

Structural variation in the 3D genome

Malte Spielmann¹, Darío G. Lupiáñez² and Stefan Mundlos^{3,4*}

Abstract | Structural and quantitative chromosomal rearrangements, collectively referred to as structural variation (SV), contribute to a large extent to the genetic diversity of the human genome and thus are of high relevance for cancer genetics, rare diseases and evolutionary genetics. Recent studies have shown that SVs can not only affect gene dosage but also modulate basic mechanisms of gene regulation. SVs can alter the copy number of regulatory elements or modify the 3D genome by disrupting higher-order chromatin organization such as topologically associating domains. As a result of these position effects, SVs can influence the expression of genes distant from the SV breakpoints, thereby causing disease. The impact of SVs on the 3D genome and on gene expression regulation has to be considered when interpreting the pathogenic potential of these variant types.

Structural variation (SV). Genetic variation that includes all structural and quantitative chromosomal rearrangements, that is, deletions and duplications, as well as copy-number-neutral aberrations, such as inversions, insertions and translocations.

The application of molecular karyotyping and, more recently, next-generation sequencing (NGS) has revealed the enormous structural complexity of the human genome^{1–3}. Structural and quantitative chromosomal rearrangements, collectively referred to as structural variation (SV), include deletions, duplications, inversions, insertions and translocations and make up the majority of varying nucleotides among human genomes⁴. Unbalanced rearrangements, also known as copy number variation (CNV), alter the diploid status of DNA by changing the copy number of chromosomes or chromosomal regions⁵, whereas balanced rearrangements such as inversions, reciprocal translocations or copy-number-neutral insertions do not result in loss or gain of genetic material. Germline SVs ranging from whole-chromosome abnormalities, such as trisomy 13, 18 or 21, to recurrent microdeletion and microduplication syndromes are a common cause of congenital disease^{6–10}, and high levels of somatic SVs are a key signature of human cancer genomes^{1,11}. High-throughput sequencing technologies have dramatically accelerated the discovery and characterization of SVs; however, the medical interpretation of SVs and the prediction of phenotypic consequences remain unsatisfactory¹².

The discovery that SVs can be pathogenic without changing coding genome sequences first indicated that SVs can have regulatory effects, which were explained by so-called position effects^{13,14}. Indeed, 98% of the genome is non-coding¹⁵, and a large part of the non-coding genome is, in one way or another, involved in gene regulation. The complexity of assigning function to non-coding regions makes accurate estimates difficult; the fraction of the genome implicated in gene regulation has been estimated to be 5%¹⁶ or 80%¹⁷ depending on the study. Nonetheless, SVs have a high probability of affecting

the position and/or function of *cis*-regulatory elements, such as promoters and enhancers. Owing to advances in 3D genome mapping technologies, it is now becoming increasingly evident that position effects are the result of much more complex alterations than just changes in the linear genome. Recent findings have shown that position effects can be understood only by taking the third dimension of chromosomes into account — the folding of chromatin in the 3D space of the nucleus.

In this Review, we first discuss how SVs can rearrange the order and number of genes and regulatory elements, resulting in human disease. We then describe in detail how SVs can modify the 3D organization of the genome by disrupting chromatin domains. We also describe the phenotypic consequences of genomic disorders resulting from reshuffling of non-coding enhancer sequences and chromatin domain boundaries with the aim of presenting possible strategies for the medical interpretation of SVs in the 3D genome. We do not review the mechanisms underlying chromosome organization, nor do we discuss methods for SV detection by chromosome conformation capture (3C) or related technologies; for recent reviews on these topics, see REFS^{4,18–22}.

Structural variation in disease

Great progress has been made in the detection of single nucleotide variants in health and disease. However, the genetic analysis of SV has been limited by technical challenges. Although array comparative genomic hybridization (array CGH) is effectively used in routine clinical diagnostics for the detection of CNVs^{5,10}, this technique has limitations owing to its fairly low resolution, its inability to detect balanced rearrangements and its low efficacy in mosaic individuals²³. Whole-genome sequencing (WGS) by short-read technology has the potential to

¹Department of Genome Sciences, University of Washington, Seattle, WA, USA.

²Epigenetics and Sex Development Group, Berlin Institute for Medical Systems Biology, Max-Delbrück Center for Molecular Medicine, Berlin-Buch, Germany.

³Max Planck Institute for Molecular Genetics, RG Development & Disease, Berlin, Germany.

⁴Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin, Berlin, Germany.

*e-mail: mundlos@molgen.mpg.de
<https://doi.org/10.1038/s41576-018-0007-0>

Copy number variation (CNV). Genetic variation that refers only to quantitative chromosomal rearrangements, such as deletions and duplications.

Microdeletion and microduplication syndromes

A group of syndromes that are caused by chromosomal microdeletions and microduplications, which make up a subset of copy number variations that are usually smaller than 5Mb. Classic examples include the 7q11 deletion (Williams–Beuren syndrome), the 15q11–15q13 deletion (Prader–Willi and Angelman syndromes) and the 17p11 deletion (Smith–Magenis syndrome).

Position effects

Effects of structural variation (SV), classically translocations, on the expression of a gene without any changes to its coding sequence or promoter region. These effects can also be observed if a gene is inserted into different regions of the genome or if SVs connect previously unconnected genes and their regulatory units. In these cases, the change in the level of gene expression is thought to result from changes in the position of the gene relative to its normal non-coding *cis*-regulatory environment.

Chromosome conformation capture

(3C). PCR-based proximity-ligation analysis method of 3D genome organization that allows a reconstruction of the native chromatin structure within the nucleus. 3C employs a PCR-based approach to confirm interactions between two previously known loci (one versus one). The introduction of a second round of digestion and ligation allows the generation of self-circularized DNA fragments (4C-seq), on which inverse PCR can be used to identify all unknown fragments interacting with a specific locus (one versus all). Multiple genomic regions can be investigated in parallel by a multiplexing approach (5C; many versus many). HiC allows the identification of every possible interaction occurring in the nucleus through the introduction of biotin labelling to pull down ligation junctions (all versus all).

detect all rearrangements but lacks the ability to detect breakpoints in repetitive regions, where many breakpoints occur²⁴. Long-read sequencing technology, for example, single molecule real time sequencing or nanopore sequencing, is expected to overcome these problems but currently remains too expensive for routine clinical application^{25,26}.

De novo structural variation

One important question in human genetics is how much variation is normally present in the human genome and how much is specific to disease. Using all currently available methodologies to define the full spectrum of human genetic variation, >30,000 SVs can be detected per human genome²⁷. In a clinical diagnostic setting, following their identification, each variant needs to be carefully evaluated according to the specific phenotype and family history of the patient, as well as in terms of the gene content and genetic context of the variant, in order to assign pathogenicity to an SV. To differentiate polymorphic from potentially disease-causing SVs, it is essential to validate whether an SV has occurred *de novo* or whether it is inherited, as *de novo* SVs are more frequently associated with disease¹⁰. Large-scale studies have established a role for *de novo* CNVs in the aetiology of many diseases, including intellectual disability^{28,29}, autism spectrum disorders (ASDs)^{6,30}, congenital heart disease⁷, sporadic schizophrenia^{8,9} and congenital limb malformation³¹. For ASDs, approximately 3.7% of affected individuals were found to carry a large (>500 kb) *de novo* CNV, in contrast to only 0.4–0.8% of their unaffected siblings³².

Given the lack of a gold standard for SV detection and the above-mentioned technical limitations, the SV *de novo* mutation rate remains unknown, although data generated to date are likely to underestimate the frequency with which these variants arise. Data from phase 3 of the 1000 Genomes Project (REF.³³) provided a structural variation map of 2,504 human genomes of healthy individuals and a general framework for SV calling from WGS data, albeit without information on SV *de novo* mutation rates¹. The Genome of the Netherlands Consortium performed WGS of 250 healthy families and reported approximately 0.16 *de novo* SVs per generation (of >20 bp in length, in contrast to most other studies, which use >50 bp as a threshold for an SV)^{3,34}. By sequencing 97 individuals and their parents, Brandler et al. calculated a *de novo* SV rate of 19%³⁵, which is close to the estimates of the Genome of the Netherlands Consortium. These numbers are surprisingly low compared with the 70–100 *de novo* single nucleotide variants per generation that have been reported^{36–38}. Another study reported 88 *de novo* SVs in 476 quartets with ASD, which equates to a mutation rate of 0.092 *de novo* SVs per generation³⁸; an earlier study reported 5 variants in ~50 quartets, that is, a mutation rate of 0.05 (REF.³⁹). Interestingly, using the same data set, another study observed an association of paternally inherited non-coding SVs with ASD⁴⁰, whereas a third study did not replicate either of these results and concluded that the contribution of non-coding SVs to ASD is limited⁴¹. The disparity in rates,

even within the same data sets, highlights the current uncertainty in both the mutation rate of *de novo* SVs and their impact on disease.

To establish disease-specific SV morbidity maps of the human genome, several large initiatives such as the 100,000 Genomes Project⁴² and the Deciphering Developmental Disorders (DDD) study⁴³ are evaluating the amount of *de novo* SVs in a variety of patient cohorts. Equivalent CNV morbidity maps have proved to be extremely useful for the clinical interpretation of CNVs^{6,7,10,31}. The most comprehensive SV database to date, the Database of Genomic Variants (DGV) provides a catalogue of SVs found in the genomes of control individuals from diverse populations⁴⁴. The current version includes >22,300 genomes.

Although *de novo* SVs are more likely to cause disease, it should be noted that a considerable number of recurrent microdeletion syndromes exhibit reduced penetrance or high clinical variability. For example, only 15% of 15q13.3 microdeletions arise *de novo*, and they present with a wide range and severity of phenotypes⁴⁵. Reduced penetrance and high clinical variability represent key challenges for the medical interpretation of SVs. A possible explanation for reduced penetrance can be underlying recessive or biallelic inheritance, which was described for thrombocytopenia absent radius syndrome⁴⁶, for example.

Gene dosage

The pathogenicity of SVs that disrupt or include coding sequences is currently interpreted on the basis of their effect on gene dosage⁴⁷. This approach has been highly successful for gene discovery in Mendelian diseases mainly based on array CGH data by mapping a ‘minimal critical region’ of overlapping CNVs and by using animal models to identify the disease gene. Most human genes are found in two copies, and many are dosage sensitive (for example, a probability of being loss-of-function intolerant >0.9 represents 17% of protein-coding genes)⁴⁸, resulting in disease by *de novo* or inherited deletions or duplications. Prototypic examples include velocardiofacial syndrome and DiGeorge syndrome, which are caused by a heterozygous deletion of chromosomal region 22q11.2. Extensive mouse studies have demonstrated that haploinsufficiency of the transcription factor *TBX1* is the major cause of the cardiovascular defects that are found in patients with these conditions^{49,50}.

The gene dosage approach has also been very successful in discerning the aetiology of several recurrent microdeletion syndromes, such as the 17q21.31 deletion syndrome, which is caused by haploinsufficiency of *KANSL1*⁵¹, the 2q23.1 microdeletion syndrome, which is caused by haploinsufficiency of *MBD5* (REF.⁵²), and the 17p11.2 microdeletion associated with Smith–Magenis syndrome, which is thought to result from haploinsufficiency of *RAI1* (REF.⁵³). Paradigmatic for gene dosage disorders are deletions and reciprocal duplications at the chromosomal region 17p12, in which duplications result in Charcot–Marie–Tooth disease type 1A⁵⁴, whereas deletions cause hereditary neuropathy with liability to pressure palsies⁵⁵.

