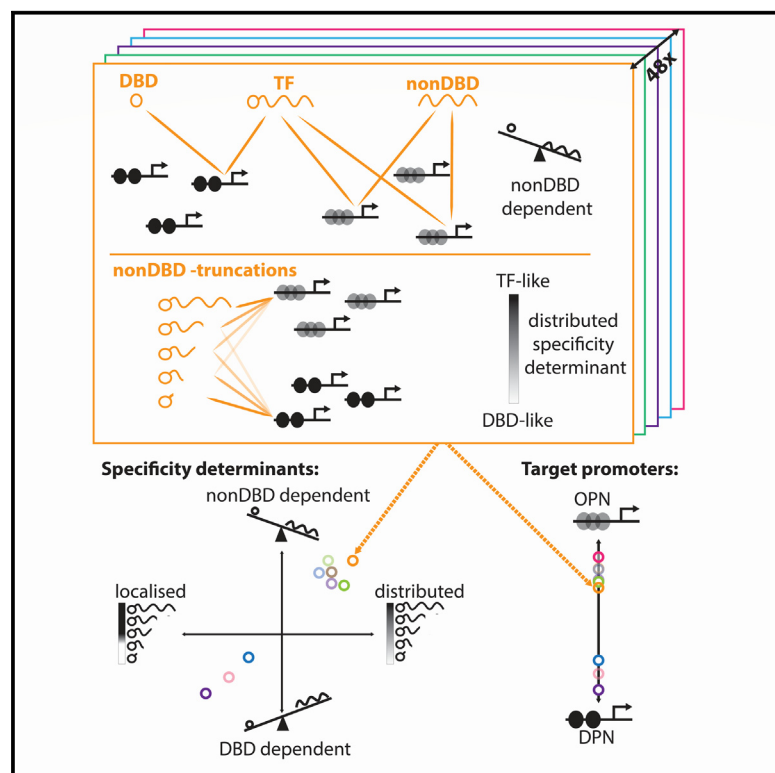


Complementary strategies for directing *in vivo* transcription factor binding through DNA binding domains and intrinsically disordered regions

Graphical abstract



Authors

Divya Krishna Kumar, Felix Jonas, Tamar Jana, Sagie Brodsky, Miri Carmi, Naama Barkai

Correspondence

naama.barkai@weizmann.ac.il

In brief

To define the molecular basis of transcription factor (TF) specificity, Kumar et al. dissected the binding determinants of 48 yeast TFs. Most TFs contained multiple weak specificity determinants distributed over long, intrinsically disordered regions outside their DNA binding domains (DBDs), which biased TF binding toward promoters with a distinct nucleosome architecture.

Highlights

- Systematic truncations and genomic mapping define TF binding specificity determinants
- Most TFs depend on many weak determinants distributed within long IDRs outside DBDs
- TFs lacking DBDs often localize to specific genes, including the right targets
- Disordered non-DBDs explain TF preference for binding promoters of fuzzy nucleosomes



Article

Complementary strategies for directing *in vivo* transcription factor binding through DNA binding domains and intrinsically disordered regions

Divya Krishna Kumar,^{1,2} Felix Jonas,^{1,2} Tamar Jana,¹ Sagie Brodsky,¹ Miri Carmi,¹ and Naama Barkai^{1,3,*}¹Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel²These authors contributed equally³Lead contact*Correspondence: naama.barkai@weizmann.ac.il<https://doi.org/10.1016/j.molcel.2023.04.002>

SUMMARY

DNA binding domains (DBDs) of transcription factors (TFs) recognize DNA sequence motifs that are highly abundant in genomes. Within cells, TFs bind a subset of motif-containing sites as directed by either their DBDs or DBD-external (nonDBD) sequences. To define the relative roles of DBDs and nonDBDs in directing binding preferences, we compared the genome-wide binding of 48 (~30%) budding yeast TFs with their DBD-only, nonDBD-truncated, and nonDBD-only mutants. With a few exceptions, binding locations differed between DBDs and TFs, resulting from the cumulative action of multiple determinants mapped mostly to disordered nonDBD regions. Furthermore, TFs' preferences for promoters of the fuzzy nucleosome architecture were lost in DBD-only mutants, whose binding spread across promoters, implicating nonDBDs' preferences in this hallmark of budding yeast regulatory design. We conclude that DBDs and nonDBDs employ complementary DNA-targeting strategies, whose balance defines TF binding specificity along genomes.

INTRODUCTION

Transcription factors (TFs) regulate gene expression by recruiting the general transcription machinery to particular genes.¹ For this, TFs bind to specific regulatory regions of their target genes. DNA binding domains (DBDs) within TFs recognize DNA sequence motifs located in target regulatory sites. However, DBD-preferred motifs are short and highly abundant in eukaryotic genomes, including many locations with no apparent regulatory roles.^{2–5} Accordingly, only a minority of motif sites are TF-bound. The challenge of TF binding specificity therefore extends beyond detecting DBD-preferred motifs but requires distinguishing the relevant motif occurrences to which TFs should bind.^{6–10}

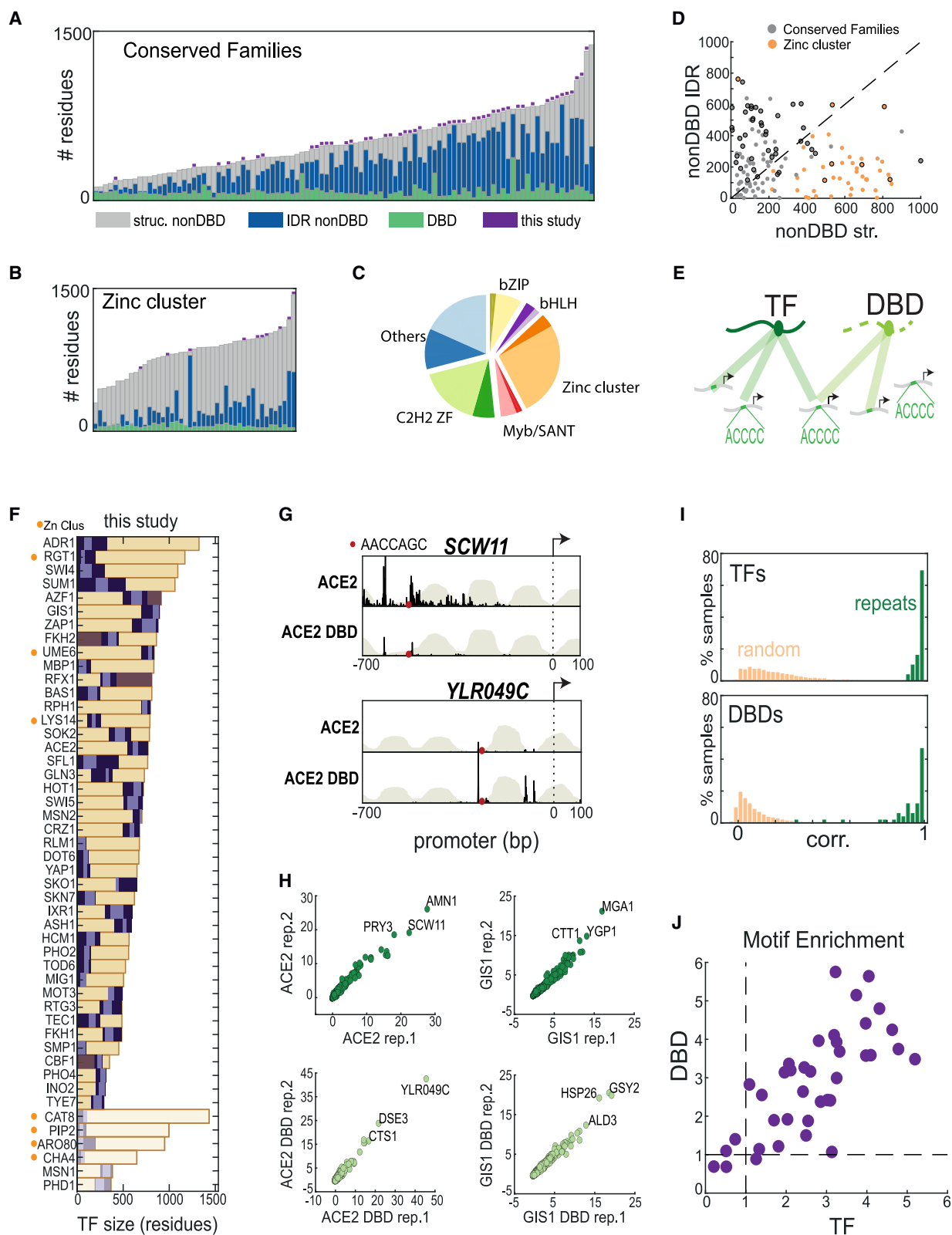
Binding specificity may result from differences in DNA accessibility that impede binding at e.g., nucleosome occupied sites. Other prevailing models include DBD preferences that extend beyond known motifs to include flanking DNA sequences or DNA shape, as well as models of combinatorial motif usage by interacting TFs. Although each of those mechanisms is supported by specific examples, at the genomic scale, the basis of specificity remains elusive.^{6–11}

Focusing on TFs themselves, binding specificity could emerge solely from the DBD or could depend on determinants within the nonDBD. Supporting the former are seminal studies showing that DBDs can activate reporter gene expression when fused

to various activation domains,^{12–15} as well as successful designs of synthetic DBDs targeting particular genomic loci.¹⁶ From an evolutionary perspective, several studies linked differential preferences of TF paralogs to variations in their DBD's sequences.^{17–20} By contrast, the abundance of TFs with similar binding locations,^{21,22} as well as DBD swapping between distant²³ or close²⁴ paralogs, points at nonDBDs as main effectors of binding preferences. Therefore, both the DBDs and the nonDBDs contribute to binding-site selection, but their relative contributions remain unclear.

Distinguishing the contribution of DBDs to TF binding specificity requires measuring their genomic binding upon the removal of respective nonDBDs. Recently, we examined this for two budding yeast DBDs²³ using chromatin endogenous cleavage with high-throughput sequencing (ChEC-seq) which provides high-resolution mapping independent of antibodies or fixation.²⁵ This mapping, coupled with subsequent TF truncation analyses, pointed at the nonDBD as the key in directing promoter preferences of those two TFs and further revealed multiple weak and partially redundant specificity determinants distributed throughout these long (>500 residues) and largely disordered regions.²³

The distribution of specificity determinants within disordered nonDBDs, and their combined action in directing TF binding, present a new paradigm for the TF-target search. To examine the generality of this paradigm among TFs, we examined here



(legend on next page)

48 (~30%) budding yeast TFs, spanning all main DBD families and including the majority of well-characterized, non-essential TFs. We compared the binding profiles of full TFs with their various mutants, including DBD-only, nonDBD truncations, and nonDBD-only mutants. Our data revealed that the majority of nonDBDs contained multiple specificity determinants that cumulatively direct TF binding across the genome. Of note, those determinants are often redundant and localize mostly to intrinsically disordered regions (IDRs). We describe the consequence of these determinants on the global organization of the yeast transcription network and discuss the distinct strategies characterizing the specificity-guiding functions of DBDs and IDR-enriched nonDBDs.

RESULTS

Comparing genomic binding profiles of full TFs to their DBD-only mutants

Models explaining TF binding preferences differ in whether the DBD alone accounts for the *in-vivo* TF binding locations or whether DBD-external regions (nonDBDs) are required. As a first step in dissecting the prevailing basis of specificity, we wished to compare systematically the binding profiles of TFs (full protein) with their corresponding DBD-only mutants. Budding yeast provides the advantage of a compact and well-annotated genome. Of its 147 TFs with well-defined DBDs, 102 belong to conserved eukaryotic families, whereas the remaining ones are of the fungal-specific zinc cluster family (Figures 1A–1C; Table S1). Most nonDBDs are highly disordered, consistent with the characteristic enrichment of IDRs within eukaryotic TFs^{26–29} (Figure 1D). The fungal-specific zinc cluster family is unique in containing a long structured domain (Figures 1B and 1D).

We selected 48 TFs including the majority of well-characterized, non-essential TFs (Figures 1A–1D). DBD-only mutants were generated using CRISPR, and their genomic binding was mapped using ChEC-seq (Figure 1E). For technical reasons, and to account for permissive DBD definitions, we typically retained sequences with varying lengths surrounding the annotated DBDs (Figure 1F; Table S2). For centrally located DBDs, we performed two truncations (from either end) and in some

included the two-sided ones. For simplicity, we still denote all mutants as DBDs.

Previous studies used isolated DBDs for quantifying DNA binding through *in-vitro* biochemical or structural analysis. Accordingly, we expected the DBD mutants to bind genomic DNA with sufficient affinity to allow profiling. This was indeed the case for the majority of DBDs, with only 6 (of 48) exceptions, including 4 of the 7 fungal-specific zinc cluster DBDs tested (Figure 1F). DBDs' binding profiles were of similar quality as full-length TFs, showing high reproducibility (Figures 1G–1I and S1) and the expected localization at their known motifs (Figure 1J). We conclude that the DBDs can bind genomic DNA at sites containing their motifs, even in the absence of nonDBD sequences.

DBDs' binding-site preferences differ from those of full TFs

Two of the prevailing models predict that DBDs are sufficient for recapitulating the TF binding profiles. Those models attribute binding specificity to differential DNA accessibility or to the recognition of motif-flanking DNA sequences or DNA shape by the DBD itself. Our data allow testing this prediction of DBD sufficiency by comparing binding profiles of DBDs with their respective full-length TFs.

First, we considered genomic sites containing the known DBD-preferred *in-vitro* motifs⁴ (Figures 2A–2D). As shown above, these motifs are enriched in the respective DBD- and TF-bound sites (cf. Figure 1J), and we observed that the relative binding signal at these sites was mostly higher for DBDs than the respective TFs (Figures 2A and 2B). Furthermore, a larger fraction of motif-containing sites was bound by DBDs when compared with full-length TFs (Figures 2C and 2D).

Focusing next on regulatory regions, we compared promoters bound by DBDs and respective TFs (Figures 2E and 2F). Only 4 of the 42 DBDs showed a greater than 80% correlation (Pearson's *r*) with their respective TFs, including AZF1, whose DBD uniquely contains four C2H2 zinc fingers, and TYE7, a glycolysis-regulating TF with a dimer-forming bHLH DBD (Figure 2E). More generally, the similarity of DBD-TF pairs was low, with 32/42 showing less than 75% and 19/39 showing less than 50% correlation, and they also differed in the number of bound targets (Figures 2E and 2F). Exemplifying these low similarities are

Figure 1. *In-vivo* binding of DNA binding domains

(A–C) Transcription factor (TF) repertoire of budding yeast: TFs were ordered by size, and the lengths of their DBD (green) as well structured (gray) and disordered (blue) nonDBD are shown, distinguishing TFs of eukaryotic-conserved families (A) and the fungal-specific zinc cluster family (B). The distribution of TFs among families is shown as a pie chart (C). TFs selected for our analysis are highlighted as purple dots (A and B) or as dark region (C). See Table S1 for full list of TFs and described regions.

(D) TFs contain high disordered content: shown is a comparison of the numbers of nonDBD residues located within intrinsically disordered regions (IDRs) (y axis) vs. structured regions (x axis). Selected TFs for analysis and those of the zinc cluster family are indicated.

(E) Experimental approach: to understand the role of DBDs in guiding binding specificity, we compared the genome-wide binding profiles of TFs with their DBD mutants.

(F) DBD-mutants: plot shows DBD-mutants generated in our study, ordered by TF lengths. DBD defined by Pfam, and engineered DBD-mutants are indicated as light and dark purple, respectively. For TFs with an alternative second-side truncation, the less effective mutant is marked by dark brown.

(G–I) DBD-mutants bind at reproducible genomic locations: shown is the binding signal of the indicated TF and DBD-mutants along the indicated promoters (G), as well as a comparison of all promoter preferences obtained in the repeated profiling of the indicated factors (H, see STAR Methods). Distributions of promoter preference correlations (Pearson's *r*) (I) compare reproducibility of biological repeats (green) with correlation between different factors (yellow).

(J) TF and DBD mutants bind preferentially at sites containing their known binding motif: DBD-binding motifs were obtained from CIS-BP, and the top 7-mers matching the *in-vitro* motif were selected. Shown is the enrichment of reads around occurrences of these 7-mers compared with a random promoter region (\log_2 , see STAR Methods). The dashed line indicates the point of 2-fold enrichment.

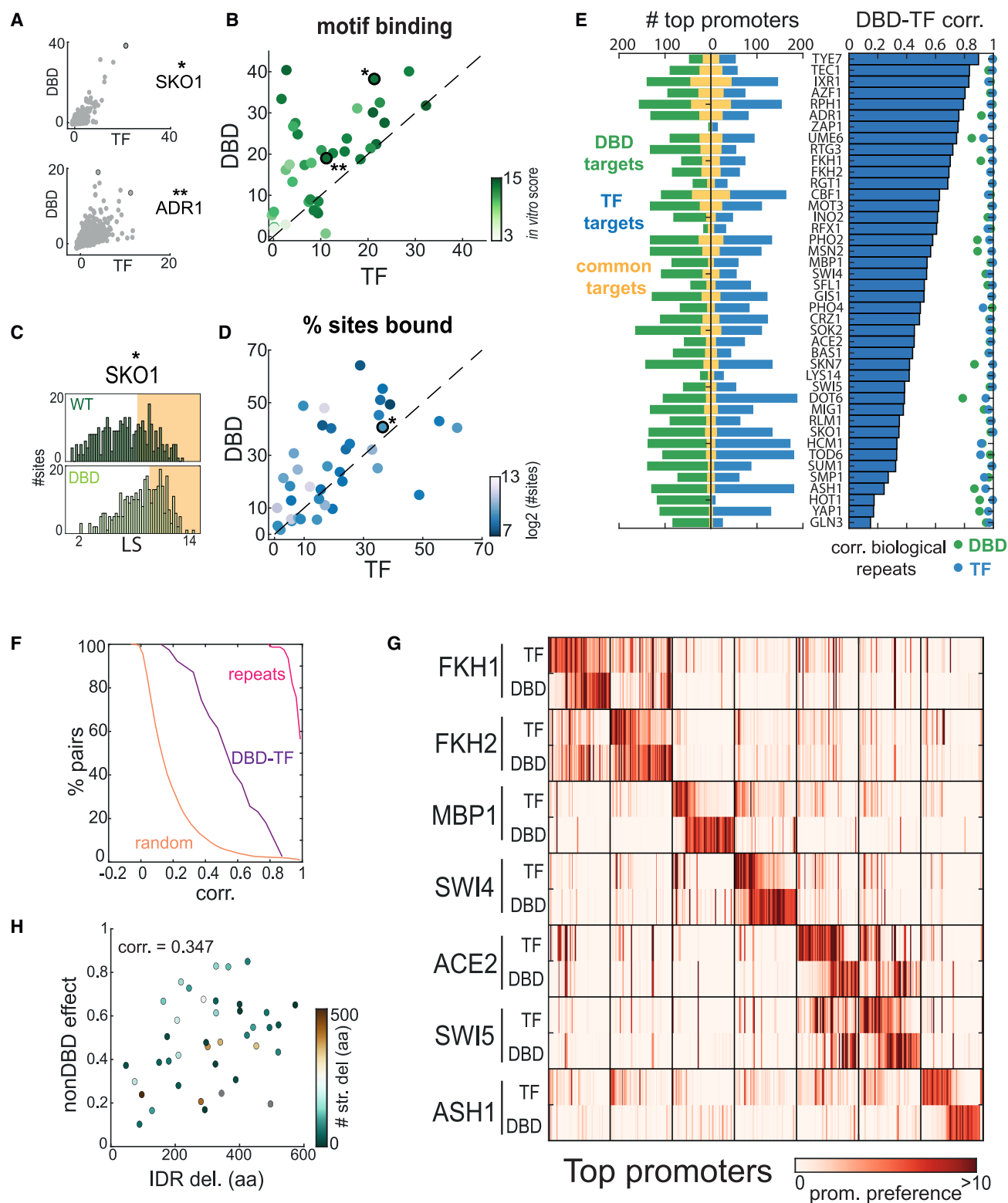


Figure 2. DBDs binding differs from their respective TFs

(A and B) Motif-selection bias is higher for DBDs than their respective TFs: for each 7-mer, the average binding signal found across all promoter occurrences was measured, and compared between DBDs and respective TFs, as indicated (A). Highlighted dot corresponds to the strongest-bound 7-mer (top 7-mer) compatible with known binding motif. Binding signal at this known motif was then compared between DBD and respective TFs of all factors in our set (B).

(legend continued on next page)

cell-cycle TFs (Figure 2G), as well as stress-related activators or repressors (Figure S2).

Given the enrichment of disordered regions within nonDBDs, and the fact that specific DBD-DNA binding involves structure-based complementarities, we asked whether IDR content correlates with DBD-TF similarity. We observed an inverse correlation: nonDBDs with higher disordered content had a stronger influence on binding specificity, when compared with those with lower disorder content ($c = 0.35$) (Figure 2H).

Multiple nonDBD regions contribute to TF binding specificity

The limited similarity between promoter preferences of DBDs and full TFs implicates nonDBDs as central effectors of TF binding specificity. This raised our interest in two principal questions. First, is it a single region within the nonDBD that determines TF binding locations, or do multiple specificity determinants spread across the nonDBD sequence? Second, do specificity determinants within the nonDBD localize to structured regions, or are they found within IDRs? We note that a localized or structured determinant indicates a lock-and-key type recognition, whereas a multiplicity of unstructured determinants is more consistent with fuzzy interactions.

To map specificity determinants outside the DBDs, we sequentially shortened the nonDBDs of our tested TFs (Figure 3A). We aimed for a 50AA resolution and achieved this to various degrees, obtaining ~2–15 successful mutants for 38 profiled TFs. To examine how promoter preferences evolve across the truncation series, we profiled the binding of those mutants, and measured their correlations (Pearson's r) with the respective DBDs and full TFs, positioning each mutant within a mutant-DBD vs. mutant-TF correlation plot (Figures 3A–3D; missing data indicated by gray points positioned at interpolated position). Mutant dynamics within this plot distinguished patterns of specificity determinants (Figure 3A): if a single, localized region confers specificity, initial truncations would all display TF-like binding and cluster on one side of the plot, whereas later ones, deleted of the specificity-causing region, would all show a DBD-like pattern and cluster at the second extreme. By contrast, if multiple nonDBD regions contribute cumulatively and equally, mutants would spread evenly along a line connecting DBD-TF.

In all truncation series, mutant locations defined a TF-DBD connecting line (Figures 3B–3D for one-sided and two-sided truncations, respectively). Furthermore, the mutant location was largely monotonic, with subsequent truncations shifting

the pattern away from the TF and toward the DBD. In some cases, the spread was rather even, indicating a similar contribution of different truncations (e.g., MOT3 and SWI5), whereas in other cases one or few truncations dominated (e.g., SKN7). Examining the associated promoters verified that this gradual change in correlation captured the behavior of individual promoters, with increasing truncations leading to gradual loss of binding signal at TF-preferred promoters and gradual increase at DBD-preferred promoters (Figure 3B).

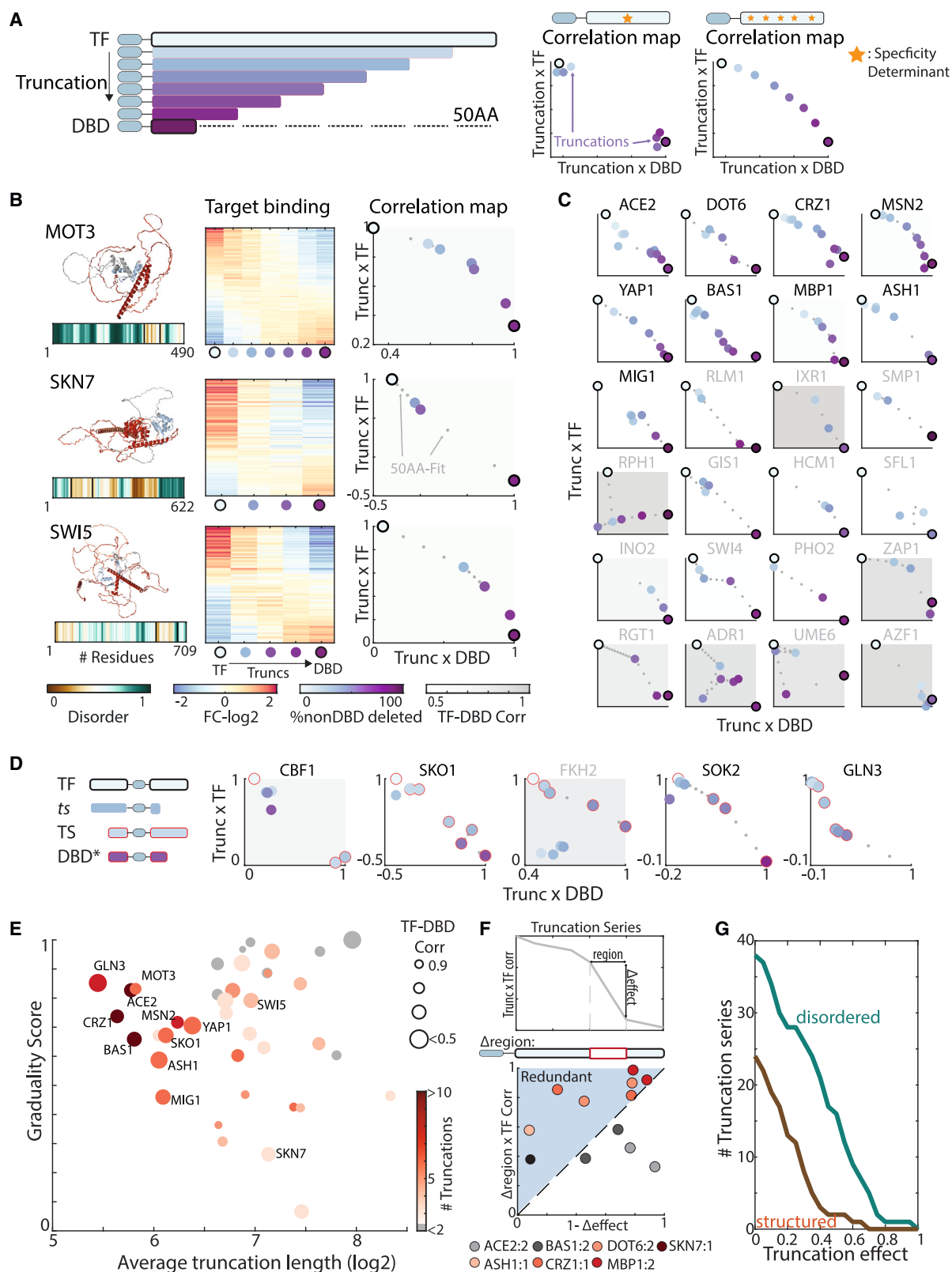
We used the incremental shift in mutant positioning in the mutant-DBD vs. mutant-TF correlation plot as a measure of the specificity effect of the additionally deleted region. Based on these measures, we quantified the distribution of truncation effects, defining a “graduality” score as the ratio of the median to the maximal effect: a graduality score of 1 corresponds to a truncation series where most truncations contribute equally, whereas zero is obtained when the majority of truncations are without effect (Figure 3E; see STAR Methods). As can be appreciated, most TFs followed the gradual design, and this was most apparent within our high-resolution series, including ACE2, CRZ1, and BAS1. Exceptions, in which few truncations dominate, include AZF1 that shows little dependence on its nonDBD (cf. Figure 2E). We note that for some series, we lack the resolution to decide confidently whether specificity is evenly distributed or not (e.g., SKN7). However, even in those cases, the cumulative effect of different truncations suggests the existence of multiple specificity determinants.

We reasoned that in cases of redundancy, our sequential truncations may underestimate the spread of determinants, as early truncations which carry redundant determinants will show no effect, whereas later ones will show the combined effect of removing the combination of redundant determinants. This was the case in our previous study of MSN2 and YAP1²³ and can be seen also in cases of two-sided truncations here, which were more effective than expected from the sum of the one-sided truncations (Figure 3D). To examine this, we engineered an additional set of 12 mutants of 7 TFs, in which we removed internal nonDBD segments of 100 residues each, corresponding to locations showing large effects on binding preference in the truncation series (Figure 3F). Of the 12 segments removed, four had no (or limited) effect and in an additional four, the effect was lower than within the truncation series, indicating partial redundancy. A notable exception was ACE2 whose two mid-nonDBD truncations led to a considerable shift in binding pattern (Figure 3F). We conclude that in the majority of TFs tested,

(C and D) DBDs binds a larger number of motif-containing sites: for each TF and its respective DBD, we measured the binding signal around each occurrence of the respective motifs. The respective distributions of those values (\log_2 normalized) are exemplified for SKO1 (C), with shaded regions indicating 3% binding threshold. This threshold was used to calculate the fraction of sites bound by each mutant, which is compared across all DBD-TF pairs in (D, see Table S4 for number of analyzed sites).

(E–G) Promoter preferences differ between DBDs and their respective TFs: depicted in (E) are the number of targets that are TF-unique, DBD-unique or common (blue, green, and yellow; E, left, see Table S5), and the correlations (Pearson's r) of promoter preferences between each TF and its DBD (E, right). Correlation values are summarized as cumulative distribution (F), and compared with correlations between biological repeats and random pairs. As an example, preferences of cell-cycle TFs and DBDs for their target promoters are also shown (G), with each line being a promoter, color-coded by preference of the indicated factor (rows, see STAR Methods). Top 60 promoters bound by each DBD-TF pair were included. Note that paralog DBDs become more similar, compared with the respective TFs (e.g., SWI5-ACE2), while DBDs of TFs acting as a complex lose their respective TFs' similarity (MBP1-SWI4).

(H) IDR content correlates with low DBD-TF similarity: shown is the nonDBD effect ($1 - C_{\text{DBD-TF}}$ where $C_{\text{DBD-TF}}$ denotes DBD-TF promoter preference correlation, Pearson's r) as a function of the number of disorder-classified residues within the deleted region. Each dot corresponds to a DBD-TF pair, color-coded by the number of structured residues in the truncated regions. The 10 fungal-specific zinc cluster TFs, carrying a characteristic long structured domain, were excluded.



(legend on next page)

nonDBDs contain multiple, partially redundant specificity determinants that direct TF binding specificity through their cumulative action.

Finally, we asked whether specificity determinants localize more often to IDRs or structured regions (Figure 3G). Examining first the three zinc cluster TFs whose DBDs were insufficient for binding, we noted that the loss of binding signal occurred once truncations entered the structured domain characterizing this TF family (not shown). By contrast, for DBDs that displayed reliable, but distinct binding preferences, specificity determinants commonly localized to IDRs (Figure 3G). We conclude that nonDBDs direct binding through multiple specificity determinants, many of which are found within IDRs.

