

## Opinion

# Speed–Specificity Trade-Offs in the Transcription Factors Search for Their Genomic Binding Sites

Tamar Jana,<sup>1,2</sup> Sagie Brodsky,<sup>1,2</sup> and Naama Barkai<sup>1,\*</sup>

**Transcription factors (TFs) regulate gene expression by binding DNA sequences recognized by their DNA-binding domains (DBDs). DBD-recognized motifs are short and highly abundant in genomes. The ability of TFs to bind a specific subset of motif-containing sites, and to do so rapidly upon activation, is fundamental for gene expression in all eukaryotes. Despite extensive interest, our understanding of the TF-target search process is fragmented; although binding specificity and detection speed are two facets of this same process, trade-offs between them are rarely addressed. In this opinion article, we discuss potential speed–specificity trade-offs in the context of existing models. We further discuss the recently described ‘distributed specificity’ paradigm, suggesting that intrinsically disordered regions (IDRs) promote specificity while reducing the TF-target search time.**

## Introduction

Cells and organisms adapt to changing requirements by reprogramming gene expression. Transcription reprogramming depends on transcription factors (TFs), which activate or repress transcription by binding to specific gene regulatory regions. Consistent with their role in diversifying cellular functions, the TF repertoire has expanded with increasing organism complexity [1,2]. For example, the human genome contains approximately threefold more genes than its simple yeast counterparts, but approximately eightfold more TF-coding genes [2]. This expanded TF repertoire is translated into a refined control of expression through the tuning of TF binding to specific subsets of genes.

TFs bind DNA through specialized **DNA binding domains (DBDs)**, see [Glossary](#)). DBDs of the same family bind DNA in a similar manner, exposing a defined subset of amino acids for direct contact with specific nucleotides. Contrasting the fast evolutionary expansion of the TF repertoire, the number of DBD families has remained limited [3]. In fact, a large fraction of eukaryotic TFs belong to just a few families [2]. In budding yeast, for example, >50% of TFs belong to the zinc-cluster (a total of 57 TFs), C2H2 (41 TFs), and bZip (15 TFs) families [4]. Interestingly, while DBD families are highly conserved within the eukaryote, bacteria, or archaea lineages, there is little overlap between the different lineages [5,6]; in a survey of 500 genomes, only a small portion of DBD families were found to be shared among all lineages [1]. The basis of this limited interlineage conservation remains unclear, as eukaryotic DBDs bind naked DNA and associate with the bacterial genome when ectopically expressed.

In this opinion article, we discuss the two mechanisms through which TFs search for their binding sites within large genomes. We focus on two key properties of this search, which are still poorly understood: the specificity of the search, namely the ability of a TF to distinguish relevant from nonrelevant binding positions, and the speed at which binding sites are detected. We argue

## Highlights

Transcription factors (TFs) rapidly detect their specific binding sites within large genomes, supporting cell adaptation to changing conditions.

Eukaryotic TF DNA binding domains bind short DNA motifs with low information content, leaving open the question of how rapid and accurate binding site detection is achieved.

Existing models fall short in addressing two facets of this TF-target search process: specificity and speed.

Intrinsically disordered regions, prevalent in eukaryotic TFs, promote binding specificity and might increase detection speed

<sup>1</sup>Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

<sup>2</sup>These authors contributed equally

\*Correspondence:  
[naama.barkai@weizmann.ac.il](mailto:naama.barkai@weizmann.ac.il)  
(N. Barkai).

that trade-offs between these two requirements (specificity and rapid detection) limit possible search mechanisms. With this view, we discuss the current existing paradigms explaining specificity, pointing at their advantages but also their possible drawbacks. We then present a new paradigm, in which the search process is guided by long **intrinsically disordered regions (IDRs)** present outside of the TF DBDs, as suggested by a recent analysis of two budding yeast TFs. Long IDRs are abundant within eukaryotic TFs. If they do have a role in the TF search process, these IDRs could explain both the specificity and the speed at which TFs locate their binding sites within the large eukaryotic genomes.

### TF Target Search *In Vivo*: The Challenge of Binding Specificity

DBDs bind preferentially to specific DNA sequences. Bacterial DBDs recognize sequences that often have sufficient information content to specify particular genomic addresses while limiting random appearances [7,8]. Some eukaryotic DBDs similarly recognize motifs of high information content, as exemplified by the general polymerase III activating factor, TFIIBA (Figure 1A). The human genome expresses ~800 multiple zinc-finger genes that could potentially encode TFs of specific targeting, but most of these remain uncharacterized. In fact, well-studied TFs, including CTCF [9], GLI1 [10], and PRDM9 [11], bind to relatively short motifs, leading to the ‘many fingers but short motif’ paradox [12]. More generally, systematic mapping of the *in vitro* binding preferences of thousands of eukaryotic DBDs revealed that long motifs supporting specific targeting is the exception rather than the rule [13–15]. For most eukaryotic DBD families, binding motifs are significantly shorter (5–11 bp) and do not contain enough information to support specific targeting [5]. This lack of sufficient information is exemplified in budding yeast, arguably the best characterized eukaryote (Figure 1B,C). Furthermore, at least for some DBD families, information content appears to decrease with increasing genome size, as exemplified by the Myb/SANT family [15]. Consistent with their low information content, DBD-recognized motifs are abundant within the eukaryotic genomes; a typical six-bp motif, for example, appears, on average, ~6000 times in the 10<sup>7</sup>-bp budding yeast genome with no apparent depletion of 6-mers acting as TF-binding sites (Figure 1D). This number of appearances grows in proportion to the genome size, becoming prohibitively large in the 10<sup>9</sup>-bp human genome.

DBD-recognized motifs are not only short, but also shared between TFs, most notably close **paralogs** or, to a lesser extent, members of the same family [13–15] (Figure 1D–F). For example, DBDs of the GATA family commonly bind to sequences containing the GATA or GATC consensus; DBDs of the C2H2 zinc family bind G-rich motifs often containing three or four consecutive guanoses; while DBDs of the fungal-like Zn(2)-C6, comprising ~28% of budding yeast TFs, bind close variations of CGGA-containing motifs.

Therefore, eukaryotic DBD-recognized sequence motifs are of low information content and poor discriminatory power. What is the relevance of these binding motifs *in vivo*? Systematic analyses [16–19] together with myriad specific studies confirmed that TFs localize preferentially at genomic sites containing their *in vitro* bound motifs. However, only a small fraction (~1% in mammalian genomes) of motif-containing sites are bound in actuality [17,20]. Therefore, while DBD-binding preferences define the sites of potentially stable TF–DNA associations, they fall short in predicting where in the genome such associations occur.