Despite the success of this strategy, the pathologies of many large microdeletion and microduplication syndromes have yet to be found. As most SVs also affect non-coding genomic regions, pathogenic mechanisms that include regulatory effects need to be considered.

From linear view to 3D perspective

Given the enormous number of regulatory sequences in the genome, most SVs can be expected to interfere with the positioning and/or copy number of regulatory elements. In a nonlinear 3D genome, such rearrangements can also be expected to alter the spatial organization at a restricted local or even chromosomal level. Therefore, understanding the basic principles underlying the spatial organization of the genome is crucial to fully comprehend the true pathogenic potential of SVs.

Chromosomes are packaged in the nucleus according to a specific 3D code that gives each chromosome its own territory⁵⁶ (BOX 1). This process requires the folding of chromatin down to nucleosome resolution^{19,20}. Chromatin organization has been classically studied and visualized through microscopy-based techniques such as fluorescence in situ hybridization, which led to the precise identification of chromosomal territories or specialized nuclear locations such as transcriptional factories or polycomb bodies^{57,58}. Although routinely used in clinical diagnosis, such methods are limited to the study of selected genomic regions in single cells. The development of the 3C technique and its derivatives has provided a toolbox that enables the systematic spatial interrogation of multiple loci or even the entire genome^{59,60}. This approach relies on the quantification of interaction frequencies between loci that lie in close spatial proximity in the nucleus independently of their linear genomic distance. Its genome-wide derivative, HiC, has the potential to reveal each interaction in the nuclear space, resulting in a 2D heat map in which each data point indicates the frequency of interaction between two points in the genome⁶¹.

Depending on their position, SVs often change the HiC profile of a locus, leaving specific signatures that can be used for further interpretation^{62,63} (FIG. 1). For example, deletions can result in novel interactions between two regions that were previously separated, whereas inversions result in a characteristic ‘bow tie’ configuration when mapped onto a reference genome. Furthermore, owing to its proximity-ligation nature, HiC is also suitable for the identification of SVs without a priori knowledge^{61,64–66} (FIG. 1).

The 3D regulatory landscape

An important finding of the past decade has been that chromatin folding not only secures the positioning of the chromosomes in the nucleus but is also a prerequisite for enhancer-driven long-range gene regulation⁶⁷. Long-range gene expression regulation usually involves two distinct types of *cis*-acting elements: the promoter, which consists of the core promoter and nearby regulatory elements, and more distal regulatory units, which include so-called enhancers or locus control regions (for a review, see REF.⁶⁸). The promoter is generally located <1 kb from the transcription start site, whereas enhancers can act over long distances — in some cases,

>1 Mb — without affecting genes that are closer in linear distance. According to the current concept, enhancers physically contact the promoter region by looping out the intervening DNA sequence, a process mediated by proteins such as transcriptional repressor CTCF, cohesin and Mediator^{69,70}. Once the contact is formed, expression is regulated, for example, by the recruitment of RNA polymerase^{60,71–73}. Large international initiatives such as the US National Institutes of Health (NIH) Roadmap Epigenomics Mapping Consortium have produced public resources of human epigenetic data that can be used to detect active, poised and repressed regions in the genome and their underlying regulatory elements¹⁶.

Most enhancer–promoter interactions occur within topologically associating domains (TADs)^{61,74,75} — megabase-long genomic regions that interact with themselves with high frequency but interact with the rest of the genome less often. TADs are separated by regions displaying abrupt changes in the directionality of interactions, which are denominated boundary regions⁷⁴ (BOX 1). This finding suggests that TADs form a genomic scaffold that facilitates regulatory interactions while insulating regulatory activity from neighbouring domains. A large fraction of TADs (60–70%) is almost invariant among different cell types and between species⁷⁴, suggestive of a role for TADs as fundamental functional units of the genome. The high degree of TAD conservation across cell types or during development enables the identification of conformational changes even when the affected cell type is not accessible. For example, 3C identified aberrant interaction across TADs in patient fibroblasts even though the genes were no longer active in these cells^{63,76}.

TADs constrain the genomic regions that an enhancer can act upon. This has been elegantly exemplified in vivo by the systematic insertion of a regulatory sensor (a *LacZ* reporter gene within a *leeping* Beauty transposon) across the mouse genome^{77,78}. Sensor activity accurately reflects the regulatory potential of the genomic environment surrounding the insertion sites, revealing the segregation of the genome into regulatory landscapes with distinct tissue-specific potential that largely overlap with TADs^{78,79}. These regions are occupied by enhancer elements with activities that largely overlap with the expression patterns of the genes sharing the same domain.

Despite our increasing knowledge of chromatin domain organization, there is still a substantial portion of the genome in which TADs cannot be detected and/or chromatin shows a more dynamic nature^{61,74}. Moreover, the identification of TADs relies heavily on computational methods (for a review, see REF.⁸⁰), which display a high degree of variation depending on the resolution and the adjustment of thresholds. Furthermore, it remains to be demonstrated whether the general principles described for TADs also operate on all loci that are called on the basis of the respective algorithm. It is also important to note that although TADs greatly limit the formation of regulatory interactions between different domains, these restrictions seem not to be absolute in certain cases⁸¹. For example, promoter–enhancer pairs that have a high affinity for each other, such as the ZRS (zone of polarizing activity regulatory sequence)

Penetrance

A measure of the proportion of people with a particular genetic change (such as a mutation in a specific gene) who exhibit signs and symptoms of a genetic disorder. If some people with the mutation do not develop features of the disorder, the condition is said to have reduced (or incomplete) penetrance.

Gene dosage

The number of copies of a particular gene, including all its regulatory regions present in a genome.

Haploinsufficiency

A state in which one copy of a gene is inactivated or deleted and expression of the remaining functional copy of the gene is not sufficient to preserve normal function.

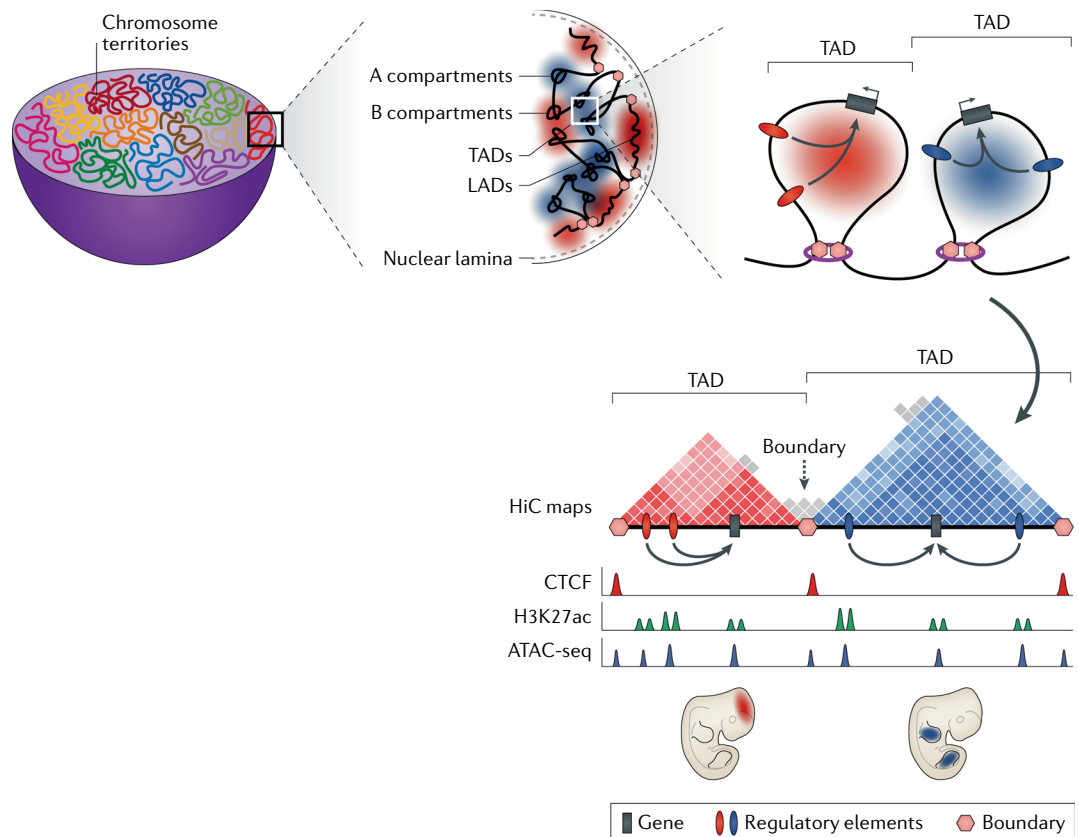
Topologically associating domains

(TADs). Genomic sequences in the range of megabases in length that are separated by boundary regions and that physically interact with themselves more frequently than with the rest of the genome.

Box 1 | Chromatin organization — from the 3D nucleus to the linear genome

At the nuclear level, chromosomes occupy specific territories (see the figure). At the chromosomal level, HiC technology has revealed a spatial segregation of open and closed chromatin into two compartments called A and B compartments^{59,60}. A and B compartments largely correlate with accessible, transcriptionally active euchromatin and compacted, transcriptionally silent heterochromatin, respectively (see the figure). B compartments predominantly locate to the periphery of the nucleus as lamina-associated domains (LADs) or to nucleoli as nucleolar-associated domains (NADs). Further studies with increased sequencing depth and resolution discovered that compartments are subdivided into regions called topologically associating domains (TADs)^{74,100,132}. TADs are identified in HiC experiments as interaction hubs, with a decrease of interactions at the boundary regions (see the figure). Inside TADs, regulatory elements interact with their cognate genes to activate transcription. Functional data sets help to identify and delineate the regulatory content of TADs. Histone H3 Lys27 acetylation (H3K27ac) and assay for transposase-accessible chromatin using sequencing (ATAC-seq) can be used to identify active regulatory elements (see the figure). The boundary regions that separate TADs are stabilized by the cohesin complex and are generally enriched in architectural proteins such as transcriptional repressor CTCF and transposable elements or housekeeping genes; boundaries seem to be critical for proper TAD function^{74,76}.

At a higher resolution, HiC data sets have revealed the existence of additional layers of organization, such as sub-TADs⁶¹, which are smaller spatial domains (around 100 kb) that display a more dynamic nature and tissue specificity¹³³. Furthermore, both TADs and sub-TADs are populated by focal points of interaction (loops) that can be visually identified in HiC maps⁶¹. Some loops are associated with enhancer–promoter interactions that are bound by Mediator and the cohesin complex, displaying a dynamic and tissue-specific nature¹³³. By contrast, other loops display an invariable nature across cells or tissues and are frequently bound by CTCF¹³⁴, an architectural protein that has a key role in chromatin folding and TAD boundary formation. CTCF binding sites, as determined by chromatin immunoprecipitation followed by sequencing (ChIP-seq), show enrichment at boundary elements (see the figure). Interestingly, CTCF-mediated interactions are facilitated when binding sites display a convergent motif orientation^{130,131}. Although still under investigation, the most accepted model for TAD formation invokes a prominent role for the CTCF–cohesin complex, which is thought to promote the extrusion of DNA through a cohesin ring until it reaches a pair of CTCF molecules in convergent orientation, where it can be retained until its dissociation^{130,131}. According to these studies, the CTCF–cohesin complex proteins play a key part in the looping process, which explains many of the features observed on a HiC interaction map. However, the depletion of either cohesin or CTCF does not abolish chromatin organization completely^{135,136}. In this case, another mode of chromatin organization emerges, which relies heavily on transcription and the organization into A and B compartments. Thus, the observed interaction maps result from a complex interplay between different biological processes that need to be carefully considered when interpreting chromosome conformation capture data.



