OXFORD

INVITED REVIEW

Looking beyond the genes: the role of non-coding variants in human disease

Malte Spielmann[1,2,3] and Stefan Mundlos[1,2,3],*

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

*To whom correspondence should be addressed at: Tel: +4930450569122; Fax: +4930450569131; Email: mundlos@molgen.mpg.de

Abstract

Over the past decades the search for disease causing variants has been focusing exclusively on the coding genome. This highly selective approach has been extremely successful resulting in the identification of thousands of disease genes, but ignores the functional and therefore disease relevance of the rest of the genome. Dropping sequencing costs and new high-throughput technologies such as ChIP-seq and chromosome conformation capture have opened new possibilities for the systematic investigation of the non-coding genome. These data have revealed the importance of non-coding DNA in fundamental processes such as gene regulation and 3D chromatin folding. Research into the principles of chromatin folding has revealed a domain structure of the genome, called topologically associated domains that provide a scaffold for enhancer promoter contacts. Non-coding mutations that affect regulatory elements can affect gene regulation by a loss of function, resulting in reduced gene expression, or a gain of function resulting in gene mis- or overexpression. Structural variations such as deletions, inversions or duplications have the potential to disturb normal chromatin folding. This may lead to the repositioning or disruption of topological associating domains and the relocation of enhancer elements with consecutive gene misexpression. Several recent studies highlight this as important disease mechanisms in developmental disorders and cancer. Therefore, the regulatory landscape of the genome has to be taken into consideration when investigating the pathology of human disease. In this review, we will discuss the recent discoveries in the field of non-coding variation, gene regulation, 3D genome architecture, and their implications for human genetics.

Introduction

Beyond the exome

Medical genetics is being transformed by next-generation sequencing (NGS) technologies enabling simultaneous investigation of all relevant disease genes, all protein-coding genes, and even the entire genome (1,2). Due to dropping sequencing costs and constant improvements of sequencing technologies, it is likely that in the near future whole-genome sequencing (WGS) will be feasible and affordable in a diagnostic setting (3,4). Pilot studies have shown that WGS can detect a broader range of genetic variation than other sequencing approaches, not only single nucleotide variants (SNVs) and insertion or deletions (indels), but also structural variants such as copy number variants (CNVs), inversions and translocations (3). However, if WGS is to be implemented as a "one test for all" diagnostic strategy, fundamental obstacles remain. The "regulatory code" of the non-coding genome, determining whether and how a given genetic variant affects the function of a regulatory element, remains poorly understood (5). To address this challenge, the

Encyclopaedia of DNA Elements (ENCODE) consortium was established in 2004, with the aim of systematically investigating and annotating the 98.5% of the genome that is non-coding (6). The collective findings of the ENCODE consortium suggested that around 80% of the genome contains elements linked to some biochemical function (6,7). Moreover, the space between genes is filled with *cis*-regulatory elements such as enhancers, silencers, promoters, and numerous previously overlooked regions containing untranslated RNA transcripts that have a regulatory role (8,9). These findings are directly relevant to human genetics, since most studies investigating the genetic basis of intellectual disability and developmental delay are focused on coding variants, but have failed to provide clear answers for over 40% of the families studied (10–12). This suggests that a large proportion of disease cases may be caused by alterations outside of the coding regions.

In this review, we will discuss recent discoveries in the field of non-coding variation, gene regulation, 3D genome architecture, and their implications for human genetics. We will focus on the latest findings and highlight the role of 3D genome architecture in human disease.

Long range control via enhancers

Gene regulation usually involves two distinct types of *cis*-acting elements, the promoter consisting of the core promoter and nearby regulatory elements, and more distal regulatory units, so called enhancers, silencers of locus control regions (9). While the promoter is generally located less than 1 kb from the transcription start site, enhancers can act over long distances, in some cases more than 1 Mb (Fig. 1) (13). Promoters and enhancers have overlapping functional properties and share many characteristics such that they have been considered a single class of regulatory elements. Both represent docking sites for transcription factors (TFs) regulating the activity of the promoter and enhancer (14). In the current concept, enhancers are primed by sequence specific pioneering TFs, providing an accessible platform for the recruitment of general transcription factors. Together with RNA polymerase II, they form the preinitiation complex which in turn recruits further proteins including the mediator complex. The large distance between enhancers and promoters requires direct or indirect mediators of communication between these elements (15). One possible scenario involves physical contact enabled by looping of the intervening DNA sequence mediated by proteins such as mediator, CTCF, and cohesins, bridging enhancers and promoters together as a chromatin structure. Once the bridge is formed, expression is regulated accordingly (16–19).

