PLEASE READ ALL 'COMMENT' TEXT BEFORE PREPARING YOUR ARTICLE. If you do not see the Comments, select View > Print Layout. Please delete them before submitting (Review > Delete All Comments in Document) so that peer reviewers see a clean copy of the manuscript.

Remember that you are writing for an interdisciplinary audience. Please be sure to discuss interdisciplinary themes, issues, debates, etc. where appropriate. Note that the WIREs are forums for review articles, rather than primary literature describing the results of original research.

If you have any questions, contact your editorial office.

WIREs Wiley Interdisciplinary Reviews

Article Title: Topologically Associated Domains: a successful scaffold for the evolution of gene regulation in animals

Article Type: Advanced Review

○ OPINION ○ PRIMER ○ OVERVIEW

○ ADVANCED REVIEW ○ FOCUS ARTICLE ○ SOFTWARE FOCUS

Authors:

Full name an affiliation; email address if corresponding author; any conflicts of interest

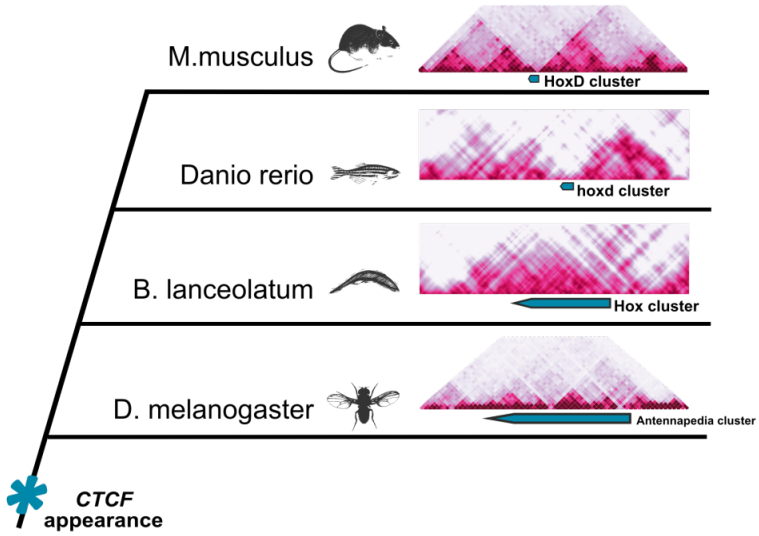
First author Rafael D. Acemel; Centro Andaluz de Biología del Desarrollo (CABD), Consejo Superior de Investigaciones Científicas/Universidad Pablo de Olavide, Carretera de Utrera Km1, Seville, Spain. rdacemel@gmail.com
Second author Ignacio Maeso; Centro Andaluz de Biología del Desarrollo (CABD), Consejo Superior de Investigaciones Científicas/Universidad Pablo de Olavide, Carretera de Utrera Km1, Seville, Spain.
Third author José Luis Gómez-Skarmeta; Centro Andaluz de Biología del Desarrollo (CABD), Consejo Superior de Investigaciones Científicas/Universidad Pablo de Olavide, Carretera de Utrera Km1, Seville, Spain. jlgomska@upo.es

Abstract The evolution of gene regulation is considered one of the main drivers causing the astonishing morphological diversity in the animal kingdom. Gene regulation in animals heavily depends upon cis-regulatory elements, discrete pieces of DNA that interact with target promoters to regulate gene expression. In the last years, Chromosome Conformation Capture experiments (4C-seq, 5C and HiC) studies in several organisms have shown that many bilaterian genomes are organized in the 3D chromatin space in compartments called Topologically Associated Domains (TADs). The appearance of the architectural protein CTCF in the bilaterian ancestor likely facilitated

the origin of this chromatin 3D organization. TADs play a critical role favoring the contact of cis-regulatory elements with their proper target promoters (that often lay within the same TAD) and preventing undesired regulatory interactions with promoters located in neighboring TADs. We propose that TAD may have had a major influence in the story of the evolution of gene regulation. They have contributed to the increment of bilaterians regulatory complexity by allowing newly evolved cis-regulatory elements to find target promoters in a range of hundreds of kilobases. In addition, they have conditioned the mechanisms of evolution of gene regulation. These mechanisms include the appearance, removal or relocation of TAD borders. Such architectural changes have been able to wire or unwire promoters with different sets of cis-regulatory elements in a single mutational event. We discuss the contribution of these architectural changes to the generation of critical genomic 3D structures required for new regulatory mechanisms associated to morphological novelties.

Keywords: Topologically associated domains; TADs; evolution; gene regulation; cis-regulatory element; HiC; 4C-seq; CTCF; enhancer

Graphical/Visual Abstract and Caption



Legend

Chromatin organization in TADs in the bilaterian lineage facilitated the evolution of cis-regulatory information

Introduction

The astonishing morphological diversity observed among the animal kingdom has been long believed to be due to gene expression differences during development of the so-called developmental genes,

a common toolkit of genes that control vast downstream regulatory networks.[1] Genes in this group are often pleiotropic, that is, required and recruited in disparate cell populations and developmental stages. In animals, this complex transcriptional behavior is mainly regulated through the combinatory interplay of discrete modules of regulatory information called *cis*-regulatory elements (CREs), which activate or repress genes in response to different cellular contexts.[2, 3] CREs are non-coding segments of DNA distributed in the vicinity of their target genes. CREs contain binding sites for a number of transcription factors, which work in combination to control the spatial and temporal specificity of CRE activity.[4] Active CREs lay in regions of open chromatin[5, 6] and display common histone modifications signatures[7–9], features that can be assayed genome-wide. Evolutionarily, these CRE hallmarks seem to be already present in the last common ancestor of bilaterians and cnidarians.[10]

Remarkably, the genomic regions containing developmental gene CREs are often very large, extending more than 1 Mb in some mammalian examples.[11–13] Within these extended regulatory landscapes, distant CREs need to physically interact with their target promoters to exert their function,[14] raising the question of how the genome is organized in the three-dimensional (3D) space to facilitate specific long-range interactions, while at the same time avoiding detrimental ones. Chromosome Conformation Capture assays (3C[15]) and their genome-wide derivatives (4C-seq[16], 5C[17] and HiC[18]) have shed light on the genomic architecture of the nucleus (Box 1). These techniques have revealed that, in some organisms, chromosomes are spatially subdivided in units called topologically associated domains (TADs). In this context, chromatin contacts within TADs are favored, while interactions between neighboring TADs are largely prevented. To date, bona fide TADs have been identified in five mammal species[19] as well as in *Drosophila*,[20, 21] although further evidence suggest that they are the rule rather than the exception in bilaterians (discussed later in this review, Figure 1). In mammals, these structures range in size from hundred of kilobases to a few megabases; in contrast, the average TAD size is rarely bigger than 100 kb in *Drosophila*, in agreement with the reduced genome size and the shorter distances between long-range CREs and their target promoters.[22, 23] Accordingly, it has been proposed that this genome compartmentalization is paralleled at the regulatory level: CREs and their target promoters are located within the same TAD, favoring their contacts. At the same time, TAD boundaries would prevent interactions between loci belonging to different TADs that could potentially lead to misregulatory effects.[24, 25] Interestingly, TAD boundaries seem to be largely stable across different cell types[26–29] (Box 1) and to be highly conserved among related organisms,[19] and recent results have shown that some of these TAD boundaries may have an even deeper evolutionary history predating the origin of vertebrates (see Figure 1b).[30]

Similarities and differences at the gene expression level have already been linked to conservation and evolution of morphological traits, within and across phyla, and are critical for body plan formation.[31–33] However, so far, most comparative studies have not taken into account the recent advances in the understanding of the 3D organization of the genome. We therefore consider that the comprehensive mapping and comparison of equivalent TADs and their CREs in different organisms will be essential for linking gene expression variation with the genomic sequence underneath. Here, in light of recent reports comparing the 3D chromatin structures in different organisms, both genome-wide and at specific loci, we evaluate how the presence of TADs could have conditioned the mechanisms of evolution of gene regulation and development in the animal kingdom. In addition, we discuss how the study of the 3D genome may impact future comparative

genomic projects. We also address the challenges and benefits of expanding the still sparse collection of chromosome conformation capture data to other evolutionary key organisms.

Sidebar title: 3C-based techniques in TAD definition and identification

TADs were defined simultaneously by two papers in 2012, which used HiC and 5C experiments in human and mouse cells.[28, 89] Both experimental approaches are part of the chromosome conformation capture techniques, which are all based on chromatin crosslinking. In brief, in a first step, nuclei are chemically crosslinked to preserve DNA-DNA, DNA-protein, and protein-protein interactions—that is, the nuclear 3D architecture. Subsequently, chromatin is digested with a restriction enzyme and ligated under diluted conditions to favor intramolecular ligations. In this way, non-consecutive restriction fragments are joined together with a frequency that is proportional to their spatial proximity. Finally, these non-consecutive ligation events are mapped and quantified, and their proportions reflects the nuclear architecture.[90] In this review, we mainly discuss results obtained either with 4C-seq or with HiC. 4C-seq is a technique that interrogate contacts established by one locus (viewpoint or bait) with the rest of the genome and is often presented as a linear profile, with peak heights being proportional to the interactions established by a given fragment with the bait. HiC interrogate at once all the contacts taking place between every restriction fragment in the genome. Therefore, the data is often shown as a heatmap that represent the contact frequencies between all the fragment combinations.

However, these approaches have a core limitation: they provide an average of the nuclear architecture of all cells assayed, thereby neglecting the cell-to-cell variability that almost certainly takes place. Even if these techniques are applied to single cells, individual nuclear structures are difficult to address, since only one ligation junction per restriction fragment is monitored. Thus, it cannot be ruled out that a particular fragment is interacting with more fragments than the one captured. Finally, the dynamics of chromatin interactions cannot be followed over time for a single cell. Since TADs were identified as self-interacting regions, with each interaction represented by a HiC- or 5C ligation event, the same limitations discussed for 3C-based methods also affects the interpretation of TAD definitions. For instance, 3C-based techniques do not give information about what extent of the contacts inside a TAD happen simultaneously, or how many mutually exclusive structures are found in the different cells of the population. In this regard, and although it cannot resolve individual nuclear architectures, single-cell HiC has revealed contact variability even within the same cell type.[91] The majority of the contacts observed differed from cell to cell yet shared that they did not cross over the TAD boundaries identified by regular HiC. Moreover, overall average chromatin structures and TAD boundaries in particular are rather stable across different cell types and throughout development.[22, 28, 29] Therefore, irrespective of our limited understanding of how TADs may appear at the single-cell level and their temporal dynamics, we can assume that the majority of CREs regulating the expression of a target gene must be found within the same TAD, even when this TAD has been identified in a cell type in which the CRE is not active and even when a very heterogeneous sample with many different cell types, such as a whole embryo, has been used.