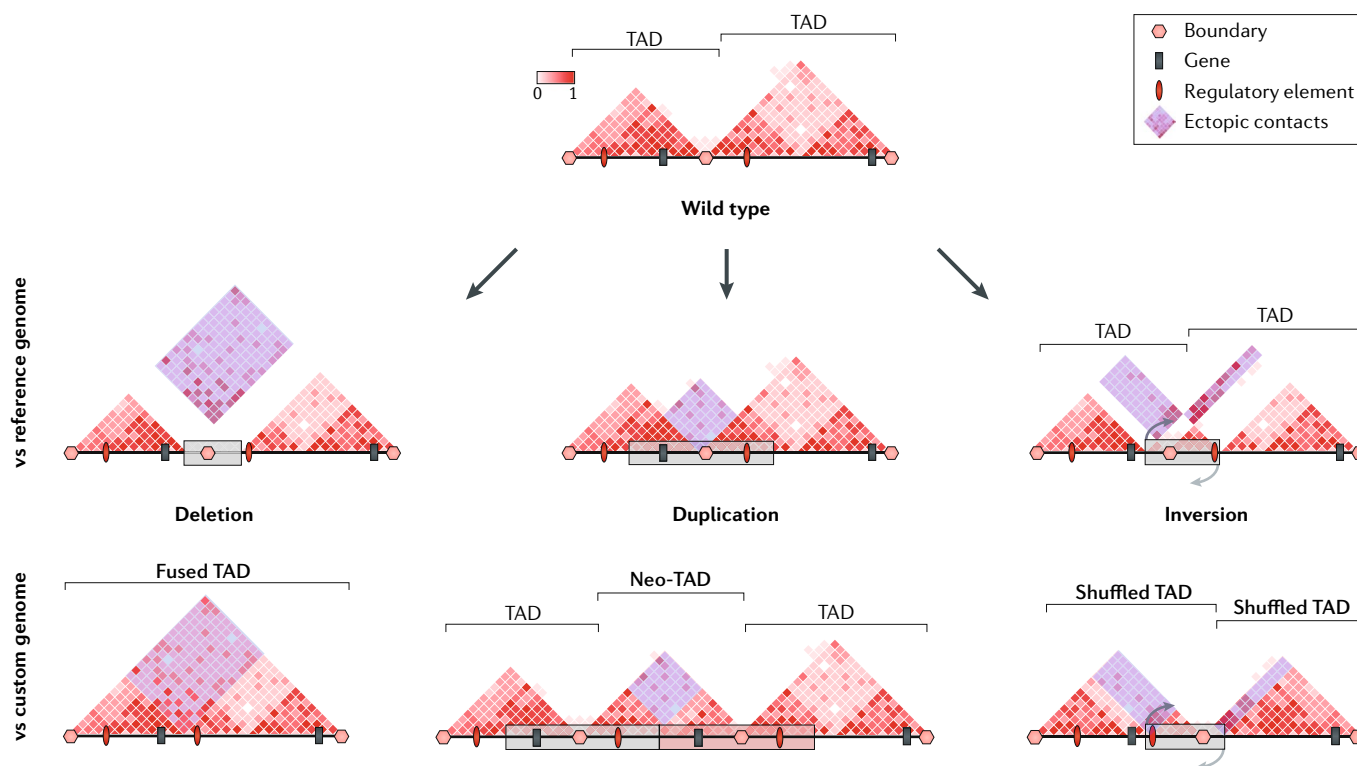


Fig. 1 | HiC as a tool to study the effect of structural variation on chromatin folding. Structural variation can induce dramatic changes in chromatin organization, thus creating specific signatures that are noticeable by visual inspection of interaction maps. The high degree of topologically associating domain (TAD) conservation across cell types or during development facilitates the identification of such conformational changes even when the affected cell type is not accessible. The schematics display HiC maps and their corresponding representation of the locus, showing diverse genomic elements (boundaries, genes and regulatory elements). By comparison against a wild-type reference HiC map (top panel), ectopic interactions resulting from alterations in spatial proximity become apparent (represented in blue); these changes can be identified even in the heterozygous state. However, this approach requires that the underlying sequence is relatively unique in the genome. The changes provide specific information about the nature of the variant (that is, whether it is a deletion, duplication, inversion or translocation; marked by grey rectangles) as well as the position of the breakpoints at nucleotide resolution. Furthermore, by mapping HiC data against a genome that considers the identified variant (bottom panels), potential pathomechanisms can be observed, such as TAD fusion (deletion), neo-TAD formation (duplication) or TAD shuffling (inversion).

enhancer of *SHH*, have been shown to functionally interact across a TAD boundary element when their genomic distance is reduced beneath a certain threshold⁸².

Enhancer redundancy

Genome-wide studies as well as studies at individual loci suggest that developmental genes are usually controlled by multiple enhancer elements with tissue-specific activity. Single enhancers can be responsible for controlling expression patterns in a specific tissue and can display phenotypical effects by loss of function or misexpression upon mutation^{83,84}. However, such regulation seems to be the exception, and there is growing evidence that tissue-specific transcription is generally controlled by multiple enhancer elements that work in cooperation to achieve the necessary precision of gene expression⁸⁵. Accordingly, the activity of enhancers overlaps in parts or completely, a phenomenon termed enhancer redundancy or 'shadow enhancers' (REF.⁸⁶). The term 'super enhancers' was further used to describe clusters of redundant enhancers with high transcriptional activity that seem to control important genes and that are

associated with cancer⁸⁷. Different types of cooperation between enhancers have been described on the basis of their impact on gene expression, including additive, synergistic, competitive or even repressive functions (reviewed in REF.⁸⁸). However, the in vivo importance of enhancer redundancy is still a matter of debate.

A recent in-depth analysis of the regulatory landscape of *IHH* identified nine enhancer elements that showed overlapping regulatory activities in the finger tips, the growth plates and the sutures of the skull⁸⁹. Deletions of these enhancer elements in transgenic mice resulted in variable loss of function and corresponding phenotypes, as previously observed in an *Ihh* complete knockout model⁹⁰. By contrast, duplications led to tissue-specific upregulation and misexpression of *Ihh*, causing syndactyly and malformation of the skull. This study shows that enhancers are not single independent elements that together add up to the expression pattern of a target gene but rather form a complex and redundant regulatory unit that provides robust and precise regulation of expression⁸⁹. Depending on the degree of redundancy and sensitivity, alterations of the composition and dosage of parts

or the entire unit can result in alterations of gene expression. Redundancy has also been observed in vivo at the α -globin or the *Wap* locus, although with distinct functional interplays between enhancers^{91,92}. Furthermore, CRISPR interference (CRISPRi) assays have been used to dissect the *cis*-regulatory logic of genes associated with human disease, such as *GATA1*, which encodes erythroid transcription factor, or the proto-oncogene *MYC*⁹³. At these loci, regulatory elements act redundantly as deduced from their individual disruption, affecting both gene expression and cell proliferation. It should be considered that the relevance of enhancer redundancy can be easily overseen in animal models, especially if perturbations have subtle effects on fitness, which are usually difficult to measure in a laboratory setting⁹⁴. Nevertheless, recent studies strongly support enhancer redundancy as a widespread phenomenon that provides developmental robustness to mammalian gene expression⁹⁵.

Of note, enhancers are capable of discriminating between different types of promoters, thus adding an additional layer of regulation to regulatory landscapes^{76,96}. Overall, this delineates a complex scenario whereby genes are controlled by regulatory elements with their own affinity and functional relationships between them. These properties are further constrained by their specific arrangement in the linear genome, as well as the overall 3D organization of chromatin.

The impact of SVs on gene regulation

Although the interpretation of SVs, in particular of CNVs, has been focused on the coding part of the genome, it is important to note that besides changing gene dosage, deletions and duplications can also change the dosage and position of *cis*-regulatory elements, as well as the higher-order chromatin organization of a locus. Similarly, copy-number-neutral SVs, such as inversions and translocations, have the potential to disrupt coding sequences or create fusion transcripts, but these types of variants can also disrupt or create new enhancer landscapes and chromatin domains, resulting in regulatory loss or gain of function. Clinical examples of SVs that change the 3D architecture of the genome are shown in FIG. 2.

Intra-TAD SVs: effects on enhancer dosage

SVs affecting *cis*-regulatory elements can have an effect on genes located several hundred kilobases away. Examples include duplications of the *SHH* limb enhancer element ZRS, which cause polydactyly^{97,98}, and deletions in the regulatory landscape of *SOX9*, which are associated with Pierre Robin sequence⁹⁹. The main challenge has been to identify the potential disease-associated gene in the surrounding regulatory landscape of the SV. With the discovery of TADs, it became evident that regulatory elements within a domain can, in principle, regulate any gene within the domain^{74,100}. These findings have direct implications for the medical interpretation of SVs located within regulatory domains (intra-TAD SVs), as the search for potential target genes of an SV can be limited to the genes within the domain. Similar to the concept of gene dosage, intra-TAD SVs of regulatory elements can change the enhancer dosage and may result in loss or gain of function of their endogenous target gene (FIGS. 2a, 3).

The most common strategy to identify the responsible regulatory elements is to map overlapping CNV or SV breakpoints and establish a minimal critical region^{101,102}. Most developmental genes show complex expression patterns and are regulated by several enhancer elements^{67,89}. Disruption of specific enhancers by SVs within TADs usually results in tissue-specific loss of function, resembling selective phenotypic sub-features of coding mutations³¹. Examples of intra-TAD SVs that cause regulatory loss of function have been described at several genomic loci, including *SOX9* (REF.⁹⁹), *DLX5* and *DLX6* (REFS^{103–105}), *ATOH7* (REF.¹⁰⁶) and *PAX6* (REF.¹⁰⁷). However, given that most enhancers are highly redundant, the loss of individual elements is likely to often be tolerated, in particular when the loss is heterozygous.

In contrast to deletions, intra-TAD duplications of regulatory elements can result in tissue-specific overexpression and misexpression of the endogenous target genes and disease^{97,99,108,109}. One example of enhancer dosage is a set of duplications involving regulatory elements within the regulatory domain of *IHH* that are associated with highly tissue-specific phenotypes, including craniosynostosis and synpolydactyly¹¹⁰. In this case, craniosynostosis strictly associates to *IHH* overexpression, whereas the observed synpolydactyly is the result of tissue-specific misexpression⁸⁹. Further examples are duplications of an enhancer region that lead to increased expression of *CTSB*, the gene encoding cathepsin B, resulting in the skin disorder keratolytic winter erythema¹¹¹.

