Enhancers, spatial chromosome structuring and pathological changes: towards a better understanding of complex genome alterations

Éric Soler, Institut de génétique moléculaire de Montpellier, université de Montpellier, CNRS, Montpellier, France Laboratory of Excellence GR-Ex, université de Paris, France

Tirés à part : &. Soler
eric.soler@igmm.cnrs.fr

Liens d'intérêt: L'auteur déclare n'avoir aucun lien d'intérêt en rapport avec cet article.

Enhancers, structuration spatiale des chromosomes et remaniements pathologiques : vers une meilleure compréhension des altérations complexes du génome

Abstract

Enhancers are central elements in the control of mammalian gene expression. They are frequently altered in haematological disorders. However, their distal locations within the non-coding genome and their complex modes of actions complicate their functional analysis. This review aims to provide an update on enhancer activity in normal and pathological situations, with a special focus on the three-dimensional (3D) organisation of the genome. Pathological alterations of spatial genome organisation will be discussed with the goal of providing a framework for better understanding complex genome alterations in human disease.

Résumé

Les enhancers représentent des composants centraux impliqués dans le contrôle de l'expression de nos gènes, qui sont fréquemment dérégulés dans les hémopathies tant bénignes que malignes. Leur localisation distale, au sein des régions non codantes du génome, leur dynamique et leur complexité d'action rendent cependant leur étude particulièrement difficile. Cette revue a pour objet de livrer des notions clés au sujet des enhancers, de leur dérégulation en situation pathologique et, de façon plus générale, de leur mode de fonctionnement, qui est intimement lié à l'organisation tridimensionnelle (3D) du génome. À travers la description de cas pathologiques impliquant des remaniements spatiaux de la chromatine, un des buts est de sensibiliser la communauté à l'importance de la considération du génome 3D dans les activités diagnostiques pour éclairer ou (ré)évaluer, à la lumière de ces dernières découvertes, des cas génétiques complexes ou restés obscurs ou mal compris.

Cellular functioning is based on the balance of numerous biological parameters, including biochemical and metabolic, structural, genetic and epigenetic parameters. The ability of mammalian cells to coordinate the transcriptional activity of hundreds or even thousands of genes at the same time gives them the adaptive capacity to respond to their environment or to engage in developmental or differentiation processes during ontogeny. The gene expression profile of a given cell is therefore representative of its state of differentiation, as well as its physiological

state (stress, quiescence, *etc.*), and represents its (epi)genetic "identity card". Many diseases are characterised by a disruption of this genetic "harmony", with the aberrant expression of genes that profoundly alter tissue and cell physiology, as in the case of cancers, through the activation of oncogenes or the inactivation of tumour suppressor genes, which, through a cascade reaction, permanently alters gene expression programmes.

Whereas in prokaryotes and lower eukaryotes such as yeast, gene control elements are located proximally (*e.g.* proximal promoters located immediately in the 5′ area of the genes), the organisation of the genetic elements involved in higher eukaryotes (enhancer-type, see below) is fundamentally different, and involves an additional dimension: the spatial organisation of chromatin [1-4].

The mechanisms of spatial organisation of the genome necessary for the proper functioning of gene expression are increasingly well characterised, and with the identification of new genomic structures, models have recently emerged of their dynamics and their role in normal and pathological cell function. The purpose of this article is to provide an overview of the latest advances in this field and to highlight the implications of these discoveries for our understanding of the (epi) genetic basis of many human diseases that have not yet been explained by "classic" genetics.

Enhancers: historical concepts and physiological importance

Enhancers are essential elements in the regulation of our gene expression [5]. They are characterised by short DNA sequences (a few dozen to a few hundred base pairs) scattered throughout our chromosomes. These sequences are enriched with motifs recognised by transcription factors and have the property of recruiting multi-protein complexes capable of modifying the structure of chromatin and influencing the level of transcription of the surrounding gene(s). Historically, the first enhancers, which were discovered in the 1980 s, were derived from viral sequences with the property of stimulating the expression of model genes in artificial genetic constructs (plasmids). The presence of endogenous enhancers naturally present in our genomes was soon revealed. The key role of these elements has been revealed in particular by transgenic studies which has shown that the use of promoter sequences of genes is often insufficient to allow the expression of a transgene integrated in a stable and random manner into the mammalian genome. In the majority of cases, these transgenes are weakly expressed, do not recapitulate the tissue specificity of the endogenous gene, or are aberrantly expressed in exogenous tissues. These experiments, which were particularly conducted with globin genes, highlighted the fact that additional elements were missing and led to the demonstration that enhancers play an essential role in achieving correct expression and tissue specificity [6, 7].

The major role that enhancers play in the spatiotemporal control of gene expression – i.e. tissue or cell specificity and timing of expression during development, or as a function of a stimulus or stress – has been demonstrated, among other things, in targeted inactivation studies (knock out [KO]) studies in mice. The *SHH* gene, encoding the sonic hedgehog morphogen, has long been used as a model. *SHH* has more than ten different enhancers, each conferring expression of *SHH* in a particular tissue (e.g. the central nervous system, epithelial mucosa or limb budding) at a specific stage of embryonic development. The individual inactivation of *SHH* enhancers affects developmental expression profiles only in the target tissue without altering expression in other embryonic sites, demonstrating the fundamental role of these elements in imposing precise spatiotemporal control of SHH necessary for proper embryonic development [3, 8]. The *GATA2* gene, which encodes an essential regulator for haematopoietic stem cells (HSCs) and the myeloid part of haematopoiesis, also contains numerous

enhancers. Their inactivation in vivo showed that they function in a non-redundant manner, being individually involved in the activation of *GATA2* in HSCs during myelopoiesis or in regulating its silencing during commitment to the erythroid lineage [9]. Beyond these examples focused on selected elements at model loci, larger studies have shown the general importance of these genetic elements using systematic approaches, notably in embryonic stem cells, for the maintenance of the pluripotency programme, or in different cancer models [10, 11]. Epigenomic analysis of the haematopoietic system, including stem cell and immature progenitor populations, and the various lymphoid and myeloid progenitors, has shown that enhancer activation precedes the transcriptional activation of target genes, reflecting the importance of these elements for proper gene function during cell differentiation [12]. A more recent study has shown that the self-renewal mechanism of differentiated macrophages involves the transient and punctuated activation of a set of pluripotency gene-associated enhancers, coordinated by MAF proteins [13]. Interestingly, this transient activation of pluripotency genes involves enhancers which are distinct from those used in embryonic stem cells to activate a similar self-renewal programme (centred around the *Myc* and *Klf2* genes), which indicates that the same gene may be under the control of different enhancers depending on the cell type. This observation also applies to genes showing rather general expression patterns that use sets of tissue-specific regulatory elements, even if their expression is comparable between different cell types (*e.g. Myc, Smad7, etc.*) [14]. This probably reflects different uses of regulatory factors (binding enhancers) as well as distinct regulatory needs within different tissues (*e.g.* transient induction of the pluripotency network in self-renewing macrophages, or fine modulations of expression levels in response to fluctuations in the cellular environment).

"Enhanceropathies": disruption of enhancers in human pathology

While mutations in the coding regions of genes (exons) have long been considered to be the major cause of genetic diseases in humans, it is now increasingly clear that alterations in enhancers, whether genetic, structural or epigenetic in origin, represent major aetiological components of many human diseases. Such diseases, which originate in the deregulation of enhancers, form a group of pathologies grouped under the concept of "enhanceropathies".