### TF Target Search *In Vivo*: The Challenge of Rapid Detection

The limited size of binding motifs and their high abundance in genomes challenge not only the specificity of TF binding, but also the speed by which the relevant genomic binding sites can be detected within the genome. The rate at which a TF encounters a binding site depends on the effective interaction volume, corresponding principally to the size of the site. The number of

### Glossary

**1D diffusion:** DNA-binding proteins that slide along the DNA in search of their actual binding sites.

**3D diffusion:** random movement of TFs in 3D space that can lead to collisions with the DNA.

**Base skipping:** while sliding along the DNA, TFs can hop on and off, resulting in faster scanning at the cost of nucleotide skipping.

**DNA-binding domain (DBD):** protein domain in TFs that allows them to bind specific DNA sequences. This binding is facilitated by the formation of direct bonds between the amino acids comprising the DBD and the DNA nucleotides. These domains are highly conserved in evolution, and TF families are often classified by them.

**Facilitated diffusion:** kinetic model describing the search of TFs for their genomic targets. According to this model, TFs can combine 3D diffusion and 1D sliding along the DNA, to speed up the search for their binding sites.

#### Intrinsically disordered regions

**(IDRs):** protein regions that are not predicted to form a fixed 3D structure. Instead, these regions tend to be flexible and can be found in various orientations. Such regions are prevalent in TFs.

#### Molecular recognition features

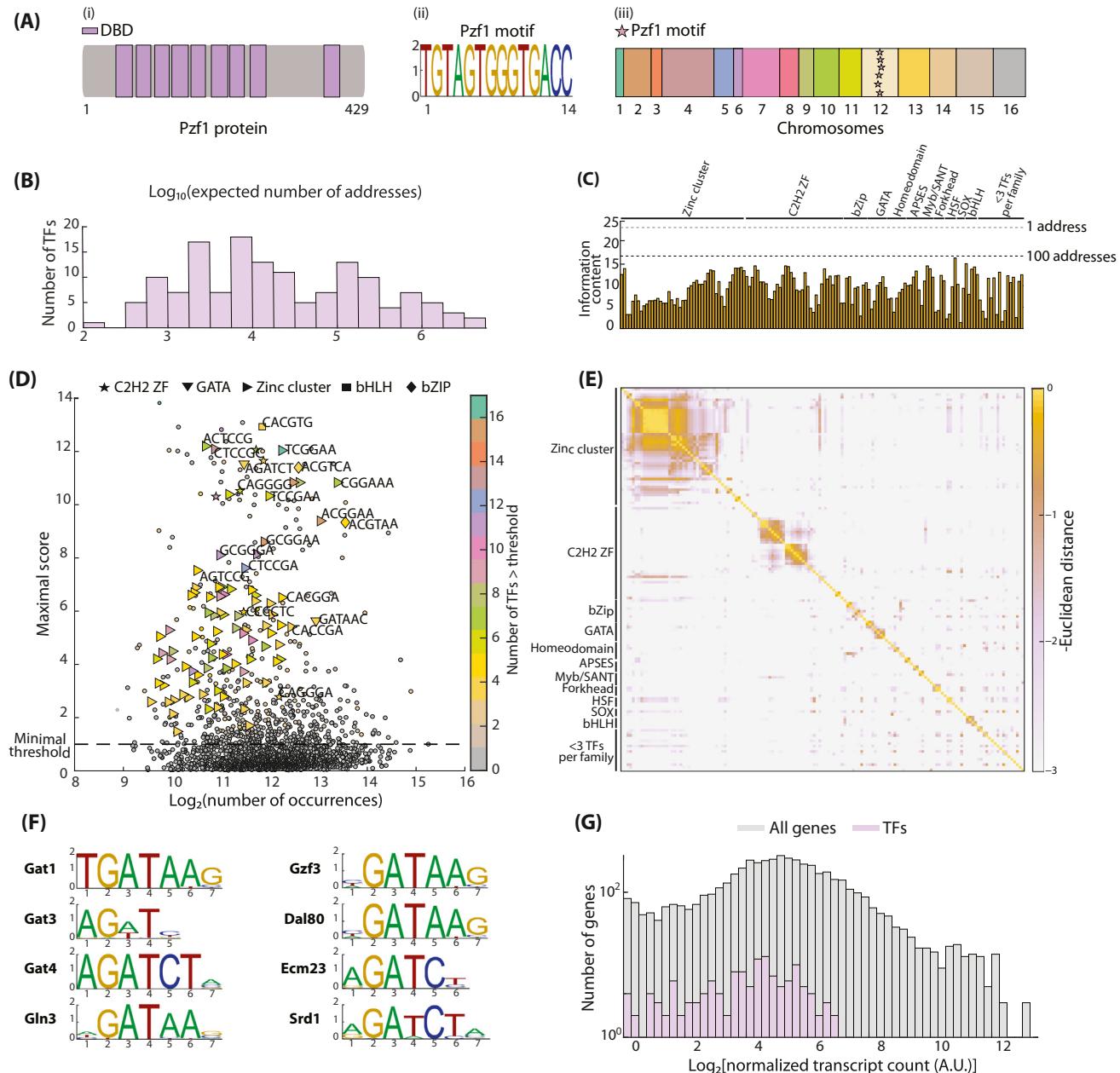
**(MoRFs):** short IDRs in proteins that undergo disorder-to-order transitions (induced folding) upon binding.

**Paralogs:** two related genes found in the same organism that originated from a single ancestral gene that was duplicated. Paralogs may preserve their original function and act similarly or diverge with time.

#### Phase-separated condensates:

cellular compartments that are not surrounded by membrane and which contain a concentrated mixture of biopolymers, such as protein and nucleic acids.

**Superenhancers:** DNA regions in mammalian genomes containing groups of enhancers that are bound by multiple TFs, promoting the expression of genes controlling cell identity.



**Figure 1. Transcription Factor (TF) Binding Motifs Are of Low Information Content and Are Shared between TFs from the Same DNA-Binding Domain (DBD) Family.** (A) Pzf1, a TF with nine C2H2 zinc fingers, binds a high information content motif. (i) The Pzf1 protein, annotations by SMART [134]. (ii) The Pzf1 binding motif [15]. (iii) Scheme of the *Saccharomyces cerevisiae* genome. Indicated are the Pzf1 motif positions, appearing only six times in the genome in a 30-kb region of chromosome 12 containing rDNA genes. (B,C) TF-binding motifs are of low information content. (B) The expected number of genomic addresses of the *S. cerevisiae* TFs. (C) The information content of the respective DBDs. Information required for having only one or 100 addresses in the genome is shown by broken lines, following [5]. TFs are sorted by families and further ordered by binding similarities. (D–F) Binding motifs are not under-represented in the genome and are shared between multiple TFs from the same family. (D) Shown is the maximal TF position weight matrix (PWM)-based score calculated for each possible 6-mer as a function of the number of genomic occurrences (see the supplemental information online for more details). The number of TFs with score >1 is indicated by color. 6-mers preferred by multiple TFs from the same family are marked by shapes. (E) PWM-based motif-binding similarities. TFs are clustered by families, and further ordered based on binding similarity [order as in (C)]. (F) Motif-binding preferences of the indicated *S. cerevisiae* GATA TFs. (G) TFs are expressed at low levels. Histogram showing the normalized mRNA levels of the *S. cerevisiae* genes. Data from [135].

competing sites where transient TF trapping occurs also influences the detection rate and can slow down the search. Considering these challenges, the speed at which activated TFs reprogram gene expression is remarkable. For example, budding yeast activates stress-responsive genes within <5 min of stress exposure [21,22], during which relevant TFs are activated, translocated to the nucleus, detect their binding sites, and recruit the general transcriptional machinery. In the *Drosophila* salivary gland, the TF HSF localizes to its target promoters within 20 s of heat shock [23].

