

Enhancers, gene regulation, and genome organization

Ann Dean,¹ Daniel R. Larson,² and Vittorio Sartorelli³

¹Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda Maryland 20892, USA; ²Laboratory of Receptor Biology and Gene Expression, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA; ³Laboratory of Muscle Stem Cells and Gene Regulation, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

How transcriptional enhancers function to activate distant genes has been the subject of lively investigation for decades. “Enhancers, gene regulation, and genome organization” was the subject of a virtual meeting held November 16–17, 2020, under sponsorship of the National Cancer Institute (NCI), the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) at the National Institutes of Health (NIH). The goal of the meeting was to advance an understanding of how transcriptional enhancers function within the framework of the folded genome as we understand it, emphasizing how levels of organization may influence each other and may contribute to the spatiotemporal specification of transcription. Here we focus on broad questions about enhancer function that remain unsettled and that we anticipate will be central to work in this field going forward. Perforce, we cover contributions of only some speakers and apologize to other contributors in vital areas that we could not include because of space constraints.

Originally discovered 40 yr ago in the simian virus 40 (SV40), enhancers are DNA regions driving tissue- and stage-specific expression of the genome (Banerji et al. 1981; Long et al. 2016; Furlong and Levine 2018). They contain binding sites for lineage-regulating transcription factors and are highly cell type-specific. Typically located at long genomic distances from their target genes, enhancers may be in upstream or downstream intergenic regions, in intronic sequences, or even in introns of unrelated genes. It is estimated that the human genome hosts hundreds of thousands of enhancers, vastly outnumbering the ~20,000 genes.

Most GWAS disease and trait-associated SNPs occur in noncoding regions of the genome. Are enhancers primary

GWAS disease-associated loci? In one example, Christopher Glass and coworkers (University of California at San Diego) used fluorescence-activated nuclear sorting to systematically characterize enhancers in neurons, astrocytes, oligodendrocytes, and microglia and then intersected these genomic locations with GWAS risk alleles for Alzheimer’s disease (AD) and other neurological and psychiatric diseases (Nott et al. 2019). GWAS risk alleles associated with AD were preferentially enriched in microglia enhancers, while risk alleles associated with psychiatric diseases were preferentially enriched in neuronal enhancers. An AD risk allele associated with the *BIN1* gene resided in a putative microglia-specific enhancer. Deletion of this enhancer abolished *BIN1* expression in iPSC-derived microglia, but not iPSC-derived neurons or astrocytes.

While several studies indicate that enhancers are enriched for polymorphisms, the significance of the association between enhancers and GWAS signals has been difficult to interpret, as their cell types of action and effector transcripts remain, in most cases, unidentified. Type 2 diabetes (T2D) is a complex disease involving pancreatic β -cell dysfunction and insulin resistance of peripheral tissues such as liver, adipose, and skeletal muscle. Stephen Parker (University of Michigan) discussed the results of an integrative genome and single cell multi-omics study conducted on 287 genotyped muscle biopsies representing glucose tolerance (GT) normal, intermediate, and newly diagnosed T2D nontreated individuals. Frozen biopsies from individuals were processed for single-nucleus (sn) RNA-seq and snATAC-seq. Clusters corresponding to 13 different cell types were identified, including type 1 and type 2 myofibers, mesenchymal stem cells (MSC), and endothelial cells. Genetic variation was integrated with the snRNA profiles to identify expression quantitative trait loci (eQTL) and snATAC profiles to identify chromatin accessibility (caQTL). These cell-specific e/caQTL signals were then integrated with T2D GWAS signals. Parker and colleagues found that caQTL signals were more frequently associated with T2D GWAS signals (one order of

[Keywords: enhancers; gene regulation; genome organization]

Corresponding authors: sartorev@mail.nih.gov,

ann.dean@nih.gov, dan.larson@nih.gov

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.348372.121>.

