

Systematic analysis of low-affinity transcription factor binding site clusters in vitro and in vivo establishes their functional relevance

Received: 22 June 2022

Accepted: 25 August 2022

Published online: 07 September 2022

 Check for updates

Amir Shahein¹, Maria López-Malo ¹, Ivan Istomin ¹, Evan J. Olson¹,
Shiyu Cheng¹ & Sebastian J. Maerkl ¹ 

Binding to binding site clusters has yet to be characterized in depth, and the functional relevance of low-affinity clusters remains uncertain. We characterized transcription factor binding to low-affinity clusters in vitro and found that transcription factors can bind concurrently to overlapping sites, challenging the notion of binding exclusivity. Furthermore, small clusters with binding sites an order of magnitude lower in affinity give rise to high mean occupancies at physiologically-relevant transcription factor concentrations. To assess whether the observed in vitro occupancies translate to transcriptional activation in vivo, we tested low-affinity binding site clusters in a synthetic and native gene regulatory network in *S. cerevisiae*. In both systems, clusters of low-affinity binding sites generated transcriptional output comparable to single or even multiple consensus sites. This systematic characterization demonstrates that clusters of low-affinity binding sites achieve substantial occupancies, and that this occupancy can drive expression in eukaryotic promoters.

DNA regulatory sequences, especially in eukaryotic organisms, frequently contain clusters of proximal binding sites^{1–3}. Both theory and experiment suggest that this clustering of binding sites is necessary for effective regulatory function^{1,4,5}. Accordingly, binding site clusters, in particular those composed of binding sites for the same transcription factor (homotypic clusters) or of low-affinity sites, have received increasing attention with several functionally important examples identified to date^{6–8}. For instance, clusters of low-affinity binding sites were found to be critical for the precise temporal control of gene expression⁹, spatial control and patterning in development^{6,10,11}, as well as robustness to mutations⁶, features that are hallmarks of native biological systems and highly relevant for engineering synthetic systems. However, our understanding of how such outcomes can be encoded in eukaryotic regulatory sequences and transduced by transcription factors, co-factors, and other biological machinery into function is still in its infancy. In prokaryotes^{4,5,12,13} and eukaryotes^{14–16}, biophysical models have had success in predicting gene expression based on sequence information alone, but we are missing comparable insights where DNA regulatory sequences contain complex clusters of binding sites. In order to build a better understanding of eukaryotic

transcription, we argue that it is important to characterize and develop a quantitative understanding of how transcription factors bind to multiple proximal binding sites varying in affinity, multiplicity, and density¹⁷.

The bulk of existing research on TF-DNA interactions has revolved around single binding sites, often focusing on consensus sequence or high-affinity binding sites. We currently lack a detailed, biochemical analysis of the binding of transcription factors to binding site clusters. For instance, while several reports identified that transcription factors act synergistically when binding to clusters of low-affinity sites^{15,18}, whether this synergy is a biochemical property of the TF-DNA interaction or depends on additional in vivo complexity remains unknown. Current biochemical approaches for studying transcription factor-DNA interactions either do not provide a direct readout of the occupancy of multiple transcription factors interacting on DNA^{19,20}, or lack the throughput necessary to tackle the larger sequence space that emerges when working with multiple sites^{18,21}. Furthermore, current methods tend to be biased towards high-affinity interactions, due to the process of isolating bound molecules which can lead to dissociation of lower-affinity, transient binding events.

¹Institute of Bioengineering, School of Engineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.  e-mail: sebastian.maerkl@epfl.ch

To fill this knowledge gap, we apply the quantitative, high-throughput MITOMI assay²² to study binding site clusters. The MITOMI assay performs hundreds to thousands of parallel interaction measurements on a single microfluidic device by surface immobilizing transcription factors exposed to solution phase short DNA targets. MITOMI has been applied to the detailed quantitative characterization of protein - DNA^{22–28}, protein - RNA²⁹, and protein - protein interactions³⁰, as well as small molecule drug discovery³¹. The MITOMI method has also been extended to enable the high-throughput analysis of molecular association and dissociation kinetics^{32,33}.

We slightly modify the original MITOMI assay by inverting the assay geometry using surface immobilized 90 bp-long target DNA strands and solution-phase transcription factors in order to optimize the quantitative characterization of binding site clusters. Reconfiguring MITOMI in this way allows for regulatory sequences ranging from single binding sites to large complex clusters to be encoded on DNA, and for the equilibrium occupancy of multiple transcription factor molecules to be observed. Given the throughput of our assay, we design a library of DNA sequences to enable direct assessment of how transcription factor binding is affected by changes to cluster configuration, including site affinities, multiplicities, and densities. We assess transcription factors from two of the largest transcription factor families: zinc fingers and basic helix loop helix (bHLH).

We find that when DNA targets contain binding sites which overlap and share common basepairs, binding is only partially reduced compared to non-overlapping sites, suggesting that transcription factor molecules can withstand significant steric clash and simultaneously occupy two sites sharing basepairs. This effect, occupancy despite clash, decreases as the cluster density increases to require greater levels of steric clash in states where transcription factors bind neighboring sites, or as transcription factor occupancy on DNA becomes less energetically favorable due to lower site affinities. Binding to low-density clusters where sites do not share neighboring basepairs appears largely independent, lacking prominent levels of binding synergy.

We also show that compared to an individual high-affinity binding site, clusters containing as few as three weak binding sites each an order of magnitude lower in affinity than the consensus sequence will reach greater occupancy levels in vitro at transcription factor concentrations that likely occur in vivo. Although clusters of weak binding sites achieved high-occupancies in vitro, it remained unknown whether occupancy from low-affinity, transient binding events would translate to an in vivo setting and give rise to functional regulatory elements. For instance, many past reports in vivo have considered dwell times to be an important factor determining functional gene regulation^{34–38}. To determine the functional relevance of weak binding site clusters in vivo we generate and characterize a synthetic library of minCYC1 promoters containing multiple Zif268 binding sites, and to show that the gene expression level driven by clusters of low-affinity sites can match those achieved by high-affinity, consensus binding sites. Finally, to determine whether clusters of weak binding sites are functional in a native gene regulatory network we test weak binding site clusters in the inorganic phosphate regulatory network. We replace native binding regions containing high-affinity Pho4 and Pho2 binding sites with clusters of low-affinity Pho4 sites in the PHO5 promoter and find that in the context of this native regulatory system controlled by physiological levels of Pho4 transcription factor, gene expression is recovered with binding site clusters consisting of individual binding sites an order of magnitude lower in affinity than the consensus site.