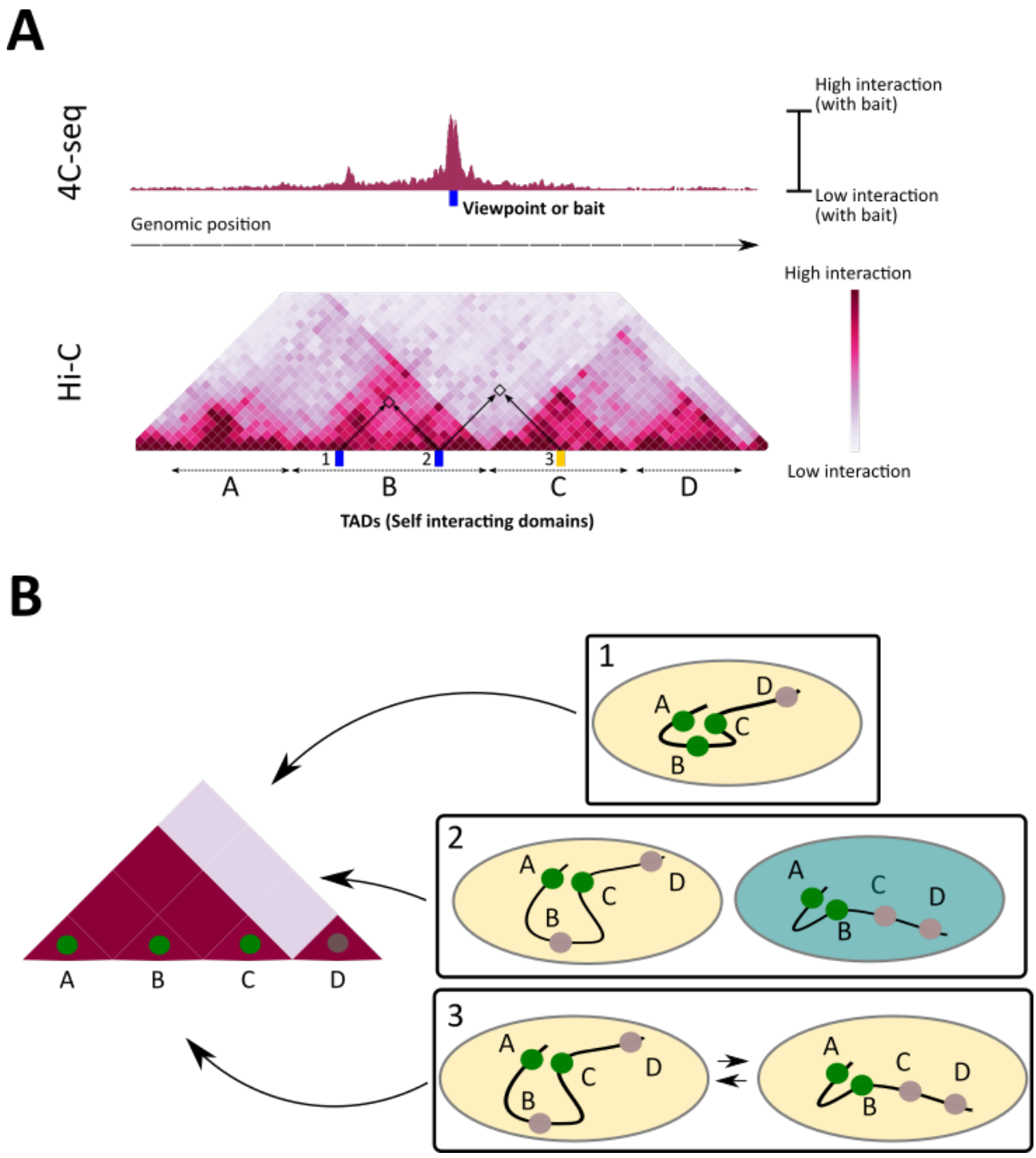


Figure Box 1: C-techniques box. **(a)** Most commonly used plots to represent 4C-seq (top) and HiC (bottom) data. In 4C-seq profiles, the X axis indicates the genomic position and the Y axis the frequency of interaction of the bait (blue square) with each genomic position. HiC and 5C heatmaps display the frequency of every pair-wise interaction in a genomic area. In the example shown, the darker the color of the square connecting two genomic regions, the higher is the frequency of interaction between them. For instance, genomic regions numbered 1 and 2 interact much more often than the region 2 with the region 3. Looking at the general picture, it is possible to distinguish several self-interacting regions (dark triangles, A-D) that define TADs. **(b)** A representation of a TAD

containing the fragments A, B and C that interact very frequently with each other (left). This TAD configuration is obtained from HiC or 5C experiments, and therefore, is an average of the genome architecture of all the cells used in these assays. In these experiments it is impossible to distinguish which of the three scenarios shown on the right (among many others) is actually taking place in each particular cell. In the first scenario, all three fragments are always interacting with each other. In the second, in one cell population fragment A interacts only with fragment B while in the other interacts only with fragment C. In the third, contacts are dynamic within the same cell populations and fragment A switches contacts between fragment B and fragment C. Despite the lack of temporal and single cell resolution, if we are interested in determining the genomic region where the CREs controlling a gene located in fragment A are contained, the average HiC profile allow us to infer that most likely these CREs will not lie beyond fragment C.

EVIDENCE OF WIDESPREAD TAD ORGANIZATION ACROSS BILATERIANS

As commented above, TADs have only been described in a handful of species. This raises the question of when this form of chromatin organization appeared in evolution, as well as its degree of evolutionary conservation across different lineages. As indicated in Box 1, TADs are revealed by those chromosome conformation capture techniques that interrogate contacts of multiple genomic regions at the same time (namely, 5C and HiC). Thus, information about how the chromatin folds genome-wide is limited to those organisms for which this kind of data is available. To date, there is HiC information available from several mammals (human,[28, 29] mouse,[28] macaque, dog, and rabbit[19]), one insect (*Drosophila melanogaster*[20, 21]), one nematode (*Caenorhabditis elegans*[34]), one plant (*Arabidopsis thaliana*[35]) two yeast species (*Saccharomyces cerevisiae*[36] and *Saccharomyces pombe*[37]) and one proteobacteria (*Caulobacter crescentus*[38])(see review by Dekker and Heard[23]). The emerging picture from these studies is that so far, a TAD chromatin compartmentalization controlling and organizing long-range transcriptional regulation has only been described in bilaterian animals (e.g., mammals, *Drosophila*, zebrafish, and amphioxus, see Figure 1a).[24] TADs are readily observed in the five studied mammals and in *D. melanogaster*. In addition, evidence of TADs in both the teleost fish *D. rerio* (zebrafish) and the cephalochordate *B. lanceolatum* (European amphioxus) have also appeared after analyzing arrays of 4C-seq experiments of the Hox genes region[39] and using this type of data to computationally reconstruct the 3D architecture of these loci.[40] Similarly, the presence of an ancient interaction border in the middle of the Six genes cluster of sea urchin, which is highly conserved with the TAD border present in the orthologous regions in vertebrates, supports the existence of TADs in the echinoderm lineage (Figure 1b).[30] In contrast, in non-animal lineages the situation is very different. No clear compartmentalization has been observed in plants or in the yeast *S. cerevisiae*, where chromatin interactions are related to linear chromosomal distance and to the presence of major chromosomal features such as the centromeres.[35, 36] Two other species, *S. pombe* and *C. crescentus*, do show some sort of sub-megabase compartmentalization. However, the non-TAD compartments of these lineages have most probably evolved independently: chromatin domains seem to be established and maintained by molecular mechanisms different from those described in bilaterians and they are not functionally related to long-range regulation.[23] Finally, *C. elegans* poses a peculiar case: despite the fact that this nematode is obviously a bilaterian animal,

most of its genome is not organized in TADs. Chromatin compartments are only present in the X-chromosome of hermaphrodites and depend on the presence of the DCC protein, a lineage-specific duplicate of cohesin required for gene dosage compensation; [34] but as discussed below, this situation is probably the result of a secondary lost of the ancestral bilaterian TAD organization.

Thus, in this review we will only consider as *bona fide* TADs the chromatin compartments of bilaterian animals and discuss TAD organization as a widespread ancestral feature of bilaterian genomes. In the lineages studied so far, TADs are maintained by a common core of conserved architectural proteins (comprehensively reviewed in [41, 42]), among which the DNA-binding protein CTCF, stands out.[43] CTCF has been shown to be fundamental for the 3D organization of the chromatin: CTCF proteins bound to very distantly located CTCF DNA binding sites that are oriented in a convergent manner are able to dimerize and generate chromatin loops. These loops are stabilized by the interactions of CTCF with other architectural proteins such as cohesin or mediator, thereby influencing loop dynamics. In this regard, CTCF has been found to be required for transient and cell-specific loops, CRE-promoter interactions, and constitutive structural chromatin contacts, such as those at TAD boundaries.[26] Indeed, CTCF binding sites arranged in a diverging orientation are characteristic signatures of TAD boundaries.[29, 44] Furthermore, the recent identification of the same diverging arrangement in a highly conserved TAD border that predates the origin of deuterostomes suggests that the crucial importance of CTCF orientation is probably an ancestral genomic feature.[30] However, it is important to note that in the case of flies, no clear association between looping and TAD borders with CTCF binding site orientation has been identified so far.[45]

But perhaps, the most interesting aspect of CTCF from an evolutionary point of view is its phylogenetic distribution. Comparative analyses of available animal genomes have shown that CTCF is only present in bilaterians. Therefore it is very tempting to speculate that the appearance of CTCF in the ancestor of bilaterians could have facilitated the evolution of long-range CRE–promoter interactions through chromatin loops in this lineage.[43] Indeed, indirect evidence indicates that distant enhancer–promoter interactions are ancestral at least to bilaterians, and are still operating in most extant bilaterian lineages.[46] Recent studies of microsynteny conservation (i.e. precise conservation of gene order) across different animal phyla have revealed the presence of hundreds of highly conserved microsyntenic pairs in bilaterians (i.e. pairs of genes maintained as immediate genomic neighbors),[47] Microsynteny conservation has been shown to be linked to regulatory constraints imposed by long-range CREs from developmental genes that are intermingled among the introns of neighboring bystander genes.[46, 48, 49] And in most cases, the genomic distances mediating between these CREs located in bystander genes and their target promoters are probably large enough to require chromatin looping[50] and thus also chromatin organization in TADs. This is in stark contrast to what has been observed in other eukaryotes such as plants or funghi, where no evidence of long-range regulation has been found yet, further suggesting that the evolution of CTCF and TADs in animals could have been concomitant and probably interconnected.[50] In addition, the integration of TADs with core nuclear processes such as DNA replication[51], lamina association,[28] and coordinated epigenetic remodelling[27] also suggest that at some point during their evolutionary history, this DNA organization became an essential element for the whole biology of the cell nucleus in animals. Notably, CTCF together with TAD compartmentalization have been secondarily lost in *C. elegans*[34] and closely-related nematodes. Accordingly, this loss resulted in a pervasive dismantling of microsynteny in this organism in comparison with other bilaterians,[46] suggesting a dramatic erosion of long-range CREs located too far away to interact with their former target promoters in the