This is a work of the US Government.

magnitude) than eQTL and were often cell type-specific. Thus, caQTL may have greater predictive power than eQTL for T2D. Half of the T2D GWAS signals that colocalized with caQTL were cell type-specific, highlighting the importance of sn-caQTL maps for GWAS functional studies. For example, in MSCs, they identified a cell type-specific caQTL for an *ARL15* intronic peak that colocalized with the T2D GWAS signal. To nominate target genes for this caQTL peak, they scanned peaks in promoters of genes in the 1-Mb neighborhood and computed coaccessibility scores that identified the gene *FST*, ~493 kb away, which encodes for follistatin, suggesting a potential regulatory function of the caQTL in question on *FST* expression. This observation is intriguing considering that *FST* increases muscle mass and reduces fat mass and insulin resistance. MSCs were not conclusively characterized and may correspond to fibro-adipogenic precursors.

Mendelian disruption of enhancers also has deleterious effects. These can result from deletions, duplications, inversions, translocation, or point mutations. Possibly the first molecularly described “enhanceropathy,” as reviewed at the meeting by Gerd Blobel (Children’s Hospital of Philadelphia), is the deletion of a large region upstream of the β -globin genes, which resulted in their silencing, although the genes themselves were intact (Driscoll et al. 1989). The discovery of this locus control region (LCR) provided an important example of regulation of genes by distant enhancers. Chromosome conformation capture (3C), which depends on proximity ligation, provided evidence to support physical interaction between this enhancer and target promoter for gene activation (Fig. 1A; Tolhuis et al. 2002). How is enhancer-gene proximity promoted or antagonized?

Blobel (Lan et al. 2020) presented a role for a novel transcription factor, ZNF410, that seems to have only one target in erythroid cells. In erythroid cells, it is known that the LCR enhancer initially interacts with fetal γ -globin genes and, later in development, switches to interact with adult type β -globin genes. Repressors BCL11A and LRF bind to the γ -globin promoter and recruit the NuRD remodeling and deacetylase complex to silence transcrip-

tion. ZNF410 uniquely activates CHD4, one of the catalytic components of NuRD. Although NuRD has numerous targets, the γ -globin genes are particularly sensitive to CHD4 reduction. This fortuitous regulatory pathway means that ZNF410 may be a druggable therapeutic target for severe hemoglobinopathies to increase γ -globin with few anticipated direct off-target effects. Whether CHD4 expression is also regulated by this single transcription factor in other cell lineages is not yet clear. Likewise, it is not known yet whether ZNF410 is as highly selective a factor in other cell types.

Although studies of the β -globin genes and other model systems seem to unequivocally support enhancer looping, imaging studies have shown distances on the order of 200 nm between active enhancers and genes. How stable or dynamic is enhancer-gene proximity and does proximity always correlate with gene activation? Using 5C and FISH assays, the laboratory of Wendy Bickmore (University of Edinburgh) had surprisingly observed greater separation between *Shh* and its enhancers upon activation during differentiation of embryonic stem cells (ESCs) to neural progenitors (Fig. 1B; Benabdallah et al. 2019). Importantly, *Shh* responded to its enhancers despite their separation, providing strong evidence that proximity is not essential for enhancer activation of this gene and may even disfavor it. In these experiments, a very early, transient enhancer-promoter proximity might have been missed. Bickmore presented new experiments using a model with a more rapid response to enhancer activation. Upon activation, responsive enhancers and target genes moved apart very fast—within 5 min—revealing separation to be an early event.

If enhancers separate from genes, how can we envision activation to occur? Bickmore suggested that PARP, whose activity was required for the separation, could contribute to a PARP-modified scaffold upon which specific transcription factors accumulate and recruit coactivators like Mediator and P300, contributing the “intelligence” of the system of gene activation within a transcription-rich environment, possibly a condensate. Nevertheless, the devotion of an enhancer to a specific target gene and