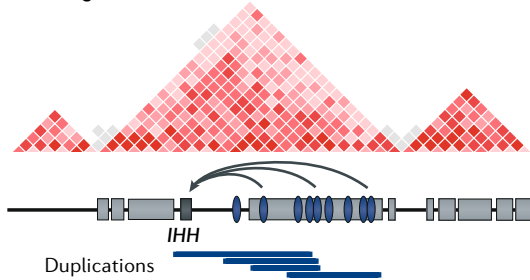
Another example of intra-TAD SVs are duplications within the regulatory domain of *SOX9* that are associated with a female to male sex reversal phenotype¹¹². Capture HiC, which combines HiC with the hybridization-based capture of targeted genomic regions to enable high-resolution mapping of genomic interactions, in transgenic

Fig. 2 | Clinical examples of structural variations in the 3D genome. a | Duplications of enhancer elements at the *IHH* locus that occur within topologically associating domains (intra-TAD) cause tissue-specific misregulation and are associated with synpolydactyly of the feet⁸⁹. For examples see REFS^{97,104,107,137,138}. **b** | Duplication of a TAD boundary at the *SOX9* locus causes neo-TAD formation and is associated with Cooks syndrome, short digits and nail aplasia⁶³. For examples see REFS^{117,121}. **c** | Deletion of a TAD boundary at the *LMNB1* locus causes enhancer adoption and adult-onset demyelinating leukodystrophy¹³⁹. For examples see REFS^{117,121,140}. **d** | Inversions of an enhancer cluster at the *EPHA4* locus cause enhancer adoption and misregulation of *WNT6* and are associated with F-syndrome, syndactyly of thumb and index finger⁷⁶. For examples see REFS^{102,114,141}. **e** | Balanced translocations at the *MEF2C* locus cause a regulator loss of function and are associated with anomalies of the brain (including callosum hypoplasia¹⁴²) and developmental delay¹¹⁶. For examples see REFS^{99,104,143}. PRS, element associated with Pierre Robin sequence; RevSex, element associated with disorders of sex development. Part **c** is adapted with permission from REF.¹¹³, Elsevier. Part **e** is adapted with permission from Shimojima, K et al. De novo microdeletion of 5q14.3 excluding *MEF2C* in a patient with infantile spasms, microcephaly, and agenesis of the corpus callosum. *Am. J. Med. Genet. Part A*, REF.¹⁴², Copyright 2011 Wiley Periodicals, Inc.

Intra-TAD SVs

Structural variation (SV) that occurs within topologically associating domains (TADs). Within a regulatory domain, SVs of regulatory elements can change the enhancer dosage and may result in a tissue-specific loss of function (deletion) or a gain of function (duplication) of their endogenous target gene, which can be located anywhere within the TAD.

a Intra-TAD gain of function



Phenotype

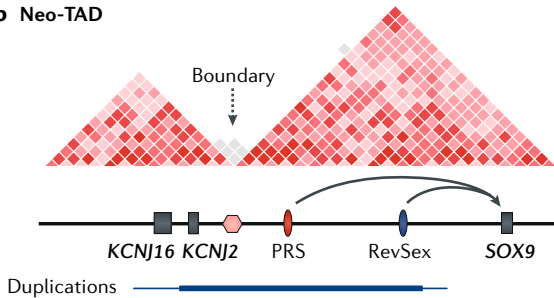


Duplications of enhancer elements cause preaxial synpolydactyly of feet

Examples

- Gain of function:**
- *SOX9* locus: duplications of gonad enhancer cause 46,XX sex reversal
 - *BCL6* locus: duplications of super enhancers cause B cell lymphomas
 - *SHH* locus: duplications of limb enhancer causes polydactyly
- Loss of function:**
- *PAX6* locus: aniridia
 - *DLX5* and/or *DLX6* loci: split hand foot malformation
 - *SOX9* locus: deletions of gonad enhancer cause 46,XY sex reversal

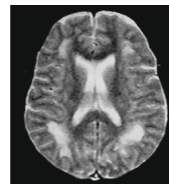
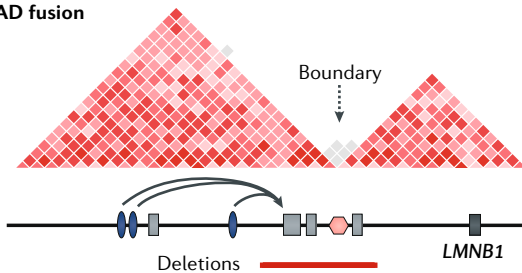
b Neo-TAD



Cooks syndrome: Duplications of TAD boundary, *KCNJ2* and *KCNJ16* cause aplasia of nails and short digits

- *FGF2* locus: colorectal cancer
- *PRDM6* locus: medulloblastoma

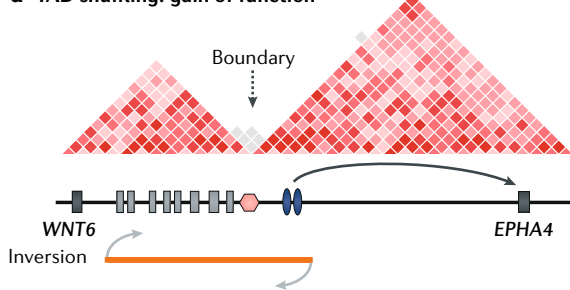
c TAD fusion



Adult-onset demyelinating leukodystrophy

- *GFI1* locus: medulloblastoma
- *TAL1* and *LMO2* loci: T cell acute lymphoblastic leukaemia
- *IRS4* locus: lung squamous carcinoma, sarcoma and cervical squamous carcinoma
- *SOX4* locus: mesomelic dysplasia

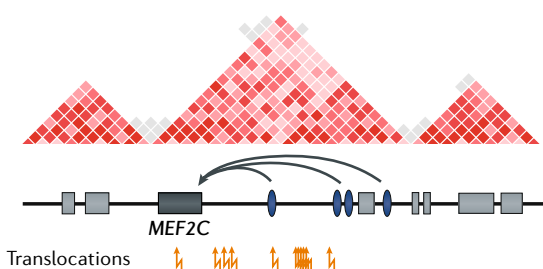
d TAD shuffling: gain of function



F-syndrome: syndactyly

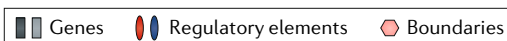
- *SHH* locus: inversion of enhancer causes short digits (Dsh mouse model)
- *SHH* locus: inversion of enhancer causes polysyndactyly
- *GFI1* locus: medulloblastoma
- Translocation at the *PITX1* locus: Liebenberg syndrome

e TAD shuffling: loss of function



Hypoplastic corpus callosum via loss of function of *MEF2C* at 5q14.3

- *FOXG1* locus: atypical Rett syndrome
- *SOX9* locus: campomelic dysplasia
- *DLX5* and *DLX6* loci: split hand foot malformation



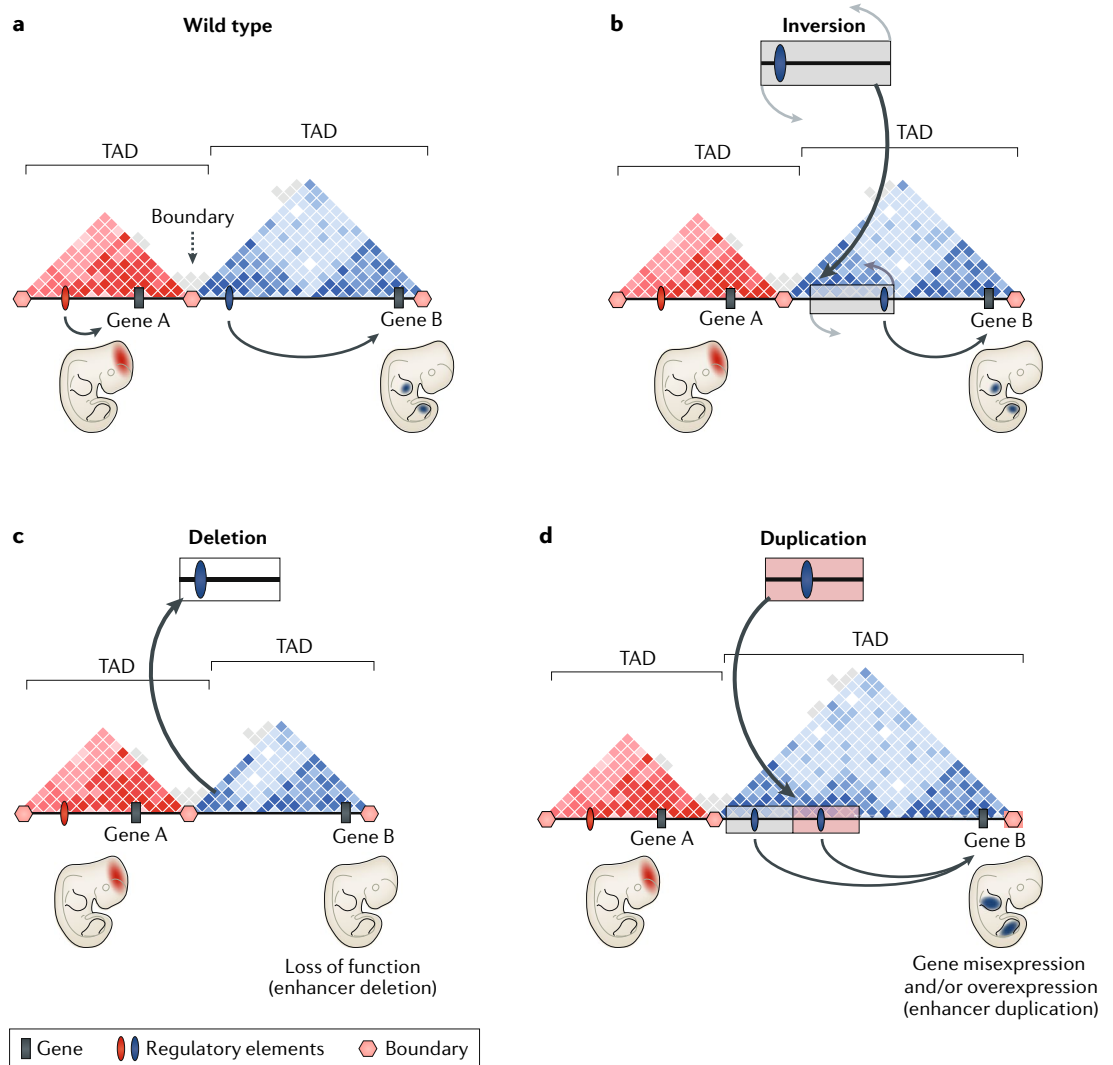


Fig. 3 | Structural variation not disrupting TAD structures. Structural variation (SV) located within topologically associating domains (intra-TAD) can change enhancer dosage and may result in loss or gain of function of their endogenous target genes without changing the overall regulatory landscape. SVs shown were selected on the basis of the presence of a known mechanism and several examples in the literature. **a** | In the schematic representation of the wild-type genomic locus, gene A is expressed in the developing brain and gene B is expressed in the developing limbs. Both genes are regulated by their own tissue-specific cis-regulatory elements (red and blue, respectively) located in different TADs separated by boundary elements. **b** | Intra-TAD inversions not disrupting a coding gene or TAD boundary probably have no major influence on long-range gene regulation, although no cases have been reported so far. **c** | The intra-TAD deletion of an enhancer element can result in the loss of a specific expression domain of gene B in the limb (blue). Such a specific loss might represent a phenotypic subset of features observed in complete loss of function, for example, by a coding mutation. **d** | Intra-TAD duplications of regulatory elements can result in dose-dependent upregulation but also tissue-specific misexpression in the limbs (blue) of gene B and disease (see also FIG. 2a).

mice carrying the corresponding duplication revealed that higher-order chromatin domains are not affected by the duplication⁶³. However, the contact between the duplicated enhancer elements and the *Sox9* promoter was increased, likely causing overexpression of *Sox9* during early ovarian development, which in turn resulted in sex reversal.