The TF target search has received intense experimental and theoretical attention [24,25]. Early *in vitro* experiments reported rapid association of the bacterial *lac* repressor with its cognate site, at a rate that is 1000-fold faster than predicted by **3D diffusion** and random collision [26]. To explain this accelerated detection, models of **facilitated diffusion** were proposed, in which the 3D diffusion is interspersed with intervals of 1D sliding along the DNA [27–30], thus, effectively, increasing the target size (a so-called ‘antenna effect’) [7,31–34]. Subsequent experiments using single-molecule tracking confirmed that TFs diffuse along the DNA, albeit at rates slower by 1–2 orders of magnitude than in 3D [24,28,35–40]. Together, these studies established facilitated diffusion as the prevailing framework for analyzing the TF–target search.

Within the facilitated diffusion model, the TF–target search depends on the diffusion rate of the TF and its residence time on nonspecific DNA. 3D diffusion defines the rate at which the TF associates with the DNA, while the nonspecific residence time, coupled with **1D diffusion** along DNA, defines the length of the DNA segments (sliding windows) scanned at each association event. For a given genome size, these parameters provide an estimate of the overall search time. Of note, within the model, increasing the 1D sliding intervals could accelerate or, conversely, slow down the search depending on quantitative parameters: Acceleration results from scanning a larger sequence window at each association event, while slowdown results from the increase in residence time at all nonspecific places, most of which do not contain the desired target. Under common assumptions, the search time is optimal when a TF associates with DNA 50% of the time [7,41–44]. In specific cases, longer residence time is beneficial, for example, when the 1D scan allows **base skipping** [39,40]. Of note, the sliding window may be mechanistically limited by the difficulty in retaining stable nonspecific DNA associations while moving along the DNA [45].

Single-molecule tracking of the *lac* repressor in live *Escherichia coli* verified the role of facilitated diffusion *in vivo*. The *lac* repressor is DNA bound 90% of the time, diffuses along DNA at a rate of  $\sim 0.05 \mu\text{m}^2/\text{sec}$  ( $5 \times 10^5 \text{ bp}^2/\text{s}$ ), and covers a sliding window of  $\sim 45 \text{ bp}$  at each millisecond-scale sliding event [36,40]. The measured kinetic parameters well explain the  $\sim 5$ -min search time of a single *lac* repressor for a single operator.

Whether facilitated diffusion can also provide the needed acceleration for the TF target search in the larger, more complex, and chromatin-packed eukaryotic genomes is unclear [7]. If sliding parameters remain the same as in *E. coli* ( $\sim 5$ -ms residence time), the search time would increase in proportion to the genome size, becoming prohibitively long in the large mammalian genomes. Indeed, kinetic parameters measured in living eukaryotic cells correspond, within the same facilitated diffusion model, to significantly longer search times. This includes the 0.8-s 1D residence time measured for budding yeast Mbp1, which implies a single Mbp1–single target detection time of  $\sim 5$  h [46], and the 0.2–1-s nonspecific residence times described for various TFs in higher eukaryotes [47–60]. Together with the  $10^5$ – $10^6 \text{ bp}^2/\text{s}$  ( $0.01$ – $0.1 \mu\text{m}^2/\text{s}$ ) 1D diffusion coefficients [37,61–63], these residence times predict sliding windows of 300–1000 bp [7,62,64–74], implying a single TF–single site detection time of  $\sim 500$  h in the  $10^9$  base-pair human genome. Clearly, these estimated times are exceedingly long. Parallel search by multiple

TF molecules would reduce the search time, but whether this multiplicity of factors is capable of explaining such a rapid search is not clear, particularly when considering the relatively low expression levels of most specific TFs (Figure 1G).

## Beyond the Core DBD-Recognized Motif: Mechanisms Contributing to the TF Target Search

### Sequences Flanking the Core Motif

Considering the theoretical difficulties discussed earlier, the specificity and speed at which eukaryotic TFs detect their binding targets *in vivo* is remarkable. An immediate question is whether the DBD-binding motifs described through the various experimental and computational analysis are accurate enough. A good model for addressing this question are TF paralogs that bind highly similar motifs *in vitro*. While some of the paralogs retained the same binding profile also *in vivo* [22], other paralog pairs bind different sets of genes [75–77]. Analysis of DNA sequences flanking described motifs revealed sequence signatures that could potentially explain the differential binding preferences of close paralogs [78–87]. If prevalent, such contributions, missed in conventional analysis, could explain binding specificity.

TF-binding motifs are typically defined by comparing DNA sequences bound by the TF and detecting the nucleotides with non-zero information content. Even though these analyses do not consider the structure of the respective DBD, the defined motifs commonly correspond to the nucleotides making direct contacts with the DBD. Therefore, the core motif size is largely constrained by the DBD family; for example, a C2H2 zinc finger makes a direct contact with three bases and, therefore, a double C2H2 DBD will form a maximum of six direct contacts. The core motif might still be longer if the DBD makes additional interactions with the DNA, independent of the zinc-finger structures; exemplifying this are the budding yeast Com2 and the *Drosophila* Hox TFs, the DBD-defined motifs of which are extended by additional A/T nucleotides that are bound by basic sequence RGRK (Com2) [84] or a disordered region (Hox) [88] proximal to the DBD. Such contributions are integral parts of the TF-binding sites and could be missed with *in vitro* assays using isolated DBDs. The contribution of such DBD-independent structures appears insufficient for explaining the conundrum of binding specificity.