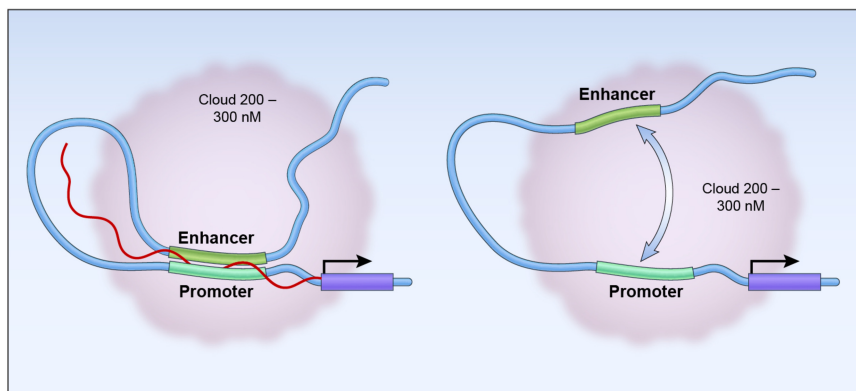


Figure 1. Enhancer-promoter communication. (Left) An enhancer and promoter are depicted as coming close together for gene activation, as captured in proximity ligation experiments, with looping away of intervening sequences. Lineage-specific transcription factors mediate this interaction proximity and serve to recruit coactivators and RNA Pol II, which then can act at the nearby promoter to activate transcription. An enhancer RNA (red line) is depicted as participating in the interaction. (Right) Enhancer and promoter are not closely associated and can even move farther apart when activating their target gene. Enhancers are still bound by specific

transcription factors but recruit coactivators and Pol II to increase the local concentration of these components within a cloud or hub, the nature of which remains unclear.

rejection of other nearby genes is still difficult to envision in this scenario.

Focusing on developmental enhancers and topologically associating domains (TADs) in the context of embryogenesis in flies, Eileen Furlong (European Molecular Biology Laboratory) has addressed when and how chromatin topology is established. Her group (Ghavi-Helm et al. 2014) previously showed that many embryonic enhancers are already in proximity to their target genes' promoter hours before the gene is expressed. Thus, enhancer proximity does not necessarily trigger transcription and, indeed, can precede it during developmental progression. Using single-cell DNA FISH on carefully staged embryos, her laboratory has now examined even earlier stages in embryogenesis and observed both TAD and long-range loop formation at the zygotic genome activation (ZGA) stage. Notably, this occurs at loci before the transcription of the genes initiates, and TAD formation happens even when transcription is inhibited. Interestingly, transcription factors are known to be engaged at early embryonic enhancers at this period before gene activation, suggesting a potential role for these factors in promoting or stabilizing the loops. Are eRNAs produced at these enhancers before gene activation? The Ghavi-Helm et al. (2014) study found that many genes with preformed enhancer-promoter loops have paused RNA polymerase at their promoters. How is gene activation prevented? Is there a role for Polycomb or other repressors in counteracting transcriptional elongation at these target genes?

According to proximity ligation experiments, enhancers interact with other enhancers as well as target promoters. Moreover, genetic variations in enhancers can affect the activity of other enhancers. How are enhancer-enhancer interactions established and maintained, and what is their function? A new approach to probe these questions was described by Susanne Mandrup (University of Denmark). Her laboratory devised an enhancer-capture HiC method based on the logic of promoter capture HiC. Examining enhancer repertoires at time points after hMSCs are induced to differentiate to adipocytes, they observed fundamentals of enhancer-enhancer interactions. Enhancer-enhancer interactions were at least as dynamic as enhancer-promoter interactions, and interacting enhancers tended to be coregulated and to be occupied by the same transcription factors. In fact, what Mandrup referred to as cross-interaction stabilization (CIST) tended to increase transcription factor occupancy at interacting enhancers bound by the same transcription factor, regardless of motif strength but dependent on the number of interactions bound by the factor. Highly connected enhancers (more than eight interactions) were enriched for CTCF occupancy. Mandrup described "enhancer communities" that interact more with each other than with other enhancers and showed that highly connected enhancer communities appear to drive lineage-specific genes. Interestingly, highly connected enhancer communities governing many different lineages appear to coexist in the undifferentiated hMSCs. Upon differentiation, interactions are modulated in a lineage-specific manner. One caveat is that, absent an advance-

ment to the single-cell level, enhancer capture HiC is unable to reveal whether the multiple enhancer contacts that are observed occur at the same time in one cell or in different cells.

Enhancers and target genes are predominately located within the same TAD. Do TADs influence interaction between the enhancers and target genes or do enhancer-gene interactions influence TAD formation? Eileen Furlong's group has manipulated TAD structure in *cis*, as well as depleted different insulator proteins in *trans*, to understand how TADs are formed and what their role is in enhancer function. Doing these experiments at the ZGA stage indicates that TADs can still be established in early embryogenesis when any one insulator protein is depleted. However, although there is little global change, chromatin structure at some individual loci is affected. Similarly, when TAD structure is changed by genomic rearrangements in *cis* (using deletions and inversions in highly rearranged "balancer chromosomes"), the expression of the majority of genes is unaffected. However, some genes are affected, and this appears to be due to the formation of ectopic contacts by enhancers. Why some enhancers are constrained or impacted by the presence of a TAD boundary and others are not is a really interesting question. Perhaps these enhancers (or the boundaries) have different properties.