No pathogenic intra-TAD inversion has been described so far. Because this class of variation remains largely uncharacterized, it is difficult to predict what the effects may be. However, according to current concepts, inversions that do not disrupt a coding gene or TAD boundary are not expected to have a major

influence on long-range gene regulation. With the broader introduction of clinical WGS, we can expect to find more cases of small intra-TAD SVs that disturb the fragile composition of enhancer landscapes and cause disease.

Inter-TAD SVs: effects on 3D architecture

CNVs and copy number-neutral SVs, such as inversions, insertions and translocations, have the potential to disrupt higher-order chromatin organization, thereby rewiring the complex 3D chromatin organization of a locus¹¹³. This may lead to the repositioning of

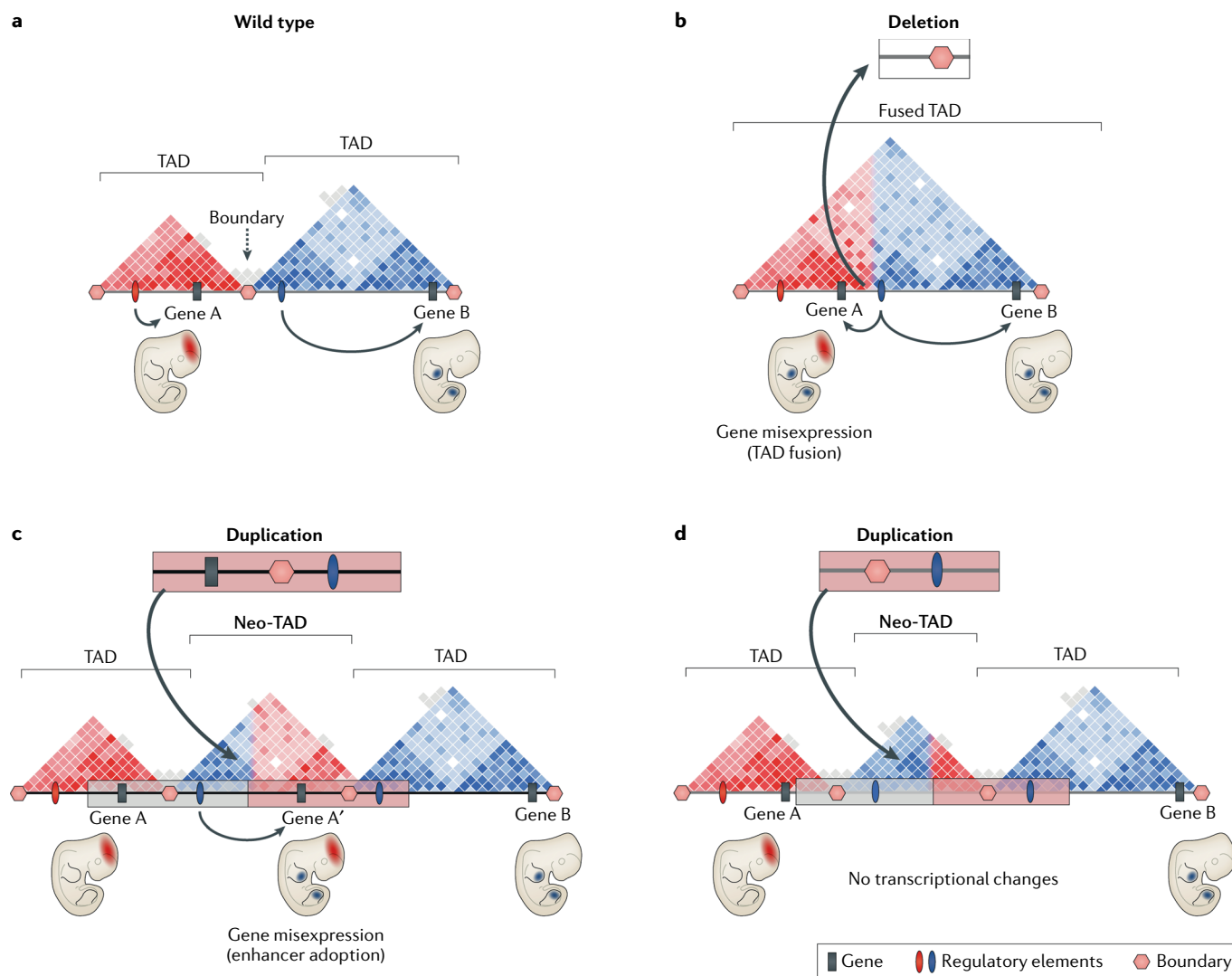


Fig. 4 | TAD disruption: deletions and duplications. Structural variations that cross topologically associating domain (TAD) boundaries (inter-TAD) have the potential to disrupt chromatin domains and create new regulatory units. **a** | In the schematic representation of the wild-type genomic locus, gene A is expressed in the developing brain and gene B in the developing limbs. Both genes are regulated by their own tissue-specific cis-regulatory elements (red and blue, respectively) located in different TADs separated by boundary elements. **b** | An inter-TAD deletion of a boundary element can cause TAD fusion and enhancer adoption; the relocation of enhancer elements into a neighbouring TAD causes misexpression and disease. Through the deletion of the boundary, the enhancer of gene B (blue) is free to act on gene A, driving ectopic expression in the developing limbs (see also FIG. 2c). **c** | Duplications of the boundary and the flanking gene A (inter-TAD duplication) change the overall chromatin architecture of the locus, creating a new chromatin domain (neo-TAD). In the neo-TAD, the duplicated gene A (gene A') is regulated under the control of the duplicated enhancer of gene B, driving ectopic expression in the developing limbs (see also FIG. 2a). **d** | Inter-TAD duplications that do not include gene A can also result in neo-TAD formation but have not been associated with any transcriptional changes⁸².

Enhancer adoption

A phenomenon that refers to a regulatory gain of function mutation in which ectopic expression of a gene is driven by an enhancer that normally regulates another gene located in a different regulatory domain; also known as enhancer hijacking.

TAD boundaries and/or the relocation of enhancer elements into other compartments, causing misexpression and disease. These position effects on gene regulation have been known since the early days of karyotyping and have recently been termed enhancer adoption¹¹⁴ or enhancer 'hijacking' (REF.¹¹⁵) (FIGS. 4, 5). In addition, inversions and translocations can also disrupt existing chromatin domains by removing enhancer elements from their endogenous target genes, resulting in regulatory loss of function¹¹⁶. Chromatin domain disruption has been reported as an important disease mechanism for developmental disorders and cancer¹¹⁷⁻¹¹⁹.

TAD fusion. SVs that cross TAD boundaries (inter-TAD SVs) have the potential to disrupt chromatin domains (FIGS. 2c, 4). One of the most prominent examples linking TADs to human disease is a large deletion including CTCF-associated boundary elements at the *EPHA4* locus. This deletion results in the ectopic interaction of an enhancer cluster and genes that are normally separated, causing misexpression and congenital limb malformation⁷⁶. Circular chromatin conformation capture (4C), also called 4C-seq, which allows the unbiased detection of all genomic regions that interact with a particular region of interest, confirmed that deletions that

include the boundary led to ectopic chromatin contacts, whereas similar deletions that left the boundary intact prevented this interaction, resulting in a normal phenotype⁷⁶. Analysis of the SVs by HiC revealed a fusion of two neighbouring TADs upon deletion of the boundary but the presence of two separate TADs when the boundary was

intact⁶². Thus, the presence of the CTCF-associated boundary element prevents ectopic contacts and insulates the TAD from neighbouring enhancers. Several subsequent studies have shown TAD disruption to be a common phenomenon in cancer development. For example, in T cell acute lymphoblastic leukaemia, recurrent microdeletions

