Molecular Cell

CellPress
OPEN ACCESS

Review

Deciphering the multi-scale, quantitative *cis*-regulatory code

Seungsoo Kim[1,2,3,4] and Joanna Wysocka[1,2,3,4,*]
[1]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA
[2]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA 94305, USA
[3]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305, USA
[4]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA
*Correspondence: wysocka@stanford.edu
https://doi.org/10.1016/j.molcel.2022.12.032

SUMMARY

Uncovering the *cis*-regulatory code that governs when and how much each gene is transcribed in a given genome and cellular state remains a central goal of biology. Here, we discuss major layers of regulation that influence how transcriptional outputs are encoded by DNA sequence and cellular context. We first discuss how transcription factors bind specific DNA sequences in a dosage-dependent and cooperative manner and then proceed to the cofactors that facilitate transcription factor function and mediate the activity of modular *cis*-regulatory elements such as enhancers, silencers, and promoters. We then consider the complex and poorly understood interplay of these diverse elements within regulatory landscapes and its relationships with chromatin states and nuclear organization. We propose that a mechanistically informed, quantitative model of transcriptional regulation that integrates these multiple regulatory layers will be the key to ultimately cracking the *cis*-regulatory code.

INTRODUCTION

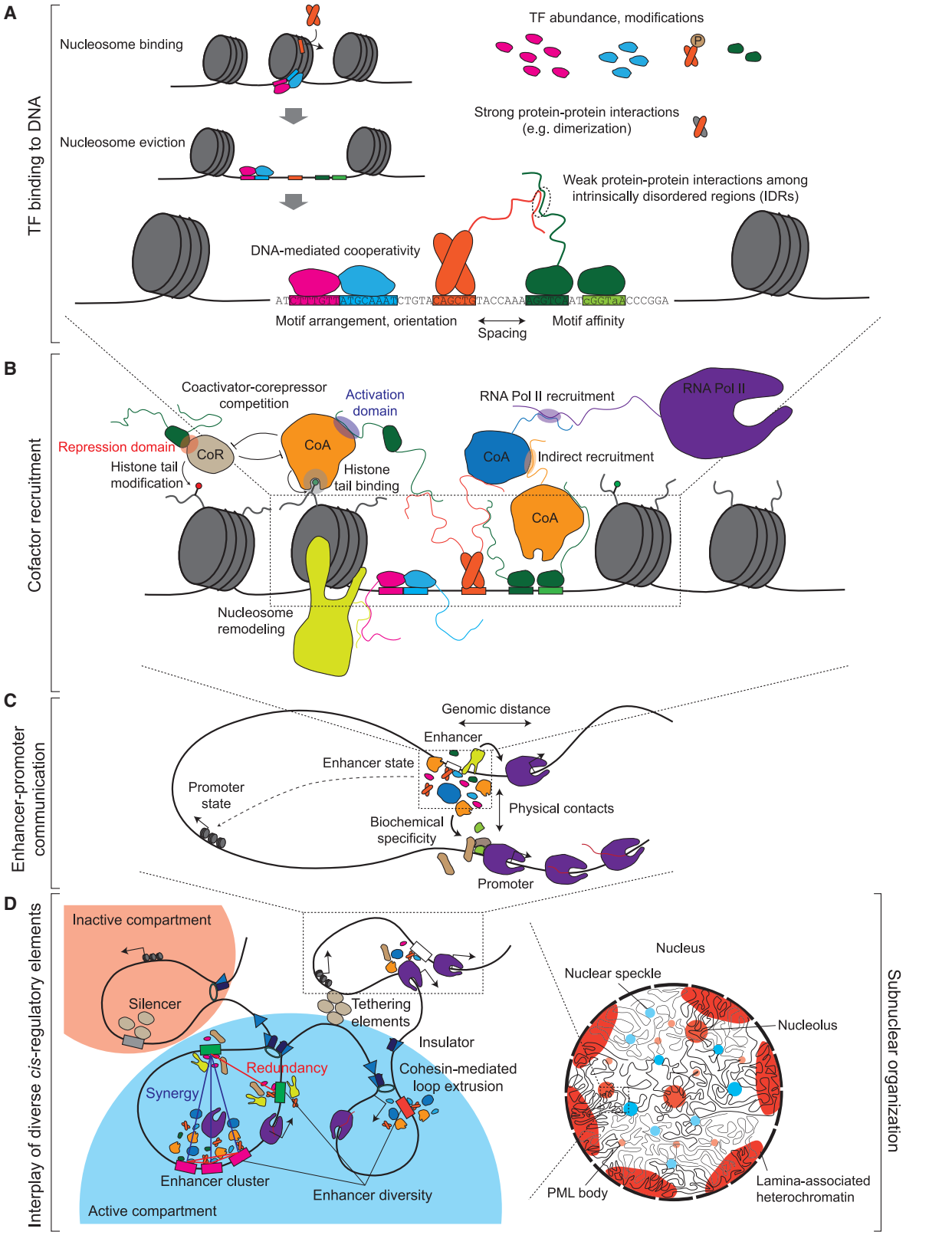
One of the fundamental goals of biology is to understand how organisms decode genetic information into phenotypes—that is, unique biological characteristics and functions. This decoding occurs over distinct steps, as Francis Crick hypothesized in his Central Dogma,[1] beginning with the linear sequence information in DNA being transferred into RNA and then into protein. The most archetypal of these steps is encapsulated by the genetic code, which describes how each sequence of nucleotides in a messenger RNA molecule is translated into a string of amino acids.[2]

Since the seminal discovery that the protein-coding regions of the genome remain largely conserved between humans and chimpanzees, it has long been postulated that phenotypic divergence between and within closely related species arises in large part from quantitative changes to gene expression.[3] These changes are to a great extent genetically encoded, as evidenced by high heritability of many phenotypic traits in different species and within human populations, suggesting the existence of another DNA-based code that dictates them.[4] This code is often referred to as the "*cis*-regulatory code" because gene expression is controlled by the regulatory DNA elements, which typically act in "*cis*," meaning at the same chromosomal allele (and, as discussed below, usually within the same chromosomal neighborhood) as the gene they regulate. In contrast to the genetic code, however, the *cis*-regulatory code remains ill-defined and difficult to crack. In this review, we compare and contrast the genetic and *cis*-regulatory codes, discuss features of the *cis*-regulatory code that make it challenging to comprehensively decipher, outline our current understanding of these features, and consider approaches moving forward.

WHAT IS THE *CIS*-REGULATORY CODE?

The term "code" implies conversion of one form of information to another, akin to the mapping between nucleotides to amino acids in the genetic code. Broadly, the *cis*-regulatory code can be described as the mapping between DNA sequence and gene expression level. (For simplicity, in this review, we will equate the gene expression level with transcriptional activity; nonetheless, it is important to note that post-transcriptional regulatory mechanisms also play key roles in gene regulation and in phenotypic divergence between and within species.) A solution to the code should be predictive; as the genetic code allows one to predict the amino acid sequence given a messenger RNA sequence, the *cis*-regulatory code should allow one to predict transcription levels from DNA sequence. However, this definition remains vague: how much DNA sequence is needed to achieve a comprehensive prediction of gene expression? Is it adequate to predict transcription of one gene at a time? And transcription in what cellular contexts? This ambiguity reflects three major challenges of the *cis*-regulatory code, which contrast three fortuitous features of the genetic code that facilitated efforts to decipher it: the genetic code is *universal*, *modular*, and *qualitative*. In contrast, as we discuss below, the *cis*-regulatory code is *context-dependent*, in part modular, but with *complex* and poorly understood relationships between modules, and *quantitative*.

A single, universal genetic code is used in all cell types in an organism and among virtually all species, with minor variations in organelles and prokaryotes, reflecting the early evolution of translational machinery.[2] The *cis*-regulatory code is instead



(legend on next page)

highly dependent on cellular context because the transcription factors (TFs) that directly read DNA sequences differ from one cell type to another,[5] as do epigenetic marks such as DNA methylation that can modulate TF binding to DNA.[6] Though it can be tempting to treat the hundreds of major cell types as separate identities, single-cell studies have demonstrated how they are linked by uncountable transient intermediate states that can occur naturally in development or only in diseased or experimentally perturbed states. Ultimately, a solved *cis*-regulatory code would be integrated across the continuum of cell states and be able to predict how perturbations to cell state affect transcription in development and disease.

A second challenge of the *cis*-regulatory code is the size of the input DNA sequence. In the genetic code, each RNA molecule and the triplet codons within each open reading frame are "modular"—they are independent (one codon does not affect any others, except stop codons) and non-overlapping. However, in the *cis*-regulatory code, is the entire genome necessary to predict the expression level of all genes? Many have focused on the more operational prediction of the impact of short DNA sequences on the expression of one reporter gene,[7] as this is experimentally feasible. Indeed, many short *cis*-regulatory elements such as the 480-bp minimal even-skipped stripe 2 enhancer[8] are capable of individually regulating transcription and can often be treated as modular units.[9] However, such fragments can fail to fully explain the activity of larger genomic regions,[10] as many *cis*-regulatory elements interact physically or epistatically, across megabases of DNA (e.g., at the *SOX9*[11] and *MYC* loci[12]) and even across chromosomes (e.g., during olfactory receptor selection).[13] Thus, a complete solution to the *cis*-regulatory code must be able to predict how large genomic regions, each containing tens or hundreds of *cis*-regulatory elements with complex relationships among each other, control the expression of multiple genes.

A third twist to the *cis*-regulatory code is that unlike the qualitative genetic code governing the identity of translated peptides, the *cis*-regulatory code aims to describe the quantitative levels of transcription. Genes are not simply "on" or "off" but exhibit precise levels of transcription controlled by the frequency of RNA polymerase molecules traversing the gene.[14] Transcription occurs in "bursts" whose frequency and amplitude are subject to regulation, but which are inherently stochastic.[15] As a result of this discontinuous process, the two gene alleles within the same cell (even if genetically identical) can have distinct transcriptional status at any given moment in time, whereas within a population of cells of the same "type," transcript levels can vary widely between individual cells. Yet, even relatively subtle changes in gene expression (~2-fold or less) can dramatically impact development or increase risk of disease. For example, modest downregulation of *SOX9* causes the craniofacial syndrome Pierre Robin sequence,[11] while upregulation of *PDGFRA*[16] or *MYC*[17] accelerate cell proliferation and can drive cancer. These examples underscore the importance of quantitatively precise regulation of transcription. Thus, the ultimate goal of solving the *cis*-regulatory code requires quantitatively accurate predictions of transcription levels.

Given the aforementioned context-dependent, complex, and quantitative nature of the *cis*-regulatory code, can we understand how precise transcription levels of more than 20,000 genes across essentially uncountable cell states are encoded in the 3 billion bases of the human genome? Although this goal still remains elusive, the field has made enormous progress in understanding the broad rules by which DNA sequences encode transcriptional outputs. For the purpose of this review, we divide these mechanisms into four levels at which they operate, from smallest to largest in genomic scale: (1) the reading of DNA sequence by TFs, (2) the modular function of individual *cis*-regulatory elements, (3) interactions and compatibility between distal *cis*-regulatory elements and their target promoters, and (4) relationships among diverse *cis*-regulatory elements within a complex regulatory landscape (Figure 1). Below, we discuss these four regulatory levels in turn.

FROM DNA SEQUENCE TO TF BINDING

TFs bind a subset of their motifs

The minimal unit of the *cis*-regulatory code—paralleling the codon of the genetic code—is the TF-binding site (TFBS). TFs contain typically structured and evolutionarily conserved DNA-binding domains (DBDs) that recognize and bind a 6–12-bp DNA sequence called a TF "motif."[5] TF motifs are often depicted as a sequence logo or position weight matrix (PWM) that represents the degeneracy in TF binding specificity. The deep conservation of DBDs and high-throughput methods for measuring TF sequence specificities *in vitro* have enabled the cataloging of roughly 1,600 TFs in humans (though more may exist) and characterizing binding motifs for a large fraction of these known TFs.[5]

Yet, TF-binding motifs are insufficient to predict DNA binding *in vivo* for most TFs. While most TFBSs contain at least an imperfect match for their preferred motifs, most TFs bind only a small fraction of their motif matches in the genome.[18] While efforts to better encapsulate TF specificity by including nucleotides flanking the core motif[19] or using more complex representations of sequence preference (e.g., dinucleotide motifs and DNA shape[20]) can improve predictions of genomic binding for some TFs, the best predictor of *in vivo* binding for most TFs is chromatin

Figure 1. Four layers of the *cis*-regulatory code

(A) TF binding to DNA depends on the sequence recognition and ability to overcome the nucleosomal barrier through direct association with nucleosome and/or cooperative binding of multiple TFs in turn facilitating nucleosome eviction. TF occupancy is dependent on TF levels, post-translational modifications, protein-protein interactions (which can either be structured or mediated by weak-affinity interactions among the IDRs) and is regulated DNA-mediated cooperativity with other TFs, which is itself governed by *cis*-regulatory features such as motif arrangement, spacing, and affinity.

(B) TFs function at modular *cis*-regulatory elements by recruiting cofactors with diverse and sometimes competing functions such as coactivators (CoAs) vs. corepressors (CoRs). Many interactions between TFs and coactivators are mediated by IDRs.

(C) Enhancers and other distal *cis*-regulatory elements selectively regulate promoters, depending on multiple features such as genomic distance, enhancer and promoter state, biochemical specificity between enhancer- and promoter-associated proteins and physical contacts. See also Figure 2.

(D) Diverse regulatory elements, including enhancers, promoters, silencers, insulators, and tethering elements interact with each other physically and/or epistatically (e.g., redundancy or synergy) in the context of local chromosomal neighborhood and a spatially organized nucleus.

accessibility as measured by DNase-seq[21,22] or assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq).[23] This observation has largely been attributed to the occupancy of DNA by nucleosomes that must be outcompeted or evicted either by so-called "pioneer" factors that possess a special capacity to bind nucleosomal DNA[24] or combinations of co-bound TFs.[25,26]

TFs bind DNA in nucleosomal context

The idea of pioneer factors was conceived as an explanation for TFs that bound enhancer sequences in a closed chromatin state preceding subsequent activation during development.[27] Thus, in their original definition pioneer factors are able to both access nucleosome-bound DNA and facilitate enhancer competence for binding of other TFs.[28] Consequently, pioneer factors are associated with TF cocktails capable of inducing cellular reprogramming upon overexpression, such as the OSKM (Oct4, Sox2, Klf4, and c-Myc) pluripotency factors,[29] as their nucleosome-binding ability[30] is thought to allow them to function outside of their native cell state. TFs clearly vary in their capacity to bind nucleosomal DNA *in vitro*[31,32] and closed chromatin *in vivo*,[33] with variation even among paralogs with conserved DBDs.