From a different angle, Effie Apostolou (Weill Cornell Medical College) reported on the formation of genome topology in mouse pluripotent stem cells (PSCs). The transcriptional program of PSCs is rapidly reset upon exit from mitosis, and this requires mitotic bookmarking by H3K27ac. Using in situ HiC, Apostolou showed that early changes in transcription are associated with large-scale topological changes after exiting mitosis, including A-B compartmentalization and TAD boundary insulation that are overall rapidly re-established. However, intra-TAD connections and chromatin loops reformed in a more gradual and asynchronous manner. The first interactions that re-emerged were relatively small in size and were enriched for H3K27ac bookmarking and for transcription factors, RNA Pol II, and Mediator binding, suggesting that they represent active regulatory loops (e.g., enhancer-promoter contacts). In contrast, loops that formed more gradually or late, referred to as structural loops, which could represent subTADs or neighborhoods, were significantly larger and were enriched for CTCF and cohesin but not H3K27ac, Mediator, or transcription factors. These experiments argue that enhancer-promoter interactions form prior to and potentially independently of neighborhoods/structural loops and that neighborhoods might even coalesce around enhancer-promoter loops.

Another point of view emerged from experiments carried out by Denis Duboule and Chase Bolt (EPFL). Genes of the HoxD locus lie between two TADs. Enhancers of Hox genes expressed early in the proximal forelimb are located on an adjacent downstream TAD, while enhancers of genes expressed later in the distal forelimb are located in an adjacent upstream TAD. At early developmental times, the enhancers of the proximal genes are active, while the enhancers of the distal genes are silent, which

is reversed at later times. Duboule described moving a distal gene enhancer into the proximal gene enhancer region. Under these new conditions, the translocated distal enhancer lost the distal specificity that is otherwise fully penetrant when integrated at random genomic positions. This is at least consistent with some aspect of TADs playing a specific role in the function of enhancers within their borders. The context seems to be important for the enhancer function. However, enhancer-promoter interactions were not specifically examined to see whether they changed.

That transcription causes locus decondensation has been known for many years (Janicki et al. 2004), possibly indicating multiple conformations of enhancer-regulated transcription where spatial proximity is important (Chen et al. 2018; Alexander et al. 2019) but variable over the course of transcriptional activation. How do enhancers up-regulate gene transcription without achieving proximity? Bickmore had proposed that clusters or condensates embrace the two elements still at some detectable distance, possibly driven by transcription factors occupying the elements and recruiting coactivators (Fig. 2). Imaging provided further insight as reported by Michael Levine (Princeton University) in studies of coregulation of two genes, *Scylla* and *Charybda*, by the same enhancer in fly embryos. When activating the two genes, the enhancer and gene signals do not merge but maintain “social separation,” a theme of our COVID-19 times, which Levine attributed to molecular travels within a cloud or cluster (called a hub) of transcriptional activity. In this case, there is tethering of the enhancer and genes using separate elements enriched for GAGA factor. Deletion of the tethering sequences delays but does not prevent the onset of transcription of the genes. The lack of close enhancer-gene proximity, as well as the ability of an enhancer to activate more than one gene, is in contrast to enhancer behavior observed at other genes, as in the β -globin model.

Richard Young (Massachusetts Institute of Technology) reported on transcriptional condensates formed at superenhancers. Biomolecular condensates are clusters of proteins and nucleic acids that assemble and compartmentalize diverse molecules with shared functions, forming due to dynamic multivalent interactions in the

intrinsically disordered regions (IDRs) of proteins. Superenhancers, which drive expression of key cell identity genes and oncogenes, are bound by a high density of transcription factors and cofactors with domains enriched in IDRs. The properties of transcription factor activating domains required for gene activation are also proposed to be required for the formation of transcriptional condensates in vivo. The transcriptional condensates formed at superenhancers are ~ 300 nm in diameter, contain >100 molecules of Mediator and RNA Pol II, form and dissolve dynamically, and display characteristic liquid-like behaviors.