The physical presence of looped chromatin has been demonstrated by fluorescent in situ hybridization (FISH) experiments, in which the distance between two given chromosome fragments can be measured. The development of chromosome conformation capture (3C) and its high-throughput derivatives (e.g. 4C, 5C, ChIA-PET and Hi-C) (19,20), have made it possible to explore in more detail the 3D architecture of the genome by quantifying chromatin looping via a proximity ligation assay. For example, it was shown at the *SHH* locus that, even when its limb enhancer element is deleted, loop formation is still observed (21). In another interesting study, Deng and colleagues showed that chromatin looping can also be induced using artificial zinc fingers, with subsequent activation of developmentally silenced genes (22).
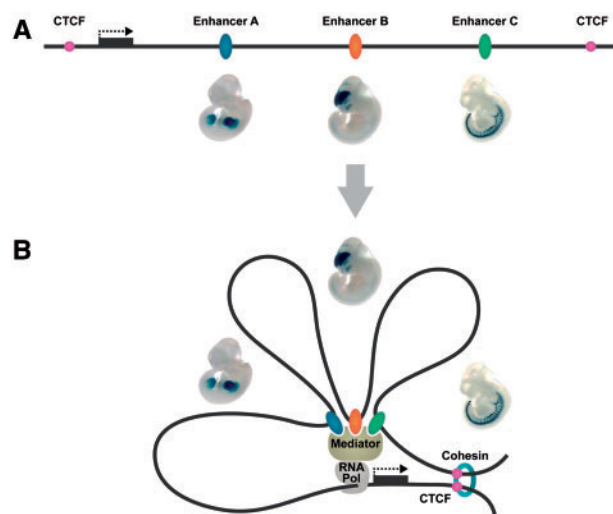

Figure 1. Enhancer promotor interaction in the nucleus via DNA-looping. The human genome is often depicted in a linear fashion, despite the fact that the two meters of DNA needs to be tightly folded in order to be packed into the nucleus. The large distance between enhancers and promoters requires communication between these elements. One current concept involves looping of the intervening DNA sequence with proteins such as the mediator complex, CTCF, cohesins and others, bridging enhancers and promoters together as a chromatin structure. Once the bridge is formed, tissue specific expression is regulated accordingly (16–19). Enhancers data from the VISTA Enhancer Browser: http://enhancer.lbl.gov/(75).

Topologically associating domains (TADs) and the 3D structure of chromatin

The human genome is often depicted in a linear fashion, despite the fact that the two meters of DNA needs to be tightly folded in order to be packed into the nucleus. Recent data indicate that chromatin folding, and thus the 3D organization of the genome in the nucleus, is directly linked to central aspects of gene regulation. Major insights have been gained by Hi-C, an expansion of the 3C technologies, using a purification of ligation products followed by massively parallel sequencing (17,23). Deep sequencing of Hi-C libraries has revealed sub-orders of chromosome organization at the megabase scale, designated as topologically associating domains or TADs (24–26) (Fig. 2). Interestingly, these structures are conserved among species, cell types, and tissues (24,25). They appear to form a regulatory scaffold for the genome restricting the contacts an enhancer may have, thereby preventing promiscuous enhancer activity (24,26–28). An example is shown in Figure 2 (Fig. 2).

TADs appear to be structures that both promote contacts within a domain and at the same time prevent contacts between neighbouring domains. Important for the separation of neighbouring activities are so called boundary regions, initially inferred from Hi-C data sets by measuring abrupt changes in the directionality of contacts (24). Boundaries are strongly enriched in the architectural proteins CTCF (CCCTC-binding factor) and cohesin (29). CTCF seems to be crucial for boundary function, as its ablation affects TAD organization by decreasing intra-domain and increasing inter-domain contacts (30), although its presence is not exclusive to boundaries (24). Furthermore, a striking correlation was found between CTCF motif orientation and looping, taking place in more than 90% of the cases in a convergent manner (27,31). CRISPR-Cas9 genome editing experiments have further supported the role of CTCF in TAD boundary organization (25,32).
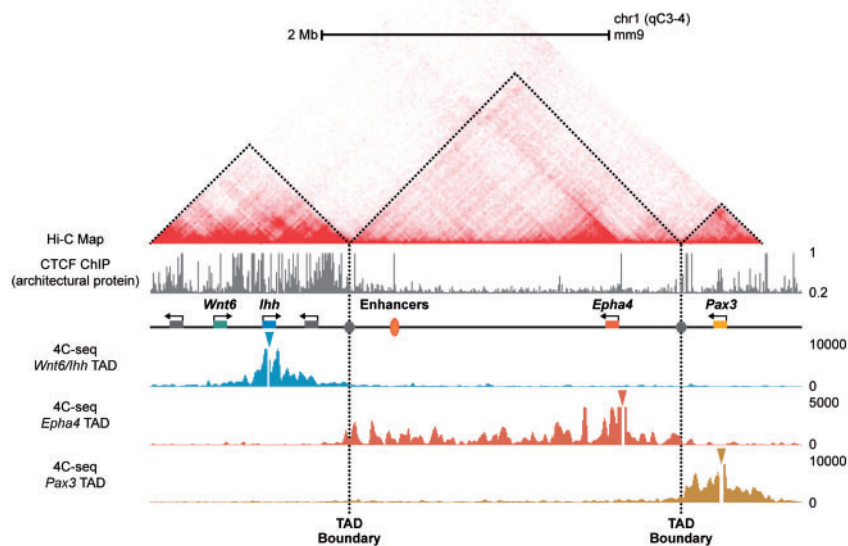
Figure 2. The topological-associated domain (TAD) architecture of the *Epha*4 locus. The genomic landscape of the *Epha*4 locus: Hi-C interactions from mouse ES cell are shown in a heat map in which each dot reflects two interaction pairs of DNA. The resulting interaction profile shows the formation of triangles (dotted lines) that represent individual TADs. There is a high degree of interaction within each TAD but little contact between TADs. The *Epha*4 locus appears as one large TAD with transitions on either side demarcating boundary regions where the interactions diminish and orientate to the other direction. The binding profile of CTCF transcription factor is shown below. There is an enrichment of CTCF binding at the boundaries and at gene promoters. The 4C-seq profiles of the viewpoints *Ihh*, *Epha*4, and *Pax3* are depicted below. Note that the interaction profiles are restricted to the respective TADs (24,34,76).