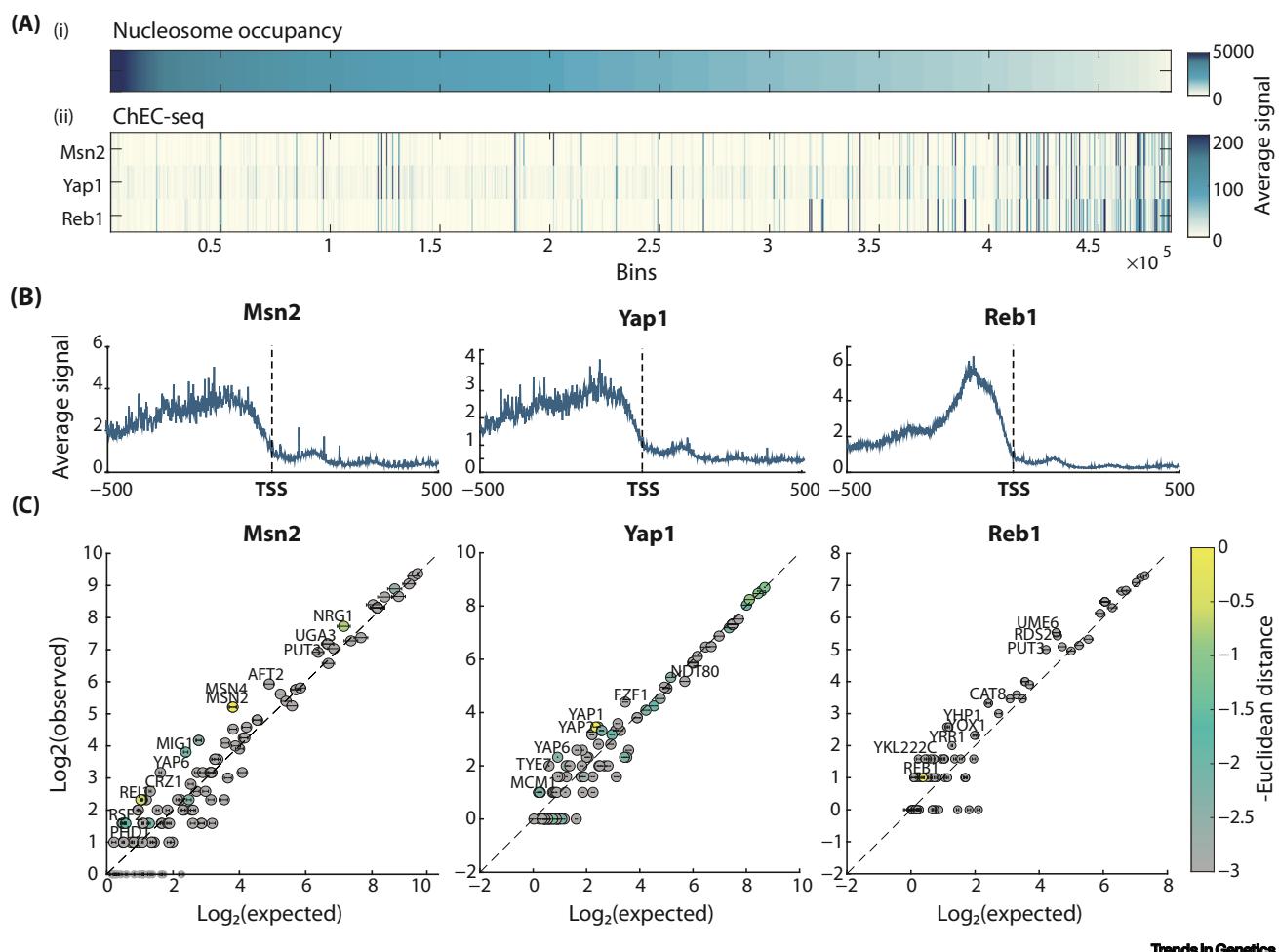
Motif-searching methods, both *in vivo* and *in vitro*, may miss contributions of DNA sequences located away from the core motif [79]. Such contributions were suggested by targeted analysis that specifically examined DNA sequences outside the core [78–87]. For example, *in vitro* analysis of the budding yeast TF Pho4 revealed distantly positioned flanking DNA sequences that affect binding affinity as much as the core motif itself [81]. How such sequences can contribute to binding without directly contacting the DBD is less clear. One compelling hypothesis is that flanking sequences modulate the 1D sliding on DNA and, in this way, increase the effective target site [79]. However, strong binding to flanking regions could trap the TF and limit its arrival at the desired target sites *in vivo*. In addition, the consistency between the *in vitro* and *in vivo* preferences for sequences flanking the core appears low [81], in contrast to the high consistency between the *in vivo* and *in vitro* core-motif preferences.

### DNA Accessibility

In addition to their intrinsic DBD preferences, TF binding to DNA *in vivo* depends on the cellular environment. In particular, the ability of regulatory factors to access DNA varies along the chromatin-packed eukaryotic genomes [89]. The wrapping of DNA within nucleosomes may seclude potential binding sites with regions of heterochromatin being of particularly low accessibility. In terms of the search kinetics, differential DNA accessibility could accelerate the TF target search by reducing the effective genome size, but might also slow it down by shortening the

sliding windows scanned at each association event [7]. Quantitative estimates using measured parameters of the budding yeast TF Mbp1 suggested that limiting the search to nucleosome-free regions only would slow down, rather than accelerate, binding site detection [46].

Experimentally, TFs are depleted from nucleosome-associated DNA [17,90–92] (Figure 2A). However, this depletion is only partial, with some TFs effectively competing with nucleosomes for DNA binding [93–98]. Furthermore, nucleosome depletion is often the consequence, rather than the cause, of TF binding [99], making it difficult to discern the contribution of nucleosomes to binding specificity. Notable also is the depletion of TFs from coding sequences, which may be due to high nucleosome occupancy, or, in highly expressed genes, to the density of propagating polymerases (Figure 2B).



**Figure 2. Transcription Factors (TFs) Are Generally Depleted from Nucleosome-Associated DNA and Their Binding Sites Are Rarely Linked to other TF-Binding Sites.** (A) TFs are generally depleted from nucleosome-associated DNA. The *Saccharomyces cerevisiae* genome was divided into evenly sized bins (25 bp). (i) Average nucleosome occupancy in these bins, ordered based on the average signal measured by MNase-seq. (ii) Binding signal of the indicated TFs, measured by chromatin endogenous cleavage (ChEC)-seq [bins are ordered as in (i)]. (B) TFs are generally depleted from coding sequences. Meta-genes showing the average binding signal detected for the indicated TFs over all *S. cerevisiae* promoters. Broken line indicates the transcription start site (TSS). Data from [131]. (C) TF-binding sites are rarely linked together: The number of events in which a motif of the indicated TFs was found in proximity to that of another TF is shown as a function of the number of such events expected by random (data are presented in log scale, error bars represent standard deviation; see supplemental information online for further details). Color indicates the PWM similarities of the different factors to that of the indicated TF (same measure as in Figure 1E).