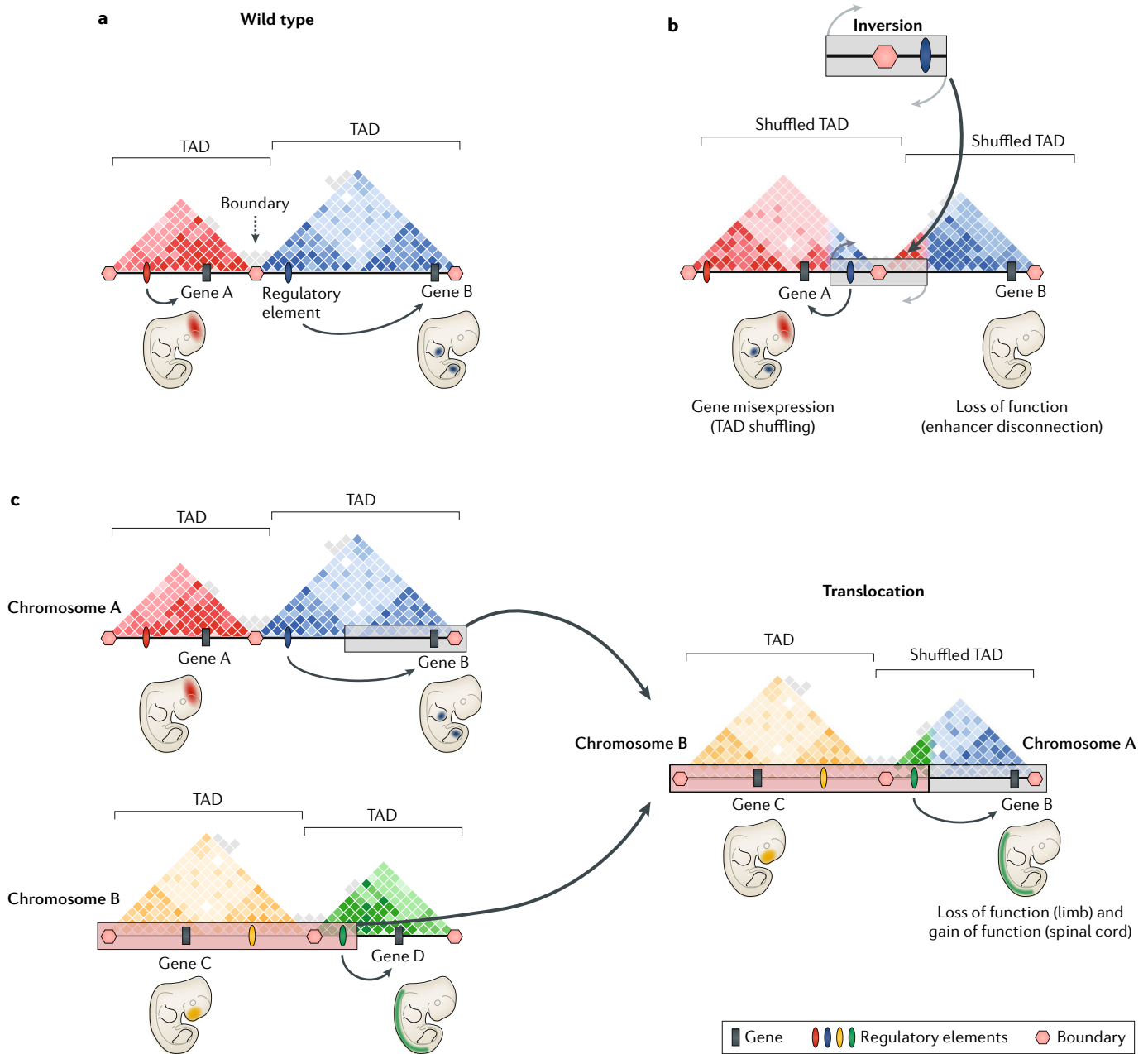


Fig. 5 | TAD shuffling: inversions and translocations. **a** | Balanced structural variations that cross topologically associating domain (TAD) boundaries (inter-TAD) have the potential to cause regulatory gain and/or loss of function. In the schematic representation of the wild-type genomic locus, gene A is expressed in the developing brain and gene B in the developing limbs. Both genes are regulated by their own tissue-specific cis-regulatory elements (red and blue, respectively) located in different TADs separated by boundary elements. **b** | The inter-TAD inversion of an enhancer element (blue) can cause enhancer adoption, causing ectopic expression of gene A in the limbs. At the same time, the inversion removes the enhancer element (blue) from gene B, resulting in regulatory loss of function in the limbs (see also FIG. 2d). **c** | Balanced translocations can reshuffle the TAD architecture of the genome by replacing the limb enhancer of gene B with a spinal cord enhancer of gene D. This results in the loss of expression of gene B in the limbs and ectopic expression in spinal cord. The loss or gain of gene D is not shown. Alternatively, this type of rearrangement can result only in loss of function or gain of function, depending on the breakpoint and the flanking genes. Translocations can also be unbalanced.

remove CTCF-associated domain boundaries, allowing proto-oncogene activation¹²⁰. CRISPR–Cas9-induced deletions of these boundaries in non-malignant cells were sufficient to activate proto-oncogenes. However, it should be pointed out that the deletion of only the boundary including its CTCF sites had no major effect at other loci (namely, *Sox9*)⁶³, indicating that the stability of TADs does not depend solely on the presence or absence of boundaries. Another study in patients with medulloblastoma identified recurrent somatic SVs that relocate the proto-oncogenes *GFI1* and *GFI1B* in close proximity to active super enhancers, initiating oncogenic activity¹¹⁵. More recently, the *cis*-expression structural alteration mapping (CESAM) algorithm was developed as a framework to systematically predict cancer-related gene overexpression resulting from TAD disruption¹²¹. The CESAM algorithm combines genome-wide expression and whole-genome SV data and analyses them in the context of chromatin domains. This approach identified a subset of recurrent SVs at the *SNCAIP* locus in 491 sequenced medulloblastoma samples that results in enhancer hijacking and subsequent overexpression of *PRDM6* (REF.¹¹⁷).

Neo-TADs. SVs also have the potential to create new regulatory units. For example, intra-TAD duplications of non-coding DNA at the *SOX9* locus cause female to male sex reversal¹¹² (FIG. 3d). By contrast, overlapping duplications that extend over the next boundary into the neighbouring regulatory domain including the next flanking gene, *KCNJ2*, are associated with isolated distal digit abnormalities (Cooks syndrome)¹²² (FIGS. 2b, 4c). Capture HiC and 4C-seq analyses in transgenic mice and human cells indicated that inter-TAD duplications change the overall chromatin architecture of the locus, creating new chromatin domains (neo-TADs) that are insulated from the rest of the genome⁶³. In this case, the neo-TAD included the duplicated *Kcnj2* gene, causing misexpression of *Kcnj2* under the control of the duplicated *Sox9* enhancers, which results in limb malformations. By contrast, slightly smaller deletions that do not include the *Kcnj2* gene also resulted in neo-TAD formation but did not cause misexpression, as no gene was located within the regulatory domain⁶³ (FIG. 4d). The phenomena of neo-TAD formation and enhancer hijacking have also been observed in medulloblastoma cell lines¹¹⁷.

TAD shuffling. Inversions that cross TAD boundaries or translocations can result in the fusion of two regulatory domains that do not naturally belong together ('TAD shuffling'), causing enhancer adoption (FIGS. 2d, 5). Owing to the difficulties in detecting inversions at the submicroscopic level with current technologies, very few of such events have been described. One case is an inversion at the *EPHA4* locus in which an enhancer cluster, normally controlling *EPHA4* expression, is relocated into the vicinity of *WNT6* (REF.⁷⁶). This change results in the activation of *WNT6* in an *EPHA4*-like pattern and, consequently, in abnormal limb development. At the same time, the inversion disconnects *EPHA4* from its native enhancers, resulting in gene loss of function.

The disconnection of regulatory regions as a cause of developmental disorders is an increasingly recognized

mechanism. A recent study systematically mapped the breakpoints of 273 balanced translocations at nucleotide resolution by WGS in individuals with congenital anomalies. Although approximately 34% of balanced translocations resulted in gene disruption, >7% disrupted TADs encompassing known syndromic loci¹¹⁶. Eight patients carried balanced translocations separating the *MEF2C* gene from its endogenous enhancer elements, suggestive of a regulatory loss of function (FIG. 2e). Subsequent functional experiments in patient-derived lymphoblastoid cell lines demonstrated that the disruption of the chromatin domain was associated with decreased *MEF2C* expression. In a follow-up study¹²³, the same group computationally predicted the effects of balanced translocations in 17 individuals with developmental delay and neurological conditions. By combining phenotypic data from the Human Phenotype Ontology with WGS, HiC and enhancer data, it was possible to predict the position effect on several disease-associated target genes and validate transcriptional changes in patient-derived cell lines. An earlier study applied a similar approach to deletions¹²⁴, thus highlighting the relevance of TAD disruption as a mechanism of human disease.

Interpretation of SVs in the 3D genome

The study of chromatin domain disruption has several important implications for a meaningful medical interpretation of SVs. First, breakpoint mapping at nucleotide resolution is essential to accurately assess which genomic feature — TAD boundaries, regulatory elements, CTCF sites or genes — are affected by the aberration. Second, to predict the phenotypic effect of intra-TAD and inter-TAD SVs, the traditional minimal critical region approach is insufficient, as shown, for example, in the case of *SOX9* overlapping duplications, which show completely unrelated phenotypes depending on the size, content and position of the aberration⁶³. Third, TAD shuffling may lead to a gain of function via enhancer adoption or a regulatory loss of function, but in many cases, it can also be a combination of both. Considering the multiple effects that SVs can have, we propose a systematic approach when interpreting SVs in a clinical context (FIG. 6).

While it is difficult to predict how frequent 3D position effects are, recent studies indicate that this might be phenotype specific. Two studies suggest that the activation of cancer genes by enhancer hijacking is a fairly common process that may be comparable to recurrent in-frame gene fusions^{117,121}. In neurodevelopmental disorders, approximately 7% of balanced translocations cause TAD disruption¹¹⁶. By contrast, in congenital limb malformation, 57% of disease-associated CNVs were shown to result in *cis*-regulatory position effects³¹.

Conclusions and future developments

The technological developments of recent years have dramatically accelerated the discovery and characterization of SVs, expanded our knowledge of the 3D nature of chromatin folding and assigned functionality to large parts of the non-coding genome. The integration of all these layers of information is crucial to understand the true pathogenic potential of SVs in the

Inter-TAD SVs

Structural variations (SVs) that occur between topologically associating domains (TADs). Such SVs affect several regulatory domains and have the potential to disrupt and rearrange the complex 3D chromatin organization of a locus by repositioning TAD boundaries, genes and enhancer elements. These position effects may lead to TAD fusion (deletion), neo-TAD formation (duplications) or TAD shuffling (inversion and/or translocation).

TAD shuffling

The reordering of topologically associating domains (TADs) within the genome. Inversions that cross TAD boundaries or translocations can result in the fusion of two regulatory domains that do not naturally belong together, causing enhancer adoption. At the same time, these structural variations can result in regulatory loss of function by removing enhancer elements from their target genes.