These data demonstrate that TADs and their boundaries play a key role in controlling gene expression. Structural variations such as deletions, inversions or duplications thus have the potential to interfere with the TAD structure by disrupting or repositioning boundaries. Recent studies highlight this as an important disease mechanism (33–35).
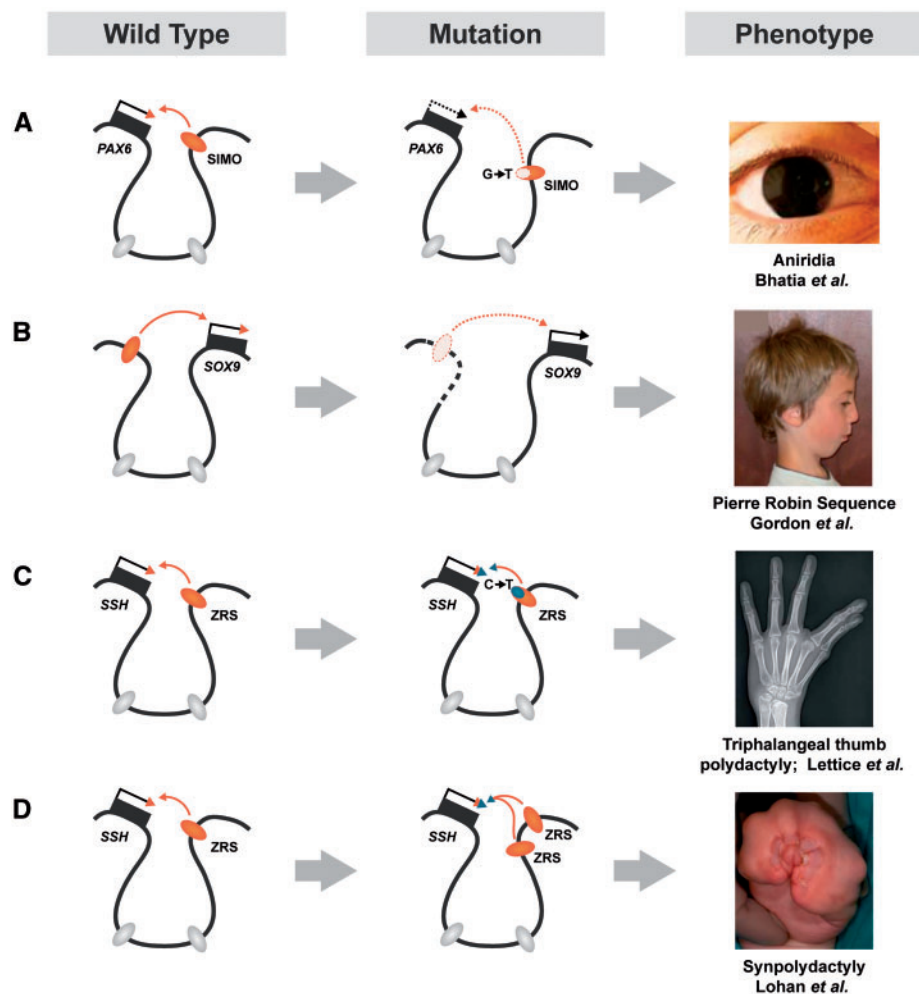
Loss of function mutations in regulatory elements

Similar to coding mutations, single nucleotide variants in enhancer elements can result in a in a loss of function or a gain of function of the target gene (Fig. 3). Given the temporal and spatial specificity of enhancers and the complexity of expression patterns of many developmental genes, the loss of a specific regulatory element might result in the loss of a specific expression domain, provided that it is not compensated by other regulatory elements. Such a specific loss can result in a phenotype consisting of a subset of features observed with a complete loss of the gene's function, e.g. by a coding mutation. An example is isolated pancreatic agenesis which is caused by mutations in a highly conserved pancreas specific enhancer located 25 kb downstream of *PTF1A* (36). Mutations in *PTF1A* itself result in syndromic form of pancreatic agenesis, featuring severe neurological symptoms, whereas PTF1A enhancer mutations result in an isolated pancreatic anomaly (37). A similar example was described for *PAX6* in patients with aniridia (Fig. 4A) (38). Regulatory loss of function point mutations have also been reported for *SOX9*. Mutations in *SOX9* cause campomelic dysplasia, a lethal skeletal dysplasia with male to female sex reversal (39,40). Deletions of the proposed testis enhancer of *SOX9* result in sex reversal but no skeletal phenotype, whereas deletions and point mutations further upstream cause Pierre-Robin syndrome, a condition with growth defects of the
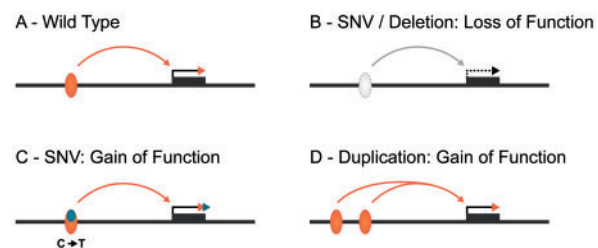
Figure 3. The effects non-coding mutations on gene expression. Similar to coding mutations, single nucleotide variants (SNV) in enhancer elements can results in a in a loss of function and a gain of function of the target gene. A loss of a regulatory element by a point mutation or a deletion is expected to result in the loss of a specific expression domain, while a gain of function mutation or a duplication of an enhancer element can increase the frequency of enhancer-promotor interactions causing tissue specific gain of expression and misexpression.

cranial skeleton, but normal sexual development (Fig. 4B). Further examples have been reported for *SHH* (41) and *TBX5* (42).