Results

iMITOMI development and characterization

We developed an inverted MITOMI assay (iMITOMI), by adapting the original MITOMI platform²² to generate quantitative measurements of

transcription factors binding at equilibrium to longer DNA targets containing binding site clusters of up to 6 distinct binding sites (Fig. 1A–H, Supplementary Note 2 for platform development) (thesis: Istomin I., EPFL, 2020). We also tested and characterized these binding site clusters in vivo in the eukaryotic model organism *S. cerevisiae* (Fig. 1I, J).

In vitro characterization of binding site clusters

We applied our method to characterize the binding of Pho4 from the bHLH family and Zif268 from the zinc finger family to DNA targets composed of multiple binding sites, ranging from from 1 to 6 sites for Zif268 (Fig. 2A–E) and from 1 to 5 sites for Pho4 (Fig. 2F–J). We characterize compositions of weak sites (W) and very-weak sites (V), in the range of one, or two orders of magnitude lower affinity than the consensus site, respectively.

For each transcription factor, the bound transcription factor signal (normalized by DNA fluorescence) at saturating transcription factor concentrations (which we refer to as the saturation fluorescence, B_{\max}) was similar for single strong, medium, and weak binding sites (Supplementary Fig. 3). As the number of weak binding sites in a DNA target is increased, we observe a linear increase in the saturation fluorescence (R^2 of 0.99 for Zif268 and 0.95 for Pho4), with a step size corresponding to the saturation fluorescence for single binding site targets (Supplementary Fig. 4). This suggests that the binding sites are saturating with transcription factor, and that the step increase in B_{\max} (for Zif268, the slope of the line in Supplementary Fig. 4E) corresponds to the fluorescence signal resulting from an occupancy of 1 TF bound per DNA molecule (units of RFU per TF molecule bound on average). Therefore normalizing by this value converts our bound transcription factor signal (RFU) into a measure of mean occupancy, $\langle N \rangle$, which is simply the average number of transcription factors bound per molecule of DNA (Fig. 2C–E, H–J). In the Pho4 DNA library, the fourth binding site we introduced was a very-weak site (Supplementary Fig. 3), and thus not fully saturated in the assay. Accordingly, in this case we fit our linear regression to the targets ranging from 1 to 3 binding sites to obtain the step increase in fluorescence at saturation. We chose to use a simple model with only two parameters, K_d and B_{\max} , rather than a model that accounts for the different affinities of each of the individual binding sites.

Our results show that binding to clusters of sites an order of magnitude weaker than the consensus binding sequence can be saturated, and that all binding sites in even closely-spaced clusters can be concurrently occupied by transcription factors, resulting in a maximum occupancy (N_{\max}) that consistently increases by 1 with each additional site in a DNA target (Fig. 2E, J, upper panels). In contrast, the measured affinity (K_d) is determined by the individual binding site affinities present in the target cluster (Fig. 2E, J, middle panels).

Having shown that clusters of weak and very-weak binding sites can be fully occupied at high transcription factor concentrations of 1 μM , we next asked whether weak and very-weak clusters also lead to sufficient occupancy at lower transcription factor concentrations. The in vivo nuclear concentration of Pho4 in *S. cerevisiae* has previously been estimated to be approximately 355 nM based on single-cell fluorescence experiments¹⁴. Estimates for zinc-fingers in yeast range from 538 to 3334 copies per cell, which translates to a nuclear concentration of roughly 213 nM to 1322 nM^{39,40}. As single consensus sequence targets are known to be physiologically functional in a cellular milieu, we determined the concentrations at which low-affinity binding site clusters exhibit similar occupancy levels as the corresponding single consensus sequence (Fig. 2E, J, lower). In other words, we plotted the concentrations at which the low-affinity cluster saturation binding curves cross the single consensus target saturation binding curve. It can be readily seen that similar occupancies can be reached by all low-affinity clusters, with cross-over occurring at

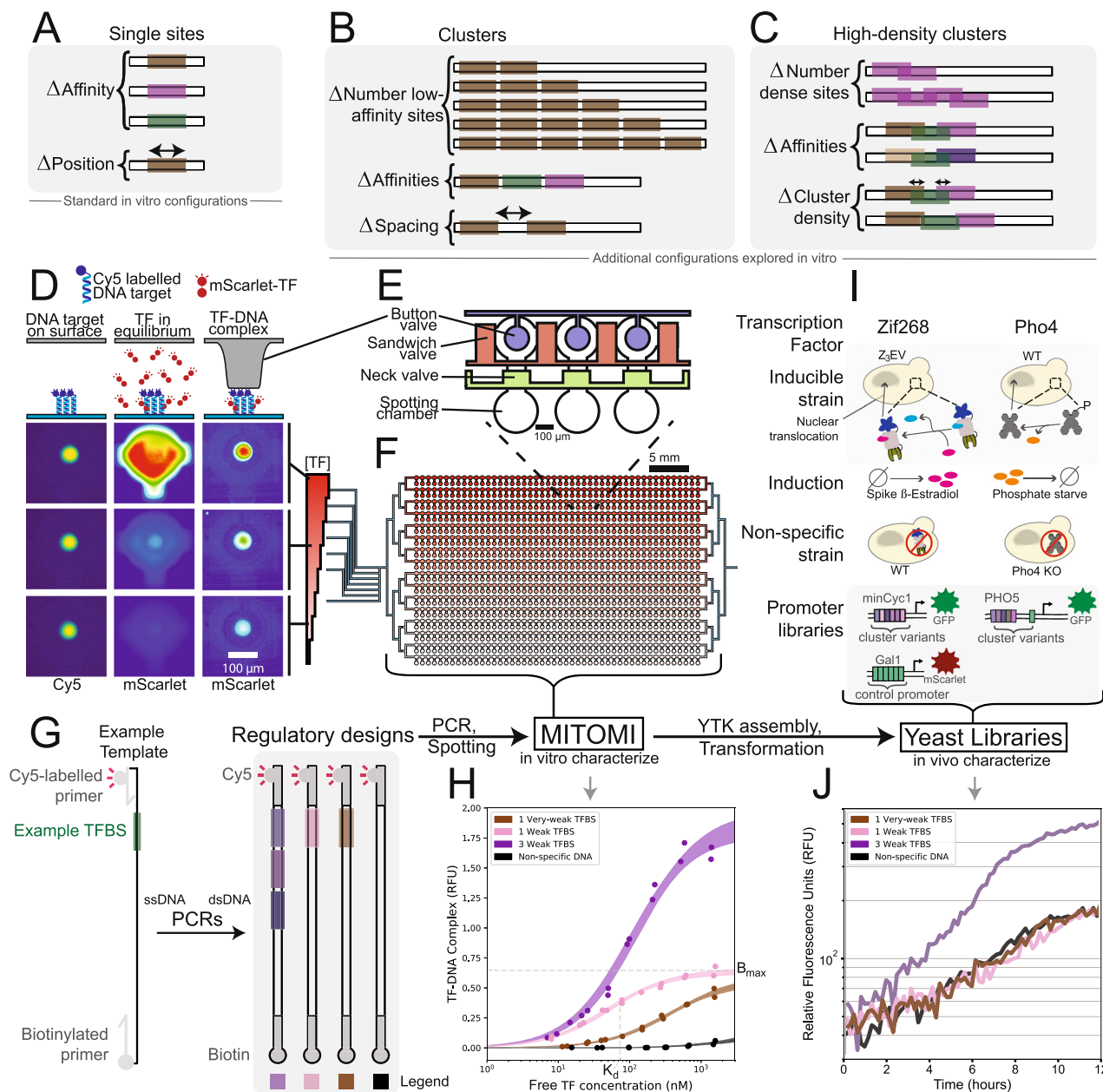


Fig. 1 | Description of the in vitro and in vivo measurements. A–C DNA libraries and parameters explored in vitro using iMITOMI, including clusters of multiple binding sites (**B**), and high-density clusters of overlapping binding sites (**C**).

