autosomal-dominant adult-onset demyelinating leukodystrophy (ADLD), a rare neurological disorder in which overexpression of *Lamin B1* (*LMNB1*) causes progressive central nervous system demyelination [77,78]. Using 4C the authors were able to show that a deletion upstream of the *LMNB1* gene eliminates parts of the LMNB1 TAD and the neighboring TAD together with the intervening boundary. This resulted in ectopic interactions between the two TADs and at least three forebrain enhancers with the *LMNB1* promoter (Figure 3B). They confirmed lamin B1 overexpression in a postmortem brain sample of the frontal lobe.

For the interpretation of genomic rearrangements that change the regulatory landscape of a locus, it is important to map the exact breakpoints of the aberration, as exemplified by a study investigating deletions at the *SRY-box 4* (*SOX4*) locus [79]. The authors identified three unrelated patients with mesomelic dysplasia (severe shortening of the middle segment of the lower limbs) and identified overlapping *de novo* microdeletions encompassing four known genes: *Membrane bound* O-*acyltransferase domain containing 1* (MBOAT1), *E2F transcription factor 3* (E2F3), *CDK5 regulatory subunit associated protein 1-like 1* (CDKAL1), and *SOX4*. Interestingly, they found a fourth patient without any skeletal changes carrying an overlapping, slightly larger *de novo* deletion also encompassing the flanking gene *Inhibitor of DNA binding 4* (*ID4*). An analysis of the genomic region showed that the deletions span three TADs and two boundaries, bringing several potential limb enhancers into close proximity with *ID4* (Figure 3C). Thus, the deletions are likely to result in the aberrant activation and misexpression of *ID4* in the limb bud, thereby causing the mesomelic dysplasia. In the fourth family without a phenotype, however, the target gene *ID4* was also deleted and therefore no skeletal abnormalities were present.

TAD reorganization is also expected to play a role in the pathogenesis of cancer. Some examples are AML/MDS, the MonoMac/Emberger syndromes, and medulloblastoma, where enhancer adoption or enhancer hijacking has been identified as the major mechanism of disease [80,81]. Extensive remodeling of TAD structures and boundaries is intuitively thought to occur in these genomic rearrangements, thus allowing ectopic enhancer–promoter associations.

Concluding Remarks

The results from large-scale chromosome conformation capture studies using Hi-C have shown that metazoan genomes are subdivided into functional units that have been designated TADs. TADs are fundamental regulatory units of our genome that link higher-order chromatin structure with gene regulation and function. DNA folding and regulatory activity within one TAD is separated from neighboring TADs by boundary elements. These regions of DNA are ill defined and cannot be identified by their sequence alone, but the transcription factor CTCF has been shown to be functionally associated with boundary regions. We are just starting to understand the nature of TADs and their boundaries and therefore many questions remain unanswered (see Outstanding Questions).

Structural variations have the potential to interfere with the TAD structure of the genome by shifting regulatory elements between domains and/or by interfering with the position of boundaries. For example, deletions that include a boundary element can result in ectopic enhancer–promoter interactions thereby inducing gene misexpression and disease. This should be taken into consideration when evaluating structural variations and their implication in disease.

Outstanding Questions

What is the minimal critical region needed to confer boundary function?

How many types of boundary exist and which elements are required for their formation?

Can the genomic context or epigenetic factors affect boundary function?

To what extent are TADs influenced by the contacts occurring within them?

What, besides TADs, determines enhancer–promoter specificity?

Studies at the *Epha4* locus demonstrated that only a subset of all accessible genes shows responsiveness to a cluster of enhancers on genomic rearrangement and TAD remodeling. Thus, in contrast to earlier findings that enhancers are promiscuous (i.e., nonselective), these findings indicate a degree of specificity that differentiates between responsive and nonresponsive genes, probably driven by biochemical compatibility between promoters and enhancers [82]. Genome-wide studies in *Drosophila* using STARR-seq technology have provided some of the principles underlying this compatibility, separating housekeeping and developmental enhancers as different subclasses of regulatory elements [83]. As illustrated with the beta globin locus control region, mammalian regulatory elements might be subject to similar rules [84,85]. Systematic assays like those performed in *Drosophila* might help us to decipher the principles of this specificity. These data, combined with more comprehensive 3D genome interaction maps and the possibilities that CRISPR/Cas technologies offer, will be key to ultimately understanding and predicting the effects of genomic rearrangements and TAD disruption in mammalian genomes.