### Cooperative Binding

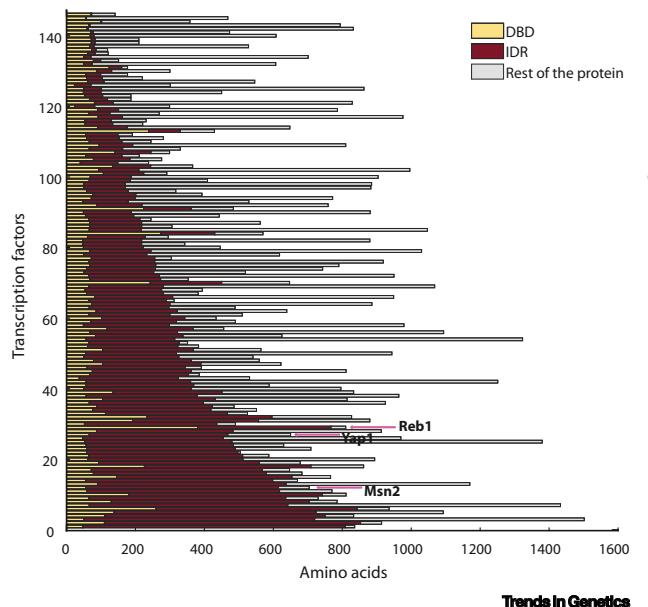
The limited information content provided by each individual DBD can be increased *in vivo* by the formation of multi-TF complexes that preferentially bind to composites of their individual motifs. Such composite motifs characterize the obligatory dimers basic leucine zipper (bZip) and basic helix-loop-helix (bHLH) DBD families, although the motifs remain short in these cases. Longer composite motifs include the budding yeast 31-bp regulatory sequence that is bound by the TFs Mcm1 and MF alpha-2 upstream of mating genes [100–102] and the 11-bp composite FOX:ETS binding element found in endothelial-specific enhancers in humans [103]. Therefore, cooperative binding of interacting TFs is a theoretically compelling solution explaining *in vivo* binding specificity. However, genomic analyses have not yet provided evidence of the widespread prevalence of composite motifs (Figure 2C), although these may be difficult to discern if allowing flexible spacing. Furthermore, regulatory regions experience rapid motif turnover and rearrangements [104–107] and tolerate shifts in motif positions or incorporation of new binding sites [104,108,109], contrasting the expected stringent constraints associated with the co-binding of TFs.

The apparent lack of evolutionary constraints on the orientation, order, and distance between motifs questions the prevalence of TF co-binding as a principal specificity mechanism. Therefore, TFs may interact more often following DNA binding. For example, the interaction between the mammalian TFs Sox2 and Oct4 occurs following Sox2 binding to DNA, and acts to stabilize this binding [56]. Similarly, cooperation between the budding yeast TFs Mcm1 and Rap1 is explained by their mutual interaction with the general transcription factor TFIID [110]. More generally, TFs could cooperate indirectly by displacing nucleosomes [111] or bending the DNA [112,113] in a manner that promotes TF binding to adjacent sites. Such indirect cooperation may explain the increased tendency of some motifs to co-occur in the same promoters [114], the localization of multiple TFs to the same enhancer regions [115–118], and the prevalence of multiple, single-TF footprints within DNase-hypersensitive regions [18].

Cooperative binding can promote specificity regardless of whether it results from co-binding of TF complexes or from the independent arrivals of individual TFs. However, the consequences of such interactions for the TF search speed differ greatly. Co-binding of interacting TFs will have limited effect on the search kinetics, perhaps slowing it down by reducing diffusion rates and increasing the number of possible nonspecific genomic traps. Postbinding recruitment could promote specificity and perhaps the detection speed of the recruited TF, but will leave open the question of how the recruiting TF reaches its position. To promote binding specificity of all interacting TFs, stable binding would require the interacting TFs to be simultaneously present within the same regulatory region. In this case, the search time will greatly increase, because the probability that two independently searching TFs will reach an adjacent binding motif within a limited time frame appears exceedingly low, raising the question of whether postbinding stabilization is a plausible specificity mechanism. Rigorous analysis of these possible scenarios is needed to resolve their compatibility with the *in vivo* search kinetics.

### A Role for Long IDRs in the TF Target Search

As discussed earlier, eukaryotic TFs differ from their bacterial counterparts by using DBDs that bind DNA motifs of low information content. An additional distinctive feature of eukaryotic TFs is the enrichment of long IDRs outside their DBDs, which, again, is not seen in bacteria or archaea [119–125] (Figure 3). These IDRs contribute to the formation of **phase-separated condensates** [125–128], which concentrate the transcriptional apparatus at specific **superenhancers** [129,130]. A recent study explored whether IDRs contribute to TF-target recognition.



**Figure 3. Transcription Factors (TFs) Are Enriched with Intrinsically Disordered Regions (IDRs).** The number of amino acids comprising the DNA-binding domains (DBDs) and IDRs is shown for all *Saccharomyces cerevisiae* TFs. TFs are ordered based on the length of their IDR.

Msn2 and Yap1 are budding yeast TFs containing >500 amino acid (aa)-long IDRs [131]. Msn2 is a zinc-finger TF activated by general stresses, while Yap1 is a member of the bZip family activated in response to oxidative stress. Unexpectedly, it was found that both TFs use their IDRs to recognize their target promoters. A series of observations led to this conclusion. First, it was found that the DBDs do not account for the *in vivo* binding specificity, because the subset of motif-containing sites bound by the DBDs, on their own, were largely different from those bound by the intact TFs. Conversely, TF mutants that lacked the respective DBDs still localized to most of the promoters preferred by the intact TFs, although losing the preference for binding at the DBD-bound motif. This promoter selection was invariant to the deletion of the few structured regions embedded within the otherwise disordered non-DBD or to deletion of 200-aa segments spanning the full non-DBD. Furthermore, swapping the non-DBDs between distant orthologs retained the binding profile, despite little (alignment-based) sequence similarity. Most notably, gradually shortening the IDRs beyond 200 residues led to a gradual shift in the promoter binding profile.

These results suggest a novel paradigm, explaining specificity through distributed specificity. In this paradigm, promoter binding depends on multiple, weak, and partially redundant specificity determinants distributed throughout an extended IDR. The detection of target sites then follows a two-step process: the IDR recognizes broad DNA regions, within which the DBD detects its preferred, short binding motif.

A central question is how IDRs recognize specific genomic loci. One possibility is that IDRs bind DNA directly, with a preference for specific promoters. Similar to the (short) disordered AT hook domains that specifically recognize minor-groove A/T-rich DNA [119,132], extended IDRs could include a multiplicity of peptide motifs, such as the TF prevalent **molecular recognition features (MoRFs)**, which may recognize specific DNA sequences or geometrical features. Most notable are MoRFs predicted to form a helical structure upon binding to nucleic acid or protein partners ( $\alpha$ -MoRFs [119,133]). IDRs could also interact with chromatin or be recruited by other

DNA-binding proteins. The latter proposal, which initially appeared most promising, was tested, but deletion of known interacting partners of both Msn2 and Yap1 had no detectable effect on their binding profiles. Furthermore, sequence motifs in IDR-bound promoters that could serve as binding sites of potential recruiting TFs were searched for, but no evidence for such potential collaborating DNA binding partners was detected. Further analysis should help define the molecular basis of IDR-based promoter recognition.

What are the potential consequences of IDR-based promoter recognition on the speed of target-site detection? We propose that IDR-based promoter recognition limits the search space in large genomes by increasing the association rate, or residence time at a specific subset of promoters. In principle, this could occur through two different mechanisms. First, IDRs contribute to the formation of phase-separated condensates, which also include co-activators and components of the general transcription machinery [125–130]. If a TF is attracted to such potential condensates, the search for its targets could be accelerated. Alternatively, association with DNA through long IDRs may allow recognition of broad DNA regions surrounding the core sequence motif, thereby increasing the effective target size. The local search of the DBD for its preferred motif may be further facilitated through the flexible IDR, thereby overcoming the difficulty of retaining a DNA association while sliding along the promoter [45]. Through a similar effect, long IDRs may resolve the speed–stability paradox limiting the sliding speed in the rough energy landscapes required for stable binding [7,44]. However, further experiments are required to examine these potential contributions.

### Concluding Remarks

In this opinion article, we discussed two challenges limiting our current understanding of the eukaryotic transcription logic: the highly specific binding of TFs to a subset of their potential binding sites *in vivo*, and the rapid detection of these selected sites within large genomes. Specificity and detection speed are intimately linked, because both are limited by the low information content and discriminative power of the DBD-recognized sequence motifs. We argued that mechanisms proposed for promoting binding specificity should be viewed not only with regards to their ability to distinguish potential binding sites, but also concerning their impact on the detection speed and highlighted potential trade-offs between them. Specificity–speed trade-offs appear particularly limiting within the prevailing model of specificity through cooperative TF binding, calling for a rigorous theoretical and experimental analysis of the multi-TF target search process. Finally, based on a recent study of two budding yeast TFs, we presented the hypothesis that long and flexible IDRs, characterizing a large fraction of eukaryotic TFs, may promote both specificity and high detection speed. Future experiments will test this new paradigm and define its generality and implications (see Outstanding Questions).

### Supplemental Information

Supplemental information associated with this article can be found online at <https://doi.org/10.1016/j.tig.2020.12.001>.

### References

- Charoensawan, V. et al. (2010) Lineage-specific expansion of DNA-binding transcription factor families. *Trends Genet.* 26, 388–393
- Lambert, S.A. et al. (2018) The human transcription factors. *Cell* 172, 650–665
- Charoensawan, V. et al. (2010) Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucleic Acids Res.* 38, 7364–7377
- de Boer, C.G. and Hughes, T.R. (2012) YeTFaSCo: a database of evaluated yeast transcription factor sequence specificities. *Nucleic Acids Res.* 40, D169–D179
- Wunderlich, Z. and Miry, L.A. (2009) Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet.* 25, 434–440
- Wilson, D. et al. (2008) DBD–taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res.* 36, D88–D92
- Miry, L. et al. (2009) How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A Math. Theor.* 42, 434013
- Eckweiler, D. et al. (2018) PRODORIC2: the bacterial gene regulation database in 2018. *Nucleic Acids Res.* 46, D320–D326

### Outstanding Questions

What is the contribution of sequences flanking the core DBD in promoting TF binding specificity, and what are the effects on target detection speed?

What is the prevalence and consequences of TF co-binding on the *in vivo* TF-target search?

How widespread is the use of IDRs for promoting binding specificity?

What are the effects of long TF IDRs on target detection speed?

What is the molecular basis behind the specific promoter recognition of TF IDRs?