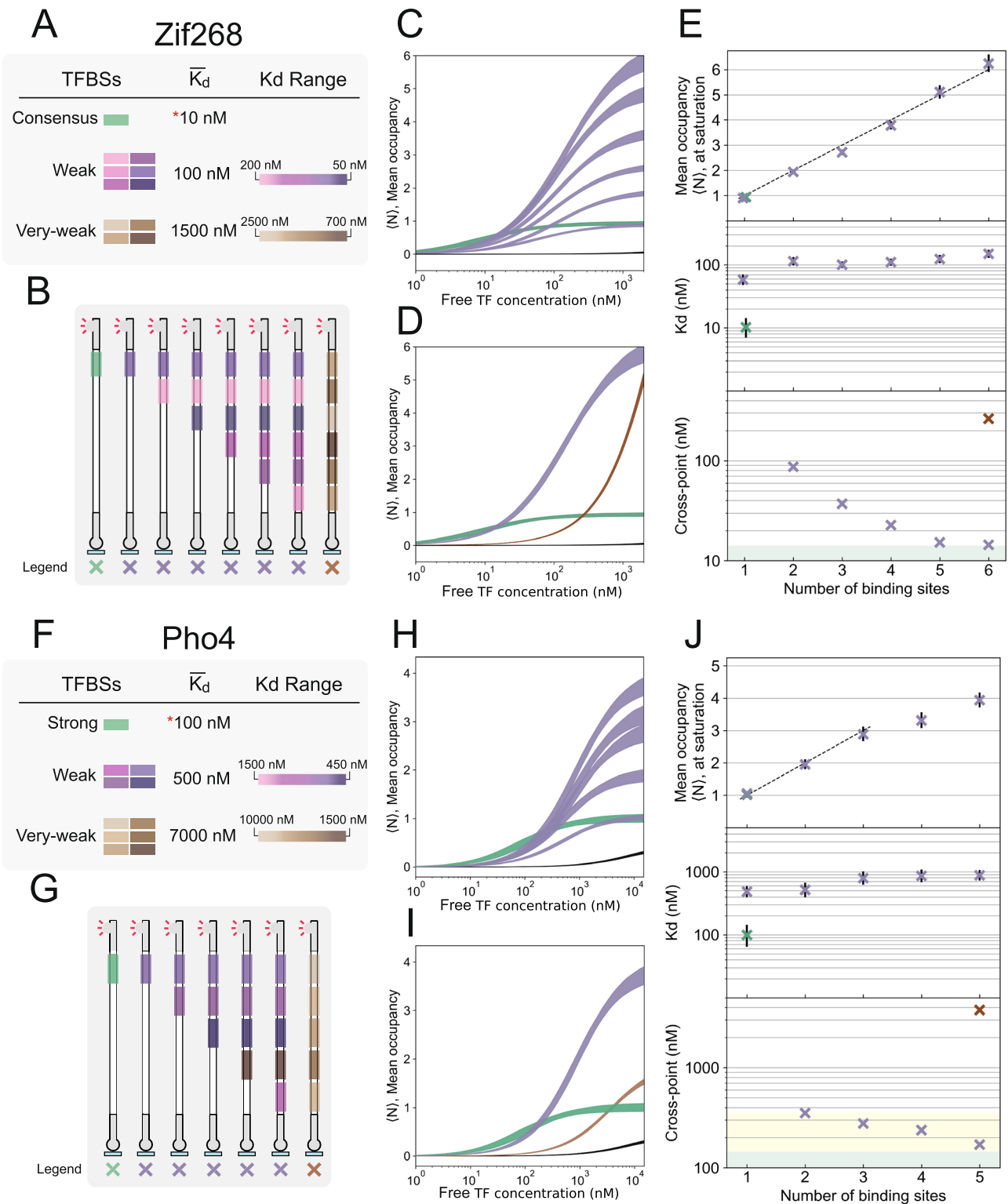
D–F Schematic of the high-throughput, in vitro iMITOMI method. The iMITOMI chip consists of two-layers, where microvalves in the control layer (**E**) control the flow of fluid in a flow-layer containing 768 programmable reaction chambers (**F**). **D** Image rows represent different TF concentrations, columns represent stages of the assay. DNA is immobilized on the chip's surface, in each assay chamber, under the button valves (left column). Fluorescently-tagged transcription factor molecules are flowed at different concentrations (red color gradient) into different rows of the chip. The free (middle column) and bound (right column) transcription factor signal is quantified at equilibrium. **G** To obtain the dsDNA targets, 90 bp-long

single-stranded templates are amplified with biotinylated and Cy5-tagged primers. **H** Example 2-parameter saturation binding curves for DNA targets containing from 0 to 3 binding sites spanning a wide affinity-range. Shaded regions represent the 5 to 95% confidence intervals identified by exploring model parameter space with a Markov Chain Monte Carlo sampler. **I** In vitro-characterized clusters are assembled into expression cassettes, and chromosomally integrated into *S. cerevisiae* for characterization (**J**). Zif268 regulatory sequence designs were characterized in the minCYC1 promoter using an inducible Z3EV system, where a wildtype (WT) strain enabled quantification of the non-specific response (**I**, left). Pho4 regulatory sequence designs were characterized in the PHO5 promoter, in a WT strain induced by phosphate starvation, where a Pho4 knockout strain enabled quantification of the non-specific response (**I**, right).

transcription factor concentrations of as low as 14 nM for a cluster of 6 low-affinity Zif268 binding sites, and 170 nM for a cluster of 5 low-affinity Pho4 binding sites. These values are well within the range of physiological concentration estimates and indicate that low-affinity binding site clusters may achieve comparable occupancy levels as a single high-affinity consensus site in vivo.