Some forms of thalassaemia, for example, are caused by deletions that remove enhancers of the globin genes, causing under-expression, and imbalance of α/β chains, inefficient erythropoiesis and anaemias in individuals with these mutations. More generally, the importance of enhancer sequences has been highlighted by the numerous GWAS (genome-wide association studies), which have been carried out in the past few years. These studies demonstrate that the vast majority of sequence variants naturally present in our genomes and associated with phenotypic manifestations or predispositions to develop certain pathologies in humans, such as cancers, are observed at high frequency outside the coding sequences of genes [15-18] and are highly enriched in presumed enhancer regions – *i.e.,* not yet functionally validated as enhancers but possessing the epigenetic characteristics. Sequence variations located ∼335 kb upstream of the MYC gene are notably associated with an increased susceptibility to developing colorectal cancers [15, 19]. These variants create a new binding site for the TCF4/TCF7L2 (transcription factor 4/transcription factor 7-like 2) complex, involved in the WNT signalling pathway, at an enhancer of the *MYC* oncogene. In line with the prominent role of this enhancer, its deletion in mice shows a remarkable phenotype and confers protection against colorectal cancer – which otherwise develops spontaneously in the *Apc*^min mouse model [20]. In addition, numerous studies have linked the deregulation of enhancers to the development of acute leukaemia in humans. For example, the presence of the oncogenic fusion protein, ETO2-GLIS2,

caused by chromosome 16 inversion in acute megakaryoblastic leukaemias, deregulates the activity of (super)enhancer regions, conferring an increased self-renewal capacity to leukaemic blasts [21, 22]. In T-acute lymphoblastic leukaemia (T-ALL), duplication of a 1.5 Mb enhancer region downstream of the *MYC* gene, targeted by the NOTCH pathway, is directly involved in the pathogenicity and development/maintenance of the disease [23] (*figure 1*). In another subtype of T-ALL (TAL1+), micro-insertions of upstream sequences of the *TAL1* oncogene generate a *de novo* enhancer sequence that aberrantly activates the expression of *TAL1* in T lymphoid progenitors [24, 25]. Even more surprisingly, single base variations in DNA can reactivate oncogene expression – as illustrated by a C>T variant on chromosome 1 at the *TAL1* locus, which is able to create a binding site for the YY1 factor and reactivate TAL1 expression in various cases of T-ALL by establishing a new enhancer region [26]. In the myeloid compartment, a particularly aggressive subtype of acute myeloblastic leukaemia (AML) is associated with recurrent t(3;3) (q21;q26.2) or inv(3) (q21q26.2) translocations that reposition the *GATA2* gene to the vicinity of the *EVI1* oncogene [27]. These alterations lead to both haploinsufficiency of the GATA2 gene as well as the activation of the EVI1 oncogene which together contribute to leukaemic development (*figure 1*). Additional cases of AML with 3q26 alterations have been reported, the majority with hyperactivation of the *EVI1* gene due to translocation of enhancers of myeloid origin near this gene [28]. These examples illustrate that amplification, deletion or creation of "neo-enhancers", or repositioning of enhancers in the genome, are drivers of leukaemic transformation. The functional characterisation and consideration of these genetic elements are, therefore, fundamental and often unavoidable elements in the molecular characterisation of human (genetic) diseases and cancers.

Enhancers, the non-coding genome and long-distance regulation

Although deregulation of enhancers can pose major pathological problems, our understanding of the mechanisms of action of enhancers and the control they exert on gene expression remains very patchy. Indeed, their number, structural and functional diversity, and complex modes of action make their study particularly difficult. For example, the number of enhancers in the mammalian genome far exceeds the number of genes, and their sequence variability makes it almost impossible to predict them based on genomic sequence information. Indeed, enhancers show a rather poor sequence conservation – with the exception of binding sites recognised by transcription factors – but their presence is not sufficient predictive to identify enhancers. Indeed, enhancers in erythroid tissue, for example, are largely occupied by the transcription factor GATA1, which binds to WGATAR-like sequences in the genome. Of the ∼2 million sites in the genome, only a few tens of thousands are bound by GATA1 *in vivo* and represent putative erythroid enhancers [29, 30]. On the other hand, despite an apparent lack of sequence conservation, these elements have a characteristic epigenetic signature combining post-translational modifications of histones (H3K4me1 and H3K27ac in particular), local nucleosome depletion (*e.g.* regions hypersensitive to DNAse I or accessible to transposases [ATAC-Seq]), the presence of factors characteristic of enhancers (*e.g.* P300/CBP, RNA polymerase II) and/or the presence of small transcripts known as enhancer RNAs (eRNAs), although this signature is not universal and some enhancer regions escape this epigenetic definition [1, 4].

Using this signature, it has been possible to profile the landscape of enhancers in many genomes via high-throughput approaches. These analyses have shown that:

– most genes are controlled by several enhancers;
– enhancers are generally located outside the proximal promoter regions of genes;
– enhancers can be located upstream, downstream or within introns of genes;

FIGURE 1
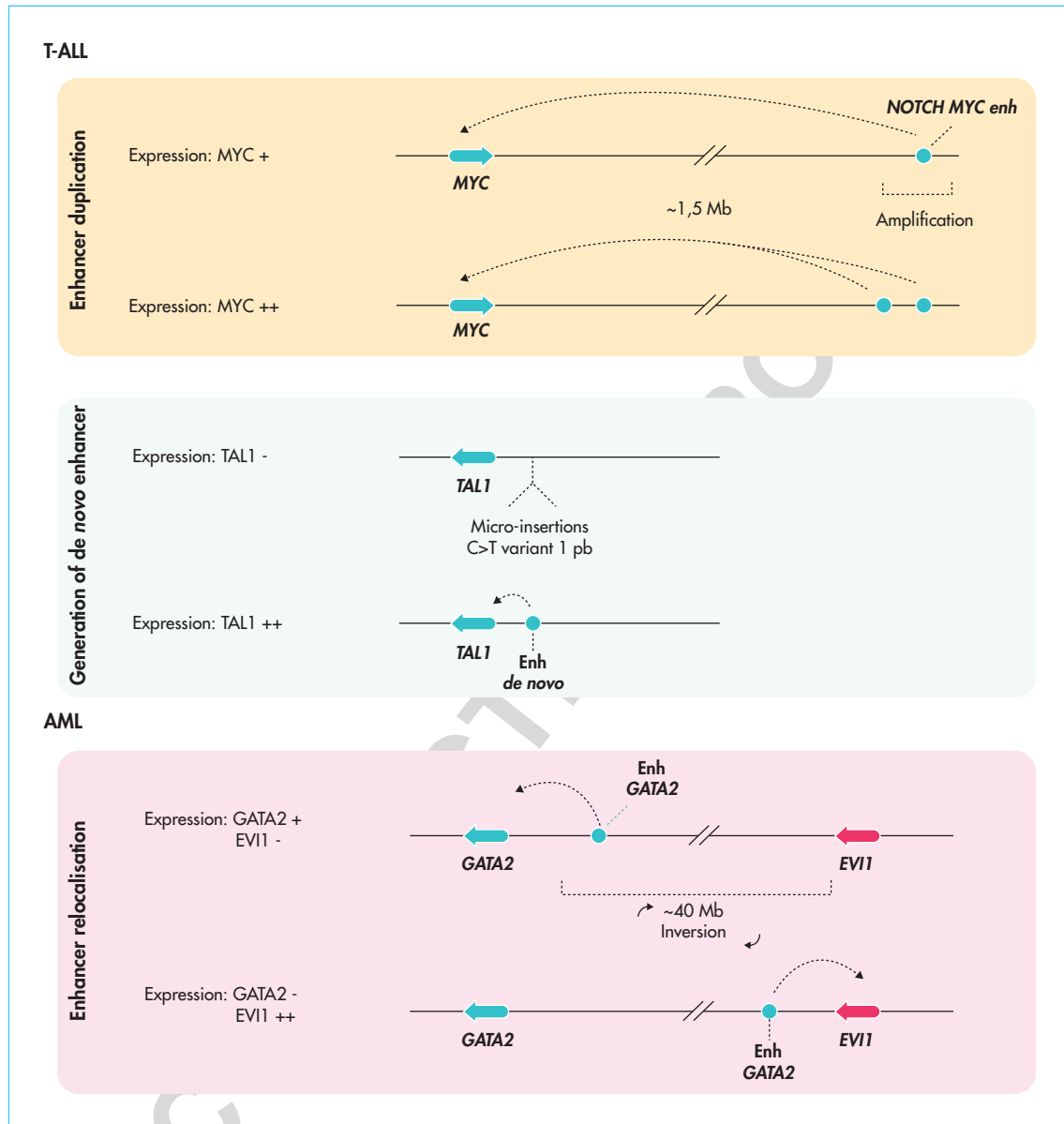
Examples of the deregulation of enhancers in acute leukaemia. Different cases of amplification, displacement or *de novo* creation of enhancers (enh) that have been implicated in the development of T-ALL or myeloid acute lymphoblastic leukaemia (AML) are represented. The result in terms of gene expression is shown for each situation.