Often, deletions remove coding sequence, but the phenotype can better be explained by the loss of a non-coding *cis*-regulatory element. A study by Birnbaum and co-workers shows that the deletion of coding exons can also result in a loss of *cis*-regulatory activity (43). Through the combination of ChIP-seq enhancer data, enhancer assays, and chromosome confirmation capture, they showed that DNA sequences can have a dual function, operating both as coding exon and as an enhancer of nearby gene. Deletions of these "exonic" enhancers (eExons) in the *DYNC1I1* gene cause a loss of function of the *DLX5/6* genes approximately 1 Mb away and have been shown to account for around 3% of all split hand and foot malformation cases (43–45).

Figure 4. Mutations and CNVs of non-coding enhancer elements. Regulatory loss of function mutations can result in a phenotype that consists of a subset of features observed with a complete loss of the gene's function by coding mutations. **(A)** An example is a point mutation in an enhancer element of *PAX6* in patients with Aniridia (38). **(B)** Deletions and point mutations in the regulatory landscape of *SOX9* have been shown to cause Pierre-Robin syndrome, while mutations in *SOX9* itself cause a lethal skeletal dysplasia (39,40). **(C).** Regulatory gain of function mutations were described at the Sonic hedgehog (*SHH*) locus. Mutations in the ZRS, the *SHH* limb enhancer element, result in a misexpression of SHH in the anterior part of the limb bud which in turn results in polysyndactyly and triphalangeal thumb (48,49). **(D)** Interestingly, very similar polysyndactyly phenotypes have been described for duplications encompassing the ZRS indicating that some enhancer elements may be dose sensitive (46,50).

Gain of function mutations in regulatory elements

An example for regulatory gain of function mutations has been described at the Sonic hedgehog (*SHH*) locus (Fig. 4C). Hedgehog genes encode for a developmentally important family of signalling molecules. In the developing limb, *SHH* is expressed at a restricted site, the zone of polarising activity, which is tightly regulated by the ZRS, a regulatory region approx. 1 Mb 5' of *SHH* (46–48). Several conditions with various types of polydactyly, including triphalangeal thumb polysyndactyly syndrome, Haass type polysyndactyly, and Werner syndrome, have been associated with point mutations in the ZRS. As shown in transgenic mice, these point mutations result in a misexpression of *SHH* in the anterior part of the limb bud, which in turn results in additional digits (polydactyly) and fusion of digits (syndactyly) (48). It was shown that ZRS mutations alter the balance of transcription factor binding sites by either creating new sites or by inactivating existing sites (49). Interestingly, very similar phenotypes have been described for duplications encompassing the ZRS, indicating that some enhancer elements may be dose

sensitive (Fig. 4D) (46,50). However, as shown for the point mutations, duplications of the ZRS do not only result in an increase of expression, but also in misexpression, as indicated by the polydactyly phenotype. Still, the precise mechanism of these mutations needs to be proven, as no animal models exist so far.

A similar example was described in patients with female-to-male sex reversal carrying tandem duplications upstream of *SOX9* (51). These duplications are thought to result in a gonad specific gain of function of *SOX9*, while overlapping deletions have been identified in individuals with male-to-female sex reversal (52). Our own lab has also described several examples of tandem duplications of non-coding *cis*-regulatory elements associated with congenital limb malformation: Duplications of a 5 kb limb enhancer element within the TAD *BMP2* cause brachydactyly type A2, a shortening of the digits (53,54). Several similar tandem duplications of non-coding *cis*-regulatory elements were described 5' of *IHH* in patients with syndactyly and craniosynostosis (55).
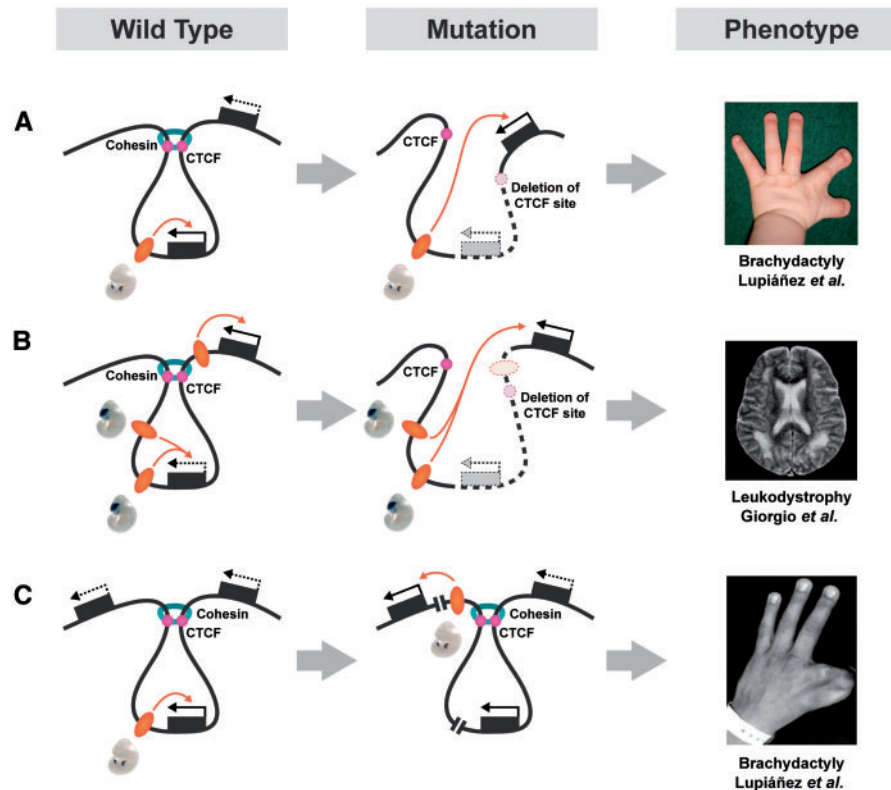
Figure 5. Disruption of TAD structures cause congenital disease. Structural variations have the potential to alter the topological domain architecture of the genome by deleting, or misplacing TAD boundaries, thereby allowing enhancers from neighbouring domains to ectopically activate genes causing misexpression and disease. (**A**) Deletions encompassing *EPHA4* were shown to result in brachydactyly (short digits). While *EPHA4* itself is not involved limb development it was shown that the deletions remove several CTCF binding sites that are part of the TAD boundary. Without this boundary a cluster of limb enhancers can act on *PAX3* causing brachydactyly (34). (**B**) A TAD boundary deletion was also identified as a cause of autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). This phenotype is usually associated with duplications of *LMNB1*. In one family a deletion of the TAD boundary was shown to result in LMNB1 misexpression due to ectopic enhancer-promotor interaction (64,65). (**C**) Also inversions can disrupt the TAD architecture of the genome. An inversion at the *EPHA4* locus was shown to locate a cluster of limb enhancers close to *WNT6* causing severe syndactyly (34). Enhancers data from the VISTA Enhancer Browser: http://enhancer.lbl.gov/(75).