Young (Klein et al. 2020) reported that drugs accumulate in specific nuclear condensates by preferentially interacting with specific condensate proteins and that this property is independent of the drugs’ canonical target. For instance, the widely used anticancer drug cisplatin selectively concentrates in MED1-containing condensates in the absence of its DNA target. Comprehensive mutagenesis of MED1 revealed that aromatic amino acids are dispensable for condensate formation but are important to concentrate drugs, either by π -stacking or π -cation interactions. Cisplatin preferentially intercalates DNA associated with SEs occupied by MED1 condensates. Additional anticancer drugs were found to concentrate selectively in one or more condensates, suggesting that this may be a general property of drug distribution in cells and might also provide a mechanism for development of drug resistance. In breast cancer cells, MED1 overexpression caused dilution of tamoxifen by its sequestration in condensates, rendering the tumor resistant to treatment. Overall, these findings have potential implications for the development of chemotherapeutic agents targeting specific cell compartments and genomic regions, improving the therapeutic index of small-molecule drugs and overcoming drug resistance.

The proliferation of vocabulary to describe fluorescent spots in the nucleus associated with enhancer-regulated transcription is one indication of our persistent lack of knowledge about their nature. These features have been called “transcription factories,” “hubs,” “foci,” “condensates,” “phases,” and “droplets” (Fig. 3). It is clear that these spots are dynamic, exchanging factors with the nucleoplasm and are related to functional output. However,

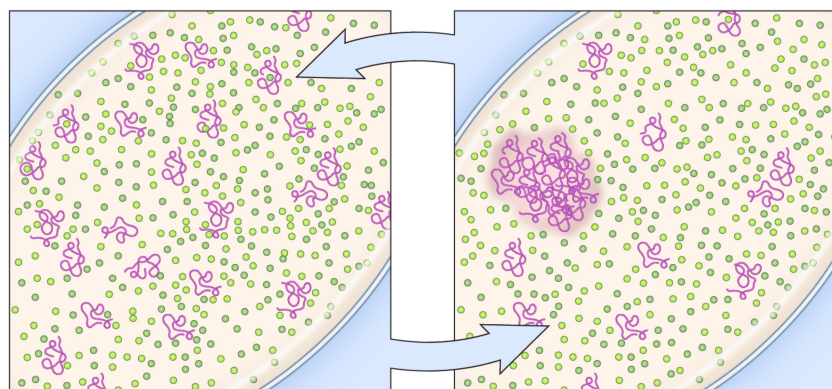


Figure 2. Phase separation. Graphic representation of a cell nuclear compartment (yellow oval) hosting two distinct molecule populations (green circles and purple squiggles). A homogeneous mixture of the two molecule populations is depicted in the *left* panel. (*Right* panel) Exogenous or endogenous cues can induce preferential concentration of one of the two molecule populations (purple squiggles) in a non-membrane-bound nuclear subregion. The arrows signify that the two states (homogeneous at the *left* and separated at the *right*) are in a dynamic relationship. Genomic DNA is not depicted.

multiple probes in living cells with nanometer spatial resolution and time resolution covering a dynamic range from milliseconds to minutes or hours. In the absence of such technology, the field is currently reliant on indirect methods such as fluorescence recovery after photobleaching (FRAP), which can have multiple interpretations (Lionnet et al. 2010); site-directed mutagenesis, which interferes with function; and perturbation with chemicals that have pleiotropic effects on chromatin (Itoh et al. 2021). As such, the nature of these foci was an area of active debate. No doubt these questions will continue to be addressed in future “enhancers, gene regulation, and genome organization” meetings.

Competing interest statement

The authors declare no competing interests.