– enhancers are usually located at considerable distances, up to more than 1 Mb from the promoters or target genes.

These discoveries have shown the richness of the so-called non-coding genome, *i.e.,* the portion of our chromosomes ($\sim$98%) distinct from the coding sequences of genes, which has been termed the "regulatory genome" by many groups [5].
In particular, these observations have shown that enhancers regulate their target genes across long genomic distances (*figure 2*A). The *KIT* oncogene has an enhancer located -114 kb upstream of its promoter [1,22] and the *MYB* oncogene is regulated by a set of enhancers spread over 125 kb, located in the intergenic region

FIGURE 2
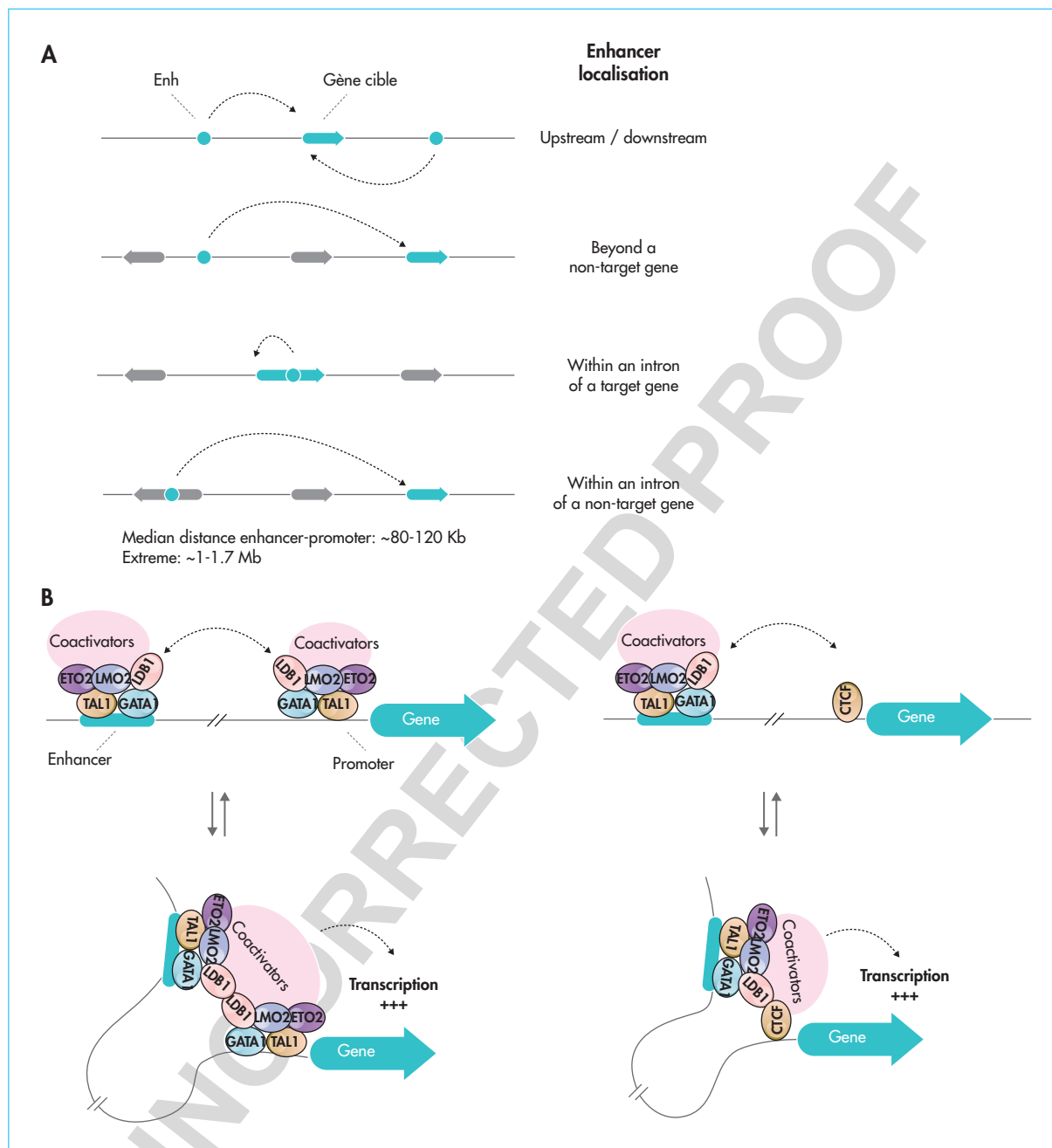
Localisation of enhancers in the human genome and mechanisms of control by chromatin folding. A) Enhancers can be identified upstream, downstream or within intronic sequences at distances of up to 1 Mb. B) Mechanism of control of distal enhancers by the LDB1 complex. The ability to dimerise (left side) or heterodimerise with CTCF (right side) allows LDB1 to bring enhancers close to their target genes (chromatin folding).

HBS1L-MYB in erythroid progenitors [4, 31-33]. More extreme cases have been described, such as the *SHH*, which has an enhancer conferring expression at limb budding during development, located 1 Mb upstream in an intron of the *LMBR1* gene and whose sequence variations are the cause of developmental syndromes such as polydactyly [8, 34, 35]. The *MYC* oncogene is further controlled in T cells

by an enhancer located ∼1.5 Mb downstream in a large intergenic region of *PVT1* [23]. In general, genome-wide analyses have estimated the median distance between enhancers and target promoters to be around 80-120 kb depending on the study [14, 36, 37]. A particularly important fact, reflecting the complexity of the organisation of our genomes, is that the gene closest to an enhancer on a linear scale (*i.e.,* distance in kilobases) is not necessarily the target gene of that enhancer, as is the case for the *SHH* limb budding enhancer present in the *LMBR1*, or certain enhancers of the *MYB* oncogene enhancers in the vicinity of the *HBS1L* [8, 31, 32]. All these observations indicate that long-distance regulation is the rule rather than the exception in mammalian genomes.

Spatial organisation of chromosomes or how to connect regulatory elements and genes in the three-dimensional space of the nucleus

This organisation of mammalian genomes raises important fundamental questions:

– how does transcriptional control of genes take place across these long genomic distances?
– how is the specificity and choice of target genes for the different enhancers ensured?
– to what extent can long-range controls and specificity be impaired in pathological situations?