9. Hashimoto, H. *et al.* (2017) Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. *Mol. Cell* 66, 711–720
10. Pavletich, N.P. and Pabo, C.O. (1993) Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* 261, 1701–1707
11. Billings, T. *et al.* (2013) DNA binding specificities of the long zinc-finger recombination protein PRDM9. *Genome Biol.* 14, R35
12. Zuo, Z. *et al.* (2019) Why do long zinc finger proteins have short motifs? *bioRxiv* Published online May 15, 2019. <https://doi.org/10.1101/637298>
13. Kribelbauer, J.F. *et al.* (2019) Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes. *Annu. Rev. Cell Dev. Biol.* 35, 357–379
14. Badis, G. *et al.* (2009) Diversity and complexity in DNA recognition by transcription factors. *Science* 324, 1720–1723
15. Weirauch, M.T. *et al.* (2014) Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 158, 1431–1443
16. Harbison, C.T. *et al.* (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* 431, 99–104
17. Wang, J. *et al.* (2012) Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* 22, 1798–1812
18. Vierstra, J. *et al.* (2020) Global reference mapping of human transcription factor footprints. *Nature* 583, 729–736
19. Meuleman, W. *et al.* (2020) Index and biological spectrum of human DNase I hypersensitive sites. *Nature* 584, 244–251
20. Neph, S. *et al.* (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 489, 83–90
21. Gasch, A.P. *et al.* (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* 11, 4241–4257
22. Chapal, M. *et al.* (2019) Resolving noise-control conflict by gene duplication. *PLoS Biol.* 17, e3000289
23. Zobbeck, K.L. *et al.* (2010) Recruitment timing and dynamics of transcription factors at the Hsp70 loci in living cells. *Mol. Cell* 40, 965–975
24. Gorman, J. and Greene, E.C. (2008) Visualizing one-dimensional diffusion of proteins along DNA. *Nat. Struct. Mol. Biol.* 15, 768–774
25. Suter, D.M. (2020) Transcription factors and DNA play hide and seek. *Trends Cell Biol.* 30, 491–500
26. Riggs, A.D. *et al.* (1970) The lac repressor-operator interaction: III. Kinetic studies. *J. Mol. Biol.* 53, 401–417
27. Berg, O.G. and Ehrenberg, M. (1982) Association kinetics with coupled three- and one-dimensional diffusion: chain-length dependence of the association rate to specific DNA sites. *Biophys. Chem.* 15, 41–51
28. Berg, O.G. *et al.* (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 20, 6929–6948
29. Winter, R.B. *et al.* (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor-operator interaction: kinetic measurements and conclusions. *Biochemistry* 20, 6961–6977
30. von Hippel, P.H. and Berg, O.G. (1989) Facilitated target location in biological systems. *J. Biol. Chem.* 264, 675–678
31. Adam, G. and Delbrück, M. (1967) Reduction of dimensionality in biological diffusion processes. In *Structural Chemistry and Molecular Biology* (Rich, A. and Davidson, N., eds), pp. 198–215, W.H. Freeman and Co
32. Richter, P.H. and Eigen, M. (1974) Diffusion controlled reaction rates in spheroidal geometry: application to repressor-operator association and membrane bound enzymes. *Biophys. Chem.* 2, 255–263
33. Berg, H.C. and Purcell, E.M. (1977) Physics of chemoreception. *Biophys. J.* 20, 193–219
34. Hu, T. *et al.* (2006) How proteins search for their specific sites on DNA: the role of DNA conformation. *Biophys. J.* 90, 2731–2744
35. Kim, J.G. *et al.* (1987) Kinetic studies on Cro repressor-operator DNA interaction. *J. Mol. Biol.* 196, 149–158
36. Elf, J. *et al.* (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316, 1191–1194
37. Gowers, D.M. *et al.* (2005) Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15883–15888
38. McKinney, K. *et al.* (2004) p53 linear diffusion along DNA requires its C terminus. *Mol. Cell* 16, 413–424
39. Marklund, E. *et al.* (2020) DNA surface exploration and operator bypassing during target search. *Nature* 583, 858–861
40. Hammar, P. *et al.* (2012) The lac repressor displays facilitated diffusion in living cells. *Science* 336, 1595–1598
41. Halford, S.E. and Marko, J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* 32, 3040–3052
42. Wunderlich, Z. and Mirny, L.A. (2008) Spatial effects on the speed and reliability of protein-DNA search. *Nucleic Acids Res.* 36, 3570–3578
43. Coppey, M. *et al.* (2004) Kinetics of target site localization of a protein on DNA: a stochastic approach. *Biophys. J.* 87, 1640–1649
44. Slutsky, M. and Mirny, L.A. (2004) Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* 87, 4021–4035
45. Halford, S.E. (2009) An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans.* 37, 343–348
46. Larson, D.R. *et al.* (2011) Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* 332, 475–478
47. Liu, Z. and Tjian, R. (2018) Visualizing transcription factor dynamics in living cells. *J. Cell Biol.* 217, 1181–1191
48. Swinstead, E.E. *et al.* (2016) Steroid receptors reprogram FoxA1 occupancy through dynamic chromatin transitions. *Cell* 165, 593–605
49. Normanno, D. *et al.* (2012) Intra-nuclear mobility and target search mechanisms of transcription factors: A single-molecule perspective on gene expression. *Biochim. Biophys. Acta (BBA) Gene Regul. Mech.* 1819, 482–493
50. van Royen, M.E. *et al.* (2011) Nuclear proteins: finding and binding target sites in chromatin. *Chromosom. Res.* 19, 83–98
51. Coulon, A. *et al.* (2013) Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat. Rev. Genet.* 14, 572–584
52. Brouwer, I. and Lenstra, T.L. (2019) Visualizing transcription: key to understanding gene expression dynamics. *Curr. Opin. Chem. Biol.* 51, 122–129
53. Mortensen, K.I. *et al.* (2010) Optimized localization analysis for single-molecule tracking and super-resolution microscopy. *Nat. Methods* 7, 377–381
54. Ball, D.A. *et al.* (2016) Single molecule tracking of Ace1p in *Saccharomyces cerevisiae* defines a characteristic residence time for non-specific interactions of transcription factors with chromatin. *Nucleic Acids Res.* 44, e160
55. Lerner, J. *et al.* (2020) Two-parameter mobility assessments discriminate diverse regulatory factor behaviors in chromatin. *Mol. Cell* 79, 677–688
56. Chen, J. *et al.* (2014) Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* 156, 1274–1285
57. Donovan, B.T. *et al.* (2019) Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting. *EMBO J.* 38, e100809
58. Louphrasithiphol, P. *et al.* (2020) Tuning transcription factor availability through acetylation-mediated genomic redistribution. *Mol. Cell* 79, 472–487
59. Presman, D.M. *et al.* (2017) Quantifying transcription factor binding dynamics at the single-molecule level in live cells. *Methods* 123, 76–88
60. Mehta, G.D. *et al.* (2018) Single-molecule analysis reveals linked cycles of RSC chromatin remodeling and Ace1p transcription factor binding in yeast. *Mol. Cell* 72, 875–887
61. Stanford, N.P. *et al.* (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites. *EMBO J.* 19, 6546–6557
62. Bonnet, I. *et al.* (2008) Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucleic Acids Res.* 36, 4118–4127
63. Porecha, R.H. and Stivers, J.T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc. Natl. Acad. Sci.* 105, 10791–10796
64. MacIsaac, K.D. *et al.* (2006) An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* 7, 113

65. Wang, Y.M. *et al.* (2006) Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys. Rev. Lett.* 97, 048302
66. Biebricher, A. *et al.* (2009) Tracking of single quantum dot labeled EcoRV sliding along DNA manipulated by double optical tweezers. *Biophys. J.* 96, L50–L52
67. Kim, J.H. and Larson, R.G. (2007) Single-molecule analysis of 1D diffusion and transcription elongation of T7 RNA polymerase along individual stretched DNA molecules. *Nucleic Acids Res.* 35, 3848–3858
68. Ricchetti, M. *et al.* (1988) One-dimensional diffusion of *Escherichia coli*/DNA-dependent RNA polymerase: a mechanism to facilitate promoter location. *Proc. Natl. Acad. Sci. U. S. A.* 85, 4610–4614
69. Harada, Y. *et al.* (1999) Single-molecule imaging of RNA polymerase-DNA interactions in real time. *Biophys. J.* 76, 709–715
70. Blainey, P.C. *et al.* (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5752–5757
71. Tafti, A. *et al.* (2008) Tumor suppressor p53 slides on DNA with low friction and high stability. *Biophys. J.* 95, L01–L03
72. Komazin-Meredith, G. *et al.* (2008) Hopping of a processivity factor on RNA revealed by single-molecule assays of diffusion. *Proc. Natl. Acad. Sci. U. S. A.* 105, 10721–10726
73. Elson, C.M. *et al.* (2010) Thioredoxin suppresses microscopic hopping of T7 DNA polymerase on duplex DNA. *Proc. Natl. Acad. Sci. U. S. A.* 107, 1900–1905
74. Lin, Y. *et al.* (2009) Using the bias from flow to elucidate single DNA repair protein sliding and interactions with DNA. *Biophys. J.* 96, 1911–1917
75. Chen, K. and Rajewsky, N. (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.* 8, 93–103
76. Hsia, C.C. and McGinnis, W. (2003) Evolution of transcription factor function. *Curr. Opin. Genet. Dev.* 13, 199–206
77. Vaquerizas, J.M. *et al.* (2009) A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* 10, 252–263
78. Shen, N. *et al.* (2018) Divergence in DNA specificity among paralogous transcription factors contributes to their differential *in vivo* binding. *Cell Systems* 6, 470–483
79. Dror, I. *et al.* (2016) How motif environment influences transcription factor search dynamics: finding a needle in a haystack. *BioEssays* 38, 605–612
80. White, M.A. *et al.* (2013) Massively parallel *in vivo* enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc. Natl. Acad. Sci. U. S. A.* 110, 11952–11957
81. Le, D.D. *et al.* (2018) Comprehensive, high-resolution binding energy landscapes reveal context dependencies of transcription factor binding. *Proc. Natl. Acad. Sci.* 115, E3702–E3711
82. Maerk, S.J. and Quake, S.R. (2007) A systems approach to measuring the binding energy landscapes of transcription factors. *Science* 315, 233–237
83. Gordán, R. *et al.* (2013) Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.* 3, 1093–1104
84. Siggers, T. *et al.* (2014) Diversification of transcription factor paralogs via noncanonical modularity in C2H2 zinc finger DNA binding. *Mol. Cell* 55, 640–648
85. Levo, M. and Segal, E. (2014) In pursuit of design principles of regulatory sequences. *Nat. Rev. Genet.* 15, 453–468
86. Afek, A. *et al.* (2014) Protein-DNA binding in the absence of specific base-pair recognition. *Proc. Natl. Acad. Sci. U. S. A.* 111, 17140–17145
87. Castellanos, M. *et al.* (2020) Eukaryotic transcription factors can track and control their target genes using DNA antennas. *Nat. Commun.* 11, 540
88. Joshi, R. *et al.* (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* 131, 530–543
89. Weintraub, H. and Groudine, M. (1976) Chromosomal subunits in active genes have an altered conformation. *Science* 193, 848–856
90. Li, X.Y. *et al.* (2011) The role of chromatin accessibility in directing the widespread, overlapping patterns of *Drosophila* transcription factor binding. *Genome Biol.* 12, R34
91. Guerlin, M.J. and Lis, J.T. (2010) Chromatin landscape dictates HSF binding to target DNA elements. *PLoS Genet.* 6, e1001114
92. John, S. *et al.* (2011) Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* 43, 264–268
93. Donovan, B.T. *et al.* (2019) Dissociation rate compensation mechanism for budding yeast pioneer transcription factors. *eLife* 8, e43008
94. Chereji, R.V. *et al.* (2019) Accessibility of promoter DNA is not the primary determinant of chromatin-mediated gene regulation. *Genome Res.* 29, 1985–1995
95. Sun, Y. *et al.* (2015) Zelda overcomes the high intrinsic nucleosome barrier at enhancers during *Drosophila* zygotic genome activation. *Genome Res.* 25, 1703–1714
96. Zhu, F. *et al.* (2018) The interaction landscape between transcription factors and the nucleosome. *Nature* 562, 76–81
97. Ballaré, C. *et al.* (2013) Nucleosome-driven transcription factor binding and gene regulation. *Mol. Cell* 49, 67–79
98. Soufi, A. *et al.* (2015) Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* 161, 555–568
99. Yan, C. *et al.* (2018) Systematic study of nucleosome-displacing factors in budding yeast. *Mol. Cell* 71, 294–305
100. Smith, D.L. and Johnson, A.D. (1992) A molecular mechanism for combinatorial control in yeast: MCM1 protein sets the spacing and orientation of the homeodomains of an alpha 2 dimer. *Cell* 68, 133–142
101. Tan, S. and Richmond, T.J. (1998) Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. *Nature* 391, 660–666
102. Johnson, A.D. and Herskowitz, I. (1985) A repressor (MAT alpha 2 Product) and its operator control expression of a set of cell type specific genes in yeast. *Cell* 42, 237–247
103. De Val, S. *et al.* (2008) Combinatorial regulation of endothelial gene expression by Ets and Forkhead transcription factors. *Cell* 135, 1053–1064
104. Hare, E.E. *et al.* (2008) Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genet.* 4, e1000106
105. Rastegar, S. *et al.* (2008) The words of the regulatory code are arranged in a variable manner in highly conserved enhancers. *Dev. Biol.* 318, 366–377
106. Dermitzakis, E.T. and Clark, A.G. (2002) Evolution of transcription factor binding sites in mammalian gene regulatory regions: conservation and turnover. *Mol. Biol. Evol.* 19, 1114–1121
107. Hess, G.D. and Riegle, G.D. (1972) Effects of chronic ACTH stimulation on adrenocortical function in young and aged rats. *Am. J. Phys.* 222, 1458–1461
108. Arnosti, D.N. and Kulkarni, M.M. (2005) Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J. Cell. Biochem.* 94, 890–898
109. Lin, Y.S. *et al.* (1990) How different eukaryotic transcriptional activators can cooperate promiscuously. *Nature* 345, 359–361
110. Sorrelles, T.R. *et al.* (2018) Intrinsic cooperativity potentiates parallel cis-regulatory evolution. *eLife* 7, e37563
111. Mirny, L.A. (2010) Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci.* 107, 22534–22539
112. Jolma, A. *et al.* (2015) DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* 527, 384–388
113. Morgunova, E. and Taipale, J. (2017) Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.* 47, 1–8
114. Sudarsanam, P. *et al.* (2002) Genome-wide co-occurrence of promoter elements reveals a cis-regulatory cassette of rRNA transcription motifs in *Saccharomyces cerevisiae*. *Genome Res.* 12, 1723–1731
115. Berman, B.P. *et al.* (2004) Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Genome Biol.* 5, R61
116. Papatsenko, D.A. *et al.* (2002) Extraction of functional binding sites from unique regulatory regions: the *Drosophila* early developmental enhancers. *Genome Res.* 12, 470–481
117. Schroeder, M.D. *et al.* (2004) Transcriptional control in the segmentation gene network of *Drosophila*. *PLoS Biol.* 2, E271

118. Hnisz, D. *et al.* (2013) Super-enhancers in the control of cell identity and disease. *Cell* 155, 934–947
119. Liu, J. *et al.* (2006) Intrinsic disorder in transcription factors. *Biochemistry* 45, 6873–6888
120. Ward, J.J. *et al.* (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* 337, 635–645
121. Minezaki, Y. *et al.* (2006) Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol.* 359, 1137–1149
122. Peng, Z. *et al.* (2014) Genome-scale prediction of proteins with long intrinsically disordered regions. *Proteins* 82, 145–158
123. Wang, C. *et al.* (2016) Disordered nucleome: abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea. *Proteomics* 16, 1486–1498
124. Skupien-Rabian, B. *et al.* (2016) Proteomic and bioinformatic analysis of a nuclear intrinsically disordered proteome. *J. Proteome* 130, 76–84
125. van der Lee, R. *et al.* (2014) Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114, 6589–6631
126. Elbaum-Garfinkle, S. *et al.* (2015) The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci.* 112, 7189–7194
127. Wang, J. *et al.* (2018) A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* 174, 688–699
128. Tompa, P. *et al.* (2015) Intrinsically disordered proteins: emerging interaction specialists. *Curr. Opin. Struct. Biol.* 35, 49–59
129. Boja, A. *et al.* (2018) Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* 175, 1842–1855
130. Sabari, B.R. *et al.* (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science (New York, N.Y.)* 361, eaar3958
131. Brodsky, S. *et al.* (2020) Intrinsically disordered regions direct transcription factor in vivo binding specificity. *Mol. Cell* 79, 459–471
132. Bewley, C.A. *et al.* (1998) Minor groove-binding architectural proteins: structure, function, and DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.* 27, 105–131
133. Oldfield, C.J. *et al.* (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* 44, 12454–12470
134. Letunic, I. and Bork, P. (2017) 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* 46, D493–D496
135. Krieger, G. *et al.* (2020) Independent evolution of transcript abundance and gene regulatory dynamics. *Genome Res.* 30, 1000–1011