NonDBDs direct TF binding toward promoters of fuzzy nucleosome architectures

Going beyond individual TFs, we asked whether, and how, nonDBDs impact the gene regulatory network design. A hallmark of budding yeast transcription is the splitting of genes into two classes: a minority whose promoters bind specific TFs and a majority that shows little TF binding.^{30,31} Of note, the nucleosome architecture is a defining property of these classes^{32–35}: promoters whose TSS-proximal nucleosome is occupied (OPN-type) show flexible expression and bind multiple TFs, whereas promoters depleted of proximal nucleosomes (“DPN-type”) are often unbound and drive stable expression.^{32–35}

Full-length TFs in our dataset localized to promoters of high OPN score, as expected from previous data³⁰ (Figure 4A). In sharp contrast, DBDs’ binding spreads across promoters, largely independent of the OPN score (Figure 4A). Therefore, the prominent preference of TFs for binding OPN promoters is not the mere consequence of differential DNA accessibility or motif occurrences but depends on determinants within the nonDBDs.

To examine more directly binding preferences of nonDBDs, we mapped binding profiles of DBD-deleted TF mutants (Figure 4B). In this, we were motivated by our previous studies

showing that the nonDBDs of MSN2 and YAP1 still localize to a large fraction of TF-bound promoters.²³ Reproducible profiles were obtained for 34 of the 49 nonDBD mutants tested. A considerable fraction of those retained similarity with their respective TFs, and these corresponded well with cases of low DBD-TF correlations (Pearson’s $r = -0.38$) (Figures 4C and 4D), with few exceptions indicating cooperation (e.g., BAS1) or redundancy (e.g., ADR1). Therefore, nonDBD and DBD exert complementary contributions guiding TF binding along the genome.

NonDBDs showed a high preference for binding OPN promoters, exceeding the preferences of full TFs (Figure 4A). Accordingly, nonDBD-bound promoters were enriched in OPN-associated properties³⁵ when compared with TF-bound ones and more so when compared with those bound by DBDs, including the presence of TATA boxes, expression plasticity, and long upstream intergenic distances (Figure 4E).

Finally, we compared the similarity in promoter preferences between different factors (Figures 4F–4H). Consistent with established results, clustering analysis revealed groups of TFs with similar functions that localize to overlapping promoters, including clusters of stress-related activators (e.g., MSN2, CRZ1, and YAP1) or repressors (e.g., SOK2, SKO1, and MGA1), co-binding TF pairs regulating same processes (e.g., STE12-TEC1, INO2-INO4, PDR1-PDR3, and MBP1-SWI4) (Figure 4F). These correlations largely disappeared when comparing DBD-only mutants, with only a few tight correlations remaining or emerging, mostly restricted to close paralogs (e.g., SWI5-ACE2, FKH1-FKH2-HCM1, RLM1-SMP1, and RPH1-GIS1). Of note, these similarities in binding profiles did not extend to other members of the same DBD family, despite similar *in-vitro* preferences.

In sharp contrast, correlations between TFs increased in the nonDBD mutants. In fact, although few nonDBDs localized at unique locations (e.g., TEC1 or SUM1), most others showed varying ranges of overlaps, with similarities ranging from near identity (SOK2-related cluster), to moderate (e.g., SOK2 and CRZ1) and low (e.g., SOK2 and SWI4) correlations (Figure 4H).

Figure 3. Multiple specificity determinants spread within intrinsically disordered regions

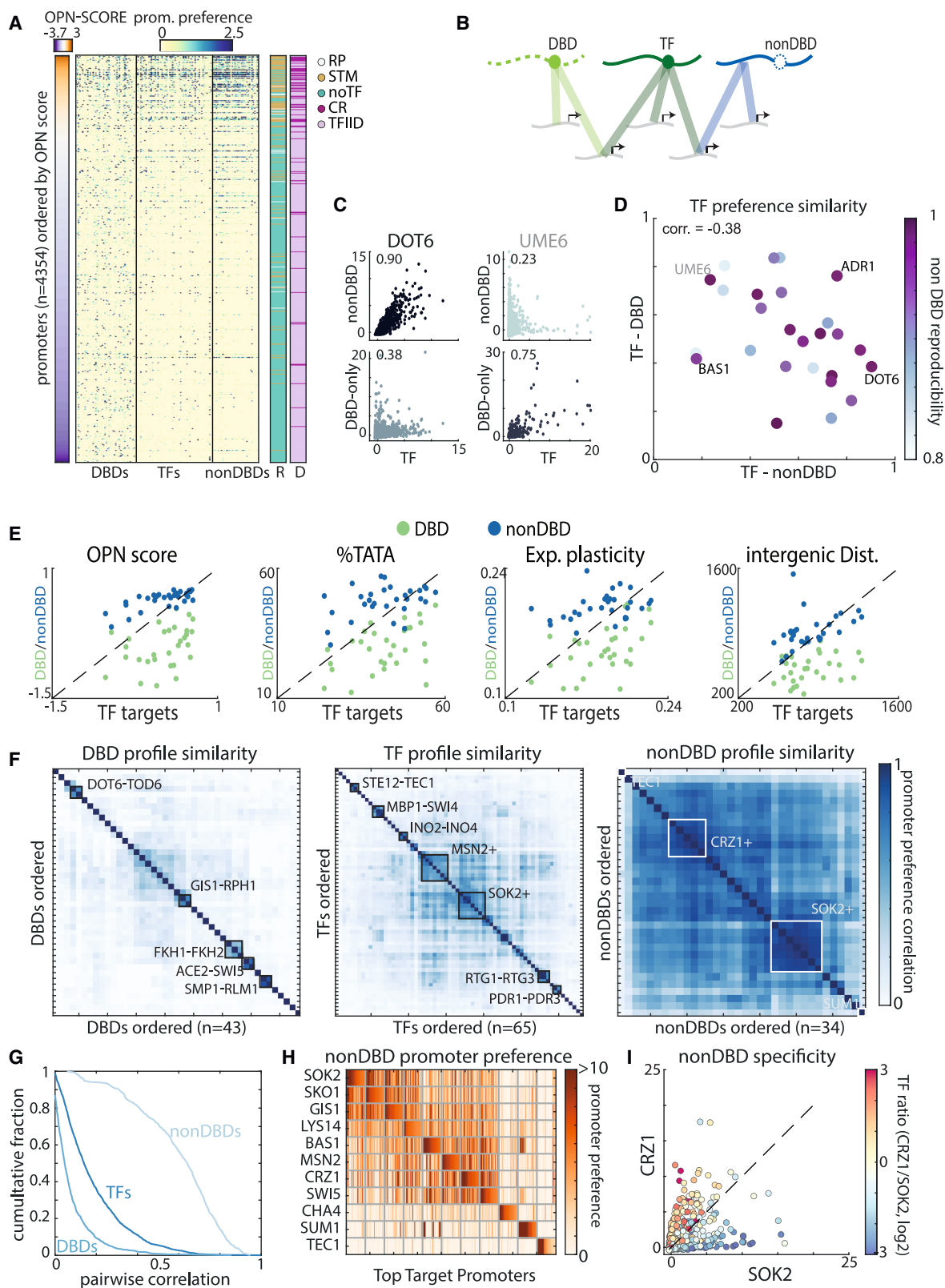
(A) Defining patterns of specificity determinant within nonDBDs—a scheme: TF sequence was gradually shortened at an attempted 50-residue resolution, and truncated mutants were profiled (left). Promoter preferences of truncation-mutants were compared with the respective TFs and DBDs, distinguishing between scenarios of localized vs. distributed specificity determinants. Simulated correlation maps (middle and right) illustrate the possible dynamics of the correlation between the truncations and the DBD or the TF in both cases. Here, dots are colored by the percentage of nonDBD removed, background color indicates DBD-TF correlation (Pearson’s r), and black circles highlight DBD and TF.

(B) Binding preferences of exemplary nonDBD truncation-mutants: distribution of specificity determinants within the indicated TFs is shown through changes in promoter preferences at the center figure, where each row is a DBD or TF target promoter color-coded by the corresponding truncation mutant’s preference. Predicted structure and disorder content are shown on the left, and the positioning of truncation mutants on the mutant-DBD vs. mutant-TF correlation map is shown on the right, as in (A). Gray dots indicate interpolated locations of missing 50-residues truncation mutants. Note the gradual loss of binding to TF-bound promoters, coupled to increased binding to DBD-bound ones.

(C–E) NonDBD truncations shift binding preferences from TF to DBD-like: correlation maps, as in (A), shown for all cases of one- (C), or two-sided (D) truncations. TFs with tight (>4) truncation series and low DBD-TF similarities (<0.5 DBD-TF correlation) are written in black. In two-sided truncations (D), red circles indicate the more effective series. Asterisk indicates cases where combined truncations were tested. A graduality score, defined as the ratio of maximal to median added effect, summarizes truncation effects (y axis) and is shown as a function of the mean truncation length, i.e., nonDBD length divided by number of truncations, (x axis) for all series in (E), with the respective number of truncations and TF-DBD correlation indicated by dot color and size.

(F) Specificity determinants are often redundant. Highly effective nonDBD regions were chosen based on the additional decrease in correlation to TF along the truncation series, and the respective internal deletion mutant was generated and profiled (scheme on top). The effect of this internal deletion, measured as Pearson’s correlation with the full TF, was then compared with the expected effect based on truncation series (bottom). Note that higher than expected correlation indicates redundancy. Each dot is a deletion mutant and color indicates the corresponding TF.

(G) Specificity determinants often localize to IDRs: the overall effect of truncating IDRs or structured regions was calculated for each series, and is shown here as (cumulative) distributions of the effects. Individual truncations were classified as disordered or structured based on the mean disorder of the removed region (see STAR Methods).



(legend on next page)

Therefore, although DBDs' profiles are clearly distinguishable, those of the nonDBDs show continuously varying levels of overlap (Figure 4G). Notably, despite these considerable overlaps, promoter preferences of nonDBDs were properly aligned with those of their respective TFs. This is exemplified by CRZ1 and SOK2, whose nonDBDs show moderately overlapping preferences (Figure 4I): although they were bound at similar promoters, comparing their binding at each promoter revealed a two-branch pattern consistent with clear specificity differences between the two nonDBDs, corresponding to the binding preferences of the full TFs. We conclude that nonDBDs direct TFs binding toward overlapping classes of fuzzy nucleosome promoters, with the DBD acting to refine this pattern through the DBD binding to specific motifs.

DISCUSSION

TF binding specificity is critical for gene regulation, making it the subject of much experimental and computational attention. Most studies approach this question from the viewpoint of the DNA: what differentiates the DNA (or chromatin) environment of TF-bound vs. unbound sites? Here, we undertook a complementary path, dissecting the TFs themselves as an entry point for distinguishing mechanisms of binding specificity. Through extensive sequence truncations, coupled with the profiling of genomic binding, we defined the relative roles of DBD and nonDBD sequences in guiding TF binding across the genome. Our results highlight specificity determinants distributed within the nonDBD sequences, their cumulative action in directing binding specificity, and their tight association with IDRs.

TF binding across the genome integrates influences of both its DBD and nonDBD (Figure 5). In some cases, the DBD dominates, as exemplified by AZF1 whose DBD uniquely includes four C2H2 zinc finger domains, whereas in other cases, the nonDBD dominates, with the DBD stabilizing binding at motifs within nonDBD-selected promoters (e.g., DOT6). The majority of TFs, however,

present intermediate situations in which promoter selection by the nonDBD is refined and stabilized by DBD-motif binding.

DBDs and nonDBDs guide TF binding using distinct strategies. DBDs bind to their short DNA sequence motifs through structural complementarity, which is highly sensitive to single-point mutations in interacting amino acids and nucleotide bases. Despite the large overlap in motif preference between DBDs of the same family, the overall promoter preferences of same-family DBDs are lower than that of the respective TFs, with the exception of close paralogs. This, perhaps unexpected, result is explained by the more permissive spread of DBDs across promoters, which contrasts the largely restricted binding of full TFs to OPN-type promoters of fuzzy nucleosome architectures.

Contrasting this digital-like encoding of DBD-motif binding, the specificity strategy encoded by nonDBDs follows an analog-like design. Rather than relying on a single determinant and structural complementarity, the majority of surveyed nonDBDs contain multiple specificity determinants that map primarily to disordered regions. The effect of each individual determinant is weak and often redundant so that only their cumulative action explains nonDBD's contribution to binding specificity. This distributed encoding of specificity determinants within nonDBDs mirrors their promoter preference patterns: rather than defining distinct and well-separated profiles, nonDBDs' binding profiles show a wide range of similarities that vary from full identity to partial or low correlation. Therefore, the wide distribution of specificity determinants within IDRs allows continuous, analog-like tuning of binding across the set of promoters with fuzzy nucleosome architectures.

A central remaining unknown is the mechanism through which IDR-localized determinants direct promoter selection.³⁶ IDRs can promote the formation of phase-separated condensates^{37–39} associated with transcription,^{40,41} and these might direct genomic binding. Studies examining for clustering of the IDR-dependent MSN2 TF, however, have failed to detect such agglomerations.⁴² Three other immediate options are interaction with other TFs, recognition of the epigenetic landscape, or direct

Figure 4. NonDBDs restrict TF binding to promoters of fuzzy nucleosomes

(A) TF binding correlates with promoter nucleosome architecture: budding yeast promoters (rows) were ordered by their OPN score, a measure associated with fuzzy nucleosome architecture (see STAR Methods). Color indicates the preference of each profiled DBD, TF, and nonDBD (columns) for every promoter. Right panels indicate previously described promoter classification into TF-bound (R, Rossi et al.³⁰; STM, SAGA, TUP, and Mediator bound; D, Donczew et al.³¹; and CR, coactivator-redundant) and TF-unbound classes, indicating also the ribosomal gene promoters (RPs).

(B) Genomic mapping of nonDBD mutants—a scheme: mutants lacking their DBD were generated and their binding profiles mapped.

(C and D) NonDBDs localize to TF-target promoters: shown in (C) are comparisons of promoter preferences between the indicated nonDBDs or DBDs and TFs. Note the tight nonDBD-TF similarity shown by DOT6, contrasting the distinct preferences for UME6. The nonDBD-TF promoter preference correlations (Pearson's *r*) are compared with respective DBD-TF correlations (D). Note the inverse relation, indicating complementary contribution of nonDBDs and DBDs to TF promoter preference.

(E) Properties of nonDBD-bound promoters: the top bound promoters of each TF-DBD-nonDBD triplet were analyzed for the indicated features (OPN score, TATA-box occurrence, Expression plasticity and intergenic distance). The scatter plots show the average value for the DBD- (green) and the nonDBD-bound (blue) promoters vs. TF-bound ones.

(F–H) nonDBDs display a range of overlapping promoter preferences contrasting DBDs' distinct binding profiles: the correlations of promoter preferences between all mutants of the indicated types are shown in clustered heatmaps (F) and as cumulative histograms (G). Factors in (F) were ordered by hierarchical clustering, applied separately to each type. Emerging clusters are outlined and exemplary members highlighted. (H) displays exemplary promoter preferences for selected nonDBDs. Here, each line is promoter, color-coded by the preference of the indicated nonDBD. Top 40 promoters bound by each nonDBD were selected and ordered by preference. Note the gradual loss of SOK2-like promoter preferences across the nonDBD series.

(I) NonDBDs preferences bias TF binding specificity. Shown are the promoter preferences of SOK2 and CRZ1 nonDBD. Each dot is a promoter with the color indicating the ratio of the binding signals of the two corresponding full-length TFs (dashed line indicates 1:1 line). Note the higher binding of each TF to the promoter preferred by its nonDBD.

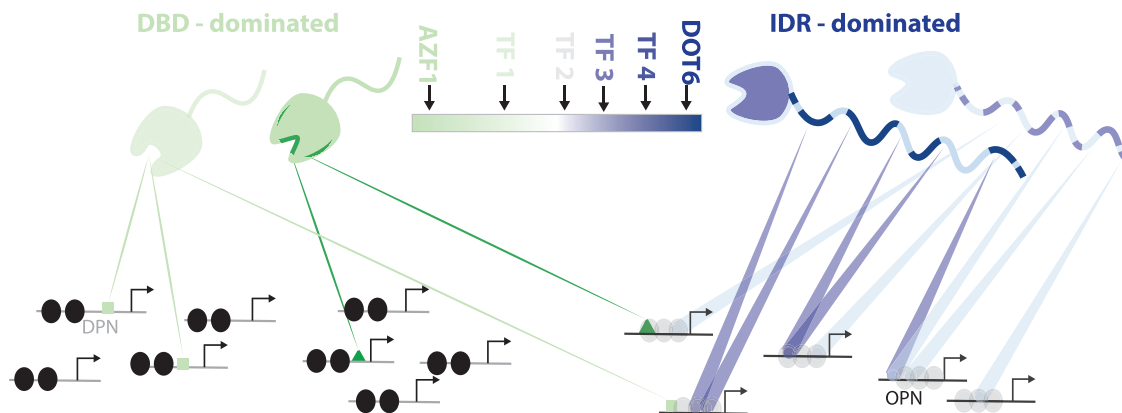


Figure 5. Strategies for directing genome binding by DBDs and IDRs

TF-binding locations across the genome integrate effects from their DBDs and nonDBDs. Those TF regions direct binding through complementary digital- and analog-like strategies. These strategies differ at three levels. First, in terms of protein-localized determinants, DBD-motif binding fully depends on particular contact residues, whereas nonDBDs include multiplicity of IDR-localized, weak specificity determinants, none of which is essential. Second, in terms of promoter selection, DBD-motif binding favors local nucleosome depletion, but is permissive concerning the overall promoter nucleosome architecture, whereas IDR-localized determinants restrict binding to promoters of fuzzy nucleosome architecture. Third, when compared across factors, DBDs show a little overlap in binding patterns, whereas IDR-enriched nonDBDs show a continuum of overlapping patterns.

binding to DNA. Of those, interactions between TFs are the most likely explanation, which we are currently exploring. A related question is the mechanism leading to analog, rather than digital control of site selection. The distribution of multiple specificity determinants within long IDRs appears optimally suited for analog control but understanding this paradigm will entail elucidating the IDR-based and promoter-based sequence grammar at the basis of this interaction.

Limitations of the study

Although the broad spectrum of evaluated TFs suggests a general principle applying to all yeast TFs, it might be that non-tested TFs or TFs from other species despite having long IDRs do not conform to it. In addition, the resolution of the performed truncations limits our ability to define the exact number and distribution of individual specificity determinants.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Yeast genome editing
 - Yeast transformation
- METHOD DETAILS
 - ChEC-seq experiments
 - NGS library preparation
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - NGS data analysis

- Promoter definition
- Promoter preference and preference correlation (Figures 2, 3, and 4)
- Normalized ChEC signal representation (Figure 1G)
- *In-vitro* 7mer score and top 7mer selection
- *In-vivo* 7mer scores
- Motif enrichment analysis (Figure 1J)
- 7mer selection bias (Figures 2A and 2B)
- Binding signal on 7mer site and fraction of bound sites (Figures 2C and 2D)
- Correlation maps (Figures 3B–3D)
- Relative binding across truncations (Figure 3B)
- Graduality score (Figure 3E)
- Defining structured and disordered regions (Figure 3G)
- Clustering of TF, DBD, and nonDBD or promoters (Figure 4F)
- Gene expression plasticity (Figure 4E)
- OPN score (Figure 4A)

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.molcel.2023.04.002>.

ACKNOWLEDGMENTS

We thank Gilad Yaakov and Offir Lupo for careful reading of the manuscript. We also thank Rotem Morag, Offir Lupo, and Tamar Gera for strains and data. We thank Tzachi Pilpel and our lab members for scientific discussions. This work was funded by the Israel Science Foundation, Horizon Europe (European Research Council), and the Minerva Foundation.

AUTHOR CONTRIBUTIONS

D.K.K., F.J., and N.B. conceived the study and designed the experiments. D.K.K. performed most of the experiments. D.K.K. and F.J. analyzed the data. All authors wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: August 25, 2022

Revised: January 17, 2023

Accepted: January 30, 2023

Published: April 27, 2023

REFERENCES

- Zaman, Z., Ansari, A.Z., Gaudreau, L., Nevado, J., and Ptashne, M. (1998). Gene transcription by recruitment. *Cold Spring Harbor Symp. Quant. Biol.* 63, 167–171. <https://doi.org/10.1101/sqb.1998.63.167>.
- Kribelbauer, J.F., Rastogi, C., Bussemaker, H.J., and Mann, R.S. (2019). Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes. *Annu. Rev. Cell Dev. Biol.* 35, 357–379. <https://doi.org/10.1146/annurev-cellbio-100617-062719>.
- Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., et al. (2009). Diversity and complexity in DNA recognition by transcription factors. *Science* 324, 1720–1723. <https://doi.org/10.1126/science.1162327>.
- Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 158, 1431–1443. <https://doi.org/10.1016/j.cell.2014.08.009>.
- Wunderlich, Z., and Mirny, L.A. (2008). Spatial effects on the speed and reliability of protein-DNA search. *Nucleic Acids Res.* 36, 3570–3578. <https://doi.org/10.1093/nar/gkn173>.
- Jana, T., Brodsky, S., and Barkai, N. (2021). Speed-specificity trade-offs in the transcription factors search for their genomic binding sites. *Trends Genet.* 37, 421–432. <https://doi.org/10.1016/j.tig.2020.12.001>.
- Inukai, S., Kock, K.H., and Bulyk, M.L. (2017). Transcription factor-DNA binding: beyond binding site motifs. *Curr. Opin. Genet. Dev.* 43, 110–119. <https://doi.org/10.1016/j.cde.2017.02.007>.
- Todeschini, A.L., Georges, A., and Veitia, R.A. (2014). Transcription factors: specific DNA binding and specific gene regulation. *Trends Genet.* 30, 211–219. <https://doi.org/10.1016/j.tig.2014.04.002>.
- Pan, Y., Tsai, C.J., Ma, B., and Nussinov, R. (2010). Mechanisms of transcription factor selectivity. *Trends Genet.* 26, 75–83. <https://doi.org/10.1016/j.tig.2009.12.003>.
- Dror, I., Rohs, R., and Mandel-Gutfreund, Y. (2016). How motif environment influences transcription factor search dynamics: finding a needle in a haystack. *Bioessays* 38, 605–612. <https://doi.org/10.1002/bies.201600005>.
- Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., and Mann, R.S. (2010). Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* 79, 233–269. <https://doi.org/10.1146/annurev-biochem-060408-091030>.
- Gill, G., and Ptashne, M. (1987). Mutants of GAL4 protein altered in an activation function. *Cell* 51, 121–126. [https://doi.org/10.1016/0092-8674\(87\)90016-X](https://doi.org/10.1016/0092-8674(87)90016-X).
- Fischer, J.A., Giniger, E., Maniatis, T., and Ptashne, M. (1988). GAL4 activates transcription in *Drosophila*. *Nature* 332, 853–856. <https://doi.org/10.1038/332853a0>.
- Gaudreau, L., Keaveney, M., Nevado, J., Zaman, Z., Bryant, G.O., Struhl, K., and Ptashne, M. (1999). Transcriptional activation by artificial recruitment in yeast is influenced by promoter architecture and downstream sequences. *Proc. Natl. Acad. Sci. USA* 96, 2668–2673. <https://doi.org/10.1073/pnas.96.6.2668>.
- Nevado, J., Gaudreau, L., Adam, M., and Ptashne, M. (1999). Transcriptional activation by artificial recruitment in mammalian cells. *Proc. Natl. Acad. Sci. USA* 96, 2674–2677. <https://doi.org/10.1073/pnas.96.6.2674>.
- Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405. <https://doi.org/10.1016/j.tibtech.2013.04.004>.
- Pérez, J.C., Fordyce, P.M., Lohse, M.B., Hanson-Smith, V., DeRisi, J.L., and Johnson, A.D. (2014). How duplicated transcription regulators can diversify to govern the expression of nonoverlapping sets of genes. *Genes Dev.* 28, 1272–1277. <https://doi.org/10.1101/gad.242271.114>.
- Del Olmo Toledo, V., Puccinelli, R., Fordyce, P.M., and Pérez, J.C. (2018). Diversification of DNA binding specificities enabled SREBP transcription regulators to expand the repertoire of cellular functions that they govern in fungi. *PLoS Genet.* 14, e1007884. <https://doi.org/10.1371/journal.pgen.1007884>.
- Pougach, K., Voet, A., Kondrashov, F.A., Voordeckers, K., Christiaens, J.F., Baying, B., Benes, V., Sakai, R., Aerts, J., Zhu, B., et al. (2014). Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. *Nat. Commun.* 5, 4868. <https://doi.org/10.1038/ncomms5868>.
- Baker, C.R., Hanson-Smith, V., and Johnson, A.D. (2013). Following gene duplication, paralog interference constrains transcriptional circuit evolution. *Science* 342, 104–108. <https://doi.org/10.1126/science.1240810>.
- Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., et al. (2010). An atlas of combinatorial transcriptional regulation in mouse and Man. *Cell* 140, 744–752. <https://doi.org/10.1016/j.cell.2010.01.044>.
- Vandel, J., Cassan, O., Lèbre, S., Lecellier, C.-H., and Bréhélin, L. (2019). Probing transcription factor combinatorics in different promoter classes and in enhancers. *BMC Genomics* 20, 103. <https://doi.org/10.1186/s12864-018-5408-0>.
- 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>.
- Gera, T., Jonas, F., More, R., and Barkai, N. (2022). Evolution of binding preferences among whole-genome duplicated transcription factors. *eLife* 11, e73225. <https://doi.org/10.7554/eLife.73225>.
- Zentner, G.E., Kasinathan, S., Xin, B., Rohs, R., and Henikoff, S. (2015). ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. *Nat. Commun.* 6, 8733. <https://doi.org/10.1038/ncomms9733>.
- Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N., and Dunker, A.K. (2006). Intrinsic disorder in transcription factors. *Biochemistry* 45, 6873–6888. <https://doi.org/10.1021/bi0602718>.
- Wang, C., Uversky, V.N., and Kurgan, L. (2016). Disordered nucleosome: abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea. *Proteomics* 16, 1486–1498. <https://doi.org/10.1002/pmic.201500177>.
- Minezaki, Y., Homma, K., Kinjo, A.R., and Nishikawa, K. (2006). Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol.* 359, 1137–1149. <https://doi.org/10.1016/j.jmb.2006.04.016>.
- Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* 337, 635–645. <https://doi.org/10.1016/j.jmb.2004.02.002>.
- Rossi, M.J., Kuntala, P.K., Lai, W.K.M., Yamada, N., Badjatia, N., Mittal, C., Kuzu, G., Bocklund, K., Farrell, N.P., Blanda, T.R., et al. (2021). A high-resolution protein architecture of the budding yeast genome. *Nature* 592, 309–314. <https://doi.org/10.1038/s41586-021-03314-8>.

31. Donczew, R., Warfield, L., Pacheco, D., Erijman, A., and Hahn, S. (2020). Two roles for the yeast transcription coactivator SAGA and a set of genes redundantly regulated by TFIID and SAGA. *eLife* 9, e50109. <https://doi.org/10.7554/eLife.50109>.
32. Hornung, G., Bar-Ziv, R., Rosin, D., Tokuriki, N., Tawfik, D.S., Oren, M., and Barkai, N. (2012). Noise-mean relationship in mutated promoters. *Genome Res.* 22, 2409–2417. <https://doi.org/10.1101/gr.139378.112>.
33. Hornung, G., Oren, M., and Barkai, N. (2012). Nucleosome organization affects the sensitivity of gene expression to promoter mutations. *Mol. Cell* 46, 362–368. <https://doi.org/10.1016/j.molcel.2012.02.019>.
34. Rosin, D., Hornung, G., Tirosh, I., Gispán, A., and Barkai, N. (2012). Promoter nucleosome organization shapes the evolution of gene expression. *PLoS Genet.* 8, e1002579. <https://doi.org/10.1371/journal.pgen.1002579>.
35. Tirosh, I., and Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* 18, 1084–1091. <https://doi.org/10.1101/gr.076059.108>.
36. Brodsky, S., Jana, T., and Barkai, N. (2021). Order through disorder: the role of intrinsically disordered regions in transcription factor binding specificity. *Curr. Opin. Struct. Biol.* 71, 110–115. <https://doi.org/10.1016/j.sbi.2021.06.011>.
37. Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C.C., Eckmann, C.R., Myong, S., and Brangwynne, C.P. (2015). The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. USA* 112, 7189–7194. <https://doi.org/10.1073/pnas.1504822112>.
38. Wang, J., Choi, J.M., Holehouse, A.S., Lee, H.O., Zhang, X., Jahnel, M., Maharana, S., Lemaître, R., Pozniakovsky, A., Drechsel, D., et al. (2018). A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* 174, 688–699.e16. <https://doi.org/10.1016/j.cell.2018.06.006>.
39. Alberti, S., Gladfelter, A., and Mittag, T. (2019). Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* 176, 419–434. <https://doi.org/10.1016/j.cell.2018.12.035>.
40. Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* 175, 1842–1855.e16. <https://doi.org/10.1016/j.cell.2018.10.042>.
41. 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>.
42. Chowdhary, S., Kainth, A.S., Pincus, D., and Gross, D.S. (2019). Heat shock Factor 1 drives intergenic association of its target gene loci upon heat shock. *Cell Rep.* 26, 18–28.e5. <https://doi.org/10.1016/j.celrep.2018.12.034>.
43. Baker Brachmann, C., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., and Boeke, J.D. (1998). Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* 14, 115–132. [https://doi.org/10.1002/\(SICI\)1097-0061\(19980130\)14:2<115::AID-YEA204>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2).
44. Meurer, M., Duan, Y., Sass, E., Kats, I., Herbst, K., Buchmüller, B.C., Dederer, V., Huber, F., Kirmaier, D., Stefl, M., et al. (2018). Genome-wide C-SWAT library for high-throughput yeast genome tagging. *Nat. Methods* 15, 598–600. <https://doi.org/10.1038/s41592-018-0045-8>.
45. Blecher-Gonen, R., Barnett-Itzhaki, Z., Jaitin, D., Amann-Zalcenstein, D., Lara-Astiaso, D., and Amit, I. (2013). High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat. Protoc.* 8, 539–554. <https://doi.org/10.1038/nprot.2013.023>.
46. Anand, R., Beach, A., Li, K., and Haber, J. (2017). Rad51-mediated double-strand break repair and mismatch correction of divergent substrates. *Nature* 544, 377–380. <https://doi.org/10.1038/nature22046>.
47. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
48. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
49. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10. <https://doi.org/10.14806/ej.17.1.200>.
50. Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H., Valen, E., et al. (2021). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* 47, W171–W174. <https://doi.org/10.1093/nar/gkz365>.
51. Gietz, R.D., and Schiestl, R.H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2, 31–34. <https://doi.org/10.1038/nprot.2007.13>.
52. Yofe, I., Weill, U., Meurer, M., Chuartzman, S., Zalckvar, E., Goldman, O., Ben-Dor, S., Schütze, C., Wiedemann, N., Knop, M., et al. (2016). One library to make them all: streamlining the creation of yeast libraries via a SWAp-Tag strategy. *Nat. Methods* 13, 371–378. <https://doi.org/10.1038/nmeth.3795>.
53. Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* 6, e21856. <https://doi.org/10.7554/eLife.21856>.
54. Policastro, R.A., Raborn, R.T., Brendel, V.P., and Zentner, G.E. (2020). Simple and efficient profiling of transcription initiation and transcript levels with STRIP-seq. *Genome Res.* 30, 910–923. <https://doi.org/10.1101/gr.261545.120>.
55. Park, D., Morris, A.R., Battenhouse, A., and Iyer, V.R. (2014). Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. *Nucleic Acids Res.* 42, 3736–3749. <https://doi.org/10.1093/nar/gkt1366>.
56. Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* 497, 127–131. <https://doi.org/10.1038/nature12121>.
57. Mészáros, B., Erdős, G., and Dosztányi, Z. (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* 46, W329–W337. <https://doi.org/10.1093/nar/gky384>.
58. Kemmeren, P., Sameith, K., van de Pasch, L.A., Benschop, J.J., Lenstra, T.L., Margaritis, T., O'Duibhir, E., Apweiler, E., van Wageningen, S., Ko, C.W., et al. (2014). Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* 157, 740–752. <https://doi.org/10.1016/j.cell.2014.02.054>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
cOmplete EDTA-free Protease Inhibitor Cocktail	Sigma Aldrich	Cat#11873580001
Proteinase K	Sigma Aldrich	Cat#P2308
RNase A	Sigma Aldrich	Cat#R4875
AMPure XP	Beckman Coulter	Cat#A63881
Glycoblue	Thermo Fisher	Cat# AM9515
Zymolase 100T	Amsbio	Cat#120493-1
IGEPAL CA-630	Sigma Aldrich	Cat#I3021
Micrococcal nuclease (MNase)	Worthington	Cat#LS004797
Digitonin	Sigma Aldrich	Cat#300410
Spermine	Sigma Aldrich	Cat# S3256-5G
Spermidine	Sigma Aldrich	Cat# S026
10x T4 DNA Ligase buffer	NEB	B0202S
T4 PNK	NEB	M0201S
T4 DNA Polymerase	Thermo Scientific	EP0061
Taq polymerase	Bioline	BIO-21040
Quick Ligase	NEB	M2200S
KAPA HiFi DNA polymerase	Roche	07958927001
Critical commercial assays		
HiYield Plasmid Mini Kit	RBC Bioscience	Cat#YPD100
Deposited data		
Raw and processed NGS data	This paper	GEO: GSE209631
Raw and proc. NGS data for individual full-length TFs	Gera et al. ²⁴	GEO: GSE179430
Raw NGS data for Msn2 and Yap2 Transcription factor	Brodsky et al. ²³	BioProject: PRJNA573518
Experimental models: Organisms/strains		
BY4741 (<i>Saccharomyces cerevisiae</i>)	Brachmann et al. ⁴³	N/A
C-terminal SWAp-Tag (C-SWAT) library (<i>S. cerevisiae</i>)	Meurer et al. ⁴⁴	N/A
Yeast strains	This study	Table S3
Oligonucleotides		
Barcoded Y-shaped adapters	Blecher-Gonen et al. ⁴⁵	N/A
Repair Oligos for CRISPR-based gene editing	This study	Table S6
Recombinant DNA		
bRA89 (plasmid)	Anand et al. ⁴⁶	RRID:Addgene_100950
pGZ108 (plasmid)	Zentner et al. ²⁵	RRID:Addgene_70231
Software and algorithms		
MATLAB	MathWorks	N/A
Bowtie 2	Langmead et al. ⁴⁷	N/A
BEDTools	Quinlan et al. ⁴⁸	N/A
cutAdapt	Martin et al. ⁴⁹	N/A
CHOPCHOP	Labun et al. ⁵⁰	N/A
Custom Analysis codes	This Study	https://doi.org/10.5281/zenodo.7780657
Growth media		
YPD	CSHP	https://doi.org/10.1101/pdb.rec12315
Synthetic Complete Medium with 2% Glucose	Gietz et al. ⁵¹	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact: Naama Barkai (naama.barkai@weizmann.ac.il).

Materials availability

All yeast strains generated in this study (see [Table S3](#)) are available upon request.

Data and code availability

- All raw sequencing data generated in this study have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the [key resources table](#).
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Yeast genome editing

We made our TF mutants by two methods of gene editing. For MNase tagging of many of our TFs, we used the C-SWAT⁵² libraries as parent strains, picked the strains in which our required TFs have been tagged with the C terminal acceptor module, and replaced this module with an MNase cassette using CRISPR. In this process, we remove the uracil auxotrophy (URA3) of the parent C-SWAT strain, and we introduce the CRISPR-plasmid-borne hygromycin resistance. We also used CRISPR to make all the TF mutants, where we deleted parts of the nonDBD and DBD by targeting specific regions of the respective TF. We designed DNA repairs that were single stranded and had a 45 bp homology to the deletion site on each side.

For CRISPR based gene editing, we used the bRA89 plasmid,⁴⁶ which encodes Cas9, target-specific guide RNA, hygromycin and ampicillin resistance. We designed the target-specific guide RNA using CHOPCHOP,⁵⁰ ligated it into the bRA89 plasmid in between BpII restriction sites. The ligated plasmids were transformed and propagated in *E. coli*. Successful transformants were selected based on PCR and plasmids were extracted using a commercial MiniPrep Kit (Real Genomics).

The other strategy we used to insert the MNase tag was homologous recombination. Here, we amplified an MNase cassette from the plasmid pGZ108,²⁵ using primers having 45 bases homology to the insertion site on each side. This cassette contains the MNase and Kanamycin resistance.

All yeast strains used in this study are listed in detail in [Table S2](#).

Yeast transformation

All yeast transformations followed the LiAc/SS DNA/PEG method.⁵¹ A single colony of the parental strain was inoculated in fresh liquid YPD and grown to saturation overnight, then diluted into fresh 5 ml YPD and grown to OD₆₀₀ of 0.5. The cells are harvested at this stage, and washed with distilled water and then with 100 mM LiAc twice. Then they were gently resuspended in the transformation mix (33% PEG-3350, 100 mM LiAc, single stranded salmon sperm DNA) and the DNA oligos (500 μmols) and plasmid (5 μg) as required for the transformation. This suspension was incubated at 30°C for 30 minutes followed by a 30-minute heat shock at 42°C. Then, the cells were transferred for recovery at 30°C, either on a YPD plate overnight or for 2 hrs in 1 mL liquid YPD. After this, they were plated on the appropriate selection plate. All strains generated were verified using PCR and gel electrophoresis, and also by Sanger DNA sequencing. For plasmid containing strains, the plasmid was lost by growing the cells in YPD until saturation. The colonies that lost the plasmid were selected by screening for the loss of the selection marker. These strains were subsequently used for experiments.

METHOD DETAILS

ChEC-seq experiments

The experiments were performed as described previously,²⁵ with some modifications²⁴: Yeast strains were freshly thawed before experiments from a frozen stock, plated on YPD plates, and grown. Single colonies were picked and grown overnight at 30°C in liquid SD (synthetic complete with dextrose) medium to stationary phase. Then, the cultures were diluted ~10³-fold into 5 mL fresh SD media and grown overnight to reach an OD₆₀₀ of 4 the following morning. Cultures were pelleted at 1500 g for 2 min and resuspended in 0.5 mL buffer A (15 mM Tris pH 7.5, 80 mM KCl, 0.1 mM EGTA, 0.2 mM spermine, 0.5 mM spermidine, 1 × cComplete EDTA-free protease inhibitors (1 tablet per 50 mL buffer), 1 mM PMSF) and then transferred to 2.2 mL 96-well plates (LifeGene). Cells were washed twice in 1 mL Buffer A. Next, the cells were resuspended in 150 μL Buffer A containing 0.1% digitonin, transferred to an 96-well PCR plate (Axygen PCR-96-flt-c) and incubated at 30°C for 5 min for permeabilization. Next, we added 120 μL of cells to 13 μL of 25 mM

CaCl₂ for a final concentration of 2 mM to activate the MNase and incubated for exactly 30 s. The MNase treatment was stopped by adding 120 μ L of stop buffer (400 mM NaCl, 20 mM EDTA, 4 mM EGTA, and 1% SDS) to the cell suspension. After this, the cells were treated with Proteinase K (0.5 mg/mL) at 55°C for 30 min. An equal volume of Phenol-Chloroform pH = 8 (Sigma-Aldrich) was added, vigorously vortexed and centrifuged at 17,000 g for 10 min to extract DNA. After phenol chloroform extraction of nucleic acids, the DNA in the aqueous phase was precipitated with 2.5 volumes (600 μ L) of cold 96% EtOH, 45 mg Glycoblue, and 24 μ L of 3M sodium acetate (=20 mM final) at -80°C for >1 hr. DNA was centrifuged (17,000 g, 4°C for 10 min), supernatant removed and the DNA pellet washed with 75% EtOH. DNA pellets were dried and resuspended in 30 μ L RNase A solution (0.33 mg/mL RNase A in Tris-EDTA [TE] buffer [10 mM Tris and 1 mM EDTA]) and treated at 37°C for 20 min. In order to enrich for small DNA fragments and remove large DNA fragments that might result from spontaneous DNA breaks, DNA cleanup was performed using SPRI beads (Ampure XP). First, a reverse SPRI cleanup was performed by adding 0.8 \times (24 μ L) SPRI beads followed by 5 min incubation at RT. Supernatant was collected and the remaining small DNA fragments purified by adding additional 1 \times (30 μ L) SPRI beads and 5.4 \times (162 μ L) isopropanol, and incubating 5 min at RT. Beads were washed twice with 85% EtOH and small fragments were eluted in 30 μ L of 0.1 \times TE buffer. For biological repeats, two yeast colonies with the same genotype were grown separately before the ChECseq experiment, or also often performed on different days.

NGS library preparation

Library preparation was performed as described in Skene and Henikoff,⁵³ with slight modifications²⁴: DNA fragments after RNase treatment and reverse SPRI cleanup served as an input to end-repair and A-tailing (ERA) reaction. For each sample 20 μ L ERA reaction (1 \times T4 DNA ligase buffer [NEB], 0.5 mM dNTPs, 0.25 mM ATP, 2.75% PEG 4000, 6U T4 PNK, 0.5U T4 DNA Polymerase and 0.5U Taq DNA polymerase) was prepared and incubated for 20 min at 12°C, 15 min at 37°C and 45 min at 58°C in a thermocycler.

After ERA reaction, reverse SPRI cleanup was performed by adding 0.5 \times (10 μ L) SPRI beads (Ampure XP). Supernatant was collected and remaining small DNA fragments purified with additional 1.3 \times (26 μ L) SPRI beads and 5.4 \times (108 μ L) isopropanol. After washing with 85% EtOH, small fragments were eluted in 17 μ L of 0.1 \times TE buffer; 16.4 μ L elution were taken into 40 μ L ligation reaction (1 \times Quick ligase buffer, 4000U Quick ligase, and 6.4 nM Y-shaped barcode adaptors with T-overhang⁴⁵) and incubated for 15 min at 20°C in thermocycler.

After incubation, ligation reaction was cleaned by performing a double SPRI cleanup: first, a regular 1.2 \times (48 μ L) SPRI cleanup was performed and eluted in 30 μ L 0.1 \times TE buffer. Then and instead of separating the beads, an additional SPRI cleanup was performed by adding 1.3 \times (39 μ L) HXN buffer (2.5 M NaCl, 20% PEG 8000) and final elution in 24 μ L 0.1 \times TE buffer; 23 μ L elution were taken into 50 μ L enrichment PCR reaction (1 \times Kappa HiFi, 0.32 μ M barcoded Fwd primer and 0.32 μ M barcoded Rev primer⁴⁵) and incubated for 45 s in 98°C, 16 cycles of 15 s in 98°C and 15 s in 60°C, and a final elongation step of 1 min at 72°C in a thermocycler.

The final libraries were cleaned by a regular 1.1 \times (55 μ L) SPRI cleanup and eluted in 15 μ L 0.1 \times TE buffer. Library concentration and size distribution were quantified by Qubit (Thermo Scientific) and TapeStation (Agilent), respectively. For multiplexed next-generation sequencing (NGS), all barcoded libraries were pooled in equal amounts, the final pool diluted to 2 nM and sequenced on NextSeq 500 (Illumina) or NovaSeq 6000 (Illumina). Sequence parameters were Read1: 51 nucleotides (nt), Index1: 8 nt, Index2: 8 nt, Read2: 51 nt, for NovaSeq or Read1: 38 nt, Read2: 37 nt for NextSeq.

QUANTIFICATION AND STATISTICAL ANALYSIS

NGS data analysis

Raw reads from ChEC-seq libraries were demultiplexed using bcl2fastq (Illumina), and adaptor dimers and short reads were filtered out using cutAdapt⁴⁹ with parameters: “-O 10 -pairfilter=any -max-n 0.8 -action=mask”. Filtered reads were subsequently aligned to the *S. cerevisiae* genome R64-1-1 using Bowtie 2⁴⁷ with the options “--end-to-end --trim-to 40 --very-sensitive”. The genome coverage of fully aligned read pairs was calculated with GenomeCoverage from BEDTools⁴⁸ using the parameters “-d -5 -fs 1”. All further processing of samples was performed using MATLAB. The signal was normalized so that the total number of reads along the nuclear genome, excluding ribosomal rRNA genes and CUP1-1/2, is 12 million.

Promoter definition

We defined transcription start sites (TSS) by comparative analysis of publicly available TSS datasets.^{54–56} Promoter region was defined from the start codon until at least 700 bp upstream of the TSS or until the closest verified ORF.

Promoter preference and preference correlation (Figures 2, 3, and 4)

For each promoter ($n=5399$, excluding promoters in sub-telomeric regions), we first sum the normalized ChEC signal over the whole promoter. Then, we define promoter preference of a mutant to a certain promoter as the z-Score normalized binding signal of this mutant at that promoter compared to all other promoters. In our analysis, we compare the promoter preferences of different samples by calculating the Pearson's correlation coefficient across all promoter preferences (unless specified to be for target promoters only). Target promoters of a certain TF or mutant are promoters with a Z-score >2.5 for this mutant.

Normalized ChEC signal representation (Figure 1G)

The plot shows the median of the normalized ChEC signal between biological repeats. The signal is scaled such that the upper limit represents 60% of the maximum signal across all promoters.

In-vitro 7mer score and top 7mer selection

If available, the PBM-derived *in-vitro* position weight matrix (PWM) for each TF was downloaded from the CIS-BP database⁴ (the actual PBM-IDs used can be found in the Table S2). After normalizing the PWM by the distribution of nucleotides in yeast promoters, we log₂ transformed the position weight matrix. For PWMs longer than 7nt, the seven most-informative consecutive positions were selected based on the variance of weights at the respective positions. We then scored each 7mer against this PWM choosing the highest score when considering forward and reverse orientation as well as different offsets. For each mutant, the top 7mer was then defined as the 7mer with the highest *in-vivo* score in the TF or its mutant (see below) among the 10 7mers with the highest *in-vitro* scores.

In-vivo 7mer scores

For the *in vivo* 7mer analysis, all possible 7mer sequences were given a numerical index (8192 indices in total, with forward and reverse complement sequence of each 7mer given the same index). Each nucleotide in the yeast genome was indexed according to the 7mer that centers on it. To measure the binding at each position, we first average ChEC-signal of 31 bp surrounding the site (-15 to +15), then subtracting the average ChEC-signal 7bp around the site (-3 to +3 bp). This subtraction is done to account for the observed ChEC signal depletion at the exact TF binding site, as the MNase only cuts in the vicinity and not at the TF binding site. The *in-vivo* 7mer score is then defined as the mean value of this calculated score at all promoter regions containing this motif.

Motif enrichment analysis (Figure 1J)

For each TF and its DBD, we defined the motif enrichment (ME) as follows:

$$ME = \log_2 (\text{mean}(BSn,m)/(\text{mean}(BSp,m)))$$

where BSn,m is the total ChECSeq signal of Variant m (DBD or TF) around all occurrences (+/- 25bp) of the Top 7mer of DBD and TF in promoters combined; and BSp,m corresponds to the binding signal on all promoter locations.

7mer selection bias (Figures 2A and 2B)

The 7mer selection bias is defined as the Z-score-normalized *in-vivo* 7mer score (see above) of the top 7mer (see above) of the TF or DBD when compared against the 7mer scores of all 8192 possible 7mers for the TF or DBD, respectively.

Binding signal on 7mer site and fraction of bound sites (Figures 2C and 2D)

For this, we consider all the sites in promoters containing the top DBD and TF 7mer (see above). The cumulative binding signal around each site (+/- 25 bp) is calculated and shown in 2D. The bound fraction for TF or DBD is then defined as the fraction of sites with signal above 3% of the maximum signal across these sites in DBD or TF, respectively.

Correlation maps (Figures 3B–3D)

Correlation maps were generated by first combining the target promoters (promoter preference > 2.5) of full-length Transcription factor and the DBD. Then the correlation between each truncation and the TF or the DBD across these target genes is calculated and plotted on the Y- and X-axis respectively. The DBD and TF correlation of the missing 50-AA truncations were interpolated from the surrounding 2 truncations using linear interpolation, where the independent variable is the number of removed residues and the dependent variable is the correlation with the TF or DBD respectively.

Relative binding across truncations (Figure 3B)

To calculate the relative binding across the truncation series on a particular target promoter, we first calculated the binding signal on this promoter (PBS) in each mutant. The relative binding log₂-FC(fold change) at each mutant was then defined as:

$$\log_2(PBS_{i,j} + 700) - \text{mean}(\log_2(PBS_{n,j} + 700)),$$

with PBS_{i,j} being the PBS of mutant *i* at promoter *j* and PBS_{n,j} being the PBS of all the nonDBD truncation mutants of that TF (*n*) at promoter *j*.

Graduality score (Figure 3E)

To calculate the graduality score of each truncation series, we first interpolated every 50AA truncation (see above) and then calculated all Euclidean distances between consecutive interpolated truncations on the correlation map, i.e. truncation effect. The graduality score (graScore) was then defined as:

$$\text{graScore} = \log_2 (\text{median 50aa truncation effect})/(\text{max 50aa truncation effect})$$

and rescaled:

$$\text{graScore Rescaled} = \text{graScore}/4+1$$

so that it is 1 when all truncations have the same effect and 0 when 20% of the truncations have 80% of the cumulative effect. [Figure 3E](#) shows the rescaled graduality score for all investigated TF series.

Defining structured and disordered regions ([Figure 3G](#))

We use the IUPRED2⁵⁷ algorithm to determine the disorder scores along the residues of all TFs. To classify a truncation into disordered or structured, we calculate the mean disorder score across all additionally removed residues compared to the last truncations. If this mean score is ≥ 0.5 we classify this truncation as disordered, and structured otherwise.

Clustering of TF, DBD, and nonDBD or promoters ([Figure 4F](#))

To analyze similarity between binding profiles of TF, DBDs or nonDBD, we first calculate the pairwise correlation distance between the individual mutants ($1 - \text{corr}(V1, V2)$) and then used this distance to generate a agglomerative hierarchical clustering tree using the build-in matlab function “linkage” with the method set to “average”. This tree was then used to optimally order the binding profiles using “optimalleaforder” and the emerging clusters manually assigned and shown on the heatmap.

To analyze similarities between combined target promoters of all TF, DBD or nonDBDs we first calculated the distance between two target promoters as their correlation across all mutants, i.e. TF, DBD or nonDBDs, and then used the same functions as above to visualize the clustering. To define clusters we used the build-in hierarchical cluster algorithm of MATLAB “cluster” with “criterion” set to “distance” and various threshold values as indicated on the plot.

Gene expression plasticity ([Figure 4E](#))

For each gene, expression plasticity was calculated as the standard deviation of this gene’s relative expression across 1484 yeast deletion mutants analysed by Kemmeren et al.⁵⁸

OPN score ([Figure 4A](#))

OPN score was calculated as described in Rosin et al.³⁴ The TSS proximal region was defined as the region from the TSS to 150bp upstream the TSS, and the TSS distal region was defined as the region between 200–400 bp upstream the TSS. Then, the average nucleosome occupancy in the proximal and the distal regions was calculated, and the DPN-score was defined as: $\log_2(\text{average nucleosome occupancy at distal TSS}) - \log_2(\text{average nucleosome occupancy at proximal TSS})$.