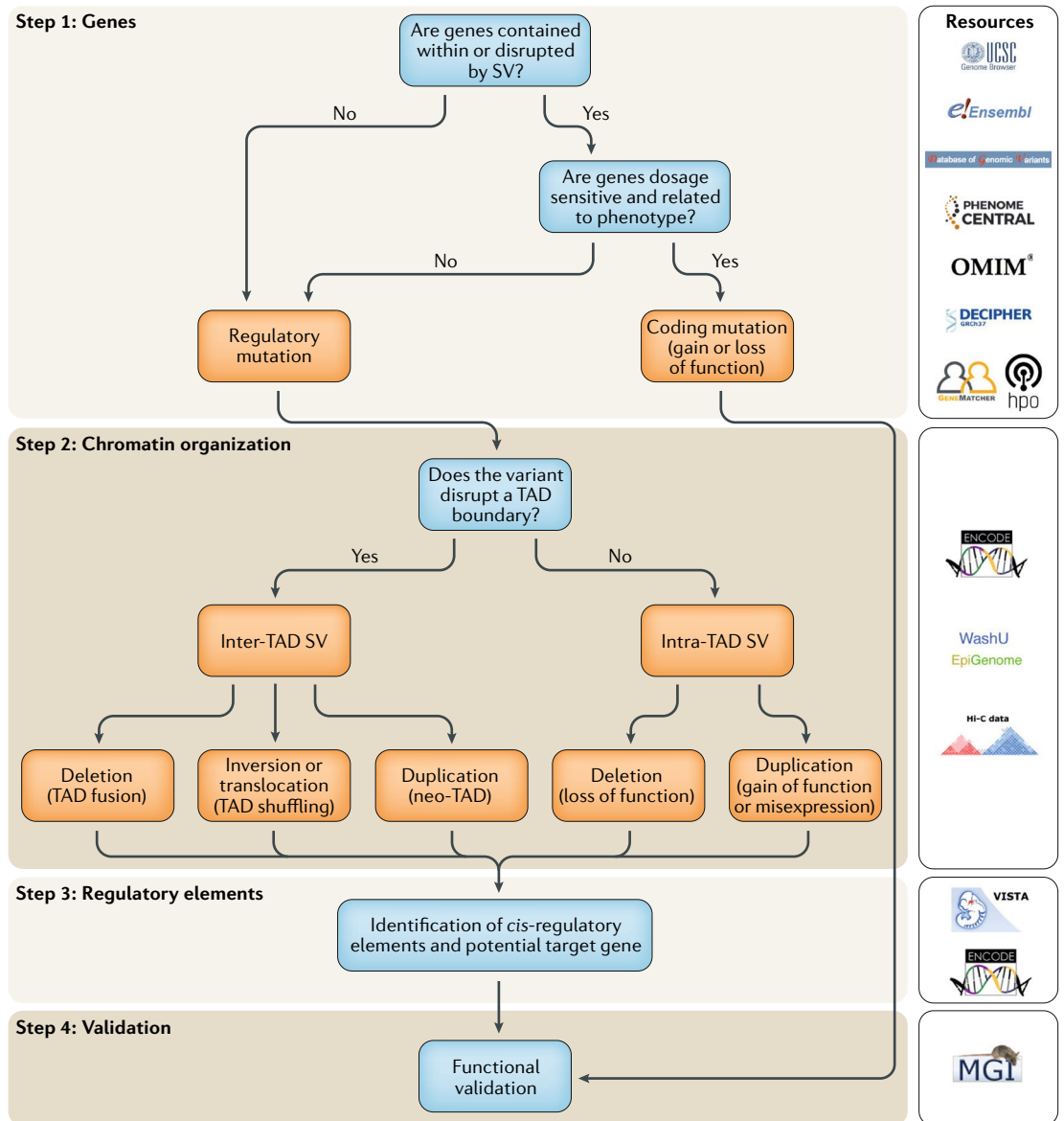


Fig. 6 | Possible strategies to interpret structural variations in the 3D genome. Identify genes within copy number variants or structural variation (SV) (step 1). If the rearrangement is balanced, identify possibly disrupted genes. Check whether genes are dosage-sensitive and/or related to the phenotype. If no (disease) genes are located in the rearrangement, or if genes are not dosage-sensitive, the SV should be positioned in relation to topologically associating domains (TADs) and boundaries on the basis of available HiC data (for example, from the 3D Genome Browser) to determine whether SVs disrupt a TAD or are located within a TAD (step 2). Classify SVs on the basis of principles shown in FIGS. 3, 4, 5. Identify potential enhancers that might drive ectopic expression of a disease gene (step 3). If possible, correlate ectopic enhancer–promoter interactions with gene expression (in tumours) or phenotype (in congenital diseases) (step 4). Interpret findings and, if necessary, perform functional validation.

3D genome and to predict their phenotypic consequences. Currently, several challenges remain: first, the detection rate of SVs and the determination of exact breakpoints remain insufficient. This is in part because breakpoints preferentially occur in highly repetitive regions of the genome, which are poorly covered by short-read WGS. Second, relevant data are scattered throughout different databases, and an interdisciplinary team is needed to evaluate individual SVs. Third, our knowledge of the non-coding regulatory genome is still very limited, a fact that makes prediction difficult.

Fourth, computational prediction tools are currently insufficient, and in-depth functional studies including animal models are necessary to assign pathogenicity to each variant. Fifth, SVs are still mainly interpreted using the gene-dosage approach, and the overall awareness of non-coding 3D position effects is low, leaving many patients without a diagnosis.

Several studies are ongoing to address the above-mentioned problems. For example, a recent study combined short-read and long-read WGS to resolve SVs in Mendelian disease at breakpoint resolution¹²⁵.

Also, new bioinformatics approaches, such as the CESAM algorithm, are being developed to systematically predict cancer-related gene overexpression resulting from chromatin domain shuffling¹²¹. The analysis framework Genomiser offers an approach for the medical identification of regulatory variants in Mendelian disease¹²⁶. Another approach is based on polymer modelling, which has recently been introduced to study the 3D organization of chromatin^{127–131}. The polymer physics-based approach (PRISMR) algorithm can successfully model 3D chromatin folding and predict enhancer–promoter contacts observed in HiC data. More importantly, PRISMR can be used to predict interactions in silico, thereby providing a tool for analysing the disease-causing potential of SVs in the absence of experimental data⁶². However, in silico interpretations rely on detailed knowledge of biological mechanisms. Here, major gaps remain that limit our ability to predict the effect of rearrangements. Pressing questions include the identification of factors and mechanisms that control enhancer–promoter specificity, a functionally

relevant definition of TADs, the identification of all relevant factors that govern 3D chromatin folding and the prediction of enhancers and their specific activity and importance.

Thus, there are substantial hurdles that need to be overcome before the effect of SVs can be reliably predicted in the clinic. The introduction of third-generation sequencing approaches, which enable the mapping of SVs at single base pair resolution even in repetitive regions, the measurement of their effect within the 3D nucleus, the analysis of their transcriptional consequences and a better overall understanding of gene regulation in a genomic context will pave the way to reach this goal. On the basis of how quickly NGS technologies have transformed the field of cytogenetics in combination with the recent advances in computational modelling, we expect that it will be possible to accurately predict the transcriptional effects of SVs in the 3D genome in the near future.

Published online 24 April 2018

- Sudmant, P. H. et al. An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).
This study provides a comprehensive structural variation map of >25,000 healthy human genomes.
- Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **46**, 818–825 (2014).
- Kloosterman, W. P. et al. Characteristics of de novo structural changes in the human genome. *Genome Res.* **25**, 792–801 (2015).
- Weischenfeldt, J., Symmons, O., Spitz, F. & Korbel, J. O. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat. Rev. Genet.* **14**, 125–138 (2013).
- Zarrei, M., MacDonald, J. R., Merico, D. & Scherer, S. W. A copy number variation map of the human genome. *Nat. Rev. Genet.* **16**, 172–183 (2015).
- Sebat, J. et al. Strong association of de novo copy number mutations with autism. *Science* **316**, 445–449 (2007).
- Soemedi, R. et al. Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. *Am. J. Hum. Genet.* **91**, 489–501 (2012).
- Xu, B. et al. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat. Genet.* **40**, 880–885 (2008).
- Walsh, T. et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539–543 (2008).
- Cooper, G. M. et al. A copy number variation morbidity map of developmental delay. *Nat. Genet.* **43**, 838–846 (2011).
- Yang, L. et al. Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* **153**, 919–929 (2013).
- de Leeuw, N. et al. Diagnostic interpretation of array data using public databases and internet sources. *Hum. Mutat.* **33**, 930–940 (2012).
- Biederman, B. & Bowen, P. Balanced translocations involving chromosome 12: report of a case and possible evidence for position effect. *Ann. Genet.* **19**, 257–260 (1976).
- Hecht, F. & Kaiser-McCaw, B. Position effect in 8;14 translocation in Burkitt's lymphoma. *N. Engl. J. Med.* **304**, 174–175 (1981).
- Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
- Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
- The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Carvalho, C. M. & Lupski, J. R. Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* **17**, 224–238 (2016).
- Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.* **16**, 245–257 (2015).
- Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 772 (2016).
- Krijger, P. H. & de Laat, W. Regulation of disease-associated gene expression in the 3D genome. *Nat. Rev. Mol. Cell Biol.* **17**, 771–782 (2016).
- Alkan, C., Coe, B. P. & Eichler, E. E. Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **12**, 363–376 (2011).
- Wright, C. F., FitzPatrick, D. R. & Firth, H. V. Paediatric genomics: diagnosing rare disease in children. *Nat. Rev. Genet.* **19**, 253–268 (2018).
- Gilissen, C. et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344–347 (2014).
This landmark study uses whole-genome sequencing to identify complex de novo SV in families with intellectual disability, including single exon deletions and insertional duplications.
- Cretu Stancu, M. et al. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat. Commun.* **8**, 1326 (2017).
- Huddleston, J. et al. Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Res.* **27**, 677–685 (2017).
- Chaisson, M. J. P. et al. Multi-platform discovery of haplotype-resolved structural variation in human genomes. Preprint at *bioRxiv* <https://doi.org/10.1101/193144> (2017).
- de Vries, B. B. et al. Diagnostic genome profiling in mental retardation. *Am. J. Hum. Genet.* **77**, 606–616 (2005).
- Talkowski, M. E. et al. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* **149**, 525–537 (2012).
- Marshall, C. R. et al. Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82**, 477–488 (2008).
- Flottmann, R. et al. Noncoding copy-number variations are associated with congenital limb malformation. *Genet. Med.* <https://doi.org/10.1038/gim.2017.154> (2017).
- Sanders, S. J. et al. Insights into autism spectrum disorder genome architecture and biology from 71 risk loci. *Neuron* **87**, 1215–1233 (2015).
- The 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- Hehir-Kwa, J. Y. et al. A high-quality human reference panel reveals the complexity and distribution of genomic structural variants. *Nat. Commun.* **7**, 12989 (2016).
- Brandler, W. M. et al. Frequency and complexity of de novo structural mutation in autism. *Am. J. Hum. Genet.* **98**, 667–679 (2016).
- Goldmann, J. M. et al. Parent-of-origin-specific signatures of de novo mutations. *Nat. Genet.* **48**, 935–939 (2016).
- Lelieveld, S. H. et al. Spatial clustering of de novo missense mutations identifies candidate neurodevelopmental disorder-associated genes. *Am. J. Hum. Genet.* **101**, 478–484 (2017).
- Turner, T. N. et al. Genomic patterns of de novo mutation in simplex autism. *Cell* **171**, 710–722.e12 (2017).
This study combines six SV calling algorithms to identify all de novo SVs in 516 families with autism.
- Turner, T. N. et al. Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory DNA. *Am. J. Hum. Genet.* **98**, 58–74 (2016).
- Brandler, W. M. et al. Paternally inherited noncoding structural variants contribute to autism. Preprint at *bioRxiv* <https://doi.org/10.1101/102327> (2017).
- Werling, D. M. et al. Limited contribution of rare, noncoding variation to autism spectrum disorder from sequencing of 2,076 genomes in quartet families. Preprint at *bioRxiv* <https://doi.org/10.1101/127043> (2017).
- Siva, N. UK gears up to decode 100,000 genomes from NHS patients. *Lancet* **385**, 103–104 (2015).
- Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433–438 (2017).
- MacDonald, J. R., Ziman, R., Yuen, R. K., Feuk, L. & Scherer, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* **42**, D986–D992 (2014).
- van Bon, B. W. M., Mefford, H. C. & de Vries, B. B. A. 15q13.3 Microdeletion. *GeneReviews* <https://www.ncbi.nlm.nih.gov/books/NBK50780/> (23 July 2015).
- Klopocki, E. et al. Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome. *Am. J. Hum. Genet.* **80**, 232–240 (2007).
- Lupski, J. R. et al. Gene dosage is a mechanism for Charcot-Marie-Tooth disease type 1A. *Nat. Genet.* **1**, 29–33 (1992).
- Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
- Lindsay, E. A. et al. Tbx1 haploinsufficiency in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature* **410**, 97–101 (2001).