References

- Alexander JM, Guan J, Li B, Maliskova L, Song M, Shen Y, Huang B, Lomvardas S, Weiner OD. 2019. Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife* **8**: e41769. doi:10.7554/eLife.41769
- Austena LM, Barozzi I, Simonatto M, Masella S, Della Chiara G, Ghisletti S, Curina A, de Wit E, Bouwman BA, de Pretis S, et al. 2015. Transcription of mammalian *cis*-regulatory elements is restrained by actively enforced early termination. *Mol Cell* **60**: 460–474. doi:10.1016/j.molcel.2015.09.018
- Banerji J, Rusconi S, Schaffner W. 1981. Expression of a β -globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**: 299–308. doi:10.1016/0092-8674(81)90413-X
- Benabdallah NS, Williamson I, Illingworth RS, Kane L, Boyle S, Sengupta D, Grimes GR, Therizols P, Bickmore WA. 2019. Decreased enhancer-promoter proximity accompanying enhancer activation. *Mol Cell* **76**: 473–484.e7. doi:10.1016/j.molcel.2019.07.038
- Chen H, Levo M, Barinov L, Fujioka M, Jaynes JB, Gregor T. 2018. Dynamic interplay between enhancer-promoter topology and gene activity. *Nat Genet* **50**: 1296–1303. doi:10.1038/s41588-018-0175-z
- Driscoll MC, Dobkin CS, Alter BP. 1989. $\gamma\delta\beta$ -thalassemia due to a de novo mutation deleting the 5' β -globin gene activation-region hypersensitive sites. *Proc Natl Acad Sci* **86**: 7470–7474. doi:10.1073/pnas.86.19.7470
- Furlong EEM, Levine M. 2018. Developmental enhancers and chromosome topology. *Science* **361**: 1341–1345. doi:10.1126/science.aau0320
- Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EE. 2014. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**: 96–100. doi:10.1038/nature13417
- Henninger JE, Oksuz O, Shrinivas K, Sagi I, LeRoy G, Zheng MM, Andrews JO, Zamudio AV, Lazaris C, Hannett NM, et al. 2021. RNA-mediated feedback control of transcriptional condensates. *Cell* **184**: 207–225.e24. doi:10.1016/j.cell.2020.11.030
- Itoh Y, Iida S, Tamura S, Nagashima R, Shiraki K, Goto T, Hibino K, Ide S, Maeshima K. 2021. 1,6-hexanediol rapidly immobilizes and condenses chromatin in living human cells. *Life Sci Alliance* **4**: e202001005. doi:10.26508/lsa.202001005
- Janicki SM, Tsukamoto T, Salghetti SE, Tansey WP, Sachidanandan R, Prasanth KV, Ried T, Shav-Tal Y, Bertrand E, Singer RH, et al. 2004. From silencing to gene expression: Real-time analysis in single cells. *Cell* **116**: 683–698. doi:10.1016/S0092-8674(04)00171-0
- Klein IA, Boija A, Afeyan LK, Hawken SW, Fan M, Dall'Agnese A, Oksuz O, Henninger JE, Shrinivas K, Sabari BR, et al. 2020. Partitioning of cancer therapeutics in nuclear condensates. *Science* **368**: 1386–1392. doi:10.1126/science.aaz4427
- Lai F, Gardini A, Zhang A, Shiekhattar R. 2015. Integrator mediates the biogenesis of enhancer RNAs. *Nature* **525**: 399–403. doi:10.1038/nature14906
- Lan X, Ren R, Feng R, Ly LC, Lan Y, Zhang Z, Aborede N, Qin K, Horton JR, Grevet JD, et al. 2020. ZNF410 uniquely activates the NuRD component CHD4 to silence fetal hemoglobin expression. *Mol Cell* **81**: 239–254.e8. doi:10.1016/j.molcel.2020.11.006
- Lionnet T, Wu B, Grunwald D, Singer RH, Larson DR. 2010. Nuclear physics: quantitative single-cell approaches to nuclear organization and gene expression. *Cold Spring Harb Symp Quant Biol* **75**: 113–126. doi:10.1101/sqb.2010.75.057
- Long HK, Prescott SL, Wysocka J. 2016. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167**: 1170–1187. doi:10.1016/j.cell.2016.09.018
- Nott A, Holtman IR, Coufal NG, Schlachetzki JCM, Yu M, Hu R, Han CZ, Pena M, Xiao J, Wu Y, et al. 2019. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science* **366**: 1134–1139. doi:10.1126/science.aay0793
- Sartorelli V, Laubert SM. 2020. Enhancer RNAs are an important regulatory layer of the epigenome. *Nat Struct Mol Biol* **27**: 521–528. doi:10.1038/s41594-020-0446-0
- Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. 2002. Looping and interaction between hypersensitive sites in the active β -globin locus. *Mol Cell* **10**: 1453–1465. doi:10.1016/S1097-2765(02)00781-5