Although pioneering activity is often described as a binary trait of TFs that allows binding to any cognate motif regardless of nucleosome occupancy, multiple lines of evidence suggest that pioneering is both mechanistically diverse and highly context-dependent. First, pioneer factors span many TF families (i.e., protein folds) and can bind nucleosomal DNA with different positional preferences (e.g., near nucleosomal ends, at the dyad, or periodically along the exposed sections of the DNA major groove).[31,32,34] Second, the catalog of pioneer vs. non-pioneer TFs remains far from complete, and the precise distinction remains unclear. A recent study showed that a well-characterized pioneer factor, FOXA1, and a TF described as a non-pioneer factor, HNF4A, were equally capable of opening chromatin upon overexpression in a context in which neither is normally expressed, to a degree dependent on TF expression level.[35] While pioneering mechanisms often invoke direct functions such as bending DNA to unwrap it from histones,[34] pioneer TFs can also recruit nucleosome remodelers.[36] Whether pioneering activity is dependent on remodeler recruitment and whether remodelers might allow non-nucleosome-binding TFs to open chromatin remains largely unexplored. However, a recent *in vitro* study on the hematopoietic TF PU.1 provided an early proof-of-principle that pioneer factors can in fact facilitate activity of nucleosome remodelers at specific DNA sites—in this case, in a manner dependent not on the DBD, but on the intrinsically disordered region (IDR) of PU.1.[37] Also, third, pioneer factor activity is context-dependent: even pioneer TFs such as OCT4 or SOX2 usually bind distinct sites in different cell types, reflecting cooperativity with other TFs that is required at least for stable binding.[38,39] Thus, the mechanism and extent of nucleosomal DNA binding appears to be one of several regulatory layers controlling which DNA sequences TFs can bind.

TFs cooperate to bind DNA

A key (and importantly, non-exclusive) alternative to the pioneer factor model is that multiple TFs cooperate to outcompete nucleosomes together by binding within the same nucleosomal distance of ~150 bp.[25] While the idea of cooperativity among multiple co-bound TF molecules is not new, the precise nature of this cooperativity remains debated. Initial studies of the *INFβ* enhancer suggested that an array of different TFs must bind in tight spacing to allow protein-protein interactions that mediate cooperativity.[40] However, subsequent studies revealed that different enhancers have diverse sets of bound activators in different arrangements, in line with the more flexible nucleosome-mediated cooperativity. Even in the original enhanceosome, TFs make few protein-protein contacts[41]; instead, many TFs have been shown to cooperate via DNA-mediated cooperativity,[42] in which specific TFs bind adjacently in certain arrangements and spacings to stabilize joint DNA binding.

A major challenge for the field has been to assess the relative extents of these forms of cooperativity (nucleosome-mediated, protein-protein interactions, and DNA-mediated) among natively functioning enhancers and other *cis*-regulatory elements. Analysis of TF motif positions within nucleosome-depleted regions showed that different TFs tend to be at different positions, e.g., with known nucleosome-binding factors enriched at the edges of the nucleosome-depleted area, suggesting that TFs play distinct functional roles.[43] Recent technological advances have enabled improved inference of cooperative interactions without relying on finding enrichment of specific motif arrangements or spacings. Convolutional neural networks and other deep learning models can learn cooperative interactions *de novo*, in an interpretable manner, from training on quantitative TF binding,[44] chromatin accessibility,[45,46] or enhancer reporter assay[7] data.

These studies, which systematically tested the distance dependence of cooperativity *in silico*, suggest that TF cooperative interactions are diverse and common. TFs can prefer proximal binding with certain partners, with cooperativity decaying either sharply or slowly with distance, or instead prefer to have some distance between them.[7] Most cooperativity appears to be constrained to the ~150-bp span of nucleosomes (translational position), sometimes with periodicity at the ~10.5-bp scale of the turns of the DNA double helix (rotational position). In addition to varying cooperative modes, the extent of cooperativity can vary across TFs. TFs with generally reduced cooperativity may be more potent DNA binders or have stronger pioneering activity that enable their binding independent of cooperative partners,[47] while the weakest binders (e.g., Nanog of the often weakly binding homeodomain family) are most reliant on cooperativity.[44] These suggest that while few TFs *must* be in a certain position or orientation to function, diverse forms of TF cooperativity quantitatively modulate binding and function.

Although deep learning approaches are powerful for systematically detecting cooperativity, they are mechanistically agnostic. In some cases, the distance scales suggest certain mechanisms, e.g., ~10.5-bp periodicity suggests that TFs cobind the same face of the DNA double helix, but whether those TFs rely on protein-protein interactions requires further experiments to resolve. While some TFs can form strong protein-protein interactions, such as obligate DNA-independent dimerization among family members of basic-helix-loop-helix (bHLH), basic leucine zipper (bZIP), and activating protein 2

(AP-2) factors,[5] most other interactions are weaker and more challenging to accurately identify. Disordered regions of TFs, which are less conserved[48] and less straightforward to disrupt genetically,[49] are likely to be involved in longer-range cooperativity and more broadly DNA specificity beyond the binding motif.[50] Experiments swapping these unstructured regions between TFs suggest that they can play roles in determining loose cooperativity with other TFs (i.e., enrichments of other TFs near the core motif[47,51]) or genomic domain preference that may restrict which subset of motif instances to bind.[49]

New experimental methods are also facilitating a mechanistic understanding of TF cooperation. Single-molecule footprinting methods[26,52–55] exploit unnatural DNA methylases to determine both nucleosome and TF binding states at single-molecule resolution, which demonstrated that simultaneous co-binding to DNA can be common even for TFs that do not bind immediately adjacent positions.[26,55] Proximity labeling approaches such as BioID[56] offer alternatives to the more traditional immunoprecipitation and can detect more transient interactions.[57]

TF activity is a function of concentration and modifications

For simplicity, our discussion thus far has treated TF binding to DNA as a mostly binary event occurring at a subset of motif matches. However, unlike tRNAs, which overwhelmingly bind their cognate codon and rarely bind others, TF binding is less well defined. In addition to binding high-affinity sites that closely match their motif, TFs also bind many imperfect matches to their cognate motifs, termed low-affinity binding sites.[58] This variation in binding site affinity is thought to quantitatively affect TF occupancy (the percentage of the time the TF is at that site) and dwell time (duration of an individual binding event). Low-affinity sites can still function, either through cooperation among multiple weak binding sites[59,60] or through optimal cooperativity (spacing) with other TFBSs.[61]

In fact, the low affinity of these sites can be important for their proper specificity of function, as making them higher affinity can decrease tissue specificity of the cognate *cis*-regulatory element.[61–63] This observation supports the idea that binding site affinity is a major mechanism by which cells respond differentially to effective TF concentrations, with high-affinity sites bound by TFs even at low concentrations but low-affinity sites requiring higher TF levels. Similarly, post-translational modifications that reduce TF-DNA-binding strength (such as acetylation,[64] phosphorylation,[65] or addition of small ubiquitin-like modifiers (SUMOylation)[66]) can in some cases selectively reduce binding to low-affinity sites, counterintuitively increasing transcriptional activity at high-affinity sites.[64] This quantitative model of TF activity as a function of precise levels and states is supported by recent work experimentally titrating a TF to varying dosages, in which the most dosage-sensitive accessible genomic regions were the highest affinity binding sites.[67]

While individual cases of functional low-affinity sites have been described, it remains challenging to determine how often and in which cases they function. One potential challenge is that disrupting even high-affinity sites individually often results in marginal phenotypes[68]; thus, detecting low-affinity site function may require a whole-organism readout rather than expression levels in a single cell type. For example, systematic mutagenesis of a *Drosophila* embryonic enhancer revealed that most bases within the enhancer—including those outside the strong TF motif matches—caused a detectable phenotype upon mutation.[69] Another challenge is that creating or destroying TFBSs necessarily affects overlapping motifs, which can have stronger effects than the created or destroyed motif.[70] Weak binding sites may individually have small effects and yet collectively play important roles in gene regulation.

Note that while TF binding to DNA is the primary mechanism of *cis* regulation, sequences outside of strong TFBSs may also contribute to transcriptional regulation without directly impacting TF binding affinity. For example, sequence properties can affect intrinsic nucleosome positioning[71] or affect RNA polymerase II (RNA Pol II) elongation rate.[72]

It also remains unclear what level of TF abundance is sufficient for function (and which functions, e.g., binding to low- vs. high-affinity sites). TF expression is most often measured by RNA levels, which are further modulated by translation, protein stability, and sometimes protein localization or modification.[73] One potential scalable approach for identifying relevant TF RNA expression levels is comparing TF levels and motif enrichments in open chromatin across cell types[74]; however, this approach cannot resolve co-expressed paralogous TFs that may bind the same motif,[75] or other interaction partners that can modulate TF conformation and DNA binding.[76,77] Meanwhile, target gene-based inference of TF activity remains limited by the dearth of curated target genes.[78] Ultimately, more precise titrations of TF levels and states will be instructive as to the dependence of TF function on quantitative TF abundance.[67,79]

TF binding to DNA constitutes the first layer of the *cis*-regulatory code, with the TF-to-binding site code defining the key "words" that are arranged into clusters called *cis*-regulatory elements. The next layer of the code determines the transcriptional functions of these elements—the meaning of these "sentences."

FROM TF BINDING TO *CIS*-REGULATORY ELEMENT FUNCTION

TFBS clusters as modular *cis*-regulatory elements

The nucleosomal barrier to TF binding ensures that TF occupancy on the genome is concentrated at clusters of TFBSs where several different TFs converge.[80] These *cis*-regulatory elements are classified based on their position relative to genes (transcription start site-proximal promoters or distal elements) and their effect on transcription (enhancers that upregulate transcription, silencers that downregulate transcription, or insulators that separate regulatory domains). Promoters can be readily identified by sequencing RNA 5′ ends, but identifying distal elements is often less straightforward. Enhancers, the best studied class of distal elements, were originally defined as short sequences that autonomously drive transcription of a promoter independent of distance or orientation.[9] Nowadays, candidate enhancers active in a given cell type can be comprehensively mapped using chromatin marks[81,82] and/or massively parallel reporter assays,[68,83] but systematic genetic perturbations to validate their activity and define contributions of individual enhancers to the target gene expression remain challenging.[84]

Compared with promoters and enhancers, silencers have been much less studied. However, approaches analogous to those used to map enhancers have begun to yield catalogs of silencers. Some groups have used a chromatin mark of facultative repression, histone 3 lysine 27 trimethylation (H3K27me3), combined with 3D interactions with promoters to identify candidate silencers,[85,86] while others used high-throughput reporter assays to test open chromatin regions lacking active chromatin marks for capacity to silence a strong promoter.[87,88] Many silencers appear to function as enhancers in other cell types, at least in *Drosophila*.[89] Each *cis*-regulatory element can serve as a superimposed set of binding sites for different sets of TFs across cellular contexts. Thus, the same element might be bound by repressive TFs and function as a silencer in one cell type and be bound by activating TFs and function as an enhancer in another.

Cis-regulatory elements integrate biological information to achieve spatiotemporal specificity

The organization of TFBSs into *cis*-regulatory elements allows integration of multiple signals to achieve spatiotemporal specificity.[90] For example, an enhancer might be bound by and take input from a lineage-specific TF and a signaling effector, only becoming active in cells with both signals. From a spatiotemporal perspective, each input TF has a pattern of activity in the developing embryo; each *cis*-regulatory element combines these patterns according to the cooperativity or competition among these input TFs to generate its expression pattern. For example, an enhancer requiring three input TFs to be expressed would be active only in the intersection of the input TFs' expression ranges, while an enhancer that is inhibited by a repressor would be restricted to regions where the repressor is absent. These simplified examples imply logic based on binary absence vs. presence of TFs, but these relationships are often quantitative.[91]

Based on the TF-binding code that indicates which TFs regulate each *cis*-regulatory element, can we predict enhancer spatiotemporal specificities?[80,92] Or conversely, can we design synthetic enhancers with a desired specificity?[93] For some cell types, combinatorial binding of a few highly specifically expressed TFs is sufficient to produce cell-type-specific expression. However, predicting activity patterns from arbitrary DNA sequences[92,93] remains more challenging than classifying genomic enhancers,[80] and the difficulty of prediction depends on the cell type.[92]

TFs recruit cofactors to *cis*-regulatory elements

TFs do not modulate transcription directly but instead serve as adapters that bring coactivators (CoAs) or corepressors (CoRs) (often jointly referred to as cofactors) to their binding sites. How do these cofactors in turn modulate transcription? This topic is too extensive to cover in full and instead, we direct the reader to excellent reviews.[94–100] Here, we briefly outline major classes of cofactors and our current understanding of their functions.

Cofactors can be broadly grouped into three major and likely universal classes (1) the Mediator complex,[94] (2) nucleosome remodelers,[97] and (3) histone modifiers,[101] plus other scaffold or adapter proteins that bridge TFs and cofactors.[102] The major cofactors are typically organized in large multi-subunit complexes, some with variable composition, but have key catalytic subunits. A combination of cryoelectron microscopy,[103,104] conditional protein depletion technologies,[105] and catalytic inhibitors have revealed new insights into how these complexes function.