Figures and Captions Figure 1

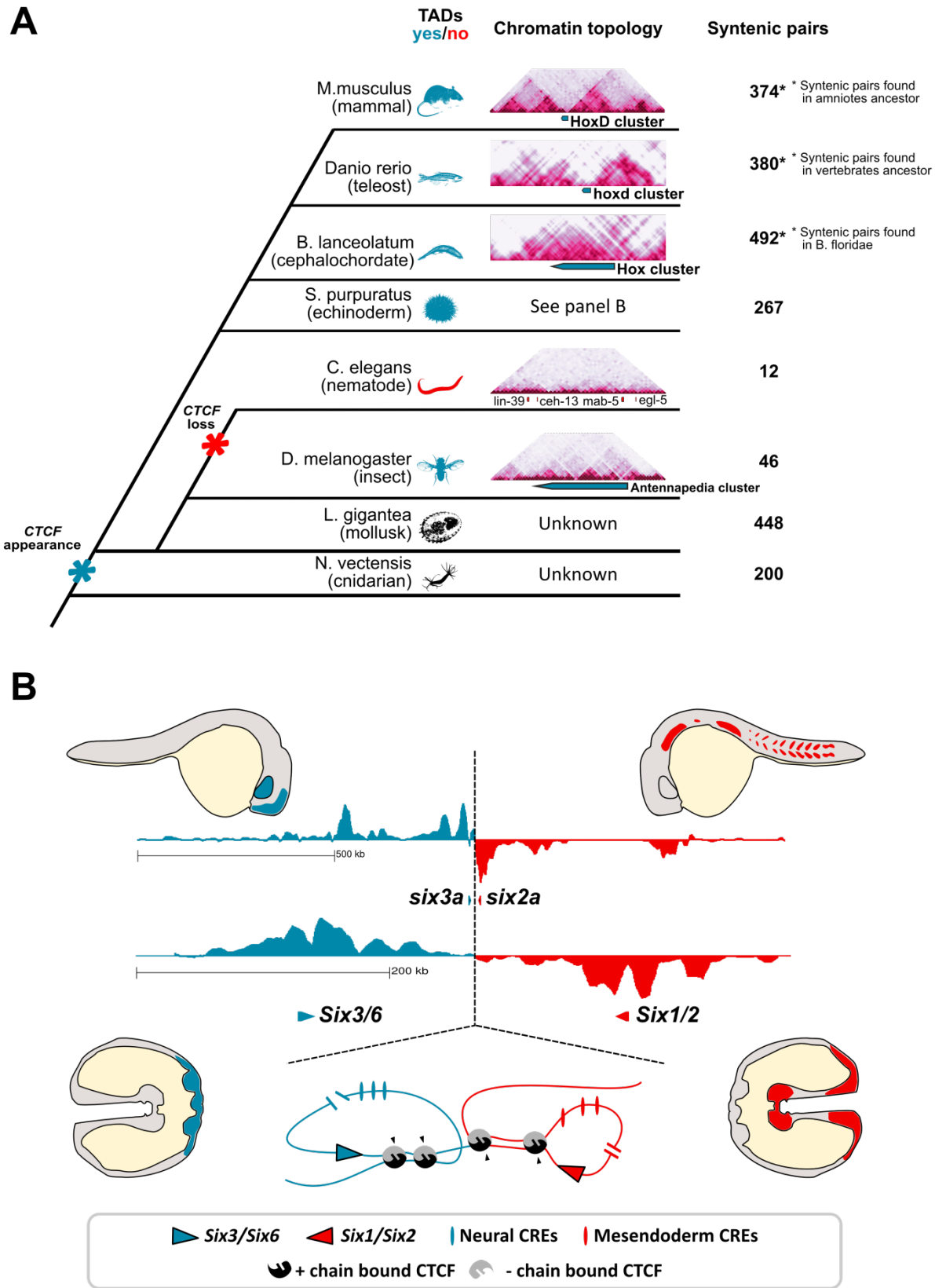


Figure 1: Evidence supporting that TADs are a widespread feature of bilaterian genomes. **(a)** Phylogenetic tree showing current TAD evidence in several representative animal species (branches are not at scale). Animal cartoons are colored in blue when there is strong evidence of TAD chromatin organization, red when the absence of TADs has been experimentally demonstrated and black when there is not enough data to support either scenario. HiC heatmaps of Hox genes regions are also shown when available. Heatmaps were plotted from public HiC matrices of the mouse HoxD cluster[28], the *Drosophila* Antennapedia cluster[20] and the region within the *lin-39* and the *egl-5* homeobox genes in *C. elegans.*[34] For zebrafish HoxDa and amphioxus Hox clusters, virtual HiC heatmaps are shown.[40] Numbers of conserved syntenic pairs were obtained from Irimia et al. 2012.[46] **(b)** Evidence of TADs in sea urchin arises from the study of the Six locus architecture.[30] A conserved topological boundary is placed bisecting Six gene clusters into two regulatory landscapes, both in zebrafish and sea urchin. This is revealed by similar 4Cseq profiles in both organisms. This shared chromatin configuration results in markedly different expression patterns of genes located on each side of the boundary, which are controlled by two different sets of CREs. The same syntax of diverging CTCF binding sites (black arrowheads) operates in both species in order to generate TAD boundaries.

absence of CTCF. This further emphasizes the close association between the presence of CTCF and the capacity of orchestrating long-range *cis*-regulatory interactions.

There is however a problem with this evolutionary scenario. In the genomes of non-bilaterian animals, such as cnidarians, placozoans and sponges, CTCF orthologs have not been identified, making it unlikely that TADs are present in these organisms and that this type of chromatin organization originated before the last common ancestor of bilaterians. And yet, these species have retained many ancestral microsyntenic pairs, way more than *C. elegans*.[46] This suggest that long-range regulation could have evolved prior to the origin of bilaterians, in the last common animal ancestor, but that initially these distant chromatin interactions depended on architectural proteins different from CTCF. Alternatively, CTCF could have been secondarily lost in non-bilaterian lineages, but maintaining long-range CREs through other means (i.e. by evolving novel lineage-specific architectural proteins such as the ones present in *Drosophila*[42, 45]).

THE INFLUENCE OF TADS IN THE EVOLUTION OF GENE EXPRESSION

TADs provide fertile territories for the appearance of new CREs

CREs can appear and be modified during evolution through different processes, ranging from a totally *de novo* CRE appearance by mutations generating functional combinations of TF binding sites, to modification of pre-existing CREs (i.e. “overprinting”), CRE duplication and transposable element insertion followed by its exaptation as a CRE (reviewed in [52, 53]) Whatever the case, it should be taken into account that CREs are not just simple collections of transcription factor binding motifs, but rather complex entities that require the ability to open the chromatin and make themselves visible to to transcription factors and the transcription machinery.[52] Therefore, the evolutionary elaboration of new CRE activities by the above mentioned mechanisms will ultimately depend on many factors.

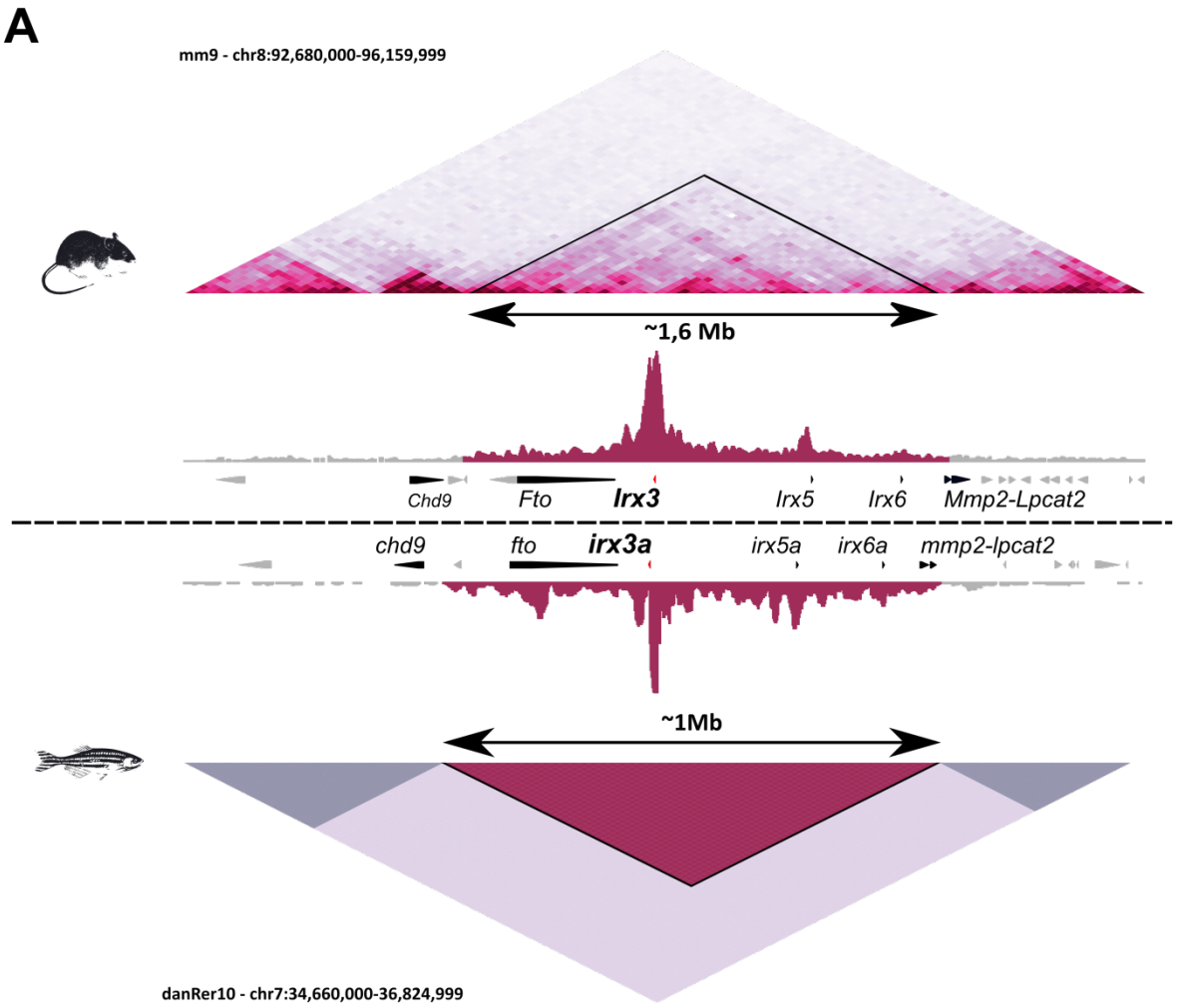
Among them, the genomic and chromatin environment in which a novel CRE could evolve is probably of particular relevance.[52] In this regard, TADs, as the basic structural and organizing units of the genome, could be considered as the self-contained genomic environments in which CRE evolution occurs. Thus, TADs could have facilitated the appearance of new CREs by allowing the interactions of these novel CREs with their potential target genes as well as with other CREs contained within the hundreds of kilobases of the same TAD. In other words, by enabling long-range interactions, TAD compartmentalization may have boosted the rate of incorporation of new CREs into existing regulatory landscapes.[12]