50. Merscher, S. et al. TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome. *Cell* **104**, 619–629 (2001).
51. Zollino, M. et al. Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype. *Nat. Genet.* **44**, 636–638 (2012).
52. Talkowski, M. E. et al. Assessment of 2q23.1 microdeletion syndrome implicates MBD5 as a single causal locus of intellectual disability, epilepsy, and autism spectrum disorder. *Am. J. Hum. Genet.* **89**, 551–563 (2011).
53. Slager, R. E., Newton, T. L., Vlangos, C. N., Finucane, B. & Elsea, S. H. Mutations in RAI1 associated with Smith-Magenis syndrome. *Nat. Genet.* **33**, 466–468 (2003).
54. Lupski, J. R. et al. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell* **66**, 219–232 (1991).
55. Chance, P. F. et al. DNA deletion associated with hereditary neuropathy with liability to pressure palsies. *Cell* **72**, 143–151 (1993).
56. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292–301 (2001).
57. Saurin, A. J. et al. The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. *J. Cell Biol.* **142**, 887–898 (1998).
58. Wansink, D. G. et al. Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J. Cell Biol.* **122**, 283–293 (1993).
59. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009). **This study presents the original description of the HiC protocol to analyse genome-wide chromatin interactions.**
60. Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* **14**, 390–403 (2013).
61. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
62. Bianco, S. et al. Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat. Genet.* <https://doi.org/10.1038/s41588-018-0098-8> (2018).
63. Franke, M. et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016). **This study shows that duplications can result in the formation of new TADs, a process that determines the pathogenicity of these rearrangements.**
64. Burton, J. N. et al. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).
65. Rickman, D. S. et al. Oncogene-mediated alterations in chromatin conformation. *Proc. Natl Acad. Sci. USA* **109**, 9083–9088 (2012).
66. Harewood, L. et al. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. *Genome Biol.* **18**, 125 (2017).
67. Montavon, T. et al. A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
68. Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
69. Kagey, M. H. et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
70. Splinter, E. et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* **20**, 2349–2354 (2006).
71. Palstra, R. J. et al. The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* **35**, 190–194 (2003).
72. de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* **26**, 11–24 (2012).
73. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
74. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
75. Nora, E. P., Dekker, J. & Heard, E. Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods? *Bioessays* **35**, 818–828 (2013).
76. Lupianez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015). **This study demonstrates that TAD fusion (deletions) and TAD shuffling (inversions) can cause congenital disease.**
77. Ruf, S. et al. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat. Genet.* **43**, 379–386 (2011).
78. Symmons, O. et al. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).
79. Marinic, M., Aktas, T., Ruf, S. & Spitz, F. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev. Cell* **24**, 530–542 (2013).
80. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin domains: the unit of chromosome organization. *Mol. Cell* **62**, 668–680 (2016).
81. Noordermeer, D. et al. Variegated gene expression caused by cell-specific long-range DNA interactions. *Nat. Cell Biol.* **13**, 944–951 (2011).
82. Symmons, O. et al. The Shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. *Dev. Cell* **39**, 529–543 (2016).
83. Lettice, L. A. et al. Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly. *Dev. Cell* **22**, 459–467 (2012).
84. Kvon, E. Z. et al. Progressive Loss of function in a limb enhancer during snake evolution. *Cell* **167**, 633–642. e11 (2016).
85. Moorthy, S. D. et al. Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome Res.* **27**, 246–258 (2017).
86. Cannavo, E. et al. Shadow enhancers are pervasive features of developmental regulatory networks. *Curr. Biol.* **26**, 38–51 (2016).
87. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
88. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167**, 1170–1187 (2016).
89. Will, A. J. et al. Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). *Nat. Genet.* **49**, 1539–1545 (2017).
90. St-Jacques, B., Hammerschmidt, M. & McMahon, A. P. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes Dev.* **13**, 2072–2086 (1999).
91. Hay, D. et al. Genetic dissection of the alpha-globin super-enhancer in vivo. *Nat. Genet.* **48**, 895–903 (2016).
92. Shin, H. Y. et al. Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat. Genet.* **48**, 904–911 (2016).
93. Fulco, C. P. et al. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
94. Dickel, D. E. et al. Ultraconserved enhancers are required for normal development. *Cell* **172**, 491–499. e15 (2018).
95. Osterwalder, M. et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243 (2018).
96. Zabidi, M. A. et al. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556–559 (2015).
97. Lohan, S. et al. Microduplications encompassing the Sonic hedgehog limb enhancer ZRS are associated with Haas-type polysyndactyly and Laurin-Sandrow syndrome. *Clin. Genet.* **86**, 318–325 (2014).
98. Klopocki, E. et al. A microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal thumb-polysyndactyly syndrome. *J. Med. Genet.* **45**, 370–375 (2008).
99. Benko, S. et al. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nat. Genet.* **41**, 359–364 (2009).
100. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012). **References 74 and 100 demonstrate the presence of TADs in the mammalian genome.**
101. Spielmann, M. & Mundlos, S. Structural variations, the regulatory landscape of the genome and their alteration in human disease. *Bioessays* **35**, 533–543 (2013).
102. Spielmann, M. et al. Homeotic arm-to-leg transformation associated with genomic rearrangements at the PITX1 locus. *Am. J. Hum. Genet.* **91**, 629–635 (2012).
103. Birnbaum, R. Y. et al. Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res.* **22**, 1059–1068 (2012).
104. Tayebi, N. et al. Deletions of exons with regulatory activity at the DYNC111 locus are associated with split-hand/split-foot malformation: array CGH screening of 134 unrelated families. *Orphanet J. Rare Dis.* **9**, 108 (2014).
105. Lango Allen, H. et al. Next generation sequencing of chromosomal rearrangements in patients with split-hand/split-foot malformation provides evidence for DYNC111 exonic enhancers of DLX5/6 expression in humans. *J. Med. Genet.* **51**, 264–267 (2014).
106. Ghiasvand, N. M. et al. Deletion of a remote enhancer near ATOH7 disrupts retinal neurogenesis, causing NCRNA disease. *Nat. Neurosci.* **14**, 578–586 (2011).
107. Bhatia, S. et al. Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am. J. Hum. Genet.* **93**, 1126–1134 (2013).
108. Dathe, K. et al. Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2. *Am. J. Hum. Genet.* **84**, 483–492 (2009).
109. Sun, M. et al. Triphalangeal thumb-polysyndactyly syndrome and syndactyly type IV are caused by genomic duplications involving the long range, limb-specific SHH enhancer. *J. Med. Genet.* **45**, 589–595 (2008).
110. Klopocki, E. et al. Copy-number variations involving the IHH locus are associated with syndactyly and craniosynostosis. *Am. J. Hum. Genet.* **88**, 70–75 (2011).
111. Ngcungu, T. et al. Duplicated enhancer region increases expression of CTSB and segregates with Keratolytic winter erythema in South African and Norwegian families. *Am. J. Hum. Genet.* **100**, 737–750 (2017).
112. Cox, J. J., Willatt, L., Homfray, T. & Woods, C. G. A. SOX9 duplication and familial 46,XX developmental testicular disorder. *N. Engl. J. Med.* **364**, 91–93 (2011).
113. Lupianez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: how alterations of chromatin domains result in disease. *Trends Genet.* **32**, 225–237 (2016).
114. Lettice, L. A. et al. Enhancer-adoption as a mechanism of human developmental disease. *Hum. Mutat.* **32**, 1492–1499 (2011). **This study describes the disease mechanism of enhancer adoption in developmental disease.**
115. Northcott, P. A. et al. Enhancer hijacking activates GF11 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014). **This study reports the phenomenon of enhancer hijacking in cancer.**
116. Redin, C. et al. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat. Genet.* **49**, 36–45 (2017). **By mapping the breakpoints of 273 balanced cytogenetic abnormalities, this study shows that TAD disruption is frequently associated with human congenital anomalies.**
117. Northcott, P. A. et al. The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311–317 (2017). **This study shows that TAD shuffling and enhancer hijacking are frequent causes of medulloblastoma, a highly malignant childhood brain tumour.**
118. Ordlu, Z. et al. Structural chromosomal rearrangements require nucleotide-level resolution: lessons from next-generation sequencing in prenatal diagnosis. *Am. J. Hum. Genet.* **99**, 1015–1033 (2016).
119. Pfeifer, D. et al. Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region. *Am. J. Hum. Genet.* **65**, 111–124 (1999).
120. Hnisz, D. et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).

121. Weischenfeldt, J. et al. Pan-cancer analysis of somatic copy-number alterations implicates *IRS4* and *IGF2* in enhancer hijacking. *Nat. Genet.* **49**, 65–74 (2017).
This paper describes a computational framework termed cis-expression structural alteration mapping (CESAM) to detect cancer-related gene overexpression resulting from TAD reorganization.
122. Kurth, I. et al. Duplications of noncoding elements 5' of *SOX9* are associated with brachydactyly-anonychia. *Nat. Genet.* **41**, 862–863 (2009).
123. Zepeda-Mendoza, C. J. et al. Computational prediction of position effects of apparently balanced human chromosomal rearrangements. *Am. J. Hum. Genet.* **101**, 206–217 (2017).
124. Ibn-Salem, J. et al. Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol.* **15**, 423 (2014).
125. Sanchis-Juan, A. et al. Complex structural variants resolved by short-read and long-read whole genome sequencing in mendelian disorders. Preprint at *bioRxiv* <https://doi.org/10.1101/281683> (2018).
126. Smedley, D. et al. A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease. *Am. J. Hum. Genet.* **99**, 595–606 (2016).
127. Duan, Z. et al. A three-dimensional model of the yeast genome. *Nature* **465**, 363–367 (2010).
128. Chiariello, A. M., Annunziata, C., Bianco, S., Esposito, A. & Nicodemi, M. Polymer physics of chromosome large-scale 3D organisation. *Sci. Rep.* **6**, 29775 (2016).
129. Giorgetti, L. et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157**, 950–963 (2014).
130. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112**, E6456–E6465 (2015).
131. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
132. Sexton, T. et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
133. Phillips-Cremins, J. E. et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
134. Andrey, G. et al. Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res.* **27**, 223–233 (2017).
135. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944.e22 (2017).
136. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
137. Benko, S. et al. Disruption of a long distance regulatory region upstream of *SOX9* in isolated disorders of sex development. *J. Med. Genet.* **48**, 825–830 (2011).
138. Chapuy, B. et al. Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777–790 (2013).
139. Giorgio, E. et al. 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 (2015).
140. Flottmann, R. et al. Microdeletions on 6p22.3 are associated with mesomelic dysplasia Savarirayan type. *J. Med. Genet.* **52**, 476–483 (2015).
141. Niedermaier, M. et al. An inversion involving the mouse *Shh* locus results in brachydactyly through dysregulation of *Shh* expression. *J. Clin. Invest.* **115**, 900–909 (2005).
142. Shimojima, K. et al. De novo microdeletion of 5q14.3 excluding *MEF2C* in a patient with infantile spasms, microcephaly, and agenesis of the corpus callosum. *Am. J. Med. Genet. Part A* **158A**, 2272–2276 (2012).
143. Goubau, C. et al. Platelet defects in congenital variant of Rett syndrome patients with *FOXC1* mutations or reduced expression due to a position effect at 14q12. *Eur. J. Hum. Genet.* **21**, 1349–1355 (2013).

Acknowledgements

The authors apologize to colleagues whose work we were unable to discuss or failed to cite owing to space constraints. We thank members of the Mundlos and the Shendure laboratories for helpful discussions. M.S. was supported by a grant from the Deutsche Forschungsgemeinschaft (SP1532/2-1). D.G.L. is supported by a grant from the Deutsche Forschungsgemeinschaft (GA2495/1-1). Work in S.M.'s laboratory is funded by the Deutsche Forschungsgemeinschaft, the Berlin Institute for Health and the Max Planck Foundation.

Author contributions

All authors contributed equally to all aspects of the manuscript.

Competing interests

The authors declare no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

RELATED LINKS

Database of Genomic Variants: <http://dgv.tcag.ca/>
3D Genome Browser: <http://promoter.bx.psu.edu/hi-c/>
VISTA Enhancer Browser: <https://enhancer.lbl.gov/>
DECIPHER: <https://decipher.sanger.ac.uk/>