The Mediator complex plays a central role in phosphorylating and activating the pre-initiation complex at most promoters,[94] while nucleosome remodelers actively maintain chromatin accessibility needed for promoter and distal enhancer activity.[106–108] Furthermore, although these core CoAs are needed for most transcription, acute depletion can exhibit surprisingly selective consequences due to feedback mechanisms.[109] For example, Mediator depletion leads to loss of transcription primarily at cell-type-specific genes due to a compensatory increase in RNA Pol II pause release at other genes.[109] By contrast, histone modifiers have competing activating and repressing roles, with CoAs depositing modifications enriched at enhancers such as H3K4 monomethylation (H3K4me1) and H3K27 acetylation (H3K27ac), and CoRs either removing those activating marks or adding repressive marks such as H3K27me3 and H3K9me3. At least in some biological contexts CoAs such as CREB binding protein (CBP) and p300, which deposit H3K27ac, are limited in abundance compared with CoRs such as histone deacetylases that remove H3K27ac.[110] This observation is consistent with the rapid genome-wide changes in H3K27ac and transcription upon acute chemical inhibition of p300/CBP catalytic activity.[111] It is also important to note that while these modifications can modulate transcription per se,[112] they often represent only a part of these cofactors' function, with a growing number of studies documenting major non-enzymatic roles of the histone-modifying CoA proteins,[101,113–115] including recruiting RNA Pol II to enhancers[114] and promoting RNA Pol II pausing at Polycomb response elements.[113] Histone modifications likely serve to stabilize the relatively weak recruitment of CoAs, many of which bind the same histone marks they deposit.[116]

Each TF can recruit multiple cofactors, sometimes even both CoAs and CoRs.[117] The TF residues involved in cofactor recruitment are often unstructured (though they can form structures upon binding),[118] and the interactions between TFs and their cofactors can be relatively weak. Furthermore, cofactors can recruit additional cofactors either directly[119] or via modifications to histone tails that modulate cofactor binding.[116] Despite these complexities, recent work has begun to make substantial progress toward dissecting TF-cofactor recruitment, largely by measuring transcriptional consequences of TF and cofactor recruitment. The core assay is the activator bypass assay,[120] in which a reporter gene is controlled by a minimal promoter with binding sites for an orthogonal DBD that in turn recruits a fused TF, cofactor, or fragments of thereof, followed by measurements of reporter activity. High-throughput variants of this assay have been used to identify minimal fragments of TFs that are necessary and sufficient for activating or repressing transcription, called activation or repression domains, respectively.[121,122]

These studies have illustrated how TFs and cofactors not only differ in the quantitative strength of activation or repression,[122] but also how their transcriptional output depends on context—i.e., presence or absence of other TFs or cofactors within the *cis*-regulatory element. TFs clustered with different cofactors in

their regulatory activity at diverse genomic contexts (i.e., DNA sequence flanking the TF/cofactor domain recruitment sites),[123] suggesting that this distinct context-dependence is mediated by each TF recruiting specific cofactors, combined with differential "need" for certain limiting CoAs in each genomic context.[124] Consistent with this model, proximity labeling of TFs and CoAs confirmed that different TFs can preferentially associate with different families of CoAs.[122] While such cofactor dependence has been thus far assayed in the reporter context, it could be potentially even more pronounced in the native genomic context—for example, certain CoA combinations may be required for enhancer function over long distances or in different chromatin environments, and they could also mediate preferential enhancer-promoter contacts, as discussed below.

Dissecting the cofactor recruitment code

If TFs each recruit distinct sets of cofactors, we would expect each cofactor to be recruited by a different class of activation or repression domain with distinguishable features. However, until recently, only a small fraction of TFs had any well-defined activation or repression domains, precluding efforts to glean generalizable rules or subclassifications into different mechanistic classes.[125]

Recent studies have vastly expanded the lexicon of activation and repression domains,[117,121,122] with now a majority of TFs and cofactors having at least one annotated effector domain, but surprisingly, activation domains lack obvious subclasses. Although activation domains have previously been noted for their different amino acid compositions, such as glutamine-rich or proline-rich domains, these enriched amino acids often are not the critical residues needed for domain function. Instead, the presence of hydrophobic residues and their interspersed position among certain other residues appears to be the main feature driving activation domain function.[117,118]

How can we reconcile the contrast between functional inference of distinct TF:CoA recruitment and the scarcity of well-defined, functionally diverse subsets of activation domains? For one, the aforementioned recruitment screens, which typically measure the effect of a single candidate regulatory domain on the reporter activity, may miss activation domains that are too weak or dependent on other CoAs to function in isolation. In addition, the tested domains were fragments of up to 80 amino acids, so if recruitment of different cofactors is mediated by multiple distributed domains, these domains would not have been detected. Consistent with this idea, weak activation domains tend to synergize and activate transcription more strongly when fused together.[126]

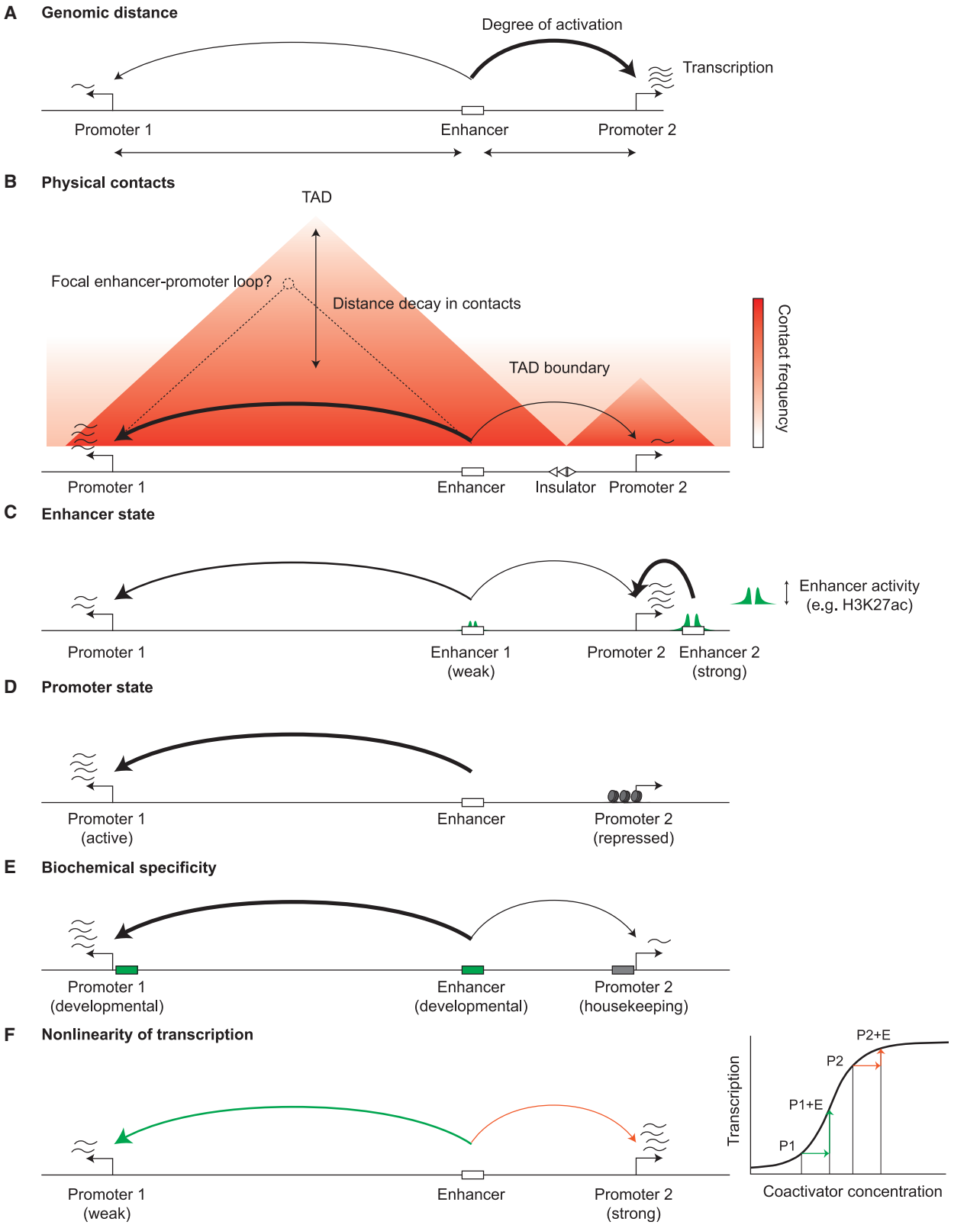
In contrast to activation domains, known repression domains appear to be more diverse. Short linear motifs known to recruit different families of CoRs are critical for many repressor domains, as are SUMOylation sites, SUMO-binding domains, and various structured domains including several DNA binding (zinc finger, homeodomain, and bHLH) or protein interaction domains.[117] Notably, some TFs may be repressive only through recruiting other repressive TFs that in turn recruit CoRs, suggesting how TFs may alter their transcriptional function depending on the cellular context (e.g., repressive TF partner abundance).

Determining total functional output of TFs and *cis*-regulatory elements

As we compile maps of cofactor recruitment domains within TFs, these maps should also enable quantitative and context-specific predictions of the overall functional output of a TF, and an understanding of how multiple TFs collectively shape *cis*-regulatory element function. How do the different effector domains in a TF "add up"? With increasing numbers of either activation or repression domains present within a given TF, one might expect a wider array of CoAs or CoRs, respectively, recruited by this TF, and thus higher activity and less context-dependent function. Recruitment screens using whole open reading frames suggests wide variation in activation strength[122]; whether this is due to greater numbers and diversity of activation domains remains to be seen. Furthermore, given the relatively weak nature of cofactor recruitment, multivalency is thought to mediate phase separation of transcriptional regulators[127] that leads to more robust activation.[128] Formation of macromolecular condensates may provide an additional layer of regulatory specificity, whereby cofactors coalesce in a manner that is dependent on multivalent interactions among specific CoAs and can be further regulated by post-translational modifications or RNAs (recently reviewed in Sabari,[129] Sharp et al.,[130] Strom and Brangwynne,[131] and Lyon et al.[129–132]).

When a TF contains both activation and repression domains, does one dominate, and under what conditions? In yeast, several TFs are known to contain both activation and repression domains, with intramolecular interactions masking one of the two depending on the presence of ligands.[133,134] In rare cases, the same short amino acid sequence can function as both an activation domain and a repression domain even at the same promoter.[117] A recent combinatorial recruitment screen found that in human cells, fusions of repressor domains with activator or dual-function domains tend to function as repressors,[126] potentially indicating a generally dominant role of repressor domains. However, further studies will be needed to resolve potential regulatory domains that might toggle activator vs. repressor function across genomic or cellular contexts.

Currently, it remains unknown how variable TF function is across cell types and binding sites. Many TFs have been reported to be both activating and repressing in different cellular or genomic contexts, but many of these observations may be indirect effects of TF loss.[105] Mutations of binding sites, e.g., in reporter assays, can resolve site-specific functions but may be confounded by alteration of binding sites for other TFs.[70] The number and affinity of motifs and presence of certain other TFs nearby have been implicated in variation in TF function across sites,[135] but the mechanism remains unclear. One could envision that a TF has both activation and repression domains of differing affinities, such that its increased occupancy at a *cis*-regulatory element leads to saturation of the stronger recruitment domain and allows the weaker domain to catch up. Another possibility is that post-translational modifications can modify both TF-binding strength and interaction with cofactors[136] such that TF function varies with binding site affinity. Or, in the "limiting CoA" model, increasing recruitment of one CoA might increase transcription up to a point, but once it is no longer limiting, it might crowd out other needed CoAs.



(legend on next page)

TF function may also be modulated by interactions among TFs at the same *cis*-regulatory element. While cooperative binding to DNA is the best characterized of such interactions, other mechanisms may exist. For example, TF-mediated repression could arise through one TF outcompeting and preventing binding by other more strongly activating TFs that share overlapping sites.[69] TFs binding the same *cis*-regulatory element could also impact each other by working together to more efficiently recruit the same CoAs. Yet, the diversity of TFs binding a *cis*-regulatory element has been proposed to be correlated with activation strength,[137] suggesting that TFs may collaborate by recruiting complementary CoAs. Such interactions might underlie parts of *cis*-regulatory grammar that cannot be explained by DNA-binding cooperativity.[7] It remains an important task for the field to decipher TF function in their cellular and genomic contexts (e.g., to understand the consequences of noncoding genetic variation), and a mechanistic framework of TF-cofactor recruitment would add useful insights toward this goal.

FROM *CIS*-REGULATORY ELEMENTS TO PROMOTER TRANSCRIPTION

While *cis*-regulatory elements such as enhancers are sequence- and cellular-context-dependent modular "units" of transcriptional regulation, the ultimate output of the *cis*-regulatory code is gene transcription. In complex genomes of multicellular organisms, enhancers often function at distances of tens or hundreds of kilobases away from their regulated promoters; in humans, there are known examples of enhancers regulating genes over a megabase away.[11,138,139] Such long-range function can result in regulatory domains—genomic regions where genes (or inserted reporter genes) have similar expression patterns.[140] Yet, in other cases, enhancers appear to exhibit promoter specificity, skipping nearby genes to regulate more distal genes.[138] Understanding how genomic distance, contact frequency between enhancers and promoters, biochemical compatibility, and chromatin state influence gene expression is another regulatory layer that has to be unpacked in the *cis*-regulatory code.

Generating a map of all the connections between the hundreds of thousands of enhancers and tens of thousands of promoters is a massive challenge. Recent efforts to systematically and experimentally determine these connections by individually perturbing enhancers with CRISPR interference (CRISPRi) and measuring changes in transcription have yielded on the order of hundreds of connections.[141,142] Given how enhancer-promoter linkages serve as a critical link for interpreting how chromatin or genetic changes affect transcription and downstream phenotypes, various heuristics for approximating these linkages have been used to improve predictions in the absence of experimentally validated connections.

Genomic distance

The simplest heuristic for associating enhancers with promoters is the linear genomic distance between the enhancer and promoter (Figure 2A). In such an approach, enhancers are assumed either (1) to regulate the closest gene promoter, or (2) to regulate all genes within a distance limit, such as 100 kb. Despite the simplicity of this approach, it is remarkably accurate—functionally validated enhancer-promoter connections are greatly enriched at the closest distances, with substantial but gradual fall-off by 100 kb.[142] Similarly, highly transcribed randomly integrated promoter constructs were enriched near enhancers.[143] A more recent study tested random local insertions of a strong enhancer around a reporter gene and found that insertions closest to the promoter drove the strongest activation,[144] highlighting the quantitative dependence of enhancer-mediated transcriptional activation on genomic distance.

Physical contacts

Nevertheless, genomic distance fails to explain how some enhancers skip genes to regulate only more distal promoters, and how some enhancers can function at very long range while others do not. Upon the discovery of sequence-specific 3D organization of the genome, physical contacts between enhancers and promoters were proposed as a mechanism for long-range enhancer action[145] (Figure 2B). Two key observations and associated hypotheses emerged. First, most enhancer-promoter connections appeared to lie within the same topologically associating domain (TAD), regions of self-interaction with insulation from neighboring regions,[146,147] suggesting that TAD boundaries might prevent inappropriate enhancer-promoter activation and that within a TAD all enhancers can activate all promoters. This notion was inspired by correlated expression of genes within the same TAD[146] and case studies that highlighted mis-regulation resulting from loss of domain boundaries.[146,148] Indeed, TADs often correspond to regions in which integrated reporter genes exhibit similar expression patterns.[140] However, this heuristic fails to account for (1) the somewhat arbitrary definition of TADs, which often contain multiple sub-TADs, and (2) the overwhelming role of genomic distance in contact probability, in which very closely located regions inevitably form frequent contacts (simply through polymer physics), at much higher rates

Figure 2. Potential mechanisms underlying enhancer-promoter specificity

(A) Enhancers tend to preferentially activate promoters at closer genomic distances.

(B) Enhancers typically activate promoters within the same topologically associating domain (TAD), though not exclusively and not necessarily to equal extents. This preference has been ascribed to the increased frequency of physical contacts of genomic regions within, as compared with between, TADs. In rare cases, focal enhancer-promoter loops are also observed on the contact frequency maps, especially at promoters and enhancers overlapping a CTCF binding site.

(C) Differences in relative enhancer activity levels (that can be estimated by quantitative levels of H3K27ac or other enhancer chromatin features) can create apparent promoter specificity as an enhancer must contribute a significant fraction of the total activation at a promoter to detectably regulate it.

(D) Promoters can be in repressed states unresponsive to enhancer activation.

(E) Promoters and enhancers can be grouped into classes (such as developmental and housekeeping) with biochemical specificity (albeit quantitative) for each other, resulting in preferential activation.

(F) The nonlinearity of transcription as a function of coactivator concentration can create apparent specificity, where weak promoters are more responsive to activation by enhancers.

than more distal regions within (or outside) of a TAD. Consistent with this, global loss of the main TAD boundary factor, CCCTC-binding factor (CTCF), leads to relatively minor changes in transcription, mostly at promoters with CTCF binding.[149,150]

Second, at least some enhancers and promoters appeared to form frequent loops, suggesting that identifying these loops might be a scalable approach to identify enhancer-promoter functional connections. Various modified versions of the chromosome conformation capture assay (capture-C,[151] promoter-capture Hi-C,[152] HiChIP,[153] and micro-C[154,155]) have been used to prioritize candidate enhancer-promoter connections but face two challenges. First, genome-wide assays such as micro-C require large amounts of sequencing to achieve the resolution to resolve individual enhancers, while targeted approaches can be challenging to normalize and distinguish focally enriched interactions. Second, due to the high background of random interactions between nearby genomic regions, these approaches are often unable to detect enhancers at <~20 kb where many enhancers lie. However, the common focus on statistically enriched interactions may be unnecessary, as apparently "random" interactions may still be functional and as a whole line up with the general decay in enhancer function with increasing distance.[144] This suggests that a diversity of mechanisms, including CTCF/cohesin-mediated loop extrusion, TF oligomerization (such as YY1[156]), Polycomb-complex-mediated clustering,[157,158] sharing the same TAD, and even simple genomic proximity all generally serve to modulate largely equivalent physical contacts.

Enhancer state

A seemingly obvious element of enhancer-promoter connectivity is that the enhancer must be in an active state to regulate any promoters. Accounting for the relative activity levels of enhancers, together with genomic distance or contact probability, can predict enhancer-promoter connections reasonably well. The activity-by-contact (ABC) model uses a simple product of enhancer activity (a geometric mean of chromatin accessibility and active enhancer mark abundance) and contact probability with each promoter to estimate each enhancer's contribution to that promoter's transcription level.[141] This mirrors the experimental setup of most functional validation, where a single enhancer is deleted or perturbed, and the resulting change in transcription must be large enough to be detectable. Thus, a promoter surrounded by many strong enhancers may not be significantly regulated by a weaker and more distal enhancer, even though that same enhancer may be the strongest and closest enhancer for another promoter lacking other enhancers (Figure 2C). Notably, the model performed nearly as well using a power law scaling of the genomic distance in place of measured contact frequencies, in line with the role of specific 3D genome conformations primarily in a subset of more distal enhancer-promoter connections.

Promoter state

Similarly, in order to be activated by any enhancers, a promoter must be in a responsive state (Figure 2D). While the idea that some promoters may be unresponsive to enhancers in certain cell states is intuitive (and was effectively incorporated into the ABC model,[141] which only considers actively transcribed gene promoters), we lack a framework for predicting which promoters are responsive and which are not—which will likely need to incorporate chromatin state as well as absolute transcription levels. Mechanistically, a promoter might be in a repressed chromatin state (e.g., H3K9me2/3-marked, minimal chromatin accessibility, and CpG methylated) that the CoAs recruited by the enhancer cannot overcome. For example, tissue-specific CpG methylation of the *Zfp42* promoter renders it unresponsive to the *Fat1* enhancer.[159] In another example, the *Hoxa* cluster requires promoter retinoic acid receptor elements (RAREs) for basal transcription levels that the enhancers then amplify; without the RAREs, there is minimal activation of *Hoxa* genes.[160]

Biochemical compatibility

Biochemical specificity or compatibility between enhancers and promoters has been proposed as another explanation for enhancers only activating certain promoters, especially for promoter-skipping (Figure 2E). Indeed, in *Drosophila*, both core promoters and enhancers can be classified as "housekeeping" or "developmental,"[108,123,161–163] with different sets of sequence motifs, cofactor dependencies, and relative positions. Housekeeping enhancers are in direct proximity to their promoters, and perhaps are not enhancers per se but should be simply considered as akin to upstream regulatory sequences (UASs) in yeast.[164] Regardless, a similar division is also believed to exist in mammals, but recent studies that combinatorically tested many enhancers and promoters in an episomal reporter assay concluded that while these groups exist, they mostly exhibit more minor quantitative differences in activation strength compared with the effect of intrinsic promoter and enhancer strength.[165–167] We note that the *Drosophila* studies were done with minimal core promoters, while the mammalian studies were done with larger promoters that may include proximal enhancers or UAS—which more closely resembles the genomic context but may be confounded by synergy/redundancy between the enhancers.

Nonlinearity of transcription

While enhancer-promoter specificity is often considered in a binary manner (an enhancer does or does not regulate a given promoter), transcription is quantitative. By default, enhancers and promoters are assumed to behave linearly (often additive in log scale) but in reality, the process of transcription is nonlinear—which may create the illusion of enhancer-promoter specificity (Figure 2F). From the perspective of an enhancer, strong promoters may be at a saturated level of transcription and therefore be unresponsive to gain or loss of a weaker enhancer, while inactive promoters may require sufficiently strong enhancers for increased transcription to be detectable.

This nonlinearity manifests in the quantitative relationship between contact frequency and transcription[144] and might be explained by a probabilistic and multi-step process.[168] Consistently, imaging studies tracking both enhancer and promoter loci plus nascent transcription showed that contacts are perhaps weakly enriched preceding transcriptional bursts,[169,170] but contacts (as measurable by microscopy) clearly have an imperfect correspondence to transcription,[171] suggesting a time delay

and probabilistic function.[168] That is, an enhancer-promoter contact may increase the probability of a cascade of events culminating in transcription[172]—but exactly what these events are remains unclear. One intriguing proposed mechanism is that enhancer recruitment of p300/CBP leads to transient acetylation and activation of TFs that can diffuse to and activate transcription at physically proximal promoters.[173] Further, the presence of multiple enhancers and promoters may contribute to additional nonlinearity, as we discuss below.

These explanations of enhancer-promoter specificity are not mutually exclusive. Instead, an integrated quantitative model (similar to the ABC model) accounting for all these features, together with more extensive quantitative measurements of enhancer-promoter relationships in diverse endogenous and synthetic contexts, should lead us toward a comprehensive enhancer-promoter code.

FROM REGULATORY CONNECTIONS TO REGULATORY LANDSCAPES

Thus far, we have focused on individual enhancer-promoter relationships, but in the genome, many genes have multiple, simultaneously active enhancers and promoters nearby that can cooperate or compete with one another. Furthermore, what constitutes an individual enhancer rather than parts of a larger enhancer? While short sequences (even <200 bp) can exhibit reporter activity,[174] histone marks associated with enhancers can stretch for 1 kb or more.[175] Similarly, as discussed above, promoters contain core promoters and UAS-like proximal regulatory sequences. Zooming out further, these enhancers and promoters function in a broader genomic context, with 3D localization at the nuclear lamina,[176] nucleoli, or nuclear speckles.[177]

Diversity of interaction modes

One common operational framework for interpreting this complexity is epistasis. Epistasis is defined by deviation from a null expectation of independence, in which perturbing one enhancer should have the same effect regardless of the presence of other *cis*-regulatory elements (Figure 3). Redundancy occurs when loss of only one of a pair of enhancers has little effect as each enhancer can confer most of the transcriptional output, whereas their combined loss dramatically reduces transcription. Conversely, synergy occurs when all such enhancers are required for full output and each individual enhancer has much less activity on its own. However, the operational definition of epistasis has ambiguities—should the expectation be additive or multiplicative? On what scale (e.g., allele-specific expression vs. total expression)? This definition also focuses on a single cellular context—but the same pair of developmental enhancers might interact in different ways across cell types (Figure 3B).

Numerous studies of interactions among enhancers, parts of enhancers, and promoters have largely failed to find universal rules. Enhancers that possess similar tissue-specific activity and regulate the same promoter are a widespread feature of metazoan genomes.[178–180] These so-called "shadow enhancers" are thought to be largely redundant. Yet, shadow enhancers are often conserved evolutionarily, suggesting that in addition to enabling *cis*-regulatory evolution by buffering the consequences of genetic variation, their overlapping activity serves important functions. Possibilities include buffering against stressful conditions[181] or transcriptional noise,[182] or generating more precise expression patterns.[91] Notably, these sharper expression patterns can involve regions where an individual enhancer activates transcription but not the shadow enhancers together, due to repressor activity.[91] In addition, the prevalence of enhancer redundancy remains debated; a recent study of enhancer epistasis found a mix of additivity and synergy, rather than prevalent redundancy.[183]

Interactions among dense groups or clusters of enhancers that lie within tens of kilobases of DNA, called super-enhancers,[184,185] stretch enhancers,[175] or enhancer clusters,[186] have garnered particular attention, even as some studies have questioned the distinction.[187] These clusters, which are typically defined by high levels of H3K27ac, Mediator binding, or chromatin accessibility within extended genomic windows, have been postulated to exhibit unique properties that distinguish them from "typical" enhancers, such as frequent 3D interactions[188] that seed or associate with phase-separated condensates of CoAs and transcriptional machinery that make them particularly potent.[189] However, whether this implies redundancy or synergy is unclear, and both have been observed[190] in addition to additivity.[187]

How do promoters interact with one another in genomic context? Do they compete for enhancers,[191] or can multiple promoters be simultaneously activated by a single enhancer?[192] Or do promoters serve as distal enhancers for other promoters?[193] All these scenarios have been observed, indicating that there is no single rule for promoter interactions, but it remains unclear when and where each condition applies. These conditions need not be mutually exclusive—for example, even when an enhancer can activate multiple promoters simultaneously, it may activate a single promoter more strongly (thus demonstrating competition).

Mechanisms underlying epistasis

Despite the caveats of interpreting epistatic interactions, epistasis reflects multi-step biological pathways and can help suggest mechanisms. For example, one pervasive observation is that within enhancers and enhancer clusters, a small portion or subset commonly drives the majority of transcription,[194–196] to the point that many regions appear individually dispensable or insufficient for autonomous enhancer activity, even though combined loss with other enhancers has measurable effects. This has often been used to invoke different classes or hierarchies of enhancers, including predominant vs. supporting sites, or "amplifier" enhancers[194,197–199] that differ in TF motifs or chromatin state features.

Given the diversity of observed epistatic interactions, is there a unifying mechanistic explanation? One potentially general principle is that different classes of enhancers may recruit different sets of CoAs that modulate different steps of the transcriptional process, such as RNA Pol II recruitment vs. phosphorylation. Thus, enhancers that only regulate non-limiting steps in the endogenous context may have little phenotype upon individual loss but may become more important when those steps become limiting. Biochemical diversity could also underlie promoter interactions, e.g., competing promoters may be limited by the same CoA.

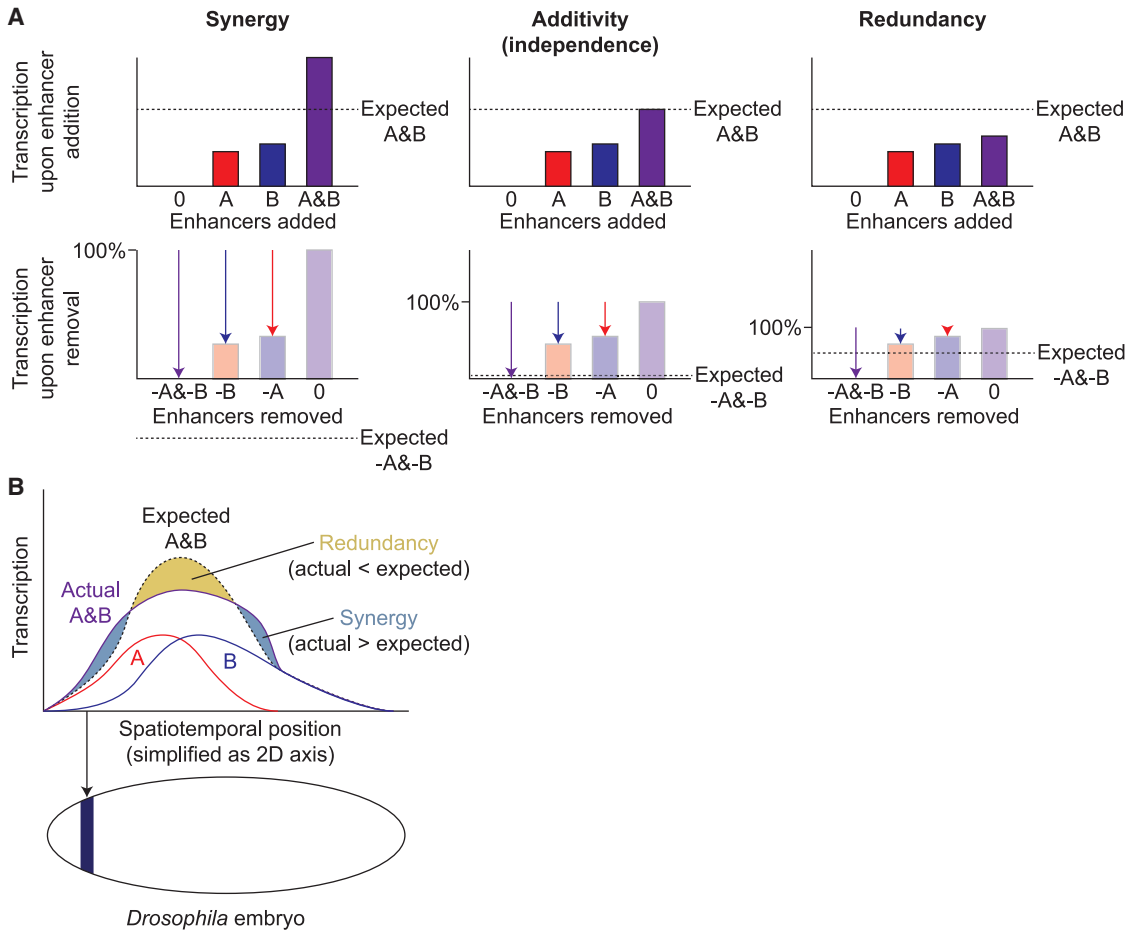


Figure 3. Enhancer epistasis

(A) Examples of synergy, additivity, and redundancy between enhancers A and B, shown both from the perspective of enhancer addition (top) and enhancer removal (bottom) for the same examples. Dashed horizontal lines indicate the expected transcriptional output of adding both enhancers A and B (top) or of removing both enhancers (bottom). Bars indicate transcriptional output upon addition of enhancers while arrows indicate the change upon removal of enhancers.
(B) Illustration of how a single pair of enhancers, A and B, can exhibit multiple types of epistatic interaction across spatiotemporal positions (e.g., *Drosophila* embryonic anterior-posterior position as shown below x axis, with dark blue color indicating one such position), simplified here as 2D axis.

Physical contacts can also mediate enhancer epistasis, but the relationship is complicated. TFs and CoAs recruited to one enhancer can spread to other enhancers in spatial proximity[199]; this spreading may allow more "dominant" enhancers to further activate nearby supporting enhancers. A recent study proposed that enhancers that are very distally located and infrequently contact each other may buffer each other's loss.[12] Yet, enhancers that form 3D hubs, including super-enhancers, have also been proposed to seed condensates or hubs that confer robust transcription of cell type identity genes.[189]

Impact of chromosome organization

In addition to specific interactions among nearby enhancers and promoters, the function of insulators and tethering elements, together with the broader organization of the domain within the nucleus, can also impact transcription.

Insulator elements—usually binding sites for the TF CTCF,[200] which can form a directional barrier for loop extrusion by cohesin—serve to insulate adjacent TADs or loop domains. As with enhancers and silencers, strong boundary elements consist of clusters of TFBSs.[201] These boundaries can impact transcription across broad domains by preventing spreading of heterochromatin.[202] Furthermore, the process of loop extrusion by cohesin shapes transcription beyond creating TAD boundaries, by increasing contacts between distal enhancers and promoters within the same TAD,[203] particularly for promoters with nearby CTCF sites,[150] or by broadly increasing contacts across the domain via promoting formation of multi-loop structures.[204] In addition to insulator elements, recently described "tethering elements" can function independently of CTCF and loop extrusion not only to link-specific enhancers and promoters[205] but also to facilitate temporally coordinated transcription.[206]

Zooming out further, chromosomes are organized into two major types of "compartments"—A compartments that tend to be enriched for active genes and B compartments that tend to contain silenced heterochromatin (each of which can be subcategorized further). These compartments tend to self-associate within their own type. The compartment-level organization tends to be broadly

similar across cell types, but with shifted boundary locations correlated with changes in chromatin state,[207] suggesting it is an emergent property shaped by chromatin state, transcription, and loop extrusion.[208] Some of these silenced regions tend to localize at the nuclear periphery,[176] while clusters of actively transcribed genes associate with nuclear speckles,[209] nuclear regions enriched in splicing factors. While the causality of these associations are tricky to untangle, experiments targeting genomic regions to nuclear bodies have demonstrated that nuclear environment can impact transcription.[210] The interplay between individual elements being influenced by domain-level effects vs. shaping the organization of the domain remains unclear, but it is important to note that these domains can be heterogeneous and dynamic.[176]

OUTLOOK

Deciphering the *cis*-regulatory code is a monumental challenge. Unlike the genetic code, which had a satisfyingly universal, modular, and deterministic solution, the *cis*-regulatory code is elusively context-specific, complex, combinatorial, and replete with redundancies and quantitative effects. It also consists of several interdependent tiers, starting from TFs binding DNA, to TFs interacting with one another and cofactors at *cis*-regulatory elements, to complex regulatory domains and landscapes. Ultimately, understanding *cis* regulation will require understanding how the thousands of proteins involved in transcription—a substantial fraction of the proteome—work with each other and much of the genome. As Sydney Brenner said of the challenge of deciphering how embryonic development is encoded in the genome, "what is going to be difficult is the immense amount of detail that will have to be subsumed."[211]

Nevertheless, our rapidly advancing tools are enabling experiments and interpretation at unprecedented scales and resolution. A complete catalog of cell types and their transcriptional profiles is within reach, with other epigenomic features to follow. We can test tens to hundreds of thousands of *cis*-regulatory elements or transcriptional regulators in reporter assays.[83,117] Even some of the largest protein complexes involved in transcription can be now structurally resolved, yielding unparalleled mechanistic insights.[103,104] Novel deep learning approaches are detecting previously unseen patterns in genomic data.[212,213] New tools for high-throughput super-resolution chromatin tracing and live-imaging of regulatory elements and transcripts provide a glimpse into genomic topology and kinetics underlying transcription at single-cell and single-allele level.[214–216] Our key remaining challenges are (1) to complement our currently scalable nucleic acid assays with improved technologies for measuring levels, localization, and dynamics of proteins at high resolution and throughput, (2) to integrate our catalogs of multi-scale measurements into a quantitative and mechanistic (interpretable) model of how transcription works at each of the regulatory levels discussed here, and (3) accurately predict gene expression patterns from DNA sequence under defined protein concentrations. By addressing these challenges, we may finally approach a complete *cis*-regulatory code and be able to readily interpret noncoding genomic sequences in the context of evolution, development, and disease.

ACKNOWLEDGMENTS

We thank K. Xue, L. Ichino, and L.-F. Chen for their comments and discussion. This work was supported by HHMI-Damon Runyon Cancer Research Foundation Fellowship (DRG-2420-21) to S.K. as well as NIH grant R35 GM131757, funding from the Howard Hughes Medical Institute, a Lorry Lokey endowed professorship, and a Stinehart Reed award to J.W.

AUTHOR CONTRIBUTIONS

S.K. and J.W. conceptualized the manuscript. S.K. wrote the first draft, and J.W. and S.K. edited subsequent versions.

DECLARATION OF INTERESTS

J.W. is paid scientific advisory board member at Camp4 and Paratus Sciences. J.W. is an advisory board member at Cell Press journals, including *Cell*, *Molecular Cell*, and *Developmental Cell*.

REFERENCES

1. Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561–563. https://doi.org/10.1038/227561a0.

2. Koonin, E.V., and Novozhilov, A.S. (2017). Origin and evolution of the universal genetic code. Annu. Rev. Genet. *51*, 45–62. https://doi.org/10.1146/annurev-genet-120116-024713.

3. King, M.C., and Wilson, A.C. (1975). Evolution at two levels in humans and chimpanzees. Science *188*, 107–116. https://doi.org/10.1126/science.1090005.

4. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195. https://doi.org/10.1126/science.1222794.

5. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The human transcription factors. Cell *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.

6. Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science *356*, eaaj2239. https://doi.org/10.1126/science.aaj2239.

7. de Almeida, B.P., Reiter, F., Pagani, M., and Stark, A. (2022). DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of synthetic enhancers. Nat. Genet. *54*, 613–624. https://doi.org/10.1038/s41588-022-01048-5.

8. Small, S., Blair, A., and Levine, M. (1992). Regulation of even-skipped stripe 2 in the Drosophila embryo. EMBO J. *11*, 4047–4057. https://doi.org/10.1002/j.1460-2075.1992.tb05498.x.

9. Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308. https://doi.org/10.1016/0092-8674(81)90413-X.

10. Crocker, J., and Stern, D.L. (2017). Functional regulatory evolution outside of the minimal even-skipped stripe 2 enhancer. Development *144*, 3095–3101. https://doi.org/10.1242/dev.149427.

11. Long, H.K., Osterwalder, M., Welsh, I.C., Hansen, K., Davies, J.O.J., Liu, Y.E., Koska, M., Adams, A.T., Aho, R., Arora, N., et al. (2020). Loss of extreme long-range enhancers in human neural crest drives a craniofacial disorder. Cell Stem Cell *27*, 765–783.e14. https://doi.org/10.1016/j.stem.2020.09.001.

12. Lin, X., Liu, Y., Liu, S., Zhu, X., Wu, L., Zhu, Y., Zhao, D., Xu, X., Chemparathy, A., Wang, H., et al. (2022). Nested epistasis enhancer networks for robust genome regulation. Science *0*, eabk3512. https://doi.org/10.1126/science.abk3512.

13. Monahan, K., Horta, A., and Lomvardas, S. (2019). LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. Nature *565*, 448–453. https://doi.org/10.1038/s41586-018-0845-0.

14. Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186–192. https://doi.org/10.1038/nature08449.

15. Rodriguez, J., and Larson, D.R. (2020). Transcription in living cells: molecular mechanisms of bursting. Annu. Rev. Biochem. *89*, 189–212. https://doi.org/10.1146/annurev-biochem-011520-105250.

16. Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suvà, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature *529*, 110–114. https://doi.org/10.1038/nature16490.

17. Cho, S.W., Xu, J., Sun, R., Mumbach, M.R., Carter, A.C., Chen, Y.G., Yost, K.E., Kim, J., He, J., Nevins, S.A., et al. (2018). Promoter of lncRNA gene PVT1 is a tumor-suppressor DNA boundary element. Cell *173*, 1398–1412.e22. https://doi.org/10.1016/j.cell.2018.03.068.

18. Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T.W., Greven, M.C., Pierce, B.G., Dong, X., Kundaje, A., Cheng, Y., et al. (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res. *22*, 1798–1812. https://doi.org/10.1101/gr.139105.112.

19. Gordân, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R., and Bulyk, M.L. (2013). Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. Cell Rep. *3*, 1093–1104. https://doi.org/10.1016/j.celrep.2013.03.014.

20. Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., and Honig, B. (2009). The role of DNA shape in protein–DNA recognition. Nature *461*, 1248–1253. https://doi.org/10.1038/nature08473.

21. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82. https://doi.org/10.1038/nature11232.

22. Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311–322. https://doi.org/10.1016/j.cell.2007.12.014.

23. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218. https://doi.org/10.1038/nmeth.2688.

24. Zaret, K.S. (2020). Pioneer transcription factors initiating gene network changes. Annu. Rev. Genet. *54*, 367–385. https://doi.org/10.1146/annurev-genet-030220-015007.

25. Mirny, L.A. (2010). Nucleosome-mediated cooperativity between transcription factors. Proc. Natl. Acad. Sci. USA *107*, 22534–22539. https://doi.org/10.1073/pnas.0913805107.

26. Sönmezer, C., Kleinendorst, R., Imanci, D., Barzaghi, G., Villacorta, L., Schübeler, D., Benes, V., Molina, N., and Krebs, A.R. (2021). Molecular co-occupancy identifies transcription factor binding cooperativity in vivo. Mol. Cell *81*, 255–267.e6. https://doi.org/10.1016/j.molcel.2020.11.015.

27. Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell *9*, 279–289. https://doi.org/10.1016/S1097-2765(02)00459-8.

28. Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241. https://doi.org/10.1101/gad.176826.111.

29. Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676. https://doi.org/10.1016/j.cell.2006.07.024.

30. Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell *161*, 555–568. https://doi.org/10.1016/j.cell.2015.03.017.

31. Fernandez Garcia, M., Moore, C.D., Schulz, K.N., Alberto, O., Donague, G., Harrison, M.M., Zhu, H., and Zaret, K.S. (2019). Structural features of transcription factors associating with nucleosome binding. Mol. Cell *75*, 921–932.e6. https://doi.org/10.1016/j.molcel.2019.06.009.

32. Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S.O., Nitta, K.R., Morgunova, E., Taipale, M., et al. (2018). The interaction landscape between transcription factors and the nucleosome. Nature *562*, 76–81. https://doi.org/10.1038/s41586-018-0549-5.

33. Desanlis, I., Kherdjemil, Y., Mayran, A., Bouklouch, Y., Gentile, C., Sheth, R., Zeller, R., Drouin, J., and Kmita, M. (2020). HOX13-dependent chromatin accessibility underlies the transition towards the digit development program. Nat. Commun. *11*, 2491. https://doi.org/10.1038/s41467-020-16317-2.

34. Michael, A.K., Grand, R.S., Isbel, L., Cavadini, S., Kozicka, Z., Kempf, G., Bunker, R.D., Schenk, A.D., Graff-Meyer, A., Pathare, G.R., et al. (2020). Mechanisms of OCT4-SOX2 motif readout on nucleosomes. Science *368*, 1460–1465. https://doi.org/10.1126/science.abb0074.

35. Hansen, J.L., Loell, K.J., and Cohen, B.A. (2022). A test of the pioneer factor hypothesis using ectopic liver gene activation. eLife *11*, e73358. https://doi.org/10.7554/eLife.73358.

36. King, H.W., and Klose, R.J. (2017). The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. eLife *6*, e22631. https://doi.org/10.7554/eLife.22631.

37. Frederick, M.A., Williamson, K.E., Fernandez Garcia, M., Ferretti, M.B., McCarthy, R.L., Donahue, G., Luzete Monteiro, E., Takenaka, N., Reynaga, J., Kadoch, C., et al. (2022). A pioneer factor locally opens compacted chromatin to enable targeted ATP-dependent nucleosome remodeling. Nat. Struct. Mol. Biol. https://doi.org/10.1038/s41594-022-00886-5.

38. Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell *14*, 838–853. https://doi.org/10.1016/j.stem.2014.04.003.

39. Lodato, M.A., Ng, C.W., Wamstad, J.A., Cheng, A.W., Thai, K.K., Fraenkel, E., Jaenisch, R., and Boyer, L.A. (2013). SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. PLoS Genet. *9*, e1003288. https://doi.org/10.1371/journal.pgen.1003288.

40. Thanos, D., and Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. Cell *83*, 1091–1100. https://doi.org/10.1016/0092-8674(95)90136-1.

41. Panne, D., Maniatis, T., and Harrison, S.C. (2007). An atomic model of the interferon-β enhanceosome. Cell *129*, 1111–1123. https://doi.org/10.1016/j.cell.2007.05.019.

42. Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., and Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature *527*, 384–388. https://doi.org/10.1038/nature15518.

43. Grossman, S.R., Engreitz, J., Ray, J.P., Nguyen, T.H., Hacohen, N., and Lander, E.S. (2018). Positional specificity of different transcription factor classes within enhancers. Proc. Natl. Acad. Sci. USA *115*, E7222–E7230. https://doi.org/10.1073/pnas.1804663115.

44. Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Fropf, R., McAnany, C., Gagneur, J., Kundaje, A., et al. (2021). Base-resolution models of transcription-factor binding reveal soft motif syntax. Nat. Genet. *53*, 354–366. https://doi.org/10.1038/s41588-021-00782-6.

45. Kim, D.S., Risca, V.I., Reynolds, D.L., Chappell, J., Rubin, A.J., Jung, N., Donohue, L.K.H., Lopez-Pajares, V., Kathiria, A., Shi, M., et al. (2021). The dynamic, combinatorial cis-regulatory lexicon of epidermal differentiation. Nat. Genet. *53*, 1564–1576. https://doi.org/10.1038/s41588-021-00947-3.

46. Minnoye, L., Taskiran, I.I., Mauduit, D., Fazio, M., Van Aerschot, L.V., Hulselmans, G., Christiaens, V., Makhzami, S., Seltenhammer, M., Karras, P., et al. (2020). Cross-species analysis of enhancer logic using deep learning. Genome Res. *30*, 1815–1834. https://doi.org/10.1101/gr.260844.120.

47. Lee, Q.Y., Mall, M., Chanda, S., Zhou, B., Sharma, K.S., Schaukowitch, K., Adrian-Segarra, J.M., Grieder, S.D., Kareta, M.S., Wapinski, O.L., et al. (2020). Pro-neuronal activity of Myod1 due to promiscuous binding to neuronal genes. Nat. Cell Biol. *22*, 401–411. https://doi.org/10.1038/s41556-020-0490-3.

48. Brown, C.J., Johnson, A.K., Dunker, A.K., and Daughdrill, G.W. (2011). Evolution and disorder. Curr. Opin. Struct. Biol. *21*, 441–446. https://doi.org/10.1016/j.sbi.2011.02.005.

49. Brodsky, S., Jana, T., Mittelman, K., Chapal, M., Kumar, D.K., Carmi, M., and Barkai, N. (2020). Intrinsically disordered regions direct transcription factor in vivo binding specificity. Mol. Cell *79*, 459–471.e4. https://doi.org/10.1016/j.molcel.2020.05.032.

50. Staller, M.V. (2022). Transcription factors perform a 2-step search of the nucleus. Genetics *222*, iyac111. https://doi.org/10.1093/genetics/iyac111.

51. Fonseca, G.J., Tao, J., Westin, E.M., Duttke, S.H., Spann, N.J., Strid, T., Shen, Z., Stender, J.D., Sakai, M., Link, V.M., et al. (2019). Diverse motif ensembles specify non-redundant DNA binding activities of AP-1 family members in macrophages. Nat. Commun. *10*, 414. https://doi.org/10.1038/s41467-018-08236-0.

52. Battaglia, S., Dong, K., Wu, J., Chen, Z., Najm, F.J., Zhang, Y., Moore, M.M., Hecht, V., Shoresh, N., and Bernstein, B.E. (2022). Long-range phasing of dynamic, tissue-specific and allele-specific regulatory elements. Nat. Genet. *54*, 1504–1513. https://doi.org/10.1038/s41588-022-01188-8.

53. Stergachis, A.B., Debo, B.M., Haugen, E., Churchman, L.S., and Stamatoyannopoulos, J.A. (2020). Single-molecule regulatory architectures captured by chromatin fiber sequencing. Science *368*, 1449–1454. https://doi.org/10.1126/science.aaz1646.

54. Abdulhay, N.J., McNally, C.P., Hsieh, L.J., Kasinathan, S., Keith, A., Estes, L.S., Karimzadeh, M., Underwood, J.G., Goodarzi, H., Narlikar, G.J., et al. (2020). Massively multiplex single-molecule oligonucleosome footprinting. eLife *9*, e59404. https://doi.org/10.7554/eLife.59404.

55. Rao, S., Ahmad, K., and Ramachandran, S. (2021). Cooperative binding between distant transcription factors is a hallmark of active enhancers. Mol. Cell *81*, 1651–1665.e4. https://doi.org/10.1016/j.molcel.2021.02.014.

56. Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. *196*, 801–810. https://doi.org/10.1083/jcb.201112098.

57. Göös, H., Kinnunen, M., Salokas, K., Tan, Z., Liu, X., Yadav, L., Zhang, Q., Wei, G.H., and Varjosalo, M. (2022). Human transcription factor protein interaction networks. Nat. Commun. *13*, 766. https://doi.org/10.1038/s41467-022-28341-5.

58. Tanay, A. (2006). Extensive low-affinity transcriptional interactions in the yeast genome. Genome Res. *16*, 962–972. https://doi.org/10.1101/gr.5113606.

59. Shahein, A., López-Malo, M., Istomin, I., Olson, E.J., Cheng, S., and Maerkl, S.J. (2022). Systematic analysis of low-affinity transcription factor binding site clusters in vitro and in vivo establishes their functional relevance. Nat. Commun. *13*, 5273. https://doi.org/10.1038/s41467-022-32971-0.

60. Crocker, J., Abe, N., Rinaldi, L., McGregor, A.P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., et al. (2015). Low affinity Binding Site clusters confer hox specificity and regulatory robustness. Cell *160*, 191–203. https://doi.org/10.1016/j.cell.2014.11.041.

61. Farley, E.K., Olson, K.M., Zhang, W., Brandt, A.J., Rokhsar, D.S., and Levine, M.S. (2015). Suboptimization of developmental enhancers. Science *350*, 325–328. https://doi.org/10.1126/science.aac6948.

62. Jindal, G.A., Bantle, A.T., Solvason, J.J., Grudzien, J.L., D'Antonio-Chronowska, A., Lim, F., Le, S.H., Larsen, R.O., Klie, A., Frazer, K.A., et al. (2022). Affinity-optimizing variants within cardiac enhancers disrupt heart development and contribute to cardiac traits. Preprint at bioRxiv. https://doi.org/10.1101/2022.05.27.493636.

63. Lim, F., Ryan, G.E., Le, S.H., Solvason, J.J., Steffen, P., and Farley, E.K. (2022). Affinity-optimizing variants within the ZRS enhancer disrupt limb development. Preprint at bioRxiv. https://doi.org/10.1101/2022.05.27.493789.

64. Louphrasitthiphol, P., Siddaway, R., Loffreda, A., Pogenberg, V., Friedrichsen, H., Schepsky, A., Zeng, Z., Lu, M., Strub, T., Freter, R., et al. (2020). Tuning transcription factor availability through acetylation-mediated genomic redistribution. Mol. Cell *79*, 472–487.e10. https://doi.org/10.1016/j.molcel.2020.05.025.

65. Jantz, D., and Berg, J.M. (2004). Reduction in DNA-binding affinity of Cys2His2 zinc finger proteins by linker phosphorylation. Proc. Natl. Acad. Sci. USA *101*, 7589–7593. https://doi.org/10.1073/pnas.0402191101.

66. Rosonina, E., Akhter, A., Dou, Y., Babu, J., and Sri Theivakadadcham, V.S. (2017). Regulation of transcription factors by SUMOylation. Transcription *8*, 220–231. https://doi.org/10.1080/21541264.2017.1311829.

67. Naqvi, S., Kim, S., Hoskens, H., Matthews, H.S., Spritz, R.A., Klein, O.D., Hallgrímsson, B., Swigut, T., Claes, P., Pritchard, J.K., et al. (2022). Precise modulation of transcription factor levels reveals drivers of dosage sensitivity. Preprint at bioRxiv. https://doi.org/10.1101/2022.06.13.495964.

68. Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.I., Cooper, G.M., et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. Nat. Biotechnol. *30*, 265–270. https://doi.org/10.1038/nbt.2136.

69. Fuqua, T., Jordan, J., van Breugel, M.E., Halavatyi, A., Tischer, C., Polidoro, P., Abe, N., Tsai, A., Mann, R.S., Stern, D.L., et al. (2020). Dense and pleiotropic regulatory information in a developmental enhancer. Nature *587*, 235–239. https://doi.org/10.1038/s41586-020-2816-5.

70. de Boer, C.G., Vaishnav, E.D., Sadeh, R., Abeyta, E.L., Friedman, N., and Regev, A. (2020). Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. Nat. Biotechnol. *38*, 56–65. https://doi.org/10.1038/s41587-019-0315-8.

71. Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., et al. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362–366. https://doi.org/10.1038/nature07667.

72. Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. eLife *3*, e02407. https://doi.org/10.7554/eLife.02407.

73. Filtz, T.M., Vogel, W.K., and Leid, M. (2014). Regulation of transcription factor activity by interconnected, post-translational modifications. Trends Pharmacol. Sci. *35*, 76–85. https://doi.org/10.1016/j.tips.2013.11.005.

74. Qiu, C., Cao, J., Martin, B.K., Li, T., Welsh, I.C., Srivatsan, S., Huang, X., Calderon, D., Noble, W.S., Disteche, C.M., et al. (2022). Systematic reconstruction of cellular trajectories across mouse embryogenesis. Nat. Genet. *54*, 328–341. https://doi.org/10.1038/s41588-022-01018-x.

75. Zhang, Y., Ho, T.D., Buchler, N.E., and Gordan, R. (2021). Competition for DNA binding between paralogous transcription factors determines their genomic occupancy and regulatory functions. Genome Res. *31*, 1216–1229. https://doi.org/10.1101/gr.275145.120.

76. Bourgeois, B., Gui, T., Hoogeboom, D., Hocking, H.G., Richter, G., Spreitzer, E., Viertler, M., Richter, K., Madl, T., and Burgering, B.M.T. (2021). Multiple regulatory intrinsically disordered motifs control FOXO4 transcription factor binding and function. Cell Rep. *36*, 109446. https://doi.org/10.1016/j.celrep.2021.109446.

77. Jen, Y., Weintraub, H., and Benezra, R. (1992). Overexpression of Id protein inhibits the muscle differentiation program: in vivo association of Id with E2A proteins. Genes Dev. *6*, 1466–1479. https://doi.org/10.1101/gad.6.8.1466.

78. Garcia-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. *29*, 1363–1375. https://doi.org/10.1101/gr.240663.118.

79. Noviello, G., Gjaltema, R.A.F., and Schulz, E. (2022). CasTuner: a degron and CRISPR/Cas-based toolkit for analog tuning of endogenous gene expression. Preprint at bioRxiv. https://doi.org/10.1101/2022.10.05.511019.

80. Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M., and Furlong, E.E.M. (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature *462*, 65–70. https://doi.org/10.1038/nature08531.

81. Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560. https://doi.org/10.1038/nature06008.

82. Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283. https://doi.org/10.1038/nature09692.

83. Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science *339*, 1074–1077. https://doi.org/10.1126/science.1232542.

84. Gasperini, M., Tome, J.M., and Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. Nat. Rev. Genet. *21*, 292–310. https://doi.org/10.1038/s41576-019-0209-0.

85. Cai, Y., Zhang, Y., Loh, Y.P., Tng, J.Q., Lim, M.C., Cao, Z., Raju, A., Lieberman Aiden, E., Li, S., Manikandan, L., et al. (2021). H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. Nat. Commun. *12*, 719. https://doi.org/10.1038/s41467-021-20940-y.

86. Huang, D., Petrykowska, H.M., Miller, B.F., Elnitski, L., and Ovcharenko, I. (2019). Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression. Genome Res. *29*, 657–667. https://doi.org/10.1101/gr.247007.118.

87. Doni Jayavelu, N., Jajodia, A., Mishra, A., and Hawkins, R.D. (2020). Candidate silencer elements for the human and mouse genomes. Nat. Commun. *11*, 1061. https://doi.org/10.1038/s41467-020-14853-5.

88. Pang, B., and Snyder, M.P. (2020). Systematic identification of silencers in human cells. Nat. Genet. *52*, 254–263. https://doi.org/10.1038/s41588-020-0578-5.

89. Gisselbrecht, S.S., Palagi, A., Kurland, J.V., Rogers, J.M., Ozadam, H., Zhan, Y., Dekker, J., and Bulyk, M.L. (2020). Transcriptional silencers in drosophila serve a dual role as transcriptional enhancers in alternate cellular contexts. Mol. Cell *77*, 324–337.e8. https://doi.org/10.1016/j.molcel.2019.10.004.

90. Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. Trends Genet. *28*, 276–284. https://doi.org/10.1016/j.tig.2012.02.008.

91. Perry, M.W., Boettiger, A.N., and Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proc. Natl. Acad. Sci. USA *108*, 13570–13575. https://doi.org/10.1073/pnas.1109873108.

92. Pennacchio, L.A., Loots, G.G., Nobrega, M.A., and Ovcharenko, I. (2007). Predicting tissue-specific enhancers in the human genome. Genome Res. *17*, 201–211. https://doi.org/10.1101/gr.5972507.

93. Taskiran, I.I., Spanier, K.I., Christiaens, V., Mauduit, D., and Aerts, S. (2022). Cell type directed design of synthetic enhancers. Preprint at bioRxiv. https://doi.org/10.1101/2022.07.26.501466.

94. Richter, W.F., Nayak, S., Iwasa, J., and Taatjes, D.J. (2022). The Mediator complex as a master regulator of transcription by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *23*, 732–749. https://doi.org/10.1038/s41580-022-00498-3.

95. Näär, A.M., Lemon, B.D., and Tjian, R. (2001). Transcriptional coactivator complexes. Annu. Rev. Biochem. *70*, 475–501. https://doi.org/10.1146/annurev.biochem.70.1.475.

96. Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707–719. https://doi.org/10.1016/j.cell.2007.01.015.

97. Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat. Rev. Mol. Cell Biol. *18*, 407–422. https://doi.org/10.1038/nrm.2017.26.

98. Brahma, S., and Henikoff, S. (2020). Epigenome Regulation by Dynamic Nucleosome Unwrapping. Trends Biochem. Sci. *45*, 13–26. https://doi.org/10.1016/j.tibs.2019.09.003.

99. Perissi, V., Jepsen, K., Glass, C.K., and Rosenfeld, M.G. (2010). Deconstructing repression: evolving models of co-repressor action. Nat. Rev. Genet. *11*, 109–123. https://doi.org/10.1038/nrg2736.

100. Chen, Y.-J.C., Koutelou, E., and Dent, S.Y.R. (2022). Now open: evolving insights to the roles of lysine acetylation in chromatin organization and function. Mol. Cell *82*, 716–727. https://doi.org/10.1016/j.molcel.2021.12.004.

101. Morgan, M.A.J., and Shilatifard, A. (2020). Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. Nat. Genet. *52*, 1271–1281. https://doi.org/10.1038/s41588-020-00736-4.

102. Matkar, S., Thiel, A., and Hua, X. (2013). Menin: a scaffold protein that controls gene expression and cell signaling. Trends Biochem. Sci. *38*, 394–402. https://doi.org/10.1016/j.tibs.2013.05.005.

103. Hanske, J., Sadian, Y., and Müller, C.W. (2018). The cryo-EM resolution revolution and transcription complexes. Curr. Opin. Struct. Biol. *52*, 8–15. https://doi.org/10.1016/j.sbi.2018.07.002.

104. Osman, S., and Cramer, P. (2020). Structural biology of RNA polymerase II Transcription: 20 Years On. Annu. Rev. Cell Dev. Biol. *36*, 1–34. https://doi.org/10.1146/annurev-cellbio-042020-021954.

105. Jaeger, M.G., and Winter, G.E. (2021). Fast-acting chemical tools to delineate causality in transcriptional control. Mol. Cell *81*, 1617–1630. https://doi.org/10.1016/j.molcel.2021.02.015.

106. Schick, S., Grosche, S., Kohl, K.E., Drpic, D., Jaeger, M.G., Marella, N.C., Imrichova, H., Lin, J.-M.G., Hofstätter, G., Schuster, M., et al. (2021). Acute BAF perturbation causes immediate changes in chromatin accessibility. Nat. Genet. *53*, 269–278. https://doi.org/10.1038/s41588-021-00777-3.

107. Iurlaro, M., Stadler, M.B., Masoni, F., Jagani, Z., Galli, G.G., and Schübeler, D. (2021). Mammalian SWI/SNF continuously restores local accessibility to chromatin. Nat. Genet. *53*, 279–287. https://doi.org/10.1038/s41588-020-00768-w.

108. Hendy, O., Serebreni, L., Bergauer, K., Muerdter, F., Huber, L., Nemčko, F., and Stark, A. (2022). Developmental and housekeeping transcriptional programs in Drosophila require distinct chromatin remodelers. Mol. Cell *82*, 3598–3612.e7. https://doi.org/10.1016/j.molcel.2022.08.019.

109. Jaeger, M.G., Schwalb, B., Mackowiak, S.D., Velychko, T., Hanzl, A., Imrichova, H., Brand, M., Agerer, B., Chorn, S., Nabet, B., et al. (2020). Selective Mediator dependence of cell-type-specifying transcription. Nat. Genet. *52*, 719–727. https://doi.org/10.1038/s41588-020-0635-0.

110. Gillespie, M.A., Palii, C.G., Sanchez-Taltavull, D., Shannon, P., Longabaugh, W.J.R., Downes, D.J., Sivaraman, K., Espinoza, H.M., Hughes, J.R., Price, N.D., et al. (2020). Absolute quantification of transcription factors reveals principles of gene regulation in erythropoiesis. Mol. Cell *78*, 960–974.e11. https://doi.org/10.1016/j.molcel.2020.03.031.

111. Narita, T., Ito, S., Higashijima, Y., Chu, W.K., Neumann, K., Walter, J., Satpathy, S., Liebner, T., Hamilton, W.B., Maskey, E., et al. (2021). Enhancers are activated by p300/CBP activity-dependent PIC assembly, RNAPII recruitment, and pause release. Mol. Cell *81*, 2166–2182.e6. https://doi.org/10.1016/j.molcel.2021.03.008.

112. Policarpi, C., Munafo, M., Tsagkris, S., Carlini, V., and Hackett, J.A. (2022). Systematic epigenome editing captures the context-dependent instructive function of chromatin modifications. Preprint at bioRxiv. https://doi.org/10.1101/2022.09.04.506519.

113. Hunt, G., Boija, A., and Mannervik, M. (2022). p300/CBP sustains Polycomb silencing by non-enzymatic functions. Mol. Cell *82*, 3580–3597.e9. https://doi.org/10.1016/j.molcel.2022.09.005.

114. Dorighi, K.M., Swigut, T., Henriques, T., Bhanu, N.V., Scruggs, B.S., Nady, N., Still, C.D., Garcia, B.A., Adelman, K., and Wysocka, J. (2017). Mll3 and Mll4 facilitate enhancer RNA synthesis and transcription from promoters independently of H3K4 monomethylation. Mol. Cell *66*, 568–576.e4. https://doi.org/10.1016/j.molcel.2017.04.018.

115. Rickels, R., Herz, H.-M., Sze, C.C., Cao, K., Morgan, M.A., Collings, C.K., Gause, M., Takahashi, Y.-H., Wang, L., Rendleman, E.J., et al. (2017). Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. Nat. Genet. *49*, 1647–1653. https://doi.org/10.1038/ng.3965.

116. Zhang, T., Cooper, S., and Brockdorff, N. (2015). The interplay of histone modifications – writers that read. EMBO Rep. *16*, 1467–1481. https://doi.org/10.15252/embr.201540945.

117. DelRosso, N., Tycko, J., Suzuki, P., Andrews, C., Aradhana, M., A., Liongson, I., Ludwig, C., Spees, K., Fordyce, P., et al. (2022). Large-scale mapping and systematic mutagenesis of human transcriptional effector domains. Preprint at bioRxiv. https://doi.org/10.1101/2022.08.26.505496.

118. Sanborn, A.L., Yeh, B.T., Feigerle, J.T., Hao, C.V., Townshend, R.J., Lieberman Aiden, E., Dror, R.O., and Kornberg, R.D. (2021). Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator. eLife *10*, e68068. https://doi.org/10.7554/eLife.68068.

119. Huang, Z.Q., Li, J., Sachs, L.M., Cole, P.A., and Wong, J. (2003). A role for cofactor–cofactor and cofactor–histone interactions in targeting p300, SWI/SNF and Mediator for transcription. EMBO J. *22*, 2146–2155. https://doi.org/10.1093/emboj/cdg219.

120. Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569–577. https://doi.org/10.1038/386569a0.

121. Tycko, J., DelRosso, N., Hess, G.T., Aradhana, B., A., Mukund, A., Van, M.V., Ego, B.K., Yao, D., Spees, K., et al. (2020). High-throughput discovery and characterization of human transcriptional effectors. Cell *183*, 2020–2035.e16. https://doi.org/10.1016/j.cell.2020.11.024.

122. Alerasool, N., Leng, H., Lin, Z.-Y., Gingras, A.-C., and Taipale, M. (2022). Identification and functional characterization of transcriptional activators in human cells. Mol. Cell *82*, 677–695.e7. https://doi.org/10.1016/j.molcel.2021.12.008.

123. Stampfel, G., Kazmar, T., Frank, O., Wienerroither, S., Reiter, F., and Stark, A. (2015). Transcriptional regulators form diverse groups with context-dependent regulatory functions. Nature *528*, 147–151. https://doi.org/10.1038/nature15545.

124. Bell, C.C., Talarmain, L., Scolamiero, L., Lam, E.Y., Ang, C.-S., Gilan, O., and Dawson, M.A. (2022). Transcription factors use a unique combination of cofactors to potentiate different promoter-dependent steps in transcription. Preprint at bioRxiv. https://doi.org/10.1101/2022.10.25.513774.

125. Soto, L.F., Li, Z., Santoso, C.S., Berenson, A., Ho, I., Shen, V.X., Yuan, S., and Fuxman Bass, J.I.F. (2022). Compendium of human transcription factor effector domains. Mol. Cell *82*, 514–526. https://doi.org/10.1016/j.molcel.2021.11.007.

126. Mukund, A.X., Tycko, J., Allen, S.J., Robinson, S.A., Andrews, C., Ludwig, C.H., Spees, K., Bassik, M.C., and Bintu, L. (2022). High-throughput functional characterization of combinations of transcriptional activators and repressors. Preprint at bioRxiv. https://doi.org/10.1101/2022.12.20.521091.

127. Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science *361*, eaar3958. https://doi.org/10.1126/science.aar3958.

128. Trojanowski, J., Frank, L., Rademacher, A., Mücke, N., Grigaitis, P., and Rippe, K. (2022). Transcription activation is enhanced by multivalent interactions independent of phase separation. Mol. Cell *82*, 1878–1893.e10. https://doi.org/10.1016/j.molcel.2022.04.017.

129. Sabari, B.R. (2020). Biomolecular condensates and gene activation in development and disease. Dev. Cell *55*, 84–96. https://doi.org/10.1016/j.devcel.2020.09.005.

130. Sharp, P.A., Chakraborty, A.K., Henninger, J.E., and Young, R.A. (2022). RNA in formation and regulation of transcriptional condensates. RNA N. Y. *28*, 52–57. https://doi.org/10.1261/rna.078997.121.

131. Strom, A.R., and Brangwynne, C.P. (2019). The liquid nucleome - phase transitions in the nucleus at a glance. J. Cell Sci. *132*, jcs235093. https://doi.org/10.1242/jcs.235093.

132. Lyon, A.S., Peeples, W.B., and Rosen, M.K. (2021). A framework for understanding the functions of biomolecular condensates across scales. Nat. Rev. Mol. Cell Biol. *22*, 215–235. https://doi.org/10.1038/s41580-020-00303-z.

133. Polish, J.A., Kim, J.-H., and Johnston, M. (2005). How the Rgt1 transcription factor of *Saccharomyces cerevisiae* is regulated by glucose. Genetics *169*, 583–594. https://doi.org/10.1534/genetics.104.034512.

134. Sze, J.-Y., Woontner, M., Jaehning, J.A., and Kohlhaw, G.B. (1992). In vitro transcriptional activation by a metabolic intermediate: activation by Leu3 depends on α-Isopropylmalate. Science *258*, 1143–1145. https://doi.org/10.1126/science.1439822.

135. White, M.A., Kwasnieski, J.C., Myers, C.A., Shen, S.Q., Corbo, J.C., and Cohen, B.A. (2016). A simple grammar defines activating and repressing cis-regulatory elements in photoreceptors. Cell Rep. *17*, 1247–1254. https://doi.org/10.1016/j.celrep.2016.09.066.

136. Marceau, A.H., Brison, C.M., Nerli, S., Arsenault, H.E., McShan, A.C., Chen, E., Lee, H.-W., Benanti, J.A., Sgourakis, N.G., and Rubin, S.M. (2019). An order-to-disorder structural switch activates the FoxM1 transcription factor. eLife *8*, e46131. https://doi.org/10.7554/eLife.46131.

137. Singh, G., Mullany, S., Moorthy, S.D., Zhang, R., Mehdi, T., Tian, R., Duncan, A.G., Moses, A.M., and Mitchell, J.A. (2021). A flexible repertoire of transcription factor binding sites and a diversity threshold determines enhancer activity in embryonic stem cells. Genome Res. *31*, 564–575. https://doi.org/10.1101/gr.272468.120.

138. Lettice, L.A., Horikoshi, T., Heaney, S.J.H., van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N., et al. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc. Natl. Acad. Sci. USA *99*, 7548–7553. https://doi.org/10.1073/pnas.112212199.

139. Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer–promoter connections with CRISPR interference. Science *354*, 769–773. https://doi.org/10.1126/science.aag2445.

140. Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. Genome Res. *24*, 390–400. https://doi.org/10.1101/gr.163519.113.

141. Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., et al. (2019). Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. Nat. Genet. *51*, 1664–1669. https://doi.org/10.1038/s41588-019-0538-0.

142. Gasperini, M., Hill, A.J., McFaline-Figueroa, J.L., Martin, B., Kim, S., Zhang, M.D., Jackson, D., Leith, A., Schreiber, J., Noble, W.S., et al. (2019). A genome-wide framework for mapping gene regulation via cellular genetic screens. Cell *176*, 377–390.e19. https://doi.org/10.1016/j.cell.2018.11.029.

143. Akhtar, W., de Jong, J., Pindyurin, A.V., Pagie, L., Meuleman, W., de Ridder, J., Berns, A., Wessels, L.F.A., van Lohuizen, M., and van Steensel, B.

(2013). Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell *154*, 914–927. https://doi.org/10.1016/j.cell.2013.07.018.

144. Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., et al. (2022). Nonlinear control of transcription through enhancer–promoter interactions. Nature *604*, 571–577. https://doi.org/10.1038/s41586-022-04570-y.

145. Popay, T.M., and Dixon, J.R. (2022). Coming full circle: on the origin and evolution of the looping model for enhancer–promoter communication. J. Biol. Chem. *298*, 102117. https://doi.org/10.1016/j.jbc.2022.102117.

146. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385. https://doi.org/10.1038/nature11049.

147. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380. https://doi.org/10.1038/nature11082.

148. Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025. https://doi.org/10.1016/j.cell.2015.04.004.

149. Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. Cell *169*, 930–944.e22. https://doi.org/10.1016/j.cell.2017.05.004.

150. Kubo, N., Ishii, H., Xiong, X., Bianco, S., Meitinger, F., Hu, R., Hocker, J.D., Conte, M., Gorkin, D., Yu, M., et al. (2021). Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. Nat. Struct. Mol. Biol. *28*, 152–161. https://doi.org/10.1038/s41594-020-00539-5.

151. Downes, D.J., Smith, A.L., Karpinska, M.A., Velychko, T., Rue-Albrecht, K., Sims, D., Milne, T.A., Davies, J.O.J., Oudelaar, A.M., and Hughes, J.R. (2022). Capture-C: a modular and flexible approach for high-resolution chromosome conformation capture. Nat. Protoc. *17*, 445–475. https://doi.org/10.1038/s41596-021-00651-w.

152. Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.-M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S.W., et al. (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res. *25*, 582–597. https://doi.org/10.1101/gr.185272.114.

153. Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat. Methods *13*, 919–922. https://doi.org/10.1038/nmeth.3999.

154. Hsieh, T.-H.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Rando, O.J., Tjian, R., and Darzacq, X. (2020). Resolving the 3D landscape of transcription-linked mammalian chromatin folding. Mol. Cell *78*, 539–553.e8. https://doi.org/10.1016/j.molcel.2020.03.002.

155. Krietenstein, N., Abraham, S., Venev, S.V., Abdennur, N., Gibcus, J., Hsieh, T.-H.S., Parsi, K.M., Yang, L., Maehr, R., Mirny, L.A., et al. (2020). Ultrastructural details of mammalian chromosome architecture. Mol. Cell *78*, 554–565.e7. https://doi.org/10.1016/j.molcel.2020.03.003.

156. Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 is a structural regulator of enhancer-promoter loops. Cell *171*, 1573–1588.e28. https://doi.org/10.1016/j.cell.2017.11.008.

157. Eagen, K.P., Aiden, E.L., and Kornberg, R.D. (2017). Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. Proc. Natl. Acad. Sci. USA *114*, 8764–8769. https://doi.org/10.1073/pnas.1701291114.

158. Kraft, K., Yost, K.E., Murphy, S.E., Magg, A., Long, Y., Corces, M.R., Granja, J.M., Wittler, L., Mundlos, S., Cech, T.R., et al. (2022). Polycomb-mediated genome architecture enables long-range spreading of H3K27 methylation. Proc. Natl. Acad. Sci. USA *119*. e2201883119. https://doi.org/10.1073/pnas.2201883119.

159. Ringel, A.R., Szabo, Q., Chiariello, A.M., Chudzik, K., Schöpflin, R., Rothe, P., Mattei, A.L., Zehnder, T., Harnett, D., Laupert, V., et al. (2022). Repression and 3D-restructuring resolves regulatory conflicts in evolutionarily rearranged genomes. Cell *185*, 3689–3704.e21. https://doi.org/10.1016/j.cell.2022.09.006.

160. Pinglay, S., Bulajić, M., Rahe, D.P., Huang, E., Brosh, R., Mamrak, N.E., King, B.R., German, S., Cadley, J.A., Rieber, L., et al. (2022). Synthetic regulatory reconstitution reveals principles of mammalian Hox cluster regulation. Science *377*, eabk2820. https://doi.org/10.1126/science.abk2820.