We exploited our platform to explore the impact of binding site proximity on binding affinity (Fig. 3). First, we set out to determine

whether proximal binding sites exhibit synergistic interaction as has been reported in vivo¹⁵. We constructed and characterized a library of targets with two low-affinity binding sites, where the position of one site was held constant, while the position of the second site was moved in intervals to adjust the spacing between sites (Δ). For Zif268, binding remained largely unchanged across the library, barring one unexplained outlier (Fig. 3A, 5 bp spacing). In the case of Pho4, we observed a minor decrease in $\langle N \rangle_{\max}$ for gap distances centered around -3 bp,



possibly indicating mild steric competition in this orientation of transcription factor molecules (Fig. 3B). The K_D values remained largely constant as a function of distance, and corresponded to the individual binding site affinities. We therefore did not observe a large degree of cooperativity between adjacent binding sites (Supplementary Fig. 4G), implying that the reported synergism may be a product of higher-order complexity in vivo. Together, our results suggest that the equilibrium binding we observe to different binding sites in targets composed of multiple sites is largely independent, possibly with mild steric effects at close distances.

Overlapping binding sites occur frequently in regulatory DNA of prokaryotes and eukaryotes^{12,13,41-44}, and are thought to influence how binding occurs among transcription factors^{15,42,45-47}, RNA polymerase, other regulatory elements^{42,48-51}, and nucleosomes²⁰. In eukaryotes, the relatively low specificity of transcription factors gives rise to regulatory sequences that contain high-density clusters of binding sites tending to overlap with one another^{12,13,42}. As a result of high binding site densities, De Boer et al. found that shifting a site by one basepair across regions of a yeast promoter impacts expression primarily through the disruption or creation of alternate transcription factor binding sites

Fig. 2 | In vitro characterization of low-affinity binding site clusters. **A, F** Legend for the Zif268 and Pho4 clusters analyzed. Binding sites were grouped into different affinity classes. Consensus or strong sites, weak sites approximately an order of magnitude lower affinity, and very-weak sites approximately two orders of magnitude lower affinity than the consensus site. **B, G** Schematic of the Zif268 and Pho4 DNA targets characterized. **C, H** Amount of TF in complex with DNA as a function of free TF concentration at equilibrium (saturation binding curves) for DNA targets containing from 1–6 (Zif268, **C**) or 1–5 (Pho4, **H**) low-affinity binding sites (purple) compared against 1 strong binding site (green). Shaded regions represent 5 to 95% confidence intervals. Binding curves with all raw datapoints, and further analysis are available in Supplementary Fig. 4. **D, I** Saturation binding curves comparing the largest cluster of very weak sites (brown) against the largest cluster of weak sites

and a single strong binding site. **E, J**, top Mean occupancies at saturating transcription factor concentrations, $\langle N \rangle_{\max}$ (R^2 of 0.99 for Zif268, R^2 of 0.95 for Pho4). **E, J**, middle) Measured affinity values (K_d). **E, J**, bottom Cross points: the transcription factor concentration at which the mean occupancy of a DNA target composed of multiple medium or weak binding sites reaches the same mean occupancy as the consensus or strong binding site target. The yellow shaded region represents a range of expected in vivo transcription factor concentrations, and the green shaded region represents the characterized K_D for the strong binding site. K_d and $\langle N \rangle_{\max}$ parameter markers are centered at the mean, while error bars represent 5 to 95% confidence intervals. Samples were analyzed in 20–84 independent chambers, across 3–6 independent experiments.

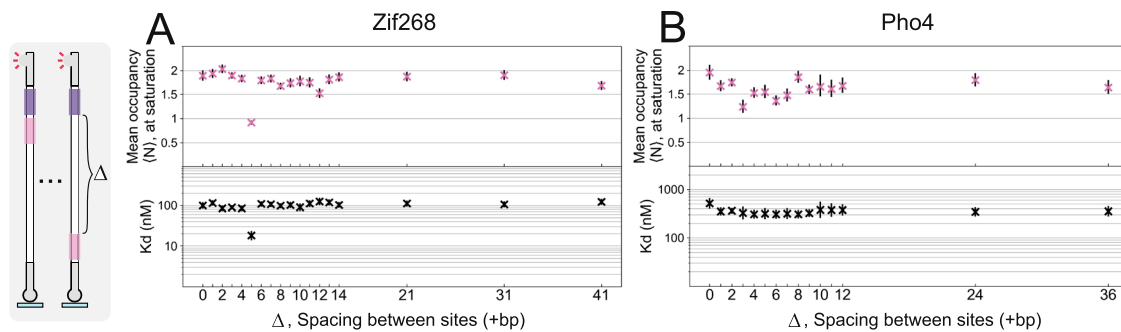


Fig. 3 | Effect of distance between two adjacent binding sites. **A, B** Affinities (K_d), and mean occupancies at saturation ($\langle N \rangle_{\max}$), for a library of DNA targets for Zif268 (**A**) and Pho4 (**B**), with different distances between 2 weak binding sites. K_d and

$\langle N \rangle_{\max}$ parameter markers are centered at the mean, while error bars represent 5 to 95% confidence intervals. Samples were analyzed in 18–34 independent chambers, across 2 independent experiments.

(ref. 13). While competition between transcription factors binding to nearby but distinct binding sites is typically thought of as a graded function of distance, as a rule of thumb overlapping binding sites are often assumed to elicit exclusive transcription factor binding due to steric clash^{46,52}, and consequently binding models tend to ignore states where two transcription factors are bound simultaneously to sites sharing common basepairs¹⁵.

To our knowledge, binding of transcription factors to high-density clusters has not been characterized in detail. Thus we designed DNA targets containing from 1 to 4 binding sites of varying affinity for Zif268, where one or three basepairs are shared between neighboring binding sites (Fig. 4A). An alignment of the protein-DNA complexes to the DNA sequences together with residue-clash prediction suggests that the protein molecules will exhibit significant steric clash as binding site motifs start to overlap and share common basepairs (Fig. 4B, C). Yet surprisingly, we discovered that the mean occupancy of transcription factor molecules binding to two binding sites that share basepairs can exceed a value of 1 at the upper range of physiologically relevant transcription factor concentrations (Fig. 4D). This demonstrates that occupancy despite clash can occur, where two transcription factor molecules bind at once and exhibit steric interference on binding sites sharing common basepairs, and that this might occur in high-density clusters in the genome depending on the degree of site saturation⁵³.