Acknowledgments

The authors thank members of the Mundlos laboratory for helpful discussions. They also thank Thomas Splettstoesser (http://www.scistyle.com) for assistance in figure design. Work in S.M.'s laboratory is funded by the Deutsche Forschungsgemeinschaft, the Berlin Institute for Health, and the Max Planck Foundation.

References

1. ENCODE Project Consortium (2012) *An integrated encyclopedia of DNA elements in the human genome* *Nature* 489, 57–74
2. Kothary, R. *et al.* (1988) A transgene containing *lacZ* inserted into the *dystonia* locus is expressed in neural tube. *Nature* 335, 435–437
3. Kothary, R. *et al.* (1989) Inducible expression of an *hsp68–lacZ* hybrid gene in transgenic mice. *Development* 105, 707–714
4. Visel, A. *et al.* (2007) VISTA Enhancer Browser – a database of tissue-specific human enhancers. *Nucleic Acids Res.* 35, D88–D92
5. de Laat, W. and Duboule, D. (2013) Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 502, 499–506
6. Bulger, M. and Groudine, M. (2011) Functional and mechanistic diversity of distal transcription enhancers. *Cell* 144, 327–339
7. Cremer, T. and Cremer, M. (2010) Chromosome territories. *Cold Spring Harb. Perspect. Biol.* 2, a003889
8. Lieberman-Aiden, E. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293
9. Bickmore, W.A. and van Steensel, B. (2013) Genome architecture: domain organization of interphase chromosomes. *Cell* 152, 1270–1284
10. Gibcus, J.H. and Dekker, J. (2013) The hierarchy of the 3D genome. *Mol. Cell* 49, 773–782
11. de Wit, E. and de Laat, W. (2012) A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* 26, 11–24
12. Dekker, J. *et al.* (2002) Capturing chromosome conformation. *Science* 295, 1306–1311
13. Palstra, R.J. *et al.* (2003) The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* 35, 190–194
14. Lettice, L.A. *et al.* (2003) A long-range *Shh* enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* 12, 1725–1735
15. Sagai, T. *et al.* (2005) Elimination of a long-range *cis*-regulatory module causes complete loss of limb-specific *Shh* expression and truncation of the mouse limb. *Development* 132, 797–803
16. Marinic, M. *et al.* (2013) An integrated holo-enhancer unit defines tissue and gene specificity of the *Fgf8* regulatory landscape. *Dev. Cell* 24, 530–542
17. Montavon, T. *et al.* (2011) A regulatory archipelago controls *Hox* genes transcription in digits. *Cell* 147, 1132–1145
18. Sanyal, A. *et al.* (2012) The long-range interaction landscape of gene promoters. *Nature* 489, 109–113
19. Amano, T. *et al.* (2009) Chromosomal dynamics at the *Shh* locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* 16, 47–57
20. Ghavi-Helm, Y. *et al.* (2014) Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* 512, 96–100
21. Nagano, T. *et al.* (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64
22. Giorgetti, L. *et al.* (2014) Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* 157, 950–963
23. Dixon, J.R. *et al.* (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380
24. Nora, E.P. *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385
25. Sexton, T. *et al.* (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148, 458–472
26. Rao, S.S. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680
27. Shen, Y. *et al.* (2012) A map of the *cis*-regulatory sequences in the mouse genome. *Nature* 488, 116–120
28. Le Dily, F. *et al.* (2014) Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev.* 28, 2151–2162
29. Symmons, O. *et al.* (2014) Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* 24, 390–400
30. Lupianez, D.G. *et al.* (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene–enhancer interactions. *Cell* 161, 1012–1025
31. Pope, B.D. *et al.* (2014) Topologically associating domains are stable units of replication-timing regulation. *Nature* 515, 402–405
32. Wijchers, P.J. *et al.* (2015) Characterization and dynamics of pericentromere-associated domains in mice. *Genome Res.* 25, 958–969
33. Nora, E.P. *et al.* (2013) Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods? *Bioessays* 35, 818–828

34. Umbarger, M.A. *et al.* (2011) The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Mol. Cell* 44, 252–264
35. Duan, Z. *et al.* (2010) A three-dimensional model of the yeast genome. *Nature* 465, 363–367
36. Feng, S. *et al.* (2014) Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in *Arabidopsis*. *Mol. Cell* 55, 694–707
37. Grob, S. *et al.* (2014) Hi-C analysis in *Arabidopsis* identifies the *KNOT*, a structure with similarities to the *flamenco* locus of *Drosophila*. *Mol. Cell* 55, 678–693
38. Lonfat, N. *et al.* (2014) Convergent evolution of complex regulatory landscapes and pleiotropy at *Hox* loci. *Science* 346, 1004–1006
39. Woltering, J.M. *et al.* (2014) Conservation and divergence of regulatory strategies at *Hox* loci and the origin of tetrapod digits. *PLoS Biol.* 12, e1001773
40. Lonfat, N. and Duboule, D. (2015) Structure, function and evolution of topologically associating domains (TADs) at *HOX* loci. *FEBS Lett.* 589, 2869–2876
41. Raab, J.R. *et al.* (2012) Human tRNA genes function as chromatin insulators. *EMBO J.* 31, 330–350
42. Hou, C. *et al.* (2012) Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* 48, 471–484
43. Van Bortle, K. *et al.* (2014) Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol.* 15, R82
44. Hadjur, S. *et al.* (2009) Cohesins form chromosomal *cis*-interactions at the developmentally regulated *IFNG* locus. *Nature* 460, 410–413
45. Hou, C. *et al.* (2010) Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3651–3656
46. Kurukuti, S. *et al.* (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to *Igf2*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10684–10689
47. Mishiro, T. *et al.* (2009) Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J.* 28, 1234–1245
48. Nativio, R. *et al.* (2009) Cohesin is required for higher-order chromatin conformation at the imprinted *IGF2-H19* locus. *PLoS Genet.* 5, e1000739
49. Splinter, E. *et al.* (2006) CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* 20, 2349–2354
50. Zuin, J. *et al.* (2014) Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U.S.A.* 111, 996–1001
51. Rubio, E.D. *et al.* (2008) CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8309–8314
52. Phillips-Cremins, J.E. *et al.* (2013) Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153, 1281–1295
53. Parelho, V. *et al.* (2008) Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 422–433
54. Wendt, K.S. *et al.* (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 796–801
55. Seitan, V.C. *et al.* (2013) Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res.* 23, 2066–2077
56. Sofueva, S. *et al.* (2013) Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* 32, 3119–3129
57. Vietri Rudan, M. *et al.* (2015) Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* 10, 1297–1309
58. Yusufzai, T.M. and Felsenfeld, G. (2004) The 5′-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc. Natl. Acad. Sci. U.S.A.* 101, 8620–8624
59. Guo, Y. *et al.* (2015) CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* 162, 900–910
60. Gomez-Marin, C. *et al.* (2015) Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7542–7547
61. Narendra, V. *et al.* (2015) Transcription. CTCF establishes discrete functional chromatin domains at the *Hox* clusters during differentiation. *Science* 347, 1017–1021
62. Tsujimura, T. *et al.* (2015) A discrete transition zone organizes the topological and regulatory autonomy of the adjacent *tfap2c* and *bmp7* genes. *PLoS Genet.* 11, e1004897
63. Gilissen, C. *et al.* (2014) Genome sequencing identifies major causes of severe intellectual disability. *Nature* 511, 344–347
64. Benko, S. *et al.* (2011) Disruption of a long distance regulatory region upstream of *SOX9* in isolated disorders of sex development. *J. Med. Genet.* 48, 825–830
65. Fernandez, B.A. *et al.* (2005) Holoprosencephaly and cleidocranial dysplasia in a patient due to two position-effect mutations: case report and review of the literature. *Clin. Genet.* 68, 349–359
66. Lauderdale, J.D. *et al.* (2000) 3′ Deletions cause aniridia by preventing *PAX6* gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 97, 13755–13759
67. Wagner, T. *et al.* (1994) Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the *SRY*-related gene *SOX9*. *Cell* 79, 1111–1120
68. Spielmann, M. and Mundlos, S. (2013) Structural variations, the regulatory landscape of the genome and their alteration in human disease. *Bioessays* 35, 533–543
69. Niedermaier, M. *et al.* (2005) An inversion involving the mouse *Shh* locus results in brachydactyly through dysregulation of *Shh* expression. *J. Clin. Invest.* 115, 900–909
70. Lettice, L.A. *et al.* (2011) Enhancer-adoption as a mechanism of human developmental disease. *Hum. Mutat.* 32, 1492–1499
71. Spielmann, M. *et al.* (2012) Homeotic arm-to-leg transformation associated with genomic rearrangements at the *PITX1* locus. *Am. J. Hum. Genet.* 91, 629–635
72. Seoighe, D.M. *et al.* (2014) A chromosomal 5q31.1 gain involving *PITX1* causes Liebenberg syndrome. *Am. J. Med. Genet. A* 164A, 2958–2960
73. Ibn-Salem, J. *et al.* (2014) Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol.* 15, 423
74. Ariani, F. *et al.* (2008) *FOXG1* is responsible for the congenital variant of Rett syndrome. *Am. J. Hum. Genet.* 83, 89–93
75. Kortum, F. *et al.* (2011) The core *FOXG1* syndrome phenotype consists of postnatal microcephaly, severe mental retardation, absent language, dyskinesia, and corpus callosum hypogenesis. *J. Med. Genet.* 48, 396–406
76. Allou, L. *et al.* (2012) 14q12 and severe Rett-like phenotypes: new clinical insights and physical mapping of *FOXG1*-regulatory elements. *Eur. J. Hum. Genet.* 20, 1216–1223
77. Giorgio, E. *et al.* (2015) A large genomic deletion leads to enhancer adoption by the lamin B1 gene: a second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). *Hum. Mol. Genet.* 24, 3143–3154
78. Brussino, A. *et al.* (2010) A family with autosomal dominant leukodystrophy linked to 5q23.2-q23.3 without lamin B1 mutations. *Eur. J. Neurol.* 17, 541–549
79. Flottmann, R. *et al.* (2015) Microdeletions on 6p22.3 are associated with mesomelic dysplasia Savarirayan type. *J. Med. Genet.* 52, 476–483
80. Groschel, S. *et al.* (2014) A single oncogenic enhancer rearrangement causes concomitant *EVI1* and *GATA2* deregulation in leukemia. *Cell* 157, 369–381
81. Northcott, P.A. *et al.* (2014) Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* 511, 428–434
82. van Arensbergen, J. *et al.* (2014) In search of the determinants of enhancer–promoter interaction specificity. *Trends Cell Biol.* 24, 695–702
83. Zabidi, M.A. *et al.* (2014) Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* 518, 556–559

84. Noordermeer, D. *et al.* (2011) Variegated gene expression caused by cell-specific long-range DNA interactions. *Nat. Cell Biol.* 13, 944–951

85. Tolhuis, B. *et al.* (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* 10, 1453–1465

86. Wang, H. *et al.* (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153, 910–918

87. Seruggia, D. and Montoliu, L. (2014) The new CRISPR–Cas system: RNA-guided genome engineering to efficiently produce any desired genetic alteration in animals. *Transgenic Res.* 23, 707–716

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

89. Mali, P. *et al.* (2013) Cas9 as a versatile tool for engineering biology. *Nat. Methods* 10, 957–963

90. Sternberg, S.H. and Doudna, J.A. (2015) Expanding the biologist's toolkit with CRISPR–Cas9. *Mol. Cell* 58, 568–574

91. Simonis, M. *et al.* (2009) High-resolution identification of balanced and complex chromosomal rearrangements by 4C technology. *Nat. Methods* 6, 837–842

92. Selvaraj, S. *et al.* (2013) Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat. Biotechnol.* 31, 1111–1118

93. Splinter, E. *et al.* (2012) Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* 58, 221–230

94. Williams, R.L., Jr *et al.* (2014) fourSig: a method for determining chromosomal interactions in 4C-seq data. *Nucleic Acids Res.* 42, e68

95. Gheldof, N. *et al.* (2013) Structural variation-associated expression changes are paralleled by chromatin architecture modifications. *PLoS ONE* 8, e79973