161. Haberle, V., Arnold, C.D., Pagani, M., Rath, M., Schernhuber, K., and Stark, A. (2019). Transcriptional cofactors display specificity for distinct types of core promoters. Nature *570*, 122–126. https://doi.org/10.1038/s41586-019-1210-7.

162. Neumayr, C., Haberle, V., Serebreni, L., Karner, K., Hendy, O., Boija, A., Henninger, J.E., Li, C.H., Stejskal, K., Lin, G., et al. (2022). Differential cofactor dependencies define distinct types of human enhancers. Nature *606*, 406–413. https://doi.org/10.1038/s41586-022-04779-x.

163. Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. Nature *518*, 556–559. https://doi.org/10.1038/nature13994.

164. Guarente, L., and Hoar, E. (1984). Upstream activation sites of the CYC1 gene of Saccharomyces cerevisiae are active when inverted but not when placed downstream of the "TATA box". Proc. Natl. Acad. Sci. USA *81*, 7860–7864. https://doi.org/10.1073/pnas.81.24.7860.

165. Bergman, D.T., Jones, T.R., Liu, V., Ray, J., Jagoda, E., Siraj, L., Kang, H.Y., Nasser, J., Kane, M., Rios, A., et al. (2022). Compatibility rules of human enhancer and promoter sequences. Nature *607*, 176–184. https://doi.org/10.1038/s41586-022-04877-w.

166. Martinez-Ara, M., Comoglio, F., van Arensbergen, J., and van Steensel, B. (2022). Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome. Mol. Cell *82*, 2519–2531.e6. https://doi.org/10.1016/j.molcel.2022.04.009.

167. Sahu, B., Hartonen, T., Pihlajamaa, P., Wei, B., Dave, K., Zhu, F., Kaasinen, E., Lidschreiber, K., Lidschreiber, M., Daub, C.O., et al. (2022). Sequence determinants of human gene regulatory elements. Nat. Genet. *54*, 283–294. https://doi.org/10.1038/s41588-021-01009-4.

168. Xiao, J.Y., Hafner, A., and Boettiger, A.N. (2021). How subtle changes in 3D structure can create large changes in transcription. eLife *10*, e64320. https://doi.org/10.7554/eLife.64320.

169. Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J.B., and Gregor, T. (2018). Dynamic interplay between enhancer–promoter topology and gene activity. Nat. Genet. *50*, 1296–1303. https://doi.org/10.1038/s41588-018-0175-z.

170. Mateo, L.J., Murphy, S.E., Hafner, A., Cinquini, I.S., Walker, C.A., and Boettiger, A.N. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. Nature *568*, 49–54. https://doi.org/10.1038/s41586-019-1035-4.

171. Alexander, J.M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S., and Weiner, O.D. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. eLife *8*, e41769. https://doi.org/10.7554/eLife.41769.

172. Reimer, A., Alamos, S., Westrum, C., Turner, M.A., Talledo, P., Zhao, J., and Garcia, H.G. (2021). Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient. Preprint at bioRxiv. https://doi.org/10.1101/2021.07.10.451524.

173. Karr, J.P., Ferrie, J.J., Tjian, R., and Darzacq, X. (2022). The transcription factor activity gradient (TAG) model: contemplating a contact-independent mechanism for enhancer–promoter communication. Genes Dev. *36*, 7–16. https://doi.org/10.1101/gad.349160.121.

174. Klein, J.C., Agarwal, V., Inoue, F., Keith, A., Martin, B., Kircher, M., Ahituv, N., and Shendure, J. (2020). A systematic evaluation of the design and context dependencies of massively parallel reporter assays. Nat. Methods *17*, 1083–1091. https://doi.org/10.1038/s41592-020-0965-y.

175. Parker, S.C.J., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., et al.; NISC Comparative Sequencing Program (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc. Natl. Acad. Sci. USA *110*, 17921–17926. https://doi.org/10.1073/pnas.1317023110.

176. Leemans, C., van der Zwalm, M.C.H., Brueckner, L., Comoglio, F., van Schaik, T., Pagie, L., van Arensbergen, J., and van Steensel, B. (2019). Promoter-intrinsic and local chromatin features determine gene repression in LADs. Cell *177*, 852–864.e14. https://doi.org/10.1016/j.cell.2019.03.009.

177. Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., et al. (2018). Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. Cell *174*, 744–757.e24. https://doi.org/10.1016/j.cell.2018.05.024.

178. Kvon, E.Z., Waymack, R., Gad, M., and Wunderlich, Z. (2021). Enhancer redundancy in development and disease. Nat. Rev. Genet. *22*, 324–336. https://doi.org/10.1038/s41576-020-00311-x.

179. Hong, J.-W., Hendrix, D.A., and Levine, M.S. (2008). Shadow enhancers as a source of evolutionary novelty. Science *321*, 1314. https://doi.org/10.1126/science.1160631.

180. Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. Nature *554*, 239–243. https://doi.org/10.1038/nature25461.

181. Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., and Stern, D.L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature *466*, 490–493. https://doi.org/10.1038/nature09158.

182. Waymack, R., Fletcher, A., Enciso, G., and Wunderlich, Z. (2020). Shadow enhancers can suppress input transcription factor noise through distinct regulatory logic. eLife *9*, e59351. https://doi.org/10.7554/eLife.59351.

183. Choi, J., Lysakovskaia, K., Stik, G., Demel, C., Söding, J., Tian, T.V., Graf, T., and Cramer, P. (2021). Evidence for additive and synergistic action of mammalian enhancers during cell fate determination. eLife *10*, e65381. https://doi.org/10.7554/eLife.65381.

184. Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319. https://doi.org/10.1016/j.cell.2013.03.035.

185. Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947. https://doi.org/10.1016/j.cell.2013.09.053.

186. Zhou, H.Y., Katsman, Y., Dhaliwal, N.K., Davidson, S., Macpherson, N.N., Sakthidevi, M., Collura, F., and Mitchell, J.A. (2014). A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. Genes Dev. *28*, 2699–2711. https://doi.org/10.1101/gad.248526.114.

187. Moorthy, S.D., Davidson, S., Shchuka, V.M., Singh, G., Malek-Gilani, N., Langroudi, L., Martchenko, A., So, V., Macpherson, N.N., and Mitchell, J.A. (2017). Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. Genome Res. *27*, 246–258. https://doi.org/10.1101/gr.210930.116.

188. Ing-Simmons, E., Seitan, V.C., Faure, A.J., Flicek, P., Carroll, T., Dekker, J., Fisher, A.G., Lenhard, B., and Merkenschlager, M. (2015). Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. Genome Res. *25*, 504–513. https://doi.org/10.1101/gr.184986.114.

189. Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A phase separation model for transcriptional control. Cell *169*, 13–23. https://doi.org/10.1016/j.cell.2017.02.007.

190. Farooq, U., Saravanan, B., Islam, Z., Walavalkar, K., Singh, A.K., Jayani, R.S., Meel, S., Swaminathan, S., and Notani, D. (2021). An interdependent network of functional enhancers regulates transcription and EZH2 loading at the INK4a/ARF locus. Cell Rep. *34*, 108898. https://doi.org/10.1016/j.celrep.2021.108898.

191. Oh, S., Shao, J., Mitra, J., Xiong, F., D'Antonio, M., Wang, R., Garcia-Bassets, I., Ma, Q., Zhu, X., Lee, J.H., et al. (2021). Enhancer release and retargeting activates disease-susceptibility genes. Nature *595*, 735–740. https://doi.org/10.1038/s41586-021-03577-1.

192. Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer control of transcriptional bursting. Cell *166*, 358–368. https://doi.org/10.1016/j.cell.2016.05.025.

193. Nguyen, T.A., Jones, R.D., Snavely, A.R., Pfenning, A.R., Kirchner, R., Hemberg, M., and Gray, J.M. (2016). High-throughput functional comparison of promoter and enhancer activities. Genome Res. *26*, 1023–1033. https://doi.org/10.1101/gr.204834.116.

194. Carleton, J.B., Berrett, K.C., and Gertz, J. (2017). Multiplex enhancer interference reveals collaborative control of gene regulation by estrogen receptor α-bound enhancers. Cell Syst. *5*, 333–344.e5. https://doi.org/10.1016/j.cels.2017.08.011.

195. Tippens, N.D., Liang, J., Leung, A.K., Wierbowski, S.D., Ozer, A., Booth, J.G., Lis, J.T., and Yu, H. (2020). Transcription imparts architecture, function and logic to enhancer units. Nat. Genet. *52*, 1067–1075. https://doi.org/10.1038/s41588-020-0686-2.

196. Hay, D., Hughes, J.R., Babbs, C., Davies, J.O.J., Graham, B.J., Hanssen, L.L.P., Kassouf, M.T., Marieke Oudelaar, A.M., Sharpe, J.A., Suciu, M.C., et al. (2016). Genetic dissection of the α-globin super-enhancer in vivo. Nat. Genet. *48*, 895–903. https://doi.org/10.1038/ng.3605.

197. Blayney, J.W., Kassouf, M., Francis, H., Camellato, B.R., Stolper, R., Higgs, D.R., Mitchell, L., and Boeke, J. (2022). Super-enhancers require a combination of classical enhancers and novel facilitator elements to drive high levels of gene expression. Preprint at bioRxiv. https://doi.org/10.1101/2022.06.20.496856.

198. Thomas, H.F., Kotova, E., Jayaram, S., Pilz, A., Romeike, M., Lackner, A., Penz, T., Bock, C., Leeb, M., Halbritter, F., et al. (2021). Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. Mol. Cell *81*, 969–982.e13. https://doi.org/10.1016/j.molcel.2020.12.047.

199. Vockley, C.M., D'Ippolito, A.M., McDowell, I.C., Majoros, W.H., Safi, A., Song, L., Crawford, G.E., and Reddy, T.E. (2016). Direct GR binding sites potentiate clusters of TF binding across the human genome. Cell *166*, 1269–1281.e19. https://doi.org/10.1016/j.cell.2016.07.049.

200. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021.

201. Huang, H., Zhu, Q., Jussila, A., Han, Y., Bintu, B., Kern, C., Conte, M., Zhang, Y., Bianco, S., Chiariello, A.M., et al. (2021). CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. Nat. Genet. *53*, 1064–1074. https://doi.org/10.1038/s41588-021-00863-6.

202. Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017–1021. https://doi.org/10.1126/science.1262088.

203. Calderon, L., Weiss, F.D., Beagan, J.A., Oliveira, M.S., Georgieva, R., Wang, Y.F., Carroll, T.S., Dharmalingam, G., Gong, W., Tossell, K., et al. (2022). Cohesin-dependence of neuronal gene expression relates to chromatin loop length. eLife *11*, e76539. https://doi.org/10.7554/eLife.76539.

204. Chen, L.-F., Long, H.K., Park, M., Swigut, T., Boettiger, A.N., and Wysocka, J. (2022). Structural elements facilitate extreme long-range gene

regulation at a human disease locus. Preprint at bioRxiv. https://doi.org/10.1101/2022.10.20.513057.

205. Pachano, T., Sánchez-Gaya, V., Ealo, T., Mariner-Faulí, M., Bleckwehl, T., Asenjo, H.G., Respuela, P., Cruz-Molina, S., Muñoz-San Martín, M., Haro, E., et al. (2021). Orphan CpG islands amplify poised enhancer regulatory activity and determine target gene responsiveness. Nat. Genet. *53*, 1036–1049. https://doi.org/10.1038/s41588-021-00888-x.

206. Levo, M., Raimundo, J., Bing, X.Y., Sisco, Z., Batut, P.J., Ryabichko, S., Gregor, T., and Levine, M.S. (2022). Transcriptional coupling of distant regulatory genes in living embryos. Nature *605*, 754–760. https://doi.org/10.1038/s41586-022-04680-7.

207. Fortin, J.-P., and Hansen, K.D. (2015). Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. Genome Biol. *16*, 180. https://doi.org/10.1186/s13059-015-0741-y.

208. Haarhuis, J.H.I., Weide, R.H. van der, Blomen, V.A., Yáñez-Cuna, J.O., Amendola, M., Ruiten, M.S. van, Krijger, P.H.L., Teunissen, H., Medema, R.H., Steensel, B. van, et al. (2017). The cohesin release factor WAPL restricts chromatin loop extension. Cell *169*, 693–707.e14. https://doi.org/10.1016/j.cell.2017.04.013.

209. Quinodoz, S.A., Jachowicz, J.W., Bhat, P., Ollikainen, N., Banerjee, A.K., Goronzy, I.N., Blanco, M.R., Chovanec, P., Chow, A., Markaki, Y., et al. (2021). RNA promotes the formation of spatial compartments in the nucleus. Cell *184*, 5775–5790.e30. https://doi.org/10.1016/j.cell.2021.10.014.

210. Wang, H., Xu, X., Nguyen, C.M., Liu, Y., Gao, Y., Lin, X., Daley, T., Kipniss, N.H., La Russa, M., and Qi, L.S. (2018). CRISPR-mediated programmable 3D genome positioning and nuclear organization. Cell *175*, 1405–1417.e14. https://doi.org/10.1016/j.cell.2018.09.013.

211. Judson, H.F. (1996). The Eighth Day of Creation: Makers of the Revolution in Biology, Expanded Edition (CSHL Press).

212. Ching, T., Himmelstein, D.S., Beaulieu-Jones, B.K., Kalinin, A.A., Do, B.T., Way, G.P., Ferrero, E., Agapow, P.-M., Zietz, M., Hoffman, M.M., et al. (2018). Opportunities and obstacles for deep learning in biology and medicine. J. R. Soc. Interface *15*, 20170387. https://doi.org/10.1098/rsif.2017.0387.

213. Li, Y., Chen, C.-Y., Kaye, A.M., and Wasserman, W.W. (2015). The identification of cis-regulatory elements: a review from a machine learning perspective. Biosystems. *138*, 6–17. https://doi.org/10.1016/j.biosystems.2015.10.002.

214. Boettiger, A., and Murphy, S. (2020). Advances in chromatin imaging at kilobase-scale resolution. Trends Genet. *36*, 273–287. https://doi.org/10.1016/j.tig.2019.12.010.

215. Hafner, A., and Boettiger, A. (2022). The spatial organization of transcriptional control. Nat. Rev. Genet. 1–16. https://doi.org/10.1038/s41576-022-00526-0.

216. Brandão, H.B., Gabriele, M., and Hansen, A.S. (2021). Tracking and interpreting long-range chromatin interactions with super-resolution live-cell imaging. Curr. Opin. Cell Biol. *70*, 18–26. https://doi.org/10.1016/j.ceb.2020.11.002.