Since a simple one-site saturation binding model fails to describe this phenomenon, we extended our description of binding to a statistical mechanical model. First we attempted to model binding to high-density clusters using an exclusive binding model (full steric occlusion) where two transcription factors cannot bind concurrently to overlapping binding sites, resulting in just three states of the system for a 2-site cluster (Fig. 4D, Supplementary Table 7). We parameterized the model using binding energies (ϵ_{bs}) derived from an independent characterization of the individual binding sites (Supplementary Fig. 3). As expected, for a two-site cluster the exclusive model fails to account for mean occupancies above 1. We incorporated an additional permissive state into the model, where two transcription factors can bind

simultaneously, and fit a floating interaction energy between them (ϵ_{clash}) (Fig. 4D, E, Supplementary Fig. 5, Supplementary Table 8). This 2-parameter (1 floating) model described the experimental data better than the exclusive model. Both the Akaike information criterion (AIC), and Bayesian information criterion (BIC) decreased significantly for the permissive model, which are measures of goodness-of-fit, and useful for conducting model selection when models differ in their number of parameters (penalizing the permissive model for its additional parameter, Supplementary Fig. 5).

As the number of basepairs shared between binding sites increases, the mean occupancy decreases (as a function of concentration), and the probability of the 2 transcription factor-bound state is reduced, although it is still higher than initially expected at a large amount of overlap (3 basepairs, or 33% of the motif, Fig. 4F). Furthermore, as we decrease site affinities, we observe a large decrease in occupancy in overlapping sites, suggesting that the transcription factor's affinity for its binding site (ϵ_{bs}) is important for compensating for the steric interference.

We further investigated binding to high-density clusters where individual sites differ in their affinities (Fig. 4G, H). In this case, binding is dominated by the stronger binding sites in a given cluster. For the clusters of three mixed-affinity binding sites, the maximum occupancies and affinities were chiefly determined by the number of strong, non-overlapping binding sites present in the cluster.

In vivo validation of low-affinity binding site cluster function in a synthetic gene regulatory system

Experimental methods and computational analyses designed to study gene regulation have traditionally been biased towards single sites of high-affinity or small clusters of high-affinity binding sites. Homotypic clusters of low-affinity sites are prevalent in eukaryotic regulatory sequences¹, yet the level of binding to these clusters in the in vivo context and the resulting impact on gene expression is uncertain. Having shown in vitro that at physiologically-relevant transcription factor concentrations an equal or greater mean transcription factor occupancy can be achieved by low-affinity clusters

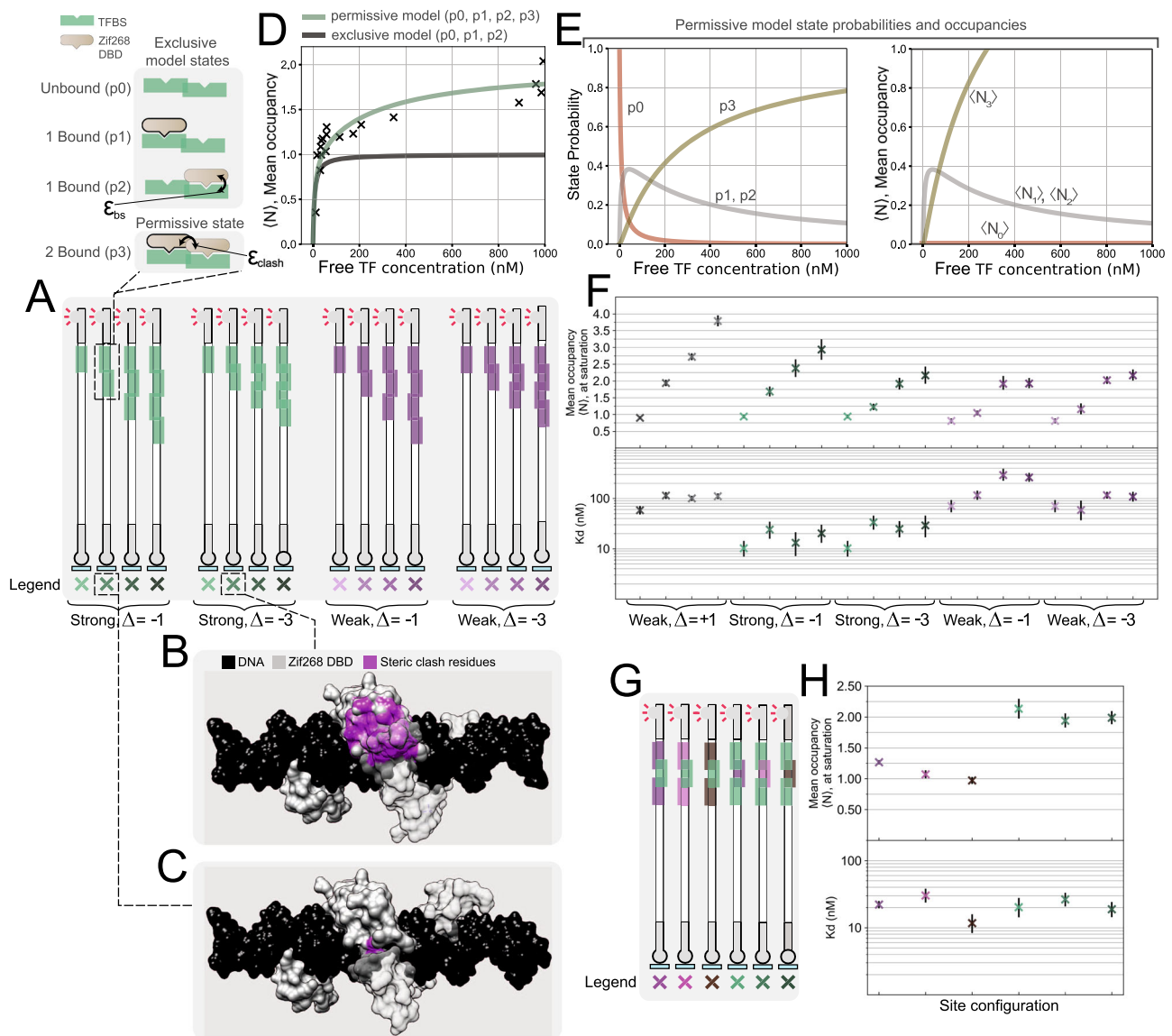


Fig. 4 | In vitro characterization of high-density binding site clusters. **A** Library of DNA targets analyzed containing Zif268 high density binding site clusters, where basepairs are shared between neighboring binding sites. Clusters contain up to four strong or weak binding sites, and neighboring sites share 1 or 3 common basepairs. **B, C** Aligned crystal structures, highlighting Zif268 residues expected to exhibit steric clash in magenta. 3 basepairs (33% of the Zif268 motif) shared in (**B**), and 1 basepair (11% of the motif) shared in (**C**). **D, E** Statistical mechanical analysis of binding to consensus binding sites sharing a common basepair. **D**, left Legend of the states corresponding to two different binding models, an exclusive binding model (top three states), and a permissive binding model (all four states). **D**, right Binding data and model predictions, illustrating that two transcription factors can

bind at once to binding sites sharing basepairs. **E** Predicted probability (left) and occupancy (right) for each permissive model state as a function of transcription factor concentration. **F** Experimental data for the full library of high-density clusters (data fit with a 2-parameter SBC similar to above). Occupancies at saturation, $\langle N \rangle_{\max}$, and affinities, K_D . **G** DNA targets containing mixed-affinity high-density clusters where 3 basepairs are shared between neighboring binding sites. **H** Corresponding experimental data for mixed-affinity clusters. K_D and $\langle N \rangle_{\max}$ parameter markers are centered at the mean, while error bars represent 5 to 95% confidence intervals. Samples were analyzed in 48–84 independent chambers, across 4–6 independent experiments.

compared to an individual consensus site, we set out to investigate whether low-affinity site clusters are functional in the context of eukaryotic promoters *in vivo*.