Non-coding variants in complex traits and diseases

Disease-associated nucleotide variants identified in genome wide association studies (GWAS) are rarely found in coding regions. Instead, most disease-associated index SNPs are located in non-coding regions of the genome (56). The recent introduction of CRISPR/Cas9 genome editing has opened new opportunities to investigate also common non-coding variants located in *cis*-regulatory elements (57,58). Claussnitzer and colleagues recently showed that the FTO allele, which shows the strongest genome wide association signal for obesity, acts a gain of function (59). Using CRISPR/Cas9 genome editing, they showed that the disease-associated single-nucleotide variant rs1421085 T-to-C disrupts a conserved motif for the *ARID5B* repressor, which unleashes a preadipocyte enhancer, leading to a doubling of IRX3 and IRX5 expression during early adipocyte differentiation. Another recent study identified a common Parkinson disease-associated risk variant in a non-coding distal enhancer element, regulating the expression of α-synuclein (*SNCA*), a key gene implicated in the pathogenesis of Parkinson disease (60). They demonstrated that the disease-associated risk variant prevents the binding of two repressive transcription factors EMX2 and NKX6-1 to an enhancer element, resulting in transcriptional upregulation of SNCA.

Another interesting example is a non-coding variant at the 8q24 locus that associates with cleft lip with or without cleft palate (61). Recent mouse studies have shown that this interval contains very remote *cis*-acting enhancers that control Myc expression in the developing face (62).

TAD disruption and position effects

The term "position effect" has long been used for balanced translocations and other larger structural rearrangements that could not be explained by the content of the variants alone. The discovery of the topological domain architecture of the genome and our increased knowledge about enhancers and chromatin-folding, enables us to better understand the mechanisms underlying potential "position effects". It is now becoming clear that structural variations have the potential to alter the topological domain architecture of the genome by deleting, or misplacing TAD boundaries, thereby allowing enhancers from neighbouring domains to ectopically activate genes causing misexpression and disease (33). This mutational mechanism was termed enhancer adoption (35).

An early example suggestive of such a disease mechanism is Liebenberg syndrome, an autosomal-dominant condition in which the arms partially acquire morphological characteristics of the legs (63). We identified several deletions of a *cis*-regulatory boundary element close to *PITX1*, an important gene for hindlimb development. Without this boundary or insulator

element, an enhancer element with fore- and hindlimb activity from a neighbouring regulatory domain is free to act on *PITX1*, thereby inducing ectopic expression of *PITX1* in the forelimb causing the phenotype.

The relevance of TADs for genomic integrity and disease was exemplarily demonstrated at the *EPHA4* locus (34). A deletion encompassing *EPHA4* at the telomeric side of the TAD was shown to result in brachydactyly (short digits) (Fig. 5A), 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 (Fig. 5C). In addition, a family with a duplication was investigated, whose breakpoints partially overlapped with a deletion present in a mouse mutant where 800 kb between the *Ihh* (*Indian hedgehog*) gene and the most centromeric part of the *Epha4* TAD are deleted. These two rearrangements, despite being of a different nature, both resulted in severe polydactyly (7 and more fingers). The results indicated that different structural variations can result in similar, but distinct malformation syndromes, and thus likely to be independent of the gene in this locus, i.e. *EPHA4*. CRISPR/Cas9 genome editing was used to re-engineer (recapitulate) the human deletions and inversions in mice (58). We could show that an enhancer cluster located within the *Epha4* TAD, which normally regulates *Epha4* expression in the limb bud, was now ectopically activating different genes, depending on the breakpoint, i.e. *Pax3* (*Paired box3*) in brachydactyly (Fig. 5A), *Wnt6* (*Wingless-type MMTV integration site family, member 6*) in syndactyly (Fig. 5C), and *Ihh* in polydactyly (34). All mutant phenotypes were caused by gene misexpression, due to the ectopic interaction of the enhancers 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 ectopic interaction and normal mice. Furthermore, a similar pathomechanism was identified causing autosomal dominant adult-onset demyelinating leukodystrophy (ADLD) (Fig. 5B) (64,65) and mesomelic dysplasia (severe shortening of the middle segment of the lower limbs) (66). It was suggested that up to 11% of all deletions in the DECIPHER database results in enhancer adoption (35).

Recent data indicate that TAD reorganization and disruption play a key role in the pathogenesis of cancer, which is not surprising since misregulation of genes is a common feature in cancer (67–69). Some examples are AML/CLL, gliomas, and medulloblastoma, where enhancer adoption or enhancer hijacking has been identified as the major mechanism of disease (70–72). In fact, recent studies show that TAD boundary-associated CTCF loops are frequently mutated in colon cancer, allowing promiscuous activity of genes (73,74).

These results suggest that non-coding mutations may contribute to a substantial number of human disease phenotypes and should thus be taken into account for the medical interpretation of mutations and copy number variants.

Conclusion for the Clinic

Recent findings show that mutations and structural variations outside of the coding genome can interfere with the normal gene regulation by regulatory loss or gain of function mutations. Structural variations have also 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. Therefore,

the *cis*-regulatory landscape and TAD architecture of the genome have to be taken into consideration when investigating the pathology of human disease.

The next challenge will be the medical interpretation of single nucleotide variants from clinical WGS data. The sheer number of non-coding variants in each individual and generation make classical functional work-up strategies impossible (3). Further knowledge about the non-coding genome and more experimentally validated non-coding variants are needed to develop computational prediction tools for the medical interpretation of non-coding mutations.

References

1. Shendure, J. (2011) Next-generation human genetics. *Genome Biol.*, **12**, 408.
2. Zemojtel, T., Kohler, S., Mackenroth, L., Jager, M., Hecht, J., Krawitz, P., Graul-Neumann, L., Doelken, S., Ehmke, N., Spielmann, M., *et al.* (2014) Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. *Sci. Transl. Med.*, **6**, 252ra123.
3. Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., *et al.* (2014) Genome sequencing identifies major causes of severe intellectual disability. *Nature*, **511**, 344–347.
4. Lelieveld, S.H., Spielmann, M., Mundlos, S., Veltman, J.A. and Gilissen, C. (2015) Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. *Hum. Mutat.*, **36**, 815–822.
5. Ritchie, G.R. and Flicek, P. (2014) Computational approaches to interpreting genomic sequence variation. *Genome Med.*, **6**, 87.
6. Consortium, E.P. (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, **306**, 636–640.
7. Consortium, E.P., Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C. and Snyder, M. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
8. Ecker, J.R., Bickmore, W.A., Barroso, I., Pritchard, J.K., Gilad, Y. and Segal, E. (2012) Genomics: ENCODE explained. *Nature*, **489**, 52–55.
9. Spielmann, M. and Klopocki, E. (2013) CNVs of noncoding cis-regulatory elements in human disease. *Curr. Opin. Genet. Dev.*, **23**, 249–256.

10. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., *et al.* (2011) A copy number variation morbidity map of developmental delay. *Nat. Genet.*, **43**, 838–846.

11. Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., *et al.* (2012) Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet*, **380**, 1674–1682.

12. de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., *et al.* (2012) Diagnostic exome sequencing in persons with severe intellectual disability. *N. Eng. J. Med.*, **367**, 1921–1929.

13. Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L. and Spitz, F. (2014) Functional and topological characteristics of mammalian regulatory domains. *Genome Res.*, **24**, 390–400.

14. Spitz, F. and Furlong, E.E. (2012) Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Gen.*, **13**, 613–626.

15. de Laat, W. and Duboule, D. (2013) Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*, **502**, 499–506.

16. Palstra, R.J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F. and de Laat, W. (2003) The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.*, **35**, 190–194.

17. de Wit, E. and de Laat, W. (2012) A decade of 3C technologies: insights into nuclear organization. *Genes Dev.*, **26**, 11–24.

18. Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002) Capturing chromosome conformation. *Science*, **295**, 1306–1311.

19. Dekker, J., Marti-Renom, M.A. and Mirny, L.A. (2013) Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.*, **14**, 390–403.

20. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.

21. Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H. and Shiroishi, T. (2009) Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell*, **16**, 47–57.

22. Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A., *et al.* (2014) Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell*, **158**, 849–860.

23. Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A. and Fraser, P. (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, **502**, 59–64.

24. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.

25. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, **485**, 381–385.

26. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A. and Cavalli, G. (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, **148**, 458–472.

27. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**(7):1665–1680.

28. Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., *et al.* (2012) A map of the cis-regulatory sequences in the mouse genome. *Nature*, **488**, 116–120.

29. Van Bortle, K., Nichols, M.H., Li, L., Ong, C.T., Takenaka, N., Qin, Z.S. and Corces, V.G. (2014) Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol.*, **15**, R82.

30. Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van, I.W.F., *et al.* (2014) Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl Acad. Sci.USA*, **111**, 996–1001.

31. Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A. and Hadjur, S. (2015) Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10**, 1297–1309.

32. Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., *et al.* (2015) CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, **162**, 900–910.

33. Spielmann, M. and Mundlos, S. (2013) Structural variations, the regulatory landscape of the genome and their alteration in human disease. *Bioessays*, **35**, 533–543.

34. Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., *et al.* (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, **161**, 1012–1025.

35. Ibn-Salem, J., Kohler, S., Love, M.I., Chung, H.R., Huang, N., Hurles, M.E., Haendel, M., Washington, N.L., Smedley, D., Mungall, C.J., *et al.* (2014) Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol.*, **15**, 423.

36. Weedon, M.N., Cebola, I., Patch, A.M., Flanagan, S.E., De Franco, E., Caswell, R., Rodriguez-Segui, S.A., Shaw-Smith, C., Cho, C.H., Lango Allen, H., *et al.* (2014) Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat. Genet.*, **46**, 61–64.

37. Sellick, G.S., Barker, K.T., Stolte-Dijkstra, I., Fleischmann, C., Coleman, R.J., Garrett, C., Gloyn, A.L., Edghill, E.L., Hattersley, A.T., Wellauer, P.K., *et al.* (2004) Mutations in PTF1A cause pancreatic and cerebellar agenesis. *Nat. Genet.*, **36**, 1301–1305.

38. Bhatia, S., Bengani, H., Fish, M., Brown, A., Divizia, M.T., de Marco, R., Damante, G., Grainger, R., van Heyningen, V. and Kleinjan, D.A. (2013) Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am. J. Human Genet.*, **93**, 1126–1134.

39. Wagner, T., Wirth, J., Meyer, J., Zabel, B., Held, M., Zimmer, J., Pasantes, J., Bricarelli, F.D., Keutel, J., Hustert, E., *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.

40. Benko, S., Gordon, C.T., Mallet, D., Sreenivasan, R., Thauvin-Robinet, C., Brendehaug, A., Thomas, S., Bruland, O., David, M., Nicolino, M., *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.

41. Jeong, Y., Leskow, F.C., El-Jaick, K., Roessler, E., Muenke, M., Yocum, A., Dubourg, C., Li, X., Geng, X., Oliver, G., *et al.* (2008) Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat. Genet.*, **40**, 1348–1353.

42. Smemo, S., Campos, L.C., Moskowitz, I.P., Krieger, J.E., Pereira, A.C. and Nobrega, M.A. (2012) Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.*, **21**, 3255–3263.

43. Birnbaum, R.Y., Clowney, E.J., Agamy, O., Kim, M.J., Zhao, J., Yamanaka, T., Pappalardo, Z., Clarke, S.L., Wenger, A.M., Nguyen, L., *et al.* (2012) Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res.*, **22**, 1059–1068.

44. Tayebi, N., Jamsheer, A., Flottmann, R., Sowinska-Seidler, A., Doelken, S.C., Oehl-Jaschkowitz, B., Hulsemann, W., Habenicht, R., Klopocki, E., Mundlos, S., *et al.* (2014) Deletions of exons with regulatory activity at the DYNC1I1 locus are associated with split-hand/split-foot malformation: array CGH screening of 134 unrelated families. *Orphanet. J. Rare Dis.*, **9**, 108.

45. Lango Allen, H., Caswell, R., Xie, W., Xu, X., Wragg, C., Turnpenny, P.D., Turner, C.L., Weedon, M.N. and Ellard, S. (2014) Next generation sequencing of chromosomal rearrangements in patients with split-hand/split-foot malformation provides evidence for DYNC1I1 exonic enhancers of DLX5/6 expression in humans. *J. Med. Genet.*, **51**:264–267.

46. Klopocki, E., Ott, C.E., Benatar, N., Ullmann, R., Mundlos, S. and Lehmann, K. (2008) A microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal thumb-polysyndactyly syndrome. *J. Med. Genet.*, **45**, 370–375.

47. Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E. and de Graaff, E. (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.

48. Lettice, L.A., Horikoshi, T., Heaney, S.J., van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N., *et al.* (2002) Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc. Natl Acad. Sci. USA*, **99**, 7548–7553.

49. Lettice, L.A., Williamson, I., Wiltshire, J.H., Peluso, S., Devenney, P.S., Hill, A.E., Essafi, A., Hagman, J., Mort, R., Grimes, G., *et al.* (2012) Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly. *Dev. Cell*, **22**, 459–467.

50. Lohan, S., Spielmann, M., Doelken, S.C., Flottmann, R., Muhammad, F., Baig, S.M., Wajid, M., Hulsemann, W., Habenicht, R., Kjaer, K.W., *et al.* (2014) Microduplications encompassing the Sonic hedgehog limb enhancer ZRS are associated with Haas-type polysyndactyly and Laurin-Sandrow syndrome. *Clin. Genet.*, **86**, 318–325.

51. Cox, J.J., Willatt, L., Homfray, T. and Woods, C.G. (2011) A SOX9 duplication and familial 46,XX developmental testicular disorder. *N. Engl. J. Med.*, **364**, 91–93.

52. Pop, R., Conz, C., Lindenberg, K.S., Blesson, S., Schmalenberger, B., Briault, S., Pfeifer, D. and Scherer, G. (2004) Screening of the 1 Mb SOX9 5' control region by array CGH identifies a large deletion in a case of campomelic dysplasia with XY sex reversal. *J. Med. Genet.*, **41**, e47.

53. Dathe, K., Kjaer, K.W., Brehm, A., Meinecke, P., Nurnberg, P., Neto, J.C., Brunoni, D., Tommerup, N., Ott, C.E., Klopocki, E., *et al.* (2009) Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2. *Am. J. Hum. Genet.*, **84**, 483–492.

54. Mundlos, S. (2009) The brachydactylies: a molecular disease family. *Clin. Genet.*, **76**, 123–136.

55. Klopocki, E., Lohan, S., Brancati, F., Koll, R., Brehm, A., Seemann, P., Dathe, K., Stricker, S., Hecht, J., Bosse, K., *et al.* (2011) Copy-number variations involving the IHH locus are associated with syndactyly and craniosynostosis. *Am. J. Hum. Genet.*, **88**, 70–75.

56. Herman, M.A. and Rosen, E.D. (2015) Making Biological Sense of GWAS Data: Lessons from the FTO Locus. *Cell Metab*, **22**, 538–539.

57. Spielmann, M., Kakar, N., Tayebi, N., Leettola, C., Nurnberg, G., Sowada, N., Lupianez, D.G., Harabula, I., Flottmann, R., Horn, D., *et al.* (2016) Exome sequencing and CRISPR/Cas genome editing identify mutations of ZAK as a cause of limb defects in humans and mice. *Genome Res.*, **26**, 183–191.

58. Kraft, K., Geuer, S., Will, A.J., Chan, W.L., Paliou, C., Borschiwer, M., Harabula, I., Wittler, L., Franke, M., Ibrahim, D.M., *et al.* (2015) Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. *Cell Reports*, pii: S2211-1247(15)00029–7.

59. Claussnitzer, M., Dankel, S.N., Kim, K.H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I.S., Beaudry, J.L., Puviindran, V., *et al.* (2015) FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N. Eng. J. Med.*, **373**, 895–907.

60. Soldner, F., Stelzer, Y., Shivalila, C.S., Abraham, B.J., Latourelle, J.C., Barrasa, M.I., Goldmann, J., Myers, R.H., Young, R.A. and Jaenisch, R. (2016) Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. *Nature*, **533**, 95–99.

61. Birnbaum, S., Ludwig, K.U., Reutter, H., Herms, S., Steffens, M., Rubini, M., Baluardo, C., Ferrian, M., Almeida de Assis, N., Alblas, M.A., *et al.* (2009) Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat. Genet.*, **41**, 473–477.

62. Uslu, V.V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N.A., Marioni, J.C. and Spitz, F. (2014) Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nat. Genet.*, **46**, 753–758.

63. Spielmann, M., Brancati, F., Krawitz, P.M., Robinson, P.N., Ibrahim, D.M., Franke, M., Hecht, J., Lohan, S., Dathe, K., Nardone, A.M., *et al.* (2012) Homeotic arm-to-leg transformation associated with genomic rearrangements at the PITX1 locus. *Am. J. Hum. Genet.*, **91**, 629–635.

64. Giorgio, E., Robyr, D., Spielmann, M., Ferrero, E., Di Gregorio, E., Imperiale, D., Vaula, G., Stamoulis, G., Santoni, F., Atzori, C., *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.

65. Brussino, A., Vaula, G., Cagnoli, C., Panza, E., Seri, M., Di Gregorio, E., Scappaticci, S., Camanini, S., Daniele, D., Bradac, G.B., *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.

66. Flottmann, R., Wagner, J., Kobus, K., Curry, C.J., Savarirayan, R., Nishimura, G., Yasui, N., Spranger, J., Van Esch, H., Lyons, M.J., *et al.* (2015) Microdeletions on 6p22.3 are associated with mesomelic dysplasia Savarirayan type. *J. Med. Genet.*, **52**, 476–483.

67. Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.

68. Lee, T.I. and Young, R.A. (2013) Transcriptional regulation and its misregulation in disease. *Cell*, **152**, 1237–1251.

69. Valton, A.L. and Dekker, J. (2016) TAD disruption as oncogenic driver. *Curr. Opin. Genet. Dev.*, **36**, 34–40.

70. Groschel, S., Sanders, M.A., Hoogenboezem, R., de Wit, E., Bouwman, B.A., Erpelinck, C., van der Velden, V.H., Havermans, M., Avellino, R., van Lom, K., *et al.* (2014) A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell*, **157**, 369–381.

71. Northcott, P.A., Lee, C., Zichner, T., Stutz, A.M., Erkek, S., Kawauchi, D., Shih, D.J., Hovestadt, V., Zapatka, M., Sturm, D., *et al.* (2014) Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature*, **511**, 428–434.

72. Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suva, M.L. and Bernstein, B.E. (2016) Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*, **529**, 110–114.

73. Katainen, R., Dave, K., Pitkanen, E., Palin, K., Kivioja, T., Valimaki, N., Gylfe, A.E., Ristolainen, H., Hanninen, U.A., Cajuso, T., *et al.* (2015) CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat. Genet.*, **47**, 818–821.

74. Ji, X., Dadon, D.B., Powell, B.E., Fan, Z.P., Borges-Rivera, D., Shachar, S., Weintraub, A.S., Hnisz, D., Pegoraro, G., Lee, T.I., *et al.* (2016) 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell*, **18**, 262–275.

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

76. Lupianez, D.G., Spielmann, M. and Mundlos, S. (2016) Breaking TADs: How alterations of chromatin domains result in disease. *Trends Genet.*, **32**(4):225–237.