CellPress
REVIEWS

Spotlight

Gene Regulation Knows Its Boundaries

Elzo de Wit[1,*]

The genome is folded nonrandomly inside the cell nucleus. How this contributes to gene regulation is an important subject of investigation. A new study by Despang *et al.* shows how the spatial segregation of genes and regulatory regions can influence developmental expression in subtle, but critical ways.

When fully stretched out, the human genome measures ~2 m in length, but has to fit in a nucleus 5–10 μm in diameter. To achieve this and maintain the required physiological functions of the genome, such as gene expression and replication, it is spatially organized in a nonrandom manner. Chromosome conformation capture methods, such as Hi-C, revealed that the genome can be subdivided into regions that show preferential self-interaction, referred to as topologically associating domains (TADs) [1,2]. Interspecies comparison has shown that the position of TADs is strongly conserved among species [3], which suggests that TADs have an important role in gene regulation. Unexpectedly, however, systemic loss of proteins important for the maintenance of these features, such as the insulator protein CTCF and the ring-shaped cohesin complex, only mildly affect gene expression in cell line systems [4,5]. How can we reconcile these seemingly contradictory observations?

Part of the answer may lie in the fact that certain crucial regulators have a specific spatiotemporal expression pattern. Expression of these factors is restricted to a limited number of cells in the embryo, where slight misexpression can cause severe developmental defects. A new study by Despang *et al.* [6] sheds light on the regulation of two of these factors, the transcription factor SOX9 and the potassium channel KCNJ2. The *Sox9* and *Kcnj2* genes are neighbors in the genome separated by a strong TAD boundary (Figure 1A). *Sox9* is flanked by a regulatory domain that is ~1 Mb in size. Loss of function of *Sox9* can lead to cleft palate, skeleton abnormalities, and sex reversal, whereas misexpression of *Kcnj2* in a *Sox9*-like pattern results in loss of claws or nails and malformed terminal phalanges, also called Cooks syndrome in humans. In their study, the authors aimed to understand the interplay between the 3D genome and gene expression of these two genes and the phenotypic consequences. To do so, they performed a series of genome-editing experiments in mice in combination with chromosome capture experiments.

The authors first deleted the TAD boundary between *Sox9* and *Kcnj2* (Figure 1B). Deletion of this boundary weakened the segregation of the two genomic loci (also known as insulation), but did not fully disrupt it. Only when the TAD boundary was deleted in conjunction with mutations in the surrounding CTCF-binding sites, did the two TADs fuse into one (Figure 1C). In the boundary-only deletion with functioning TAD separation, the expression pattern of the two genes was hardly affected. However, when the TADs fused, *Kcnj2* partially adopted the expression profile of *Sox9*, arguably because *Kcnj2* was now exposed to the regulatory environment of *Sox9*. These results are consistent with previous observations that CTCF sites and TAD boundaries act as insulator sequences that protect genes from the regulatory regions of their neighbors [7]. Interestingly, the internal TAD structure appeared to be of little importance for the expression of *Sox9* in the developing limb.

To determine whether the regulatory region is important for the expression of *Sox9*, an inversion of the regulatory region was made, including the boundary (referred to as *InvC*, Figure 1D). Chromosome capture experiments showed that this had the effect of placing the *Sox9* regulatory region in a TAD together with *Kcnj2*, but excluding *Sox9* itself from this TAD. This resulted in a strong gain of *Kcnj2* expression in a *Sox9*-like pattern and a loss of claws (Cooks phenotype) and a *Sox9* loss-of-function phenotype in the mice. *Sox9* expression was likely diminished because the *Sox9* promoter was now shielded from its own regulatory environment by the TAD boundary that was inverted along with the regulatory domain. To test this, another inversion was made, which inverted the regulatory region, excluding the TAD boundary. These embryos showed wild-type expression of *Kcnj2* and *Sox9* and were phenotypically wild-type, showing that the orientation and order of the enhancers in the regulatory domain is not important for proper gene expression. Rather, it is the TAD boundary that can direct regulatory important interactions.

Finally, the TAD boundary was deleted from the *InvC* inversion (Figure 1D), which shields *Sox9* from its own regulatory sites (Figure 1E), resulting in the formation of a single TAD encompassing both genes. *Kcnj2* expression was diminished compared with *InvC*, but was still increased over wild-type and the mice still showed a *Kcnj2* gain-of-function phenotype. Conversely, *Sox9* expression was increased compared with *InvC*, but still lower than in wild-type. There were no obvious *Sox9* loss-of-function phenotypes, except for male infertility in homozygous mice. These results suggest that, by placing the *Sox9* regulatory elements together with *Kcnj2* in a single TAD, there will be competition between

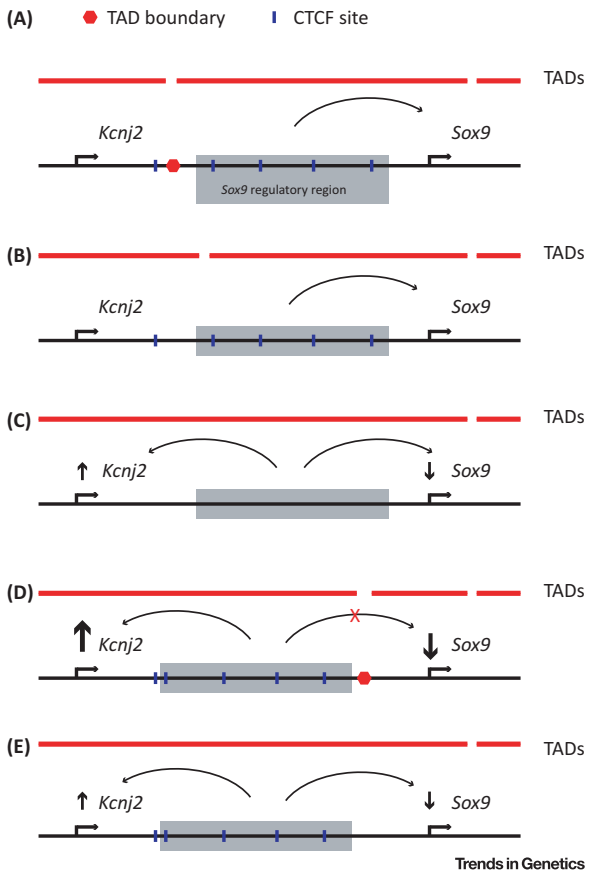

Figure 1. Reordering Topologically Associating Domain (TAD) Boundaries Affects Gene Expression

(A) Wild-type organization of the *Sox9-Kcnj2* locus showing the position of TADs, boundaries, and CTCF sites.
(B) The TAD boundary was deleted using clustered regularly interspaced palindromic repeats (CRISPR)/Cas9-mediated genome editing, which leads to a shift in the TAD boundary position.
(C) Additional mutation of the CTCF-binding sites results in fusion of the TADs. Vertical arrows in (C–E) indicate expression levels of the genes compared with wild-type. (D) Inversion of the entire *Sox9* regulatory region (*InvC*), including the boundary changes, shields *Sox9* from its own regulatory region. (E) Deletion of the boundary from *InvC* results in fusion of the two TADs and allows *Sox9* to access the regulatory region again.

Sox9 and *Kcnj2* for these elements. Furthermore, the comparison between the loss of insulation in the CTCF deletion and the inversion shows that a high-level misexpression of *Kcnj2* is only achieved by redirecting regulatory activity away from *Sox9* towards its new target *Kcnj2*.

These experiments elegantly show how the 3D genome contributes in subtle, yet important ways, to gene expression. However, they also emphasize that gene regulation occurs on various levels. Promoter-proximal transcription factor-binding sites, nucleosome positioning, and epigenetic modifications, such as DNA methylation and histone modifications, reciprocally influence each other to effectuate gene expression [8]. Distal enhancers are an additional layer in this process, although most enhancers appear to act within 20 kb of a target promoter [9]. However, certain genes, many of which encode crucial developmental regulators, have large regulatory domains and their expression depends on long-distance gene regulation. When we consider the complex spatio-temporal regulation of the neighboring *Kcnj2* and *Sox9* loci and the importance of the proper positioning of TAD boundaries, we can start to imagine why they are conserved throughout evolution. In the described experiments, TAD boundary disruption has mild effects on gene expression in a limited number of cells. However, the phenotypic outcome can be severe. This sentiment is echoed in a recent study in fruit fly embryos, in which the effect of 16 rearrangements on gene expression was assayed and found to have limited effects [10]. Although there is mounting evidence that long-range enhancers have a limited role in the regulation of most genes, they have an indispensable role for several vital regulators. Careful genetic dissection of the regulatory environment of these factors remains crucial to understand their regulation and how this may contribute to phenotypic differences.

[1]Division of Gene Regulation, Oncode Institute and Netherlands Cancer Institute, Amsterdam, the Netherlands

*Correspondence:
e.d.wit@nki.nl

https://doi.org/10.1016/j.tig.2019.09.002

References

1. Dixon, J.R. *et al.* (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380
2. Nora, E.P. *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385
3. Vietri Rudan, M. *et al.* (2015) Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep* 10, 1297–1309
4. Nora, E.P. *et al.* (2017) Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* 169, 930–944

5. Rao, S.S.P. *et al.* (2017) Cohesin loss eliminates all loop domains. *Cell* 171, 305–320
6. Despang, A. *et al.* (2019) Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.* 51, 1263–1271
7. Hanssen, L.L.P. *et al.* (2017) Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function *in vivo*. *Nat. Cell Biol.* 19, 952–961
8. Talbert, P.B. *et al.* (2019) Old cogs, new tricks: the evolution of gene expression in a chromatin context. *Nat. Rev. Genet.* 20, 283–297
9. Akhtar, W. *et al.* (2013) Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* 154, 914–927
10. Ghavi-Helm, Y. *et al.* (2019) Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* 51, 1272–1282

Spotlight

Many Ways to Build a Polyp

Detlev Arendt[1,2,*]

The freshwater cnidarian *Hydra* has been studied for centuries for its unique regenerative capacities. Whole-body single-cell transcriptomics now reveal cellular lineages and gene regulatory networks that build the *Hydra* polyp. For the first time, transcription factor signatures allow direct comparison of the polyp body plan between *Hydra* and sea anemone.

The *Hydra* polyp is a simple animal (Figure 1A). It is basically composed of an outer and an inner epithelial layer, called ectoderm and endoderm, which are separated by an extracellular matrix, the mesoglea (Figure 1B). In addition, *Hydra* has complex stinging cells, the nematocytes, and neurons – both of which reside in the interstitial spaces between the epithelial cells. Another frequent cell type, the gland cells, are integrated into the endodermal epithelium. The nervous system is composed of sensory neurons and of ganglion cells – multipolar neurons that send out axons to form a wide-meshed nerve net (Figure 1A).

All cells of the adult *Hydra* body are constantly replaced. Reflecting the simple anatomy, only three distinct cellular lineages – ectodermal, endodermal, and interstitial – build the polyp body. Ectodermal and endodermal cells remain mitotic along the body column and replace themselves. They terminally differentiate only towards the head, tentacles, or foot. This results in a constant flux of epithelial cells toward the extremities, with older cells being lost. In a separate lineage, nematocytes, neurons, and gland cells arise from the multipotent interstitial stem cells (ISCs) [1].

Benefitting from the clarity and simplicity of the *Hydra* anatomy and lineage, Siebert and colleagues have now used whole-body single-cell transcriptomics to molecularly characterize lineage trajectories and cell types of the entire *Hydra* body [2]. Collecting ~25 000 single-cell transcriptomes, their study has revealed the dynamics of gene expression that accompany cell specification and differentiation in the three *Hydra* lineages. Building on that, they have assembled the bifurcating trajectories of progenitor states that lead to the differentiating cell types; the Waddington landscape [3] of the adult *Hydra* polyp. For example, they have found that ISCs enter a single path that bifurcates towards neuronal or gland cell differentiation (Figure 1C). From this they have inferred that ISCs produce bipotential progenitors in the ectodermal layer that cross the extracellular matrix to supply the endodermal layer with neurons and gland cells.

Beyond that, the comparison of cellular transcriptomes across an entire body efficiently unravels gene sets that are uniquely coregulated within the genome. Juliano and colleagues have found ~60 of such gene sets. They then used ATACseq to identify regulatory regions for the genes of each set, detect transcription factor binding motifs that are enriched within the regions of coregulated genes, and finally match them to candidate transcription factors found in each set [2]. This yields first insights into the gene regulatory networks controlling cell type specification and differentiation in cell types and tissues. Figure 1D shows cell type-specific transcription factors selected from coregulated gene sets by the author.

Conserved Ectoderm – But How about Endoderm and Mesoderm?

The transcription factors that Siebert and colleagues have identified as specific for the ectodermal gene sets (*otx*, *six3*, *msx*, *hbn*, and *Dlx*; Figure 1D) have been shown to be active in ectodermal patterning in another cnidarian, the sea anemone *Nematostella* [4,5]. They also play well-established roles in ectodermal head patterning (otx, six3, and hbn) and mediolateral patterning (msx and Dlx) in the bilaterians [6]. This suggests overall conservation of ectodermal patterning in *Hydra* – with the exception of several factors (*Nkx6*, *PaxD*, and *rx*) that are conserved in sea anemone and bilaterians but missing from the *Hydra* genome [7].

The comparison of the endodermal gene sets sheds new light on endoderm evolution. On the one hand, specific transcription factors such as *Brachyury*, *Goosecoid*, *Pitx*, and *FoxA* (Figure 1C) make a strong signature matching the anterior endoderm of the vertebrate organizer [8]. On the other hand, *Pitx* and *FoxA* demarcate ectodermal pharynx in *Nematostella* [5], which has recently been homologized to