Although our biochemical analysis showed that low-affinity binding site clusters are able to achieve high mean occupancies, several questions remained as to whether low-affinity binding site clusters would be functional *in vivo*. First, it is not a priori known whether high mean occupancies of transcription factors alone are sufficient to recruit the necessary regulatory machinery to a promoter and give rise to transcriptional activation. Consensus binding sites are able to achieve high mean occupancies of a single transcription factor, and this occupancy is further characterized by long-dwell times of the

transcription factor on its binding site³². Low-affinity clusters on the other hand achieve a similar mean occupancy in non-saturating conditions through a time-averaged occupation of several transcription factors bound to the cluster, and each transcription factor - DNA interaction exhibits a considerably shorter dwell time. It is a matter of debate whether high occupancies or long dwell times are necessary for *in vivo* function of regulatory regions. Second, we estimated that *in vivo* transcription factor concentrations may be high enough to give rise to sufficiently high occupancies of low-affinity clusters, but this assumption remained to be tested *in vivo*.

We first addressed the question of whether low-affinity binding site clusters can be functional *in vivo* by generating synthetic minimal

promoters, driven by an engineered, exogenous Zif268 transcription factor. We inserted the 90 bp regulatory sequences that were characterized in our *in vitro* work approximately 200 bp upstream of the transcription start site in the minimal CYC1 promoter (Fig. 5A). These sequences characterized *in vitro* were originally designed into a DNA backbone selected to prevent binding to known yeast transcription factors (using R2oDNA Designer⁵⁴). We varied similar regulatory sequence parameters and maintained similar affinity-class definitions as for our low-affinity *in vitro* binding characterization.

We used the yeast toolkit (YTK) 2.0 workflow⁵⁵ to assemble integration cassettes with our engineered minCYC1 promoters upstream of GFP. We integrated these synthetic cassettes into the Lys2 locus of a modified Z₃EV-containing parental yeast strain (Fig. 5B)⁵⁶. Z₃EV is a synthetic transcriptional activator with three domains: the Zif268 DNA-binding domain, an estrogen receptor element that drives nuclear translocation in the presence of β -estradiol, and a VP16 domain that recruits transcriptional machinery to activate transcription. We also included a control promoter into the parental strain which drives mScarlet expression under the regulation of 6 consensus Zif268 binding sites in a synthetic GAL1 promoter backbone. The same parental strain was used for all strains generated, allowing us to validate the correct functioning of the inducible gene expression system to ensure consistency across strains.

Yeast strains were either induced with a high concentration of β -estradiol (200 nM), or not uninduced (0 nM), and measured in the mScarlet and GFP channels on a multi-mode plate reader (Fig. 5C) and validated by direct fluorescence microscopy imaging (Fig. 5D, Supplementary Fig. 6). All strains exhibited highly consistent growth rates and mScarlet expression from the control-promoter (Fig. 5C, Supplementary Fig. 6E–H). In contrast, the cluster variants in our engineered promoter library led to a wide range of gene expression levels (GFP) that depended on cluster configuration and binding site affinities. We normalized GFP intensities by cell densities, and averaged this normalized gene expression signal for all readings between 8.5 h and 9.5 h after induction (Fig. 5E, F). Each datapoint in Fig. 5E corresponds to an independent replicate, where generally around three replicates were characterized in each experiment, and the experiment was repeated from three to six times for each strain.

All promoters containing 1, 3, or 6 consensus binding sites gave rise to functional activation and expression of GFP. The minCYC1 promoter containing a single consensus binding site achieved only modest expression levels, whereas 3 consensus binding sites led to very high expression of GFP (Fig. 5E–H). Increasing the number of consensus binding sites from 3 to 6 led to slightly higher gene expression levels, potentially indicating that expression levels are near the upper limit of what the minCYC1 promoter can support. The output of a promoter containing no binding sites (NS) on the other hand remained low, both in the induced and uninduced conditions. We tested binding site clusters containing between 1 and 6 very weak binding sites with average K_D values 2 orders above the K_D of the consensus sequence. None of these promoters led to a measureable increase in gene expression when comparing the induced to the non-induced state, with the possible exception of a promoter containing 6 binding sites with slightly higher binding affinities (W-6). We started observing a difference in gene expression between the induced and non-induced conditions, when testing clusters of weak binding sites with average K_D values one order of magnitude above the K_D of the consensus sequence. Clusters of as few as 3 weak binding sites (W3) surpassed the single consensus binding site in its ability to activate gene expression, while clusters of 4, 5, or 6 weak binding sites achieved a similar level of gene expression compared to clusters composed of three consensus binding sites. Therefore, we were able to show in this minimal synthetic gene regulatory system, that small clusters of low-affinity binding sites can drive gene expression equivalently to clusters of consensus binding sites.

Interestingly, we found that not only did promoters vary in their induced levels of GFP-expression, but they also varied significantly in uninduced gene expression in a way that depended consistently on the promoter class. Increasing numbers of binding sites within a cluster gave rise to increasing levels of leakiness in the uninduced condition (Fig. 5I, J). The highest level of leakiness was observed for a cluster of 6 consensus sequences (S6), and the level of leakiness scaled with the number of weak binding sites in a cluster (W1-6). Even non-functional clusters of very weak binding sites gave rise to increased leakiness (V6 and W-6). The one exception to this general observation was the promoter containing 3 consensus sequences (S3) which achieved high levels of induction without any significant increase in leakiness. We reasoned that this response could either be specific and due to Z₃EV transcription-factor leak under non-inducing conditions, or non-specific, resulting from the binding of alternative endogenous transcription factors. Therefore, we characterized the entire promoter library in a wildtype BY4741 yeast strain lacking Z₃EV, and discovered that the level of expression correlated strongly with uninduced gene expression in the Z₃EV parental strain (R^2 of 0.9, Supplementary Fig. 6). This confirmed that the leakiness in the uninduced state was due to non-specific binding of one or more endogenous transcriptional activators.