TADs could have also influenced how the amount of regulatory information changes during evolution, and the addition of new regulatory regions by CRE duplications and exaptation of transposable element insertions. In this regard, TADs seem rather tolerant of size changes. For instance, comparisons of 4C-seq profiles from mouse and zebrafish syntenic regions often reveal perfect conservation of the limits of the contacts, despite the large differences in TAD sizes between the two species.[13, 39] Hence, DNA content, and likely also the amount of regulatory information it contains, can be increased or decreased without dismantling the overall 3D architecture (Figure 2). Nevertheless, the appearance of a new CRE in the same TAD together with a potential target may not be by itself sufficient to establish effective contacts that produce gene expression changes. Moreover, we cannot rule out that, within a TAD, some regions are more accessible to promoters than others due to intra-TAD structural constraints. Even so, it seems reasonable to think that TADs have been crucial for the burst of transcriptional regulatory complexity of metazoans, by allowing regulatory information to function even when located far from the genes it is regulating.

Tuning gene regulatory networks occurs mainly at the level of CREs, and the effects of connecting or disconnecting individual genes from these networks are subtle as compared to the effects of mutations in the coding sequences of transcription factors.[1, 3] However, CREs have the potential to activate or repress more than one gene, a potential that in many cases need to be tightly confined to allow only interactions with the right target promoter.[54] As mentioned above, besides favoring productive CRE-promoter interactions, TAD organization also prevents CREs from activating spurious transcription from other nearby promoters beyond the topological border.[24, 25] Therefore, TADs limit the regulatory outcome of newly- evolved CREs to changes in expression of genes that only belong to that TAD. Additionally, not all promoters are able to respond to long-range regulation,[55, 56] narrowing down further the number of genes that could be affected by the evolution of a single CRE. Thus, by restricting their target scope, the appearance of novel CREs in TAD organized genome would have less potentially harmful pleiotropic effects.

TADs could additionally provide an extra layer of control over newly-evolved CREs. TADs can be repressed in a coordinated manner during differentiation through interactions with the nuclear lamina and epigenetic repressive remodeling, resulting in a reduced number of intra-TAD connections.[27] This reversible process is known as switching from an A (active) to a B (inactive) genomic compartment.[18] The number of CREs of a TAD responding to a particular regulatory program in a cell is likely to play an important role in deciding whether or not this switch happens, although this process is still poorly understood. It is conceivable that new CREs that drive expression in similar tissues as the rest of CREs in the TAD have better odds of having a measurable impact on transcription and being subsequently selected and fixed than CREs that are activated by inputs completely different from the TAD regulatory context (see Figure 3a). This could explain the

Figures and Captions Figure 2



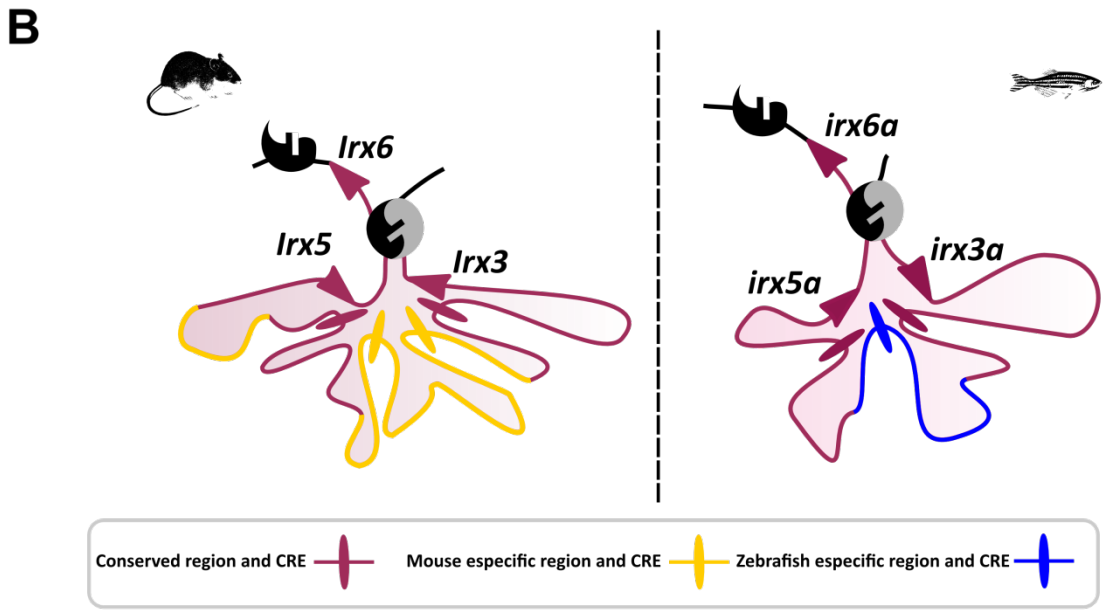


Figure 2: Examples of possible changes in gene expression that could putatively arise by the incorporation of new regulatory DNA without dismantling the architecture of a pre-existing TAD. **(a)** TADs are flexible to changes in size, as observed when comparing chromosome conformation capture data from orthologous syntenic regions in two different organisms, such as the *Irx3* and the *irx3a* loci of mouse and zebrafish respectively. HiC heatmap of the mouse *Irx3* genomic region (at the top) was obtained from public HiC matrices.[28] The 4C-seq profiles of mouse *Irx3* (top) and zebrafish *irx3a* (bottom) loci are from published data.[13] Below, a simplified cartoon of the putative HiC heatmap around the zebrafish *irx3a* locus based on 4C-seq data, is shown. Interestingly, 4C-seq profiles from both species are largely similar and contact limits are placed in equivalent conserved syntenic regions (near the *Chd9/chd9* and *Lpcat2/lpcat2* genes). Limits are compatible with the TAD border prediction in mouse[28] (black triangle in mouse HiC heatmap). However, the *Irx3* regulatory landscape in mouse is 600kb bigger than the one of *irx3a* in zebrafish. In **(b)** there is a simplified model of the two TADs sharing the overall architecture and the syntenic location of the boundaries, but differing in size and in part of the regulatory content.

common observation that redundant and shadow CREs are present in most regulatory landscapes,[57] and it may favor quantitative changes in gene expression[58] and the refinement of pre-existing expression patterns[59] over the gain of new expression domains.

Changes in TAD architecture are rare but have a strong evolutionary potential

The discovery of TADs has posed new possible mechanisms by which gene regulation can evolve; through mutations that cause chromosome reorganizations that fuse, divide, or alter in some way TAD structures. Some mutational mechanisms altering 3D structures include the emergence or erosion of CTCF binding sites and chromosomal rearrangements that encompass topological borders, such as deletions, insertions, inversions, duplications, or translocations (Figure 3b). Among these, CTCF binding site expansion through retrotransposition of SINE elements, a process still operating, is thought to have had paramount importance on the evolution of the genome architecture.[60]

From a theoretical point of view, these events could trigger drastic changes in gene expression, since a single mutational event could cause genes to either gain or lose an entire set of CREs that are already functionally assembled (Figure 3b). Indeed, in humans, this type of genomic rearrangements can cause major phenotypic consequences and are associated with several diseases.[25, 61, 62] Particularly interesting is the example of the human locus containing the *Wnt6*, *Pax3*, *Ihh*, and *EphA4* developmental genes. These four genes are located near the human chromosome 2 but are embedded in different TADs, preventing their regulatory information from acting on the other genes. *EphA4* is expressed in developing limbs under the control of a defined set of CREs confined to the *EphA4* TAD. In a recent study, several chromosomal rearrangements that disrupted either the centromeric or the telomeric *EphA4* TAD borders were shown to cause limb abnormalities, such as Brachydactyly, F-syndrome, and Polydactyly. Reproducing the same genomic rearrangements present the human genome in the mouse genome demonstrated that the *EphA4*-specific CREs

ectopically activate *Pax3*, *Wnt6*, or *Ihh* in the developing limb, causing similar morphological alterations to those observed in humans.[25] These experiments reveal not only that TADs contribute