Answers to these fundamental questions have been provided by the development of technologies for analysing the spatial organisation of the genome. The predominant model put forward to explain long-range transcriptional regulation is that of chromatin looping. Because the chromatin fibre is flexible, folding is possible, allowing distal enhancers to contact or be positioned near their target genes within the space of the nucleus. The first demonstration of this phenomenon was provided by the study of globin genes, which are controlled by a set of distal enhancers located several tens of kilobases upstream of the genes (a region commonly known as the LCR, [locus control region]). In erythroid cells, enhancers in the LCR contact globin genes via chromatin folds. The establishment of this three-dimensional structure is associated with strong activation of globin genes during terminal erythroid differentiation. Moreover, this type of structure is dynamic and evolves during development: enhancers preferentially control foetal globin genes in primitive erythroid cells and adult globin genes in definitive progenitors (adult haematopoiesis). This regulatory mechanism has since been observed for a multitude of genes in many tissues and organisms [1, 3, 4].
The mechanistic discoveries of the process of spatial approximation of enhancers to their target promoters have emerged based on haematopoietic tissue, and in particular, erythroid tissue [38-41]. Long-distance interactions are stabilised and regulated by direct or indirect DNA-binding proteins (transcription factors or chromatin-binding complexes). A limited number of factors have been directly implicated or associated with loop formation. Firstly, structural factors such as CTCF (CCCTC-binding factor) or the cohesin complex play a major role in genome architecture and in the organisation of enhancer-mediated interactions. In addition, some transcription factors are directly involved in the long-distance interactions of enhancers with their target genes, either by directly binding the enhancers such as LDB1 (LIM domain binding 1), YY1, or promoters such as ZNF143, or by being part of the molecular bypass mediating these interactions as is the case for the "mediator" complex [1, 2, 42]. In particular, the ubiquitous transcription factor LDB1 is one of the best characterised to date (*figure 2*B). Although not directly binding to DNA, LDB1 associates with major transcription factors in haematopoietic cells, such as GATA1, TAL1 (T-cell acute lymphocytic

leukaemia protein 1) or LMO2 (LIM domain only 2), to form a multiprotein complex known as LDB1 [30, 39, 41, 43], or complexes with other transcription factors in non-haematopoietic tissue [44]. The LDB1 complex binds predominantly to enhancers in haematopoietic cells [40, 43] and some promoters, such as those of the globin genes [39, 41]. Through its dimerisation domain, LDB1 is able to physically bring together complexes bound at distal enhancers and target promoters, as illustrated at the globin gene locus (*figure 2*B) [39, 41, 45, 46]. Its ability to heterodimerise with other structural factors, such as CTCF, also allows it to form loops that bring enhancers and target genes together at loci that do not contain an LDB1 complex on their promoters (*figure 2*B) [40].

In general, the control mechanisms of chromatin loops involve, in the majority of cases, protein-protein interactions. The folding increases the concentration of regulatory factors locally and creates a permissive environment for transcription (*figure 2*B). A demonstration of the instructive role of chromatin loops was also provided with the LDB1 factor, showing that the "forced" creation of loops by targeting LDB1 to foetal globin genes leads to reactivation of these genes in the adult, by re-establishing connections with enhancers upstream of the locus [45, 46]. However, it is important to note that the folding contact model is probably not universal, as in some cases, contact between enhancers and the target gene is not observed during transcriptional activation [47]. This suggests that other mechanisms may exist, but these are yet to be clarified.

Three-dimensional organisation with large-scale and topological domains: spatial constraints and specificity of long-distance contact

Despite these extensive mechanistic studies of chromatin folding control and long-distance enhancer-promoter interactions, important questions remain. How can the specificity of contact between enhancers and target genes be explained? Some theories point to a "compatibility" that may exist between an enhancer and its target promoter, thus explaining the specificity of interaction. However, the numerous cases of chromosomal translocations that juxtapose enhancers near oncogenes in solid tumours and haematological malignancies show that an enhancer is largely capable of interacting with a gene that is not its natural target, as in the cases of juxtaposition of the *MYC* oncogene with the enhancer of the immunoglobulin heavy chain locus in the t(8;14) translocations of Burkitt's lymphoma, or the repositioning of the *MYB* oncogene in the vicinity of the TCR-β enhancer in paediatric leukaemias with t(6;7) translocations [1, 4]. This shows that the specificity of enhancers for their target promoters is relatively low, and that ectopic contacts can occur and have a functional impact *(i.e.* transcriptional activation). The organisation of chromatin loops and long-distance interactions is therefore dependent on other parameters and, in particular, on spatial constraints that limit the interactions that enhancers can establish in the space of the nucleus and promote functional interactions with their target genes.

The deployment of high-throughput approaches to measure, at the genome level, all long-distance interactions (Hi-C type approaches [high chromosome contact map] or high-throughput chromosome conformation capture [48]) and super-resolution imaging technologies have revealed a novel level of spatial organisation of chromatin [36, 49-52]. Mammalian genomes are in fact organised in series of topological association domains (TADs), contiguous one after the other, varying in size from a few hundred kilobases to 1 Mb (*figure 3*). These domains reflect regions of chromatin showing increased spatial proximity, reflecting the preferential establishment of contacts within the domain compared to surrounding regions (outside the domain). The TADs are themselves separated by so-called "border" or "frontier" regions, showing little contact, and generally occupied by structural factors such as CTCF or the cohesin complex [2, 3, 53]. The organisation of

FIGURE 3
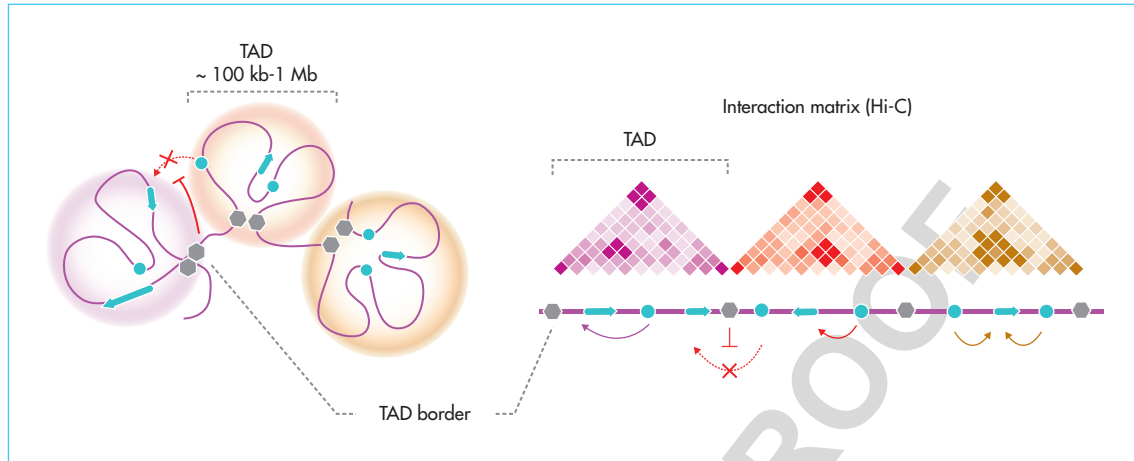
Topological domains (TADs) and spatial organisation of chromosomes. Theoretical schematic representation of TADs (left, spheres) containing spatially interacting regions of the genome (genes, enhancers, chromatin folding). The different TADs are separated by border regions (grey hexagons) enriched with structural factors (CTCF, cohesins) that isolate the TADs from each other. On the right, a Hi-C-type chromatin-chromatin interaction matrix corresponds to TADs. TADs promote intradomain gene-chancer interactions. The borders limit ectopic interactions between enhancers and genes in the surrounding domains.

chromosomes into successive TADs imposes a spatial constraint on the enhancers and target genes present in the TADs, limiting the volume and number of chromatin regions that an enhancer can `scan' to establish productive interactions. Thus, enhancer regions interact much more frequently with regions within their TAD than with regions located in the surrounding TADs. This spatial confinement therefore maximises the frequency of interactions with target genes and strongly limits ectopic interactions with non-target regions or genes, thus ensuring increased contact specificity. In summary, in this model, the borders of TAD provide space between TADs, thus limiting inter-TAD contacts and encouraging contacts within domains (intra-TAD contacts). These structures therefore play an essential role in maintaining the spatial and functional architecture of the genome. Particularly surprisingly, a large majority of the borders (with some exceptions [54]) are conserved between cell types, between different tissues or even between different species [36, 52, 55]. This observation indicates that the profound transcriptome shifts that can occur during cell development or differentiation, or in response to many stimuli, which involve spatial reconfigurations of the genome to connect enhancers to their target genes, occur primarily within preconfigured TADs [55]. This indicates that TADs and their borders are fundamental structural elements of mammalian genomes, and that gene expression dynamics are reflected by spatial configuration dynamics within TADs rather than by reconfiguration of the TADs themselves. Although the actions of enhancers are mostly restricted to the interior of TADs, it is important to note that "interdomain" contacts, *i.e.,* that encroach on a TAD border, have been observed with significant frequency from single-cell super-resolution imaging data [56]. This concept of local domain intermingling allows for control of the activity of genes located near the borders and reflects the binding dynamics of the cohesin complex and structural factors such as CTCF at the borders as well as the formation of domains. Other studies have also observed inter-TAD contacts within cells of lymphoid origin involved in antigen receptor recombination [57]. These examples underline the fact that topological domains and their borders are probably not fixed structures but exhibit structural dynamics and a certain level of permissiveness for interactions between

domains, which are visible at the single cell level but undetectable at the population level.

Disruption of topological association domains and their borders in pathology: a new approach in diagnosis

The discovery of TADs and their borders, and new concepts of the spatial organisation of the genome, have had a considerable impact on our interpretation of chromosomal changes in disease. Structural variations in the genome, such as deletions and duplications, commonly referred to as CNVs (copy number variations), or inversions and translocations, are the subject of particular attention in clinical diagnosis. In fact, CNV testing is a standard diagnostic procedure for many conditions, such as intellectual disability or congenital malformations. However, while the functional impact of many amplifications or deletions can be explained by a simple gene "dosage" effect (*e.g.* amplification of oncogenes or deletion of tumour suppressor genes), predicting the consequences of balanced rearrangements, such as inversions or CNVs restricted to non-coding regions, remains problematic. Such genetic events have the potential to alter genomic integrity, in particular the spatial organisation of chromatin, with pathological repercussions for gene expression, activation/repression timing or tissue specificity. It should be noted that such alterations remain undetectable by the exome analysis techniques commonly used in diagnosis.

The discovery of TADs and their borders represented a major turning point in our perception of genomic organisation and, above all, shed light on pathological situations that were still poorly understood (genetic diseases, developmental diseases, or cancers).

One of the first indications that TADs and their borders play a functional role in vivo in humans came from the study of several limb development syndromes (polydactyly, syndactyly and brachydactyly) [58]. These malformations are all associated with inversions, duplications or large deletions on chromosome 2q35-36 which includes loci on *EPHA4*, *PAX3*, *WNT6* and *IHH* genes (*figure 4*). These structural anomalies are varied and have few similarities. What these changes do have in common is that, in all cases, a border of the TAD containing the *EPHA4* gene is affected by either being deleted or displaced. This allows a distal enhancer of *EPHA4*, located more than 1 Mb downstream of the gene, in a genetic "desert", to make ectopic contacts with genes present in TADs adjacent to the *EPHA4* TAD, in particular with the *PAX3*, *WNT6* and *IHH* genes which leads to their aberrant expression in the developing limbs and is the cause of the developmental abnormalities observed in patients.

The effect of these abnormalities does not, therefore, come from the direct loss of genes present in the deleted regions. Importantly, the commonality between these seemingly diverse and non-identical alterations could only be identified in the light of the spatial organisation of this chromosomal region. Indeed, the pathogenicity of these structural variations is the result of a local spatial disorganisation of the genome, resulting in the displacement of a specific enhancer of limb development from the borders of the respective TAD. More generally, the pathogenicity of CNVs, especially duplications, depends on their spatial outcome [59-61]. Some duplications of intra-TAD regions lead to the deregulation of genes present within the TAD, as is the case for the *SOX9* gene in sex reversal syndrome. Some duplications may cover portions of TADs and their borders, and "recreate" a new domain, or "neo-TAD" – which in this case may be phenotypically neutral (no effect) if the duplicated region is isolated by the presence of borders while lacking the pathogenic gene. Finally, similar duplications can, conversely, lead to pathological situations in cases where the neo-TAD resulting from the duplication leads to the acquisition of a new expression profile of the duplicated gene (e.g. if it

Figure 4
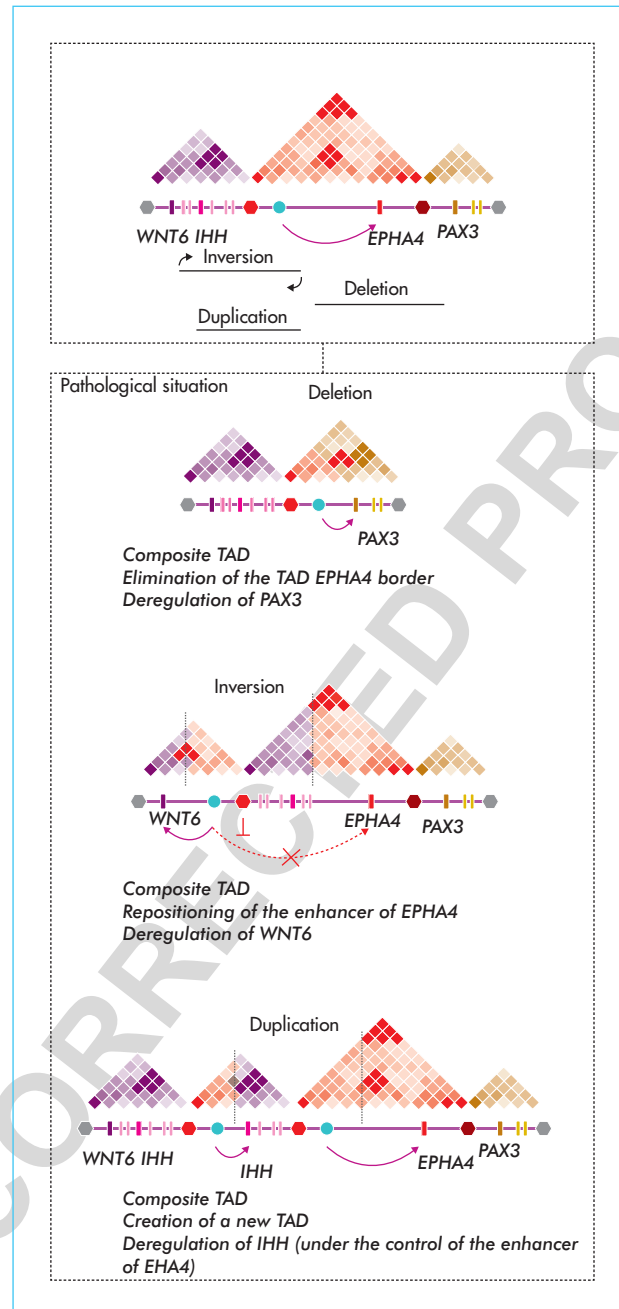
Pathological remodelling of TADs and spatial reorganisation of chromosomes. Chromosomal changes (inversions, duplications, deletions) involved in polydactyly identified in patients, and the spatial result at the level of the affected chromosomal regions. In each case, a border of the *EPHA4* locus is moved or eliminated, causing ectopic interactions between *EPHA4* and surrounding TAD genes *(WNT6, PAX3)*.

becomes the target of an ectopic enhancer), as is the case in Cooks syndrome for the *KCNJ2* gene [59] or, in colorectal cancers, for the *IGF2* gene [61] (*figure 5*). In acute leukaemia, the analysis of T-ALL genomes from lineages [62] or patient cells [63] has shown that recurrent deletions affect border sites, frequently linked by the CTCF insulator protein, and contribute to the activation of oncogenes, as is the case
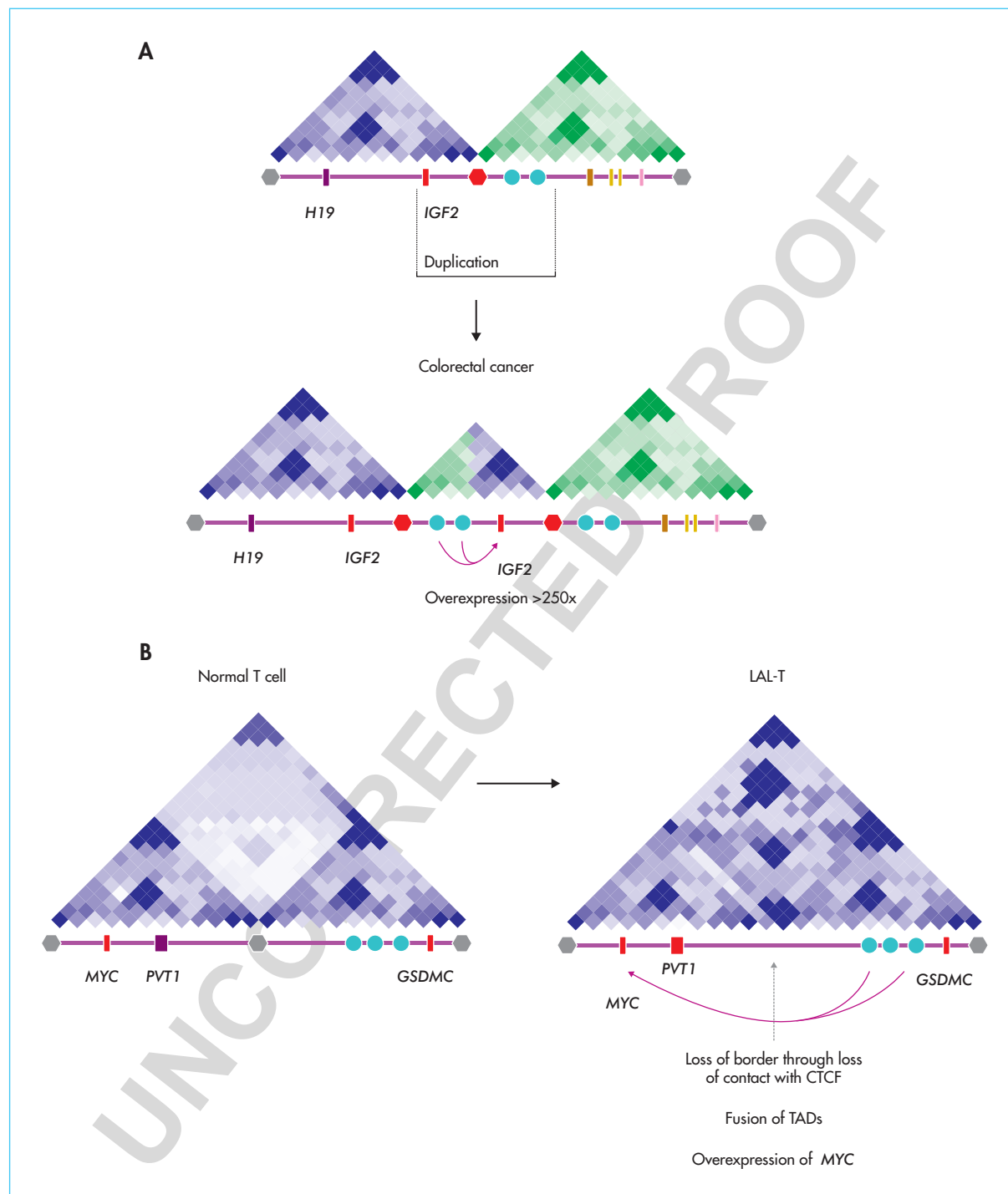
Remodelling of spatial organisation in cancer. A) Recurrent duplications resulting in the generation of a "neo-TAD" juxtaposing the *1GF2* gene to a super-enhancer region (blue dots), conferring overactivation of IGF2 in colorectal cancers. B) Example of TAD fusion at the MYC in T-ALL by loss of CTCF binding at the inner border, causing uncontrolled interaction of *MYC* with downstream enhancers. The TADs (schematic views of the Hi-C interaction matrices), the position of the genes (rectangles) as well as the borders (hexagons) are indicated.

for TAL1 and *LMO2* [62]. A fusion of TADs containing the *MYC* oncogene and promoting its interaction with a group of enhancers (super-enhancer) has also been reported in primary T-ALL [63]. The fusion results from the disappearance of a border site occupied by CTCF in normal T cells. However, the mechanisms of the loss of CTCF binding at the border remain unclear. Deletions affecting the border of the TAD containing the *FLT3* gene have also been reported in B-ALL, leading to abnormal activation of *FLT3* by an enhancer whose activity is normally restricted to the *PAN3* gene locus (present in the TAD upstream of FLT3) but which makes new contacts with *FLT3* in the absence of a border [64]. Similarly, the loss of CTCF binding due to hypermethylation of its DNA binding sites has been reported in gastrointestinal stromal tumours with mutations in the succinate dehydrogenase complex (SDH). Loss of CTCF at insulating borders leads to the establishment of ectopic interactions between groups of enhancers (super-enhancers) and amplification of KIT or *FGF3* and *FGF4* oncogene expression in these solid tumours [65]. This work highlights the commonality and causality of three-dimensional (3D) genome alterations in cancer.

Integration of three-dimensional genome data to interpret copy number variation and structural variation in patients' genomes

These recent discoveries concerning the spatial organisation of the genome and its functional impact on long-distance contacts of enhancers raise new perspectives in terms of diagnosis: do the alterations detected in certain patients involve reconfigurations of TADs? Are borders present or are they displaced by these alterations? These questions, while relevant, may be difficult to address due to the technical nature of determining the 3D genome profile of each patient, the cost of the technologies to be deployed and the bioinformatics skills required to analyse this type of data. However, there are public data systems which now offer the possibility to navigate within the 3D genome without any prior bioinformatics skills, enabling an understanding of rearrangements or mutations in the light of the spatial organisation of chromosomes. The websites listed in *table 1* offer this possibility. Indeed, the fact that TADs and their borders are largely invariant between different cell types [1-3, 35, 36] allows human 3D genome data to be used as a first approximation to determine whether mutations affect known structural regions (*e.g.,* TAD borders or structural factor binding sites). Once this initial

Table 1

Public resources for navigating and exploiting 3D genomic data.	
Websites	
3D Genome Brower	http://3dgenome.fsm.northwestern.edu
Hi-Glass	https://higlass.io/app
Capture Hi-C Plotter (CHiCP)	https://www.chicp.org
(haematopoietic tissue)	
Software	
Juicebox	https://github.com/aidenlab/Juicebox/wiki/Download

Selection of online sites allowing easy navigation within the 3D genome (visualisation of TADs and their borders). The CHiCP site focuses on haematopoietic tissue [16]. The Juicebox software developed by Erez Lieberman Aiden's group [36] makes it possible to browse and visualise a multitude of published Hi-C data, also integrating epigenetic data derived from the ENCODE database.

analysis has been carried out, confirmations of reconfigurations of the spatial organisation of cells carrying mutations can be made using specific technologies. In this framework, Hi-C type approaches allow to determine the set of genomic interactions and to identify the TADs and their respective borders [36]. This approach is comprehensive but expensive and requires significant sequencing in order to achieve good resolution. More targeted approaches, such as Capture-Hi-C, capture a predefined region to perform Hi-C, and reduce sequencing costs [66, 67]. Other approaches, such as the 4C-Seq, or the Capture-C [68-71] can, like Hi-C/Capture-Hi-C, determine the local spatial configuration of chromatin with very high resolution and at low cost (by offering the possibility of multiplexing the analyses). We have published detailed experimental procedures [69, 71] and dedicated bioinformatics analysis tools [72] for easy implementation and to make these technologies accessible to a large number of people. Finally, other approaches offer the possibility of combining both chromatin profiling (ChIP-Seq type) and chromatin immunoprecipitation assays with sequencing and genomic interaction analysis (Hi-C type). These technological tools, such as ChIA-PET (chromatin interaction analysis by pair-end tag sequencing) [73] or, more recently, the Hi-ChIP [74], which is easier to implement, offer the possibility of detecting long-distance interactions involving a factor of choice (*e.g.* CTCF or a transcription factor) or selected genomic regions (*e.g.* enhancers, using their characteristic epigenetic marks such as H3K27Ac), on a genome-wide scale. These approaches thus offer a global vision but remain focused on selected factors or genomic elements.

There are, however, limitations to this new way of understanding genome modification. Some genomes of cancers with complex karyotypes have so many alterations that it is almost impossible to define which ones are oncogenic event drivers and which are phenotypically silent or irrelevant [75]. More importantly, it has recently been shown that targeted disruption of TAD borders at the model locus of the *SHH* gene have no effect on the *SHH* expression profile nor on embryonic development in vivo [8]. This suggests that some borders are less essential than others in the genome. This observation represents a further complication in making sense of the structural changes. One of the approaches to be adopted is, therefore, to combine the recurrence of changes with those affecting expression of the surrounding genes in order to identify alterations with true functional impact.

Conclusions

The integration of the spatial component of chromosomes has shed light on disturbances caused by genetic rearrangements in humans. The development of methods for analysing the spatial conformation of the genome, now available at the single cell level [76, 77], have pushed the technological limits and scope of what is possible. These methods will undoubtedly allow us to map the structural elements of the genome and their dynamic connections with the surrounding genes in ever greater detail. This growing wealth of information provides a much better understanding of the poorly understood mechanisms of certain still unexplained gene deregulations. The deepening of this knowledge will undoubtedly lead to a better understanding of the sophisticated orchestration of the mechanisms controlling the biology of genomes.]

References

[1] Cico A, Andrieu-Soler C, Soler E. Enhancers and their dynamics during hematopoietic differentiation and emerging strategies for therapeutic action. *FEBS Lett* 2016 ; 590 : 4084-104.

[2] Dixon JR, Gorkin DU, Ren B. Chromatin domains: the unit of chromosome organization. *Mol Cell* 2016 ; 62 : 668-80.

[3] Robson MI, Ringel AR, Mundlos S. Regulatory landscaping: how enhancer-promoter communication is sculpted in 3D. *Mol Cell* 2019 ; 74 : 1110-22.

[4] van den Heuvel A, Stadhouders R, Andrieu-Soler C, Grosveld F, Soler E. Long-range gene regulation and novel therapeu-

tic applications. *Blood* 2015 ; 125 : 1521-5.

[5] Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* 2011 ; 144 : 327-39.

[6] Fraser P, Pruzina S, Antoniou M, Grosveld F. Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev* 1993 ; 7 : 106-13.

[7] Grosveld F, van Assendelft GB, Greaves DR, Kollias G. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* 1987 ; 51 : 975-85.

[8] Williamson I, Kane L, Devenney PS, *et al.* Developmentally regulated Shh expression is robust to TAD perturbations. *Development* 2019 ; 146 : dev179523.

[9] Bresnick EH, Johnson KD. Blood disease-causing and -suppressing transcriptional enhancers: general principles and GATA2 mechanisms. *Blood Adv* 2019 ; 3 : 2045-56.

[10] Korkmaz G, Lopes R, Ugalde AP, *et al.* Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol* 2016 ; 34 : 192-8.

[11] Li K, Liu Y, Cao H, *et al.* Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nat Commun* 2020 ; 11 : 485.

[12] Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, *et al.* Immunogenetics. Chromatin state dynamics during blood formation. *Science* 2014 ; 345 : 943-9.

[13] Soucie EL, Weng Z, Geirsdóttir L, *et al.* Lineage-specific enhancers activate self-renewal genes in macrophages and embryonic stem cells. *Science* 2016 ; 351 : aad5510.

[14] Kieffer-Kwon K-R, Tang Z, Mathe E, *et al.* Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* 2013 ; 155 : 1507-20.

[15] Pomerantz MM, Ahmadiyeh N, Jia L, *et al.* The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet* 2009 ; 41 : 882-4.

[16] Javierre BM, Burren OS, Wilder SP, *et al.* Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* 2016 ; 167 : Error: FPage (1369) is higher than LPage (84e19)!.

[17] van der Harst P, Zhang W, Mateo Leach I, *et al.* Seventy-five genetic loci influencing the human red blood cell. *Nature* 2012 ; 492 : 369-75.

[18] Maurano MT, Humbert R, Rynes E, *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012 ; 337 : 1190-5.

[19] Tuupanen S, Turunen M, Lehtonen R, *et al.* The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet* 2009 ; 41 : 885-90.

[20] Sur IK, Hallikas O, Vähärautio A, *et al.* Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* 2012 ; 338 : 1360-3.

[21] Thirant C, Ignacimouttou C, Lopez CK, *et al.* ETO2-GLIS2 hijacks transcriptional complexes to drive cellular identity and self-renewal in pediatric acute megakaryoblastic leukemia. *Cancer Cell* 2017 ; 31 : 452-65.

[22] Benbarche S, Lopez CK, Salataj E, *et al. Screening of clustered regulatory elements reveals functional cooperating dependencies in Leukemia.* bioRxiv. 2019. Available from: https://doi.org/10.1101/680165 (accessed on ==Q2==).

[23] Herranz D, Ambesi-Impiombato A, Palomero T, *et al.* A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat Med* 2014 ; 20 : 1130-7.

[24] Mansour MR, Abraham BJ, Anders L, *et al.* Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* 2014 ; 346 : 1373-7.

[25] Navarro J-M, Touzart A, Pradel LC, *et al.* Site- and allele-specific polycomb dysregulation in T-cell leukaemia. *Nat Commun* 2015 ; 6 : 6094.

[26] Liu Y, Li C, Shen S, *et al.* Discovery of regulatory noncoding variants in individual cancer genomes by using cis-X. *Nat Genet* 2020 ; 52 : 811-8.

[27] Gröschel S, Sanders MA, Hoogenboezem R, *et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* 2014 ; 157 : 369-81.

[28] Ottema S, Mulet-Lazaro R, Beverloo HB, *et al.* Atypical 3q26/MECOM rearrangements genocopy inv(3)/t(3;3) in acute myeloid leukemia. *Blood* 2020 ; 136 : 224-34.

[29] Fujiwara T, O'Geen H, Keles S, *et al.* Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol Cell* 2009 ; 36 : 667-81.

[30] Soler E, Andrieu-Soler C, de Boer E, *et al.* The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev* 2010 ; 24 : 277-89.

[31] Stadhouders R, Thongjuea S, Andrieu-Soler C, *et al.* Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. *EMBO J* 2012 ; 31 : 986-99.

[32] Stadhouders R, Aktuna S, Thongjuea S, *et al.* HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. *J Clin Invest* 2014 ; 124 : 1699-710.

[33] Andrieu-Soler C, Soler E. When basic science reaches into rational therapeutic design: from historical to novel leads for the treatment of b-globinopathies. *Curr Opin Hematol* 2020 ; 27 : 141-8.

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

[35] Lettice LA, Williamson I, Wiltshire JH, *et al.* Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly. *Dev Cell* 2012 ; 22 : 459-67.

[36] Rao SSP, Huntley MH, Durand NC, *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014 ; 159 : 1665-80.

[37] Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature* 2012 ; 489 : 109-13.

[38] Palstra R-J, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. The betaglobin nuclear compartment in development and erythroid differentiation. *Nat Genet* 2003 ; 35 : 190-4.

[39] Deng W, Lee J, Wang H, *et al.* Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 2012 ; 149 : 1233-44.

[40] Lee J, Krivega I, Dale RK, Dean A. The LDB1 complex co-opts CTCF for erythroid lineage-specific long-range enhancer interactions. *Cell Rep* 2017 ; 19 : 2490-502.

[41] Krivega I, Dale RK, Dean A. Role of LDB1 in the transition from chromatin looping to transcription activation. *Genes Dev* 2014 ; 28 : 1278-90.

[42] Kim S, Shendure J. Mechanisms of interplay between transcription factors and the 3D genome. *Mol Cell* 2019 ; 76 : 306-19.

[43] Stadhouders R, Cico A, Stephen T, *et al.* Control of developmentally primed erythroid genes by combinatorial corepressor actions. *Nat Commun* 2015 ; 6 : 8893.

[44] Caputo L, Witzel HR, Kolovos P, *et al.* The Isl1/Ldb1 complex orchestrates genome-wide chromatin organization to instruct differentiation of multipotent cardiac progenitors. *Cell Stem Cell* 2015 ; 17 : 287-99.

[45] Breda L, Motta I, Lourenco S, *et al.* Forced chromatin looping raises fetal hemoglobin in adult sickle cells to higher levels than pharmacologic inducers. *Blood* 2016 ; 128 : 1139-43.

[46] Deng W, Rupon JW, Krivega I, *et al.* Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* 2014 ; 158 : 849-60.

[47] Benabdallah NS, Williamson I, Illingworth RS, *et al.* Decreased enhancer-promoter proximity accompanying enhancer activation. *Mol Cell* 2019 ; 76 : Error: FPage (473) is higher than LPage (84.e7)!.

[48] Lieberman-Aiden E, van Berkum NL, Williams L, *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009 ; 326 : 289-93.

[49] Cardozo Gizzi AM, Cattoni DI, Fiche J-B, *et al.* Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms. *Mol Cell* 2019 ; 74 : Error: FPage (212) is higher than LPage (22e5)!.

[50] Szabo Q, Donjon A, Jerkovic' I, *et al.* Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat Genet* 2020 ; 52 : 1151-7.

[51] Nora EP, Lajoie BR, Schulz EG, *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 2012 ; 485 : 381-5.

[52] Dixon JR, Selvaraj S, Yue F, *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012 ; 485 : 376-80.

[53] Li Y, Haarhuis JHI, Sedeño Cacciatore Á, *et al.* The structural basis for cohesin-CTCF-anchored loops. *Nature* 2020 ; 578 : 472-6.

[54] Bonev B, Mendelson Cohen N, Szabo Q, *et al.* Multiscale 3D genome rewiring during mouse neural development. *Cell* 2017 ; 171 : Error: FPage (557) is higher than LPage (72e24)!.

[55] Dixon JR, Jung I, Selvaraj S, *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* 2015 ; 518 : 331-6.

[56] Luppino JM, Park DS, Nguyen SC, *et al.* Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. *Nat Genet* 2020 ; 52 : 840-8.

[57] Proudhon C, Snetkova V, Raviram R, *et al.* Active and inactive enhancers cooperate to exert localized and long-range control of gene regulation. *Cell Rep* 2016 ; 15 : 2159-69.

[58] Lupiáñez DG, Kraft K, Heinrich V, *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 2015 ; 161 : 1012-25.

[59] Franke M, Ibrahim DM, Andrey G, *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* 2016 ; 538 : 265-9.

[60] Melo US, Schöpflin R, Acuna-Hidalgo R, *et al.* Hi-C identifies complex genomic rearrangements and TAD-shuffling in developmental diseases. *Am J Hum Genet* 2020 ; 106 : 872-84.

[61] Weischenfeldt J, Dubash T, Drainas AP, *et al.* Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat Genet* 2017 ; 49 : 65-74.

[62] Hnisz D, Weintraub AS, Day DS, *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* 2016 ; 351 : 1454-8.

[63] Kloetgen A, Thandapani P, Ntziachristos P, *et al.* Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat Genet* 2020 ; 52 : 388-400.

[64] Yang M, Safavi S, Woodward EL, *et al.* 13q12.2 deletions in acute lymphoblastic leukemia lead to upregulation of FLT3 through enhancer hijacking. *Blood* 2020 ; 136 : 946-56.

[65] Flavahan WA, Drier Y, Johnstone SE, *et al.* Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature* 2019 ; 575 : 229-33.

[66] Tjalsma SJ, de Laat W. Novel orthogonal methods to uncover the complexity and diversity of nuclear architecture. *Curr Opin Genet Dev* 2020 ; 67 : 10-7.

[67] Denker A, de Laat W. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev* 2016 ; 30 : 1357-82.

[68] Davies JOJ, Telenius JM, McGowan SJ, *et al.* Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods* 2016 ; 13 : 74-80.

[69] Stadhouders R, Kolovos P, Brouwer R, *et al.* Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc* 2013 ; 8 : 509-24.

[70] Grob S, Cavalli G. Technical review: a hitchhiker's guide to chromosome conformation capture. *Methods Mol Biol* 2018 ; 1675 : 233-46.

[71] Brouwer RWW, van den Hout MCGN, van Ijcken WFJ, Soler E, Stadhouders R. Unbiased interrogation of 3D genome topology using chromosome conformation capture coupled to high-throughput sequencing (4C-Seq). *Methods Mol Biol* 2017 ; 1507 : 199-220.

[72] Thongjuea S, Stadhouders R, Grosveld FG, Soler E, Lenhard B. r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Res* 2013 ; 41 : e132.

[73] Fullwood MJ, Liu MH, Pan YF, *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 2009 ; 462 : 58-64.

[74] Mumbach MR, Rubin AJ, Flynn RA, *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* 2016 ; 13 : 919-22.

[75] Fagnan A, Bagger FO, Piqué-Borràs MR, *et al.* Human erythroleukemia genetics and transcriptomes identify master transcription factors as functional disease drivers. *Blood* 2020 ; 136 : 698-714.

[76] Nagano T, Lubling Y, Várnai C, *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* 2017 ; 547 : 61-7.

[77] Collombet S, Ranisavljevic N, Nagano T, *et al.* Parental-to-embryo switch of chromosome organization in early embryogenesis. *Nature* 2020 ; 580 : 142-6.

Questions à l'auteur

Q1   Please provide the document heading.

Q2   Kindly confirm on which date the website was accessed.