Our library of well-characterized regulatory sequences where mean occupancy and affinity vary independently is useful for trying to understand which binding properties predominantly drive transcription from clusters. In eukaryotes, there are limited and conflicting examples where gene expression has been explained using a reduced set of interpretable binding properties^{5,13–16,37}. We used our *in vitro* characterization data to predict the mean occupancies of transcription factor molecules binding to clusters in the Z₃EV promoter library *in vivo*. We compared these predictions to the TF-specific promoter responses (Fig. 6).

Our model suggests that the minCYC1 promoter regulates gene expression through a simple thresholded relationship with the average number of bound transcription factor molecules, where up until an average of around 1 transcription factor bound to the promoter, gene expression is largely silent, and after this threshold, expression depends primarily on the average number of transcription factor molecules bound rather than on their affinity for their binding sites. For instance, a cluster composed of 5 weak binding sites (W5) or 3 strong binding sites (S3) differ in affinity by an order of magnitude, but show similar levels of gene expression which can be explained by their predicted mean occupancies. These results do not vary qualitatively within a reasonable range of expected parameter values (Supplementary Fig. 6). Thresholding can also be observed qualitatively, since one weak site (W1), or two weak sites (W2) show no impact on expression, but each additional weak site from then on increases expression significantly (Fig. 5E). Furthermore, adding 3 very-weak sites to a very-weak cluster (moving from V-3 to V-6) does not increase transcription factor-specific expression, while appending these same 3 very-weak sites just after a cluster that has a higher starting occupancy (moving from W3 to W3V-3) causes a transcription factor-specific increase in promoter output (Fig. 5F). We observed no difference in expression level when we changed spacing between binding sites (W3 compared to W'3), and little to no difference when we changed the ordering of sites in promoters containing binding sites from different affinity classes (W3V-3 compared to W3V'-3V).

In vivo validation of low-affinity binding site cluster function in a native gene regulatory system

Having established that clusters of low-affinity binding sites are functional in a synthetic gene regulatory network, we next addressed the question of whether low-affinity binding sites clusters would also be functional in a native promoter regulated by a transcription factor expressed at physiological concentrations. We turned to the well-studied inorganic phosphate regulatory network in *S. cerevisiae*

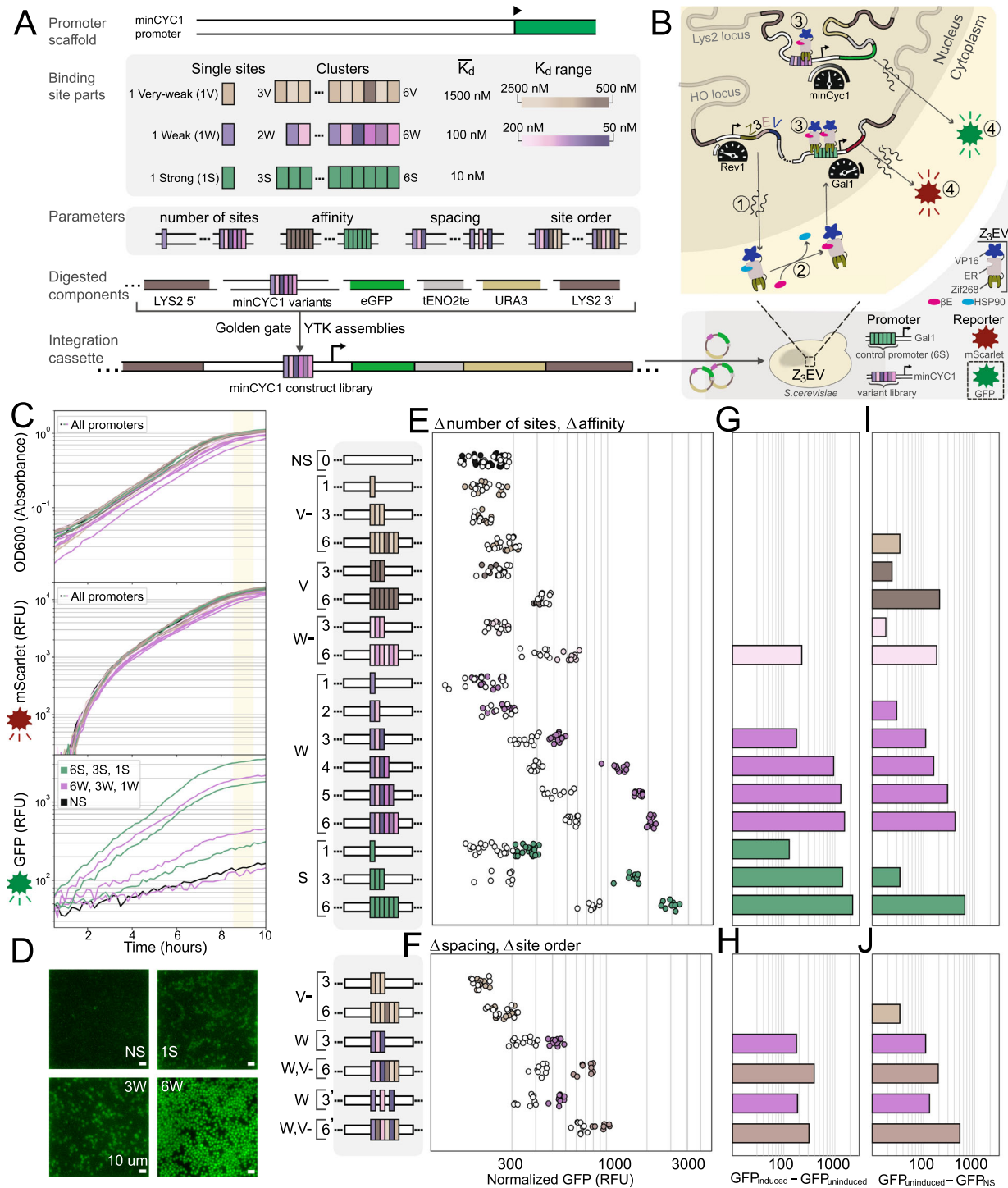


Fig. 5 | In vivo characterization in a synthetic, minimal gene regulatory system.