Figures and Captions Figure 3

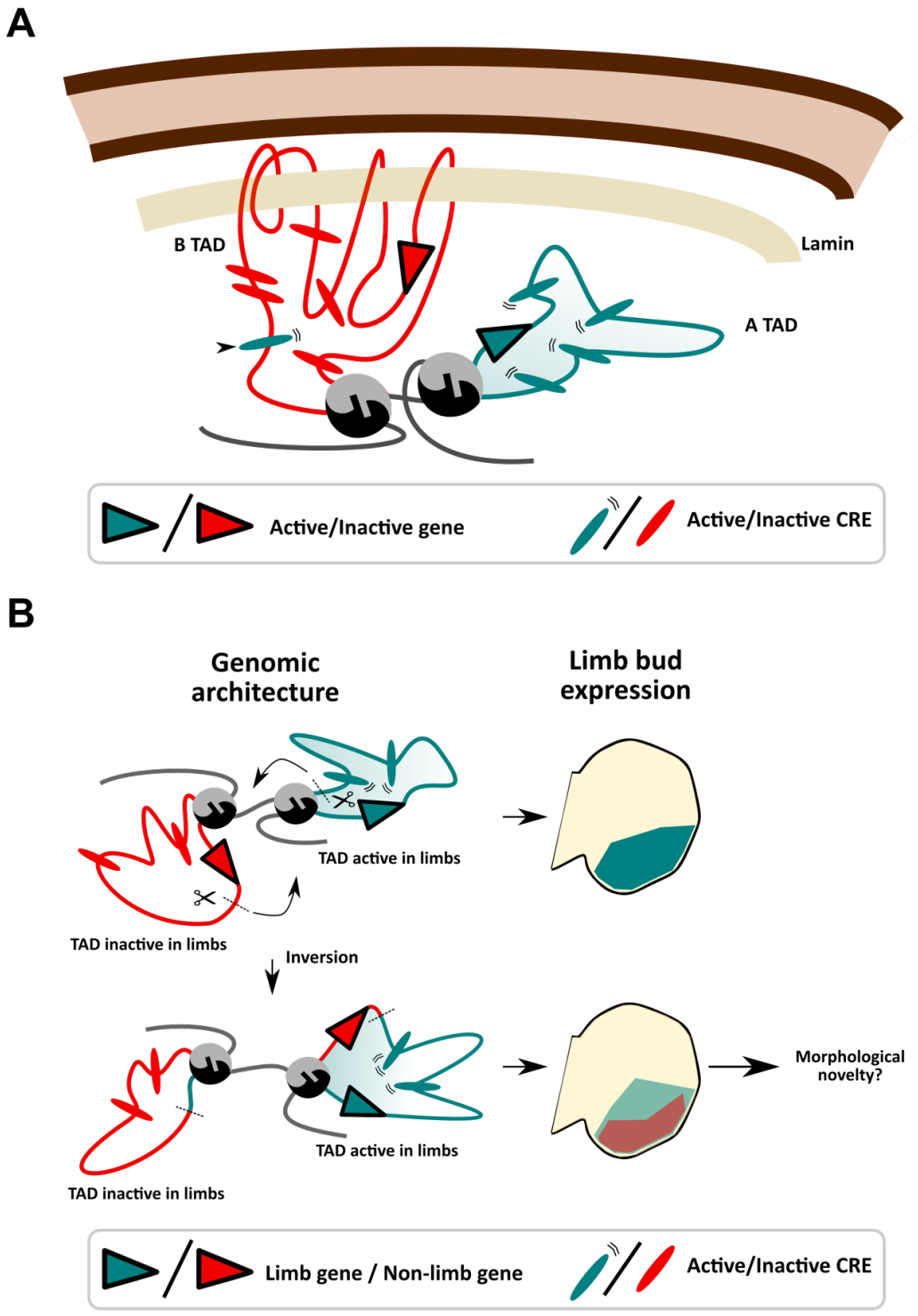


Figure 3: Genes can be exposed to new regulatory information in an unproductive and a productive manner **(a)** The appearance of a new CRE whose regulatory information is incompatible with that already present in the TAD. A new CRE (blue oval in the red TAD pointed by a black arrowhead) promotes expression in a different cell population than the rest of CREs in this TAD. In this cell population the TAD (colored in red) is in a closed B compartment associated to the nuclear lamina, which prevents the activation of the new CRE. **(b)** Modification of a TAD structure by a genomic rearrangement (in this case an inversion). This rearrangement relocates a gene into a new regulatory environment active in a different cell population. In this example, a gene that is not expressed in limbs, after an inversion, integrates in a different TAD with a set of limb CREs that activates it in this territory. This may in turn cause morphological changes, as reported recently.[25]

to preventing non-specific CRE-promoter interactions but also that changes in TAD architecture can indeed promote strong morphological changes due to alterations in gene expression. Another study that highlights the critical effects of removing topological boundaries shows that 26% of the recurrent small mutations in T-cell acute lymphoblastic leukemia overlap with boundary elements. Furthermore, they demonstrate that removing boundaries that insulate proto-oncogenes overexpressed in leukemia cells by Crispr/Cas9 is sufficient to cause their aberrant overexpression.[62] Again in this case the disruption of boundaries turns out in aberrant expression patterns of genes that are affected by CREs that were inaccessible in the presence of the boundary element.

Nonetheless, all these examples are associated to deleterious effects and accordingly the current available data indicate that genetic alterations that disrupt TADs are negatively selected during evolution. For instance, when comparing human and mouse HiC experiments, 75.9% of mouse TAD boundaries are also boundaries in human, and 53.8% of human boundaries are also boundaries in mouse.[28] This proportion is far above that expected by mere chance. Similar trends are observed when HiC results from mouse are compared with those from dog, macaque, or rabbit.[19]

Furthermore, no rearrangements disrupting a TAD structure have occurred since the divergence of mouse and dog. Rather, several examples exist in which breakpoints map precisely to topological borders, which would allow the repositioning of complete TADs to new chromosomal locations without compromising their integrity. In addition to the plausible negative selection of TAD reorganization, it has been also proposed that TAD boundaries are themselves hotspots for genomic rearrangements, due to DNA fiber stress and their sequence composition, which is often rich in repeated elements.[63] However, this theory needs to be empirically tested. Moreover, recently evolved CTCF sites (coming mainly from SINE expansion[60]) have had little impact on TAD topologies.[19] These young CTCFs seem to be instead involved in intra-TAD interactions. This finding is in line with observations that the strength of TAD borders scale up with the number of architectural proteins recruited,[64] suggesting that single CTCF binding site insertions may not be sufficient to generate a topological border. Therefore, to generate a new border, several additional nearby insertions of CTCF binding sites would be required, likely in a diverging orientation.[29, 44] These events are not often detected, at least when comparing TADs across mammalian evolution.

An illustrative example of the strong evolutionary constraints acting on certain TAD borders is found in the Six gene clusters of deuterostomes.[30] These homeobox-encoding genes, which have key developmental roles, are found together in the genome in a gene cluster whose origin predates the

last common ancestor of deuterostomes. In the sea urchin genome, this cluster is formed by three genes: *Six3/6*, *Six1/2*, and *Six4/5*. Despite their proximity, the expression domains of the *Six3/6* gene are largely different from those of the other two genes, which is likely due to the cluster partition in two TADs, as determined by 4C-seq data. This 3D configuration of the locus is conserved in the two clusters present in mammals and the four present in zebrafish, which were all generated after successive rounds of whole genomic duplications. Accordingly, in every cluster, the expression patterns of the Six genes situated on different sides of these conserved TAD borders are markedly different. Indeed, the comparison of these evolutionary conserved boundaries revealed that diverging CTCF binding sites are distinctive signatures of them, and that this code was already established at least in the common ancestor of all deuterostomes (see Figure 1b). Interestingly, Six gene clusters can be also identified in some cnidarian and lophotrocozoan genomes, although whether these clusters share the aforementioned 3D configuration, or whether CTCF contributes to the 3D organization of the cluster, remains unknown. Nevertheless, this poses an intriguing question: Why have these genes remained organized in clusters over 600 myr of evolution if they belong to separate regulatory landscapes located in two different TADs? At this point, we still do not have an answer.

Architectural changes were required for the origin of vertebrate Hox regulation

TAD rearrangements appear to be rare events, at least across relatively short evolutionary distances such as the ones spanning the mammalian radiation.[19] However, a wider evolutionary perspective suggests that in some developmental genes such changes have contributed to the origin of new 3D chromatin structures that have been critical for the origin of morphological novelties. The story of the evolution of Hox regulation in vertebrates is a paradigmatic example.[40]

Hox genes are homeobox transcription factors organized in clusters, with an ancestral function in patterning the main anteroposterior axis, a function that is conserved across bilateria.[65] Due to the two rounds of ancient genome duplications, in most jawed vertebrates there are four Hox clusters; in teleost fishes there was an extra round of whole genomic duplication and therefore the presence of eight clusters is the ancestral condition for this lineage.[66] Some of these clusters have acquired novel functions in patterning vertebrate-specific structures. For instance, HoxA and HoxD clusters are essential for the proper development of paired appendages. Interestingly, the role of HoxA and HoxD clusters in limb development is intimately linked to a special bipartite, 3D chromatin architecture, in which both clusters of genes reside precisely over a TAD boundary. This configuration facilitates genes on each side of the cluster to preferentially read regulatory information from either the anterior or the posterior genomic regions contained in the two adjacent TADs, as reviewed elsewhere.[67] Strikingly, central Hox genes have been shown to switch from one topological domain to another, reading different regulatory information depending on the developmental time and the cell population within the limb bud. This switch, together with the intricate genomic architecture that makes it possible, is essential for the proper patterning of vertebrate limbs.[68]

A recent study has shed some light on the origin of this regulatory mechanism. Comparing the synteny of the four vertebrate paralogues Hox clusters revealed that the same neighboring genes, both in the anterior and the posterior TADs, were linked to the Hox genes from the base of the

Figures and Captions Figure 4

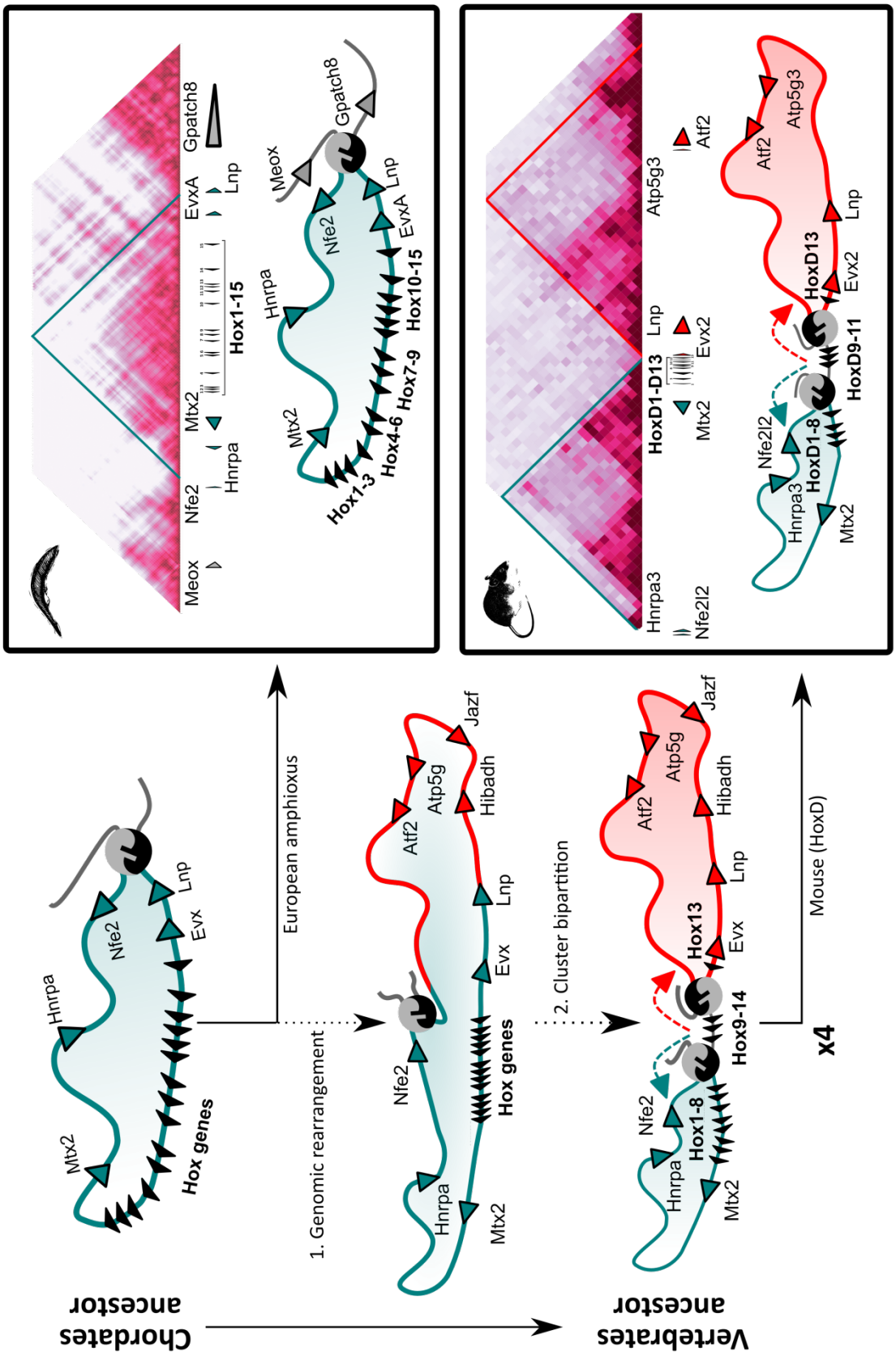


Figure 4: The elaboration of the bipartite regulation of the HoxD cluster in vertebrates required several modifications of TAD architectures between the last common ancestor of chordates and the last common ancestor of vertebrates. In vertebrates, the regulatory information governing HoxD genes expression is distributed in two adjacent TADs (see HiC heatmap for the mouse cluster at the bottom right square). HoxD genes are located precisely over the TAD boundary separating these two adjacent TADs. Genes located at the 3' end of the cluster contact preferentially CREs from the 3' TAD (in blue). Genes at the 5' end of the cluster preferentially interact with CREs from the 5' TAD (red). In the cephalochordate amphioxus, all Hox genes are embedded within the same TAD. This TAD also includes the neighboring 3' genomic region which shows conserved synteny with vertebrates (top right square). On the left, a hypothetical evolutionary scenario showing the changes in Hox clusters topologies that may have occurred after the divergence of vertebrates and amphioxus from their chordate ancestor. The scenario includes a genomic rearrangement that brought the vertebrate specific 5' TAD region to the proximity of the cluster and the partitioning of the Hox regulatory landscape into two adjacent TADs. Additional events include the formation of gene deserts in the HoxD cluster after the whole genome duplications in vertebrates and a tandem duplication of the Evx gene in the amphioxus lineage.

vertebrate lineage, before the two rounds of whole genomic duplications. Since synteny conservation is often an indirect consequence of long-range CREs of developmental genes that have colonized the introns of nearby bystander genes, it is likely that the regulatory information controlling Hox genes expression in the vertebrate ancestor was already distributed in both the 5′- and in the 3′-neighboring genomic territories flanking the cluster. If so, it is then reasonable to assume that the bipartite Hox 3D architecture was also already present in this ancestor. In contrast to vertebrates, in the genomic region around the single Hox cluster of amphioxus, a slow-evolving invertebrate chordate, gene synteny is shared with vertebrates on the 3´ side only, while in the 5′-flanking region, synteny conservation stops beyond the *Lnp* gene. This finding already suggested that Hox long-range CREs are present in the 3′-flanking region but not in the 5′-flanking one. Accordingly, using 4C-seq data to computationally reconstruct the 3D architecture of the zebrafish and amphioxus Hox clusters revealed that, in contrast to the bimodal 3D structure observed in vertebrates with two TADs bisecting the Hox cluster, the amphioxus cluster is embedded in a single TAD that includes all the Hox genes and only the 3′-flanking syntenic region. This 3D chromatin structure has functional consequences for the interaction of Hox genes with long-range CREs, as enhancer assays of amphioxus open-chromatin regions detected by ATAC-seq[6] demonstrated that this amphioxus syntenic 3′-flanking region, but not the 5′-flanking genomic territory, harbors distal Hox CREs.

Several events that modified the Hox regulatory landscapes must have happened between the last common ancestor of chordates and the last common ancestor of vertebrates (see Figure 4). One was a chromosomal rearrangement that brought a new genomic region to the 5′-end of the Hox cluster. This region became the substrate for the appearance of new Hox long-range cis-regulatory interactions, which provided these genes with new regulatory information. Finally, the cluster was bisected into two TADs, separating the anterior and posterior regulatory landscapes and introducing the possibility of a switch between two separate sets of long-range regulatory inputs, a bimodal system required for the novel Hox functions in vertebrate development. However, because of the

lack of additional information from organisms that would have diverged after the separation between amphioxus and vertebrates but before the emergence of the vertebrate lineage, the nature and sequential order of such mutational events are unknown. It is possible that further studies in lamprey may shed some light on this issue. In any case, we can conclude that chromosomal reorganization events occurred during evolution that strongly modified the 3D compartmentalization of Hox regulatory landscapes, facilitating gene expression changes linked to morphological novelties.

STRUCTURAL INFORMATION IN COMPARATIVE GENOMICS: CHALLENGES AND OPPORTUNITIES

The accessibility to the next-generation sequencing (NGS) techniques is facilitating the transition from studies focused on the evolution of particular genes or gene families, to unbiased comparative genome-wide analyses. In this regard, comparative transcriptomic profiles in multiple vertebrate animals have shown that the transcriptional programs during development reach the highest similarity between different species in a coordinate manner at the phylotypic period, when the body plan is established and at the maximum point of morphological similarity between vertebrates.[31, 69] The same high expression similarity is observed between *Drosophila* species at the arthropod´s phylotypic period.[70] These studies reinforce the hypothesis that, once a body plan is established at the phylotypic period, morphological differences observed in adult animals are due to gene expression changes.

After establishing a clear link between morphological differences and transcriptomic changes, the next challenging step will be to determine the CRE composition underlying these expression differences. With the development of the new genomic techniques also based on NGS, such as ChIP-seq, DNase-seq, and specially ATAC-seq, CREs can be mined genome-wide in almost any sequenced organism in many different contexts. However, the information provided by these experiments does not facilitate linking the identified CREs to their target promoters in a reliable manner. Although the strategy of associating them to the closest gene is useful to a certain extent for extracting general trends from genome-wide data, it has been proven to be a very gross and inaccurate approximation, especially when dealing with developmental gene enhancers that can reside up to a megabase away from their target gene.[11–13] Adding structural data from C-based techniques (especially HiC and 4C-seq) can provide critical information and will allow a definition and comparison of the complete set of CREs that govern the expression of orthologous genes in different organisms (see Figure 5a). Filtering promoters that are likely to be insensitive to long-range regulation (such as those of bystander genes) would further help to refine the potential target genes of critical CREs.[55, 56, 71] However, performing this in a genome-wide manner is still challenging, although for instance differences in CAGE-profiles could be used as signatures for promoter clustering.[55] In addition to an important improvement in CRE-promoter assignment strategy, performing 3C-based techniques in comparative genomics projects also has the advantage of easily identifying even uncommon architectural changes, such as those observed in the Hox cluster in the transition from chordates to vertebrates.[40] Such changes can have a huge regulatory and morphological impact, as discussed above.

Figures and Captions Figure 5

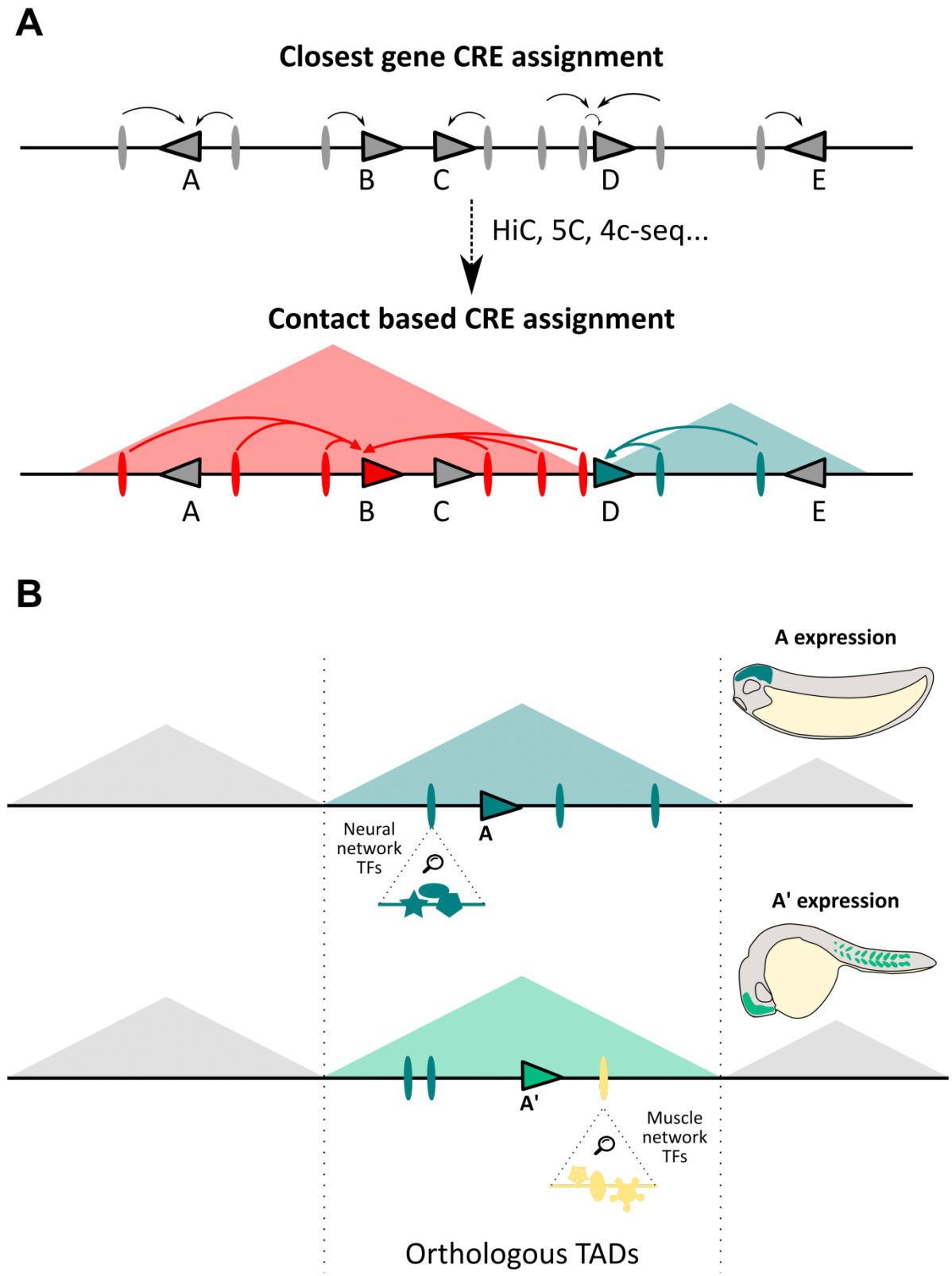


Figure 5: Benefits of including C-techniques data in comparative genomics projects. **(a)** CRE-promoter assignment based on contacts is much more reliable than association by proximity, especially for developmental genes with extended regulatory landscapes. Filtering out promoters that are likely insensitive to long-range regulation (colored in grey in the figure) may help to refine further this assignment process. **(b)** Chromatin structural information complements epigenomic data used to identify CRE elements. Among the different epigenomic data, chromatin accesibility based techniques (DNAse-seq, ATAC-seq), are able to reach base-pair resolution and reveal transcription factor footprinting. In the simplified diagram, the comparison of the CRE transcription factor binding site composition of two orthologous TADs permits to unveil the genetic determinants that allow the gene A' to gain a new expression domain in muscle.

From a technical point of view, challenges and limitations of 3C-based techniques are similar to those of other genomic approaches, such as ChIP-seq or RNA-seq. One limitation is that these techniques require an important amount of starting material. For many evolutionary relevant organisms that are considered non-classical models that are not easily cultured in the laboratory, obtaining the required amount of material can be challenging. Further, whole embryo approaches are often the only alternative to get enough material when developmental tissues are of interest, especially if the access to these embryos is limited. However, for CRE–promoter assignment purposes, TADs have been shown to be largely invariant across developmental stages and cell types (see also Box 1).[22, 28] Thus, in principle, a single HiC experiment, even if performed in a heterogenous sample, may be sufficient to unveil the most frequent and stable TAD configurations, which would be likely valid for many different cellular contexts. The poor quality of genome assemblies of many organisms, which frequently have N50 values below or around the mean sizes of a single TAD, also make it difficult to study of genome architecture and to further analyze long-range regulation. However, several methods have been recently developed to take advantage of contact data from HiC experiments, which greatly improve assembly quality, reaching even chromosome contiguity.[72–74] Thus, in these cases, HiC experiments have multiple uses: they help to generate better genome assemblies, allow the characterization of the genome-wide chromatin organization in TADs, and can be further used to link CREs and their target genes.

After linking CREs to their target genes, the next challenge is to determine how variations of CREs associated to particular genes along evolution may contribute to morphological variation. There are some examples of comparative epigenomic approaches that have already provided insights about the contribution of CREs to the evolution of human limbs[75] and facial structures[76], and in the developmental heterochronies observed between zebrafish and medaka.[77] Unfortunately, however, discerning whether a CRE is functionally conserved or species-specific is frequently a very difficult task when comparing two species. Sequence conservation has been often used as a criteria to find functionally conserved CREs. Although conserved enhancers usually drive expression to equivalent domains, there are also cases in which conservation at the sequence level does not result in an equivalent activity.[78] Moreover, most CREs are not conserved at the sequence level because sequence conservation is either low or absent except at the binding sites of the upstream regulatory transcription factors,[79] or because the rates of CRE turnover are high[53] (or a combination of both). This lack of conservation becomes a rule when comparing organisms from different lineages. Despite a lack of sequence conservation, global studies in mammals have shown that the gene networks that

operate on CREs of orthologous genes are highly conserved.[80] This is likely also the case across much more divergent species, since some cases of CRE functional equivalence in the absence of sequence conservation have been reported between mammals and teleosts[81, 82] and even between mammals and hemichordates.[83] If CREs exist that are conserved at the functional but not sequence level in different species, how can we find them? At the moment, syntenic location within the regulatory landscape of orthologous genes[46, 48, 84], bioinformatic analysis of the networks operating on CREs[80, 83, 85], and comparative functional analysis using transgenic assays[86, 87] are the best—and only—tools available to address this difficult question. We thus believe that comparative genomic projects would greatly benefit from including at least one HiC dataset, which could then be combined synergistically with other dynamic or tissue-specific epigenomic profiling.[6, 7]

From the different epigenomic approaches to detect CREs, chromatin accessibility assays such as DNase-seq[88] and ATAC-seq[6] are able to reach base-pair resolution, thereby revealing transcription factor footprinting that provides extremely useful information about networks operating upstream of the different CREs.[80] These data are particularly useful for evolutionary purposes, since the binding site compositions and upstream networks operating on inferred CREs can be used to identify equivalent CREs beyond sequence conservation (see Figure 5b).[83]

Conclusion Chromosome folding into TADs seems to be a shared feature of the majority of bilaterian animals. TADs have influenced the story of gene regulation since their first appearance, presumably in the last common bilaterian or animal ancestor. Moreover, it is very likely that this 3D organization of the DNA has been coupled to (or reinforced by) the appearance of the CTCF protein. The chromatin folding favored by TADs has a critical function: to accurately connect distant CREs to their target promoters. In the context of a TAD-compartmentalized genome, there are two main possible ways by which transcriptional regulation can be modified. First, the composition of CREs present in a given TAD can be changed without modifying the overall 3D chromatin structure of the genomic territory. TADs tolerate this kind of modification by being flexible to changes in sizes, for example due to the insertion or removal of a particular non-coding DNA region, which can have a regulatory potential. As discussed above, such events probably do not dismantle the overall architecture of the region, as long as the TAD boundaries are maintained. TAD organization is also critical for restricting the effect of newly-evolved CREs to a limited number of genes—namely, those located within the same TAD. Second, transcriptional regulation can be changed by altering TAD architecture. This may result in connecting sets of already functional CREs to new target promoters, or the opposite, in disconnecting one or several regulatory elements from the gene they were regulating. The evolutionary potential of such TAD reorganizations is huge, since a single event could cause a promoter to gain or lose multiple CREs, thereby driving the gene expression towards a completely new cellular context. Indeed, such changes have been shown to be the molecular basis of several human diseases, including some causing morphological alterations. Perhaps not surprisingly, such drastic architectural alterations seem to be negatively selected, and TAD boundaries are largely conserved in evolution, at least within mammals. Nevertheless, analyzing particular examples, such as the evolution of vertebrate Hox clusters, reveals that such rearrangements are occasionally fixed and may have a major impact in development and evolution at deeper timescales. We therefore believe that chromatin conformation data (preferentially HiC and 4C-seq) can strongly synergize with multiple comparative genomic projects by providing an accurate way to assign CREs to the genes

that they regulate and by helping to identify the regulatory basis of the evolution of animal morphology.

References

1. Carroll SB: Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell* 2008, 134:25–36.

2. Wilson MD, Barbosa-Morais NL, Schmidt D, Conboy CM, Vanes L, Tybulewicz VLJ, Fisher EMC, Tavare S, Odom DT: Species-Specific Transcription in Mice Carrying Human Chromosome 21. *Science* 2008, 322:434–438.

3. Wray GA: The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 2007, 8:206–16.

4. Visel A, Rubin EM, Pennacchio LA: Genomic views of distant-acting enhancers. *Nature* 2009, 461:199–205.

5. Thurman R, Rynes E, Humbert R, Vierstra J, Maurano M, Haugen E, Sheffield N, Stergachis A, Wang H, Vernot B, et al.: The accessible chromatin landscape of the human genome. *Nature* 2012, 489:75–82.

6. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ: Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 2013, 10:1213–8.

7. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann S, Flynn R, Wysocka J: A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 2011, 470:279–83.

8. Bogdanovic O, Fernandez-Miñán A, Tena JJ, De La Calle-Mustienes E, Hidalgo C, Van Kruysbergen I, Van Heeringen SJ, Veenstra GJC, Gómez-Skarmeta JL: Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res* 2012, 22:2043–2053.

9. Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, Phouanenavong S, Plajzer-Frick I, Shoukry M, Afzal V, et al.: Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* 2013, 155:1521–1531.

10. Schwaiger M, Schonauer A, Rendeiro AF, Pribitzer C, Schauer A, Gilles AF, Schinko JB, Renfer E, Fredman D, Technau U: Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res* 2014, 24:639–650.

11. Lettice LA, Heaney SJH, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E: A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* 2003, 12:1725–1735.

12. Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, De Laat W, Spitz F, Duboule D: A regulatory archipelago controls hox genes transcription in digits. *Cell* 2011, 147:1132–1145.

13. Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF, et al.: Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 2014, 507:371–5.

14. Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, Dean A, Blobel G: Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 2012, 149:1233–1244.

15. Dekker J, Rippe K, Dekker M, Kleckner N: Capturing chromosome conformation. *Science* 2002, 295:1306–11.

16. Splinter E, de Wit E, Nora EP, Klous P, van de Werken HJG, Zhu Y, Kaaij LJT, van Ijcken W, Gribnau J, Heard E, et al.: The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev* 2011, 25:1371–1383.

17. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, et al.: Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Res* 2006, 16:1299–1309.

18. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al.: Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009, 326:289–93.

19. Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S: Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Rep* 2015, 10:1297–1309.

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

21. Hou C, Li L, Qin ZS, Corces VG: Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. *Mol Cell* 2012, 48:471–484.

22. Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EEM: Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* 2014, 512:96–100.

23. Dekker J, Heard E: Structural and functional diversity of Topologically Associating Domains. *FEBS Lett* 2015, 589:2877–2884.

24. Nora EP, Dekker J, Heard E: Segmental folding of chromosomes: A basis for structural and regulatory chromosomal neighborhoods? *BioEssays* 2013, 35:818–828.

25. Lupiañez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al.: Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 2015, 161:1012–1025.

26. Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell JSK, Ong CT, Hookway TA, Guo C, Sun Y, et al.: Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 2013, 153:1281–1295.

27. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, et al.: Chromatin architecture reorganization during stem cell differentiation. *Nature* 2015, 518:331–336.

28. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B: Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012, 485:376–380.

29. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al.: A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014, 159:1665–1680.

30. Gomez-Marin C, Tena JJ, Acemel RD, Lopez-Mayorga M, Naranjo S, de la Calle-Mustienes E, Maeso I, Beccari L, Aneas I, Vielmas E, et al.: Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc Natl Acad Sci U S A* 2015, 112:7542–7547.

31. Irie N, Kuratani S: Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat Commun* 2011, 2:248.

32. Irimia M, Piñeiro C, Maeso I, Gómez-Skarmeta JL, Casares F, Garcia-Fernàndez J: Conserved developmental expression of Fezf in chordates and Drosophila and the origin of the Zona Limitans Intrathalamica (ZLI) brain organizer. *Evodevo* 2010, 1:7.

33. Vocking O, Kourtesis I, Hausen H: Posterior eyespots in larval chitons have a molecular identity similar to anterior cerebral eyes in other bilaterians. *Evodevo* 2015, 6:40.

34. Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, Uzawa S, Dekker J, Meyer BJ: Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* 2015, 523:240–244.

35. Feng S, Cokus S, Schubert V, Zhai J, Pellegrini M, Jacobsen S: Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. *Mol Cell* 2014, 55:694–707.

36. Duan Z, Andronescu M, Schutz K, McIlwain S, Mcllwain S, Kim YJ, Lee C, Shendure J, Fields S, Blau CA, et al.: A Three-Dimensional Model of the Yeast Genome. *Nature* 2010, 465:363–367.

37. Mizuguchi T, Fudenberg G, Mehta S, Belton J-M, Taneja N, Folco HD, FitzGerald P, Dekker J, Mirny L, Barrowman J, et al.: Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. *Nature* 2014, 516:432–5.

38. Le TBK, Imakaev MV, Mirny L, Laub MT: High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* 2013, 342:731–4.

39. Woltering JM, Noordermeer D, Leleu M, Duboule D: Conservation and divergence of regulatory strategies at Hox Loci and the origin of tetrapod digits. *PLoS Biol* 2014, 12:e1001773.

40. Acemel RD, Tena JJ, Irastorza-Azcarate I, Marletaz F, Gomez-Marin C, de la Calle-Mustienes E, Bertrand S, Diaz SG, Aldea D, Aury J-M, et al.: A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. *Nat Genet* 2016, 48:336–341.

41. Vietri Rudan M, Hadjur S: Genetic Tailors: CTCF and Cohesin Shape the Genome During Evolution. *Trends Genet* 2015, 31:651–660.

42. Cubeñas-Potts C, Corces VG: Architectural proteins, transcription, and the three-dimensional organization of the genome. *FEBS Lett* 2015, 589:2923–2930.

43. Heger P, Marin B, Bartkuhn M, Schierenberg E, Wiehe T: The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc Natl Acad Sci U S A* 2012, 109:17507–12.

44. de Wit E, Vos ESM, Holwerda SJB, Valdes-Quezada C, Verstegen MJAM, Teunissen H, Splinter E, Wijchers PJ, Krijger PHL, de Laat W: CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* 2015, 60:676–684.

45. Xu C, Corces VG: Towards a predictive model of chromatin 3D organization. *Semin Cell Dev Biol* 2016, 57:24–30.

46. Irimia M, Tena JJ, Alexis MS, Fernandez-Miñan A, Maeso I, Bogdanović O, De La Calle-Mustienes E, Roy SW, Gómez-Skarmeta JL, Fraser HB: Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Res* 2012, 22:2356–2367.

47. Simakov O, Marletaz F, Cho S-J, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo D-H, Larsson T, Lv J, Arendt D, et al.: Insights into bilaterian evolution from three spiralian genomes. *Nature* 2013, 493:526–531.

48. Kikuta H, Laplante M, Navratilova P, Komisarczuk AZ, Engström PG, Fredman D, Akalin A, Caccamo M, Sealy I, Howe K, et al.: Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res* 2007, 17:545–555.

49. Engström PG, Sui SJH, Drivenes Øyvind, Becker TS, Lenhard B: Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res* 2007, 17:1898–1908.

50. Irimia M, Maeso I, Roy SW, Fraser HB: Ancient cis-regulatory constraints and the evolution of genome architecture. *Trends Genet* 2013, 29:521–528.

51. Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, Vera DL, Wang Y, Hansen RS, Canfield TK, et al.: Topologically associating domains are stable units of replication-timing regulation. *Nature* 2014, 515:402–405.

52. Maeso I, Tena JJ: Favorable genomic environments for cis-regulatory evolution: A novel theoretical framework. *Semin Cell Dev Biol* 2015.

53. Villar D, Flicek P, Odom DT: Evolution of transcription factor binding in metazoans- mechanisms and functional implications. *Nat Rev Genet* 2014, 15:221–233.

54. Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, Weintraub AS, Schuijers J, Lee TI, Zhao K, et al.: Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. *Cell* 2014, 159:374–387.

55. Akalin A, Fredman D, Arner E, Dong X, Bryne JC, Suzuki H, Daub CO, Hayashizaki Y, Lenhard B: Transcriptional features of genomic regulatory blocks. *Genome Biol* 2009, 10:R38.

56. Zabidi M a., Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, Stark A: Enhancer–‐core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* 2014, 518:556–559.

57. Cannavò E, Khoueiry P, Garfield DA, Geeleher P, Zichner T, Gustafson EH, Ciglar L, Korbel JO, Furlong EEM: Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr Biol* 2016, 26:38–51.

58. Bothma JP, Garcia HG, Ng S, Perry MW, Gregor T, Levine M: Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. *Elife* 2015, 4:1–14.

59. El-Sherif E, Levine M: Shadow Enhancers Mediate Dynamic Shifts of Gap Gene Expression in the Drosophila Embryo. *Curr Biol* 2016, 26:1164–1169.

60. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Gonalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT: Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 2012, 148:335–348.

61. Valton AL, Dekker J: TAD disruption as oncogenic driver. *Curr Opin Genet Dev* 2016, 36:34–40.

62. Hnisz D, Weintraub AS, Day DS, Valton A, Bak RO, Li CH, Goldmann J, Lajoie BR, Fan ZP, Sigova AA, et al.: Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* 2016, 351:1454–1458.

63. Farré M, Robinson TJ, Ruiz-Herrera A: An Integrative Breakage Model of genome architecture, reshuffling and evolution: The Integrative Breakage Model of genome evolution, a novel multidisciplinary hypothesis for the study of genome plasticity. *BioEssays* 2015, 37:479–488.

64. Van Bortle K, Nichols MH, Li L, Ong C-T, Takenaka N, Qin ZS, Corces VG: Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol* 2014, 15:R82.

65. Duboule D: The rise and fall of Hox gene clusters. *Development* 2007, 134:2549–60.

66. Kuraku S, Meyer A: The evolution and maintenance of Hox gene clusters in vertebrates and the teleost-specific genome duplication. *Int J Dev Biol* 2009, 53:765–773.

67. Lonfat N, Duboule D: Structure, function and evolution of topologically associating domains (TADs) at HOX loci. *FEBS Lett* 2015, 589(20 Pt A):2869–2876.

68. Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, Trono D, Spitz F, Duboule D: A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* 2013, 340:1234167.

69. Domazet-Lošo T, Tautz D: A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* 2010, 468:815–818.

70. Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, Ohler U, Bergman CM, Tomancak P: Gene expression divergence recapitulates the developmental hourglass model. *Nature* 2010, 468:811–814.

71. Tsujimura T, Klein FA, Langenfeld K, Glaser J, Huber W, Spitz F: A discrete transition zone organizes the topological and regulatory autonomy of the adjacent tfap2c and bmp7 genes. *PLoS Genet* 2015, 11:e1004897.

72. Kaplan N, Dekker J: High-throughput genome scaffolding from in vivo DNA interaction frequency. *Nat Biotechnol* 2013, 31:1143–7.

73. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J: Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol* 2013, 31:1119–1125.

74. Marie-Nelly H, Marbouty M, Cournac A, Flot J-F, Liti G, Parodi DP, Syan S, Guillen N, Margeot A, Zimmer C, et al.: High-quality genome (re)assembly using chromosomal contact data. *Nat Commun* 2014, 5:5695.

75. Cotney J, Leng J, Yin J, Reilly SK, Demare LE, Emera D, Ayoub AE, Rakic P, Noonan JP: XThe evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell* 2013, 154:185–196.

76. Prescott SL, Srinivasan R, Marchetto MC, Grishina I, Narvaiza I, Selleri L, Gage FH, Swigut T, Wysocka J: Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimp Neural Crest. *Cell* 2015, 163:68–84.

77. Tena JJ, González-Aguilera C, Fernández-Miñán A, Vázquez-Marín J, Parra-Acero H, Cross JW, Rigby PWJ, Carvajal JJ, Wittbrodt J, Gómez-Skarmeta JL, et al.: Comparative epigenomics in distantly related teleost species identifies conserved cis-regulatory nodes active during the vertebrate phylotypic period. *Genome Res* 2014, 24:1075–1085.

78. Ariza-Cosano A, Visel A, Pennacchio LA, Fraser HB, Gomez-Skarmeta JL, Irimia M, Bessa J: Differences in enhancer activity in mouse and zebrafish reporter assays are often associated with changes in gene expression. *BMC Genomics* 2012, 13:713.

79. McGaughey DM, Vinton RM, Huynh J, Al-Saif A, Beer MA, McCallion AS: Metrics of sequence constraint overlook regulatory sequences in an exhaustive analysis at phox2b. *Genome Res* 2008, 18:252–260.

80. Stergachis AB, Neph S, Sandstrom R, Haugen E, Reynolds AP, Zhang M, Byron R, Canfield T, Stelhing-Sun S, Lee K, et al.: Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* 2014, 515:365–370.

81. Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS: Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science* 2006, 312:276–279.

82. Gehrke AR, Schneider I, de la Calle-Mustienes E, Tena JJ, Gomez-Marin C, Chandran M, Nakamura T, Braasch I, Postlethwait JH, Gomez-Skarmeta JL, et al.: Deep conservation of wrist and digit enhancers in fish. *Proc Natl Acad Sci U S A* 2015, 112:803–808.

83. Yao Y, Minor PJ, Zhao Y-T, Jeong Y, Pani AM, King AN, Symmons O, Gan L, Cardoso W V, Spitz F, et al.: Cis-regulatory architecture of a brain signaling center predates the origin of chordates. *Nat Genet* 2016, 48:575–580.

84. Maeso I, Irimia M, Tena JJ, González-Pérez E, Tran D, Ravi V, Venkatesh B, Campuzano S, Gómez-Skarmeta JL, Garcia-Fernández J: An ancient genomic regulatory block conserved across bilaterians and its dismantling in tetrapods by retrogene replacement. *Genome Res* 2012, 22:642–655.

85. Taher L, Mcgaughey DM, Maragh S, Taher L, Mcgaughey DM, Maragh S, Aneas I, Bessling SL, Miller W, Nobrega MA, et al.: Genome-wide identification of conserved regulatory function in diverged sequences Genome-wide identification of conserved regulatory function in diverged sequences. 2011:1139–1149.

86. De La Calle-Mustienes E, Feijoo CG, Manzanares M, Tena JJ, Rodríguez-Seguel E, Letizia A, Allende ML, Gómez-Skarmeta JL: A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. *Genome Res* 2005, 15:1061–1072.

87. Pennacchio L a, Ahituv N, Moses AM, Prabhakar S, Nobrega M a, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, et al.: In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 2006, 444:499–502.

88. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE: High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell* 2008, 132:311–322.

89. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al.: Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 2012, 485:381–385.

90. de Wit E, de Laat W: A decade of 3C technologies: Insights into nuclear organization. *Genes Dev* 2012, 26:11–24.

91. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P: Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 2013, 502:59–64.