A Promoter legend. In vitro characterized Zif268 binding regions were inserted into a minCYC1 minimal promoter scaffold. **B** Schematic of the inducible synthetic transcription factor system. **C** All strains displayed consistent growth (OD600, upper), and mScarlet production (middle). Strains varied in their GFP production depending on the minCYC1 promoter variant (bottom). Strains were characterized with single-cell fluorescence microscopy (scalebar represents 10 microns, experiment repeated 2 times with similar results) (**D**), with fluorescence activated cell sorting (FACS, Supplementary Fig. 8), and on a multi-mode platerreader (**E–J**). **E**, **F** Strains were characterized either without (white, uninduced), or in the presence of saturating (200 nM) β -estradiol concentrations (colored, induced), over three independent experimental measurements. **E** The number of binding sites was varied from 1–6, and six distinct

affinity classes were characterized (consensus binding sites (S), weak sites (W, 10X lower-affinity, W-, 20X lower-affinity), very-weak sites (V, 50X lower-affinity, V-, 100X lower-affinity), as well as a non-specific DNA target (NS). Promoters are organized from weakest (top) to strongest binding sites (bottom). **F** We varied the spacing between sites in clusters, characterized clusters with binding sites from mixed affinity classes, and changed the ordering of binding sites in mixed-affinity clusters. **G**, **H** A Z₃EV-specific response was quantified as the difference in fluorescence intensity between induced and uninduced conditions. **I**, **J** A non-specific response was quantified as the difference in fluorescence intensity in uninduced conditions between a given promoter strain and the NS promoter-containing strain. The non-specific response also showed strong correlation with gene expression measurements of the same promoter library integrated in a BY4741 yeast strain lacking Z₃EV (Supplementary Fig. 6).

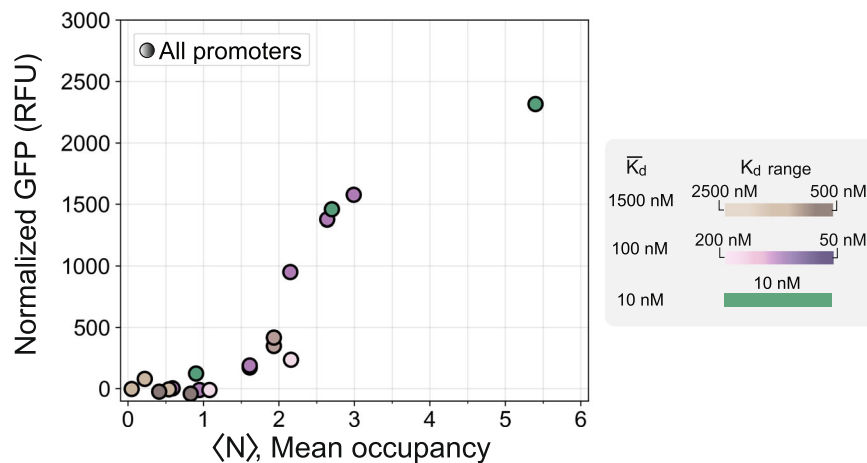


Fig. 6 | Thresholded gene activation. Thresholded relationship of gene expression (GFP normalized by OD600) as a function of predicted mean occupancy (based on *in vitro* characterization data), for the full Z₃EV strain library. The color scheme and promoters are the same as in Fig. 5.

to address this question. The PHO5 promoter is one of the best understood promoters of the inorganic phosphate regulatory network in yeast. It is activated in response to phosphate starvation, which causes the master regulator Pho4 to localize to the nucleus leading to binding to a Pho4 binding site and an overlapping Pho2 site located in a nucleosome-free region (NFR) (Fig. 7A). This binding results in the displacement of a nearby nucleosome (in the -2 position), which in turn makes a higher-affinity Pho4 site and two nearby Pho2 binding sites available for binding which are nominally in a nucleosome-occluded region (NOR). This mechanism is believed to decouple the promoter's activation threshold from its dynamic range, where as a first approximation the Pho4 and Pho2 binding sites in the NFR are considered to confer the promoter with its threshold and all binding sites in the NFR and NOR determine gene expression levels after that threshold is met^{14,57}.

Significant effort has been invested to uncover the functional relevance of the individual binding sites in the PHO5 promoter. A Pho2 binding site in the NFR has been found to recruit chromatin remodellers and to promote cooperative binding of Pho4, while the Pho2 sites in the NOR have been identified as necessary for efficient activation by Pho4^{58,59}. Ablation of Pho2 binding sites has been found to decrease gene expression up to 10-fold^{14,58}, and a Pho2 knockout strain exhibits close to no PHO5 promoter expression compared to a wildtype strain⁶⁰. However, Pho4 overexpression has also been shown to compensate partially for the loss of expression due to Pho2 binding site ablation⁵⁸, or even for the loss of expression in Pho2 knockout strains^{60–62}, presumably through an increase in occupancy of the native Pho4 binding sites. Based on this picture, and since weak clusters of Pho4 binding sites attain similar to greater occupancies than the individual consensus or native sites at relevant transcription factor concentrations *in vitro* (Fig. 2), we hypothesized that weak binding site clusters might achieve sufficiently high Pho4 occupancies in the PHO5 promoter and reproduce part of the functionality conferred by the native regulatory sequence, both in the NFR and NOR.

We replaced the native Pho4 binding sites together with their neighboring Pho2 sites, in the NFR or NOR, with clusters of 5 low-affinity Pho4 binding sites (weak or very-weak) (Fig. 7A–C) that we had characterized *in vitro* (Fig. 2). We conducted this replacement either in a native promoter scaffold (WT), or a scaffold entirely depleted of known Pho2 binding sites (delPho2), in order to decouple the Pho2-specific effect from the effect of Pho4 occupancy. PHO5 promoters were assembled into a cassette driving GFP reporter expression (Fig. 7C), and integrated into the LYS2 locus of either a wildtype yeast strain (BY4741) or a Pho4 knockout strain lacking Pho4 (Fig. 7D). All strains were sequence-verified, and exhibited consistent growth

irrespective of their PHO5 promoter variant (Fig. 7E). The positioning of nucleosomes in the PHO5 promoter has been characterized extensively and found to be robust to promoter context, such that it remains consistent when the promoter is located in an alternate locus (e.g., the LYS2 locus) and for a large degree of modification to the promoter sequence^{14,62}.

Our results on deleting native Pho4 binding sites in the NFR and NOR regions are in concordance with previous characterizations of the PHO5 promoter^{14,57} (Fig. 7E–J, Supplementary Fig. 7). Indeed, we observed greater expression for the WT promoter with a native NFR but an ablated NOR Pho4 site (WT N-A), compared to the promoter where both NFR and NOR Pho4 sites were ablated (WT A-A). This level of expression is then scaled with the addition of the native NOR Pho4 site (WT N-N). On the other hand, the promoter with an ablated Pho4 site in the NFR and a native Pho4 site in the NOR (WT A-N), resulted in similar expression to the ablated-ablated promoter (A-A), presumably because the nucleosome was not displaced effectively upon induction.

Compared to the wildtype A-N promoter, replacing the Pho4 binding site in the NFR (which also deleted the Pho2 binding site) with a cluster of low-affinity Pho4 binding sites (WT C-N) recovered Pho4-specific expression. Furthermore, compared to the N-A promoter, replacing the NOR (which also deleted the Pho2 binding sites) with a cluster of low-affinity sites (WT N-C) also improved expression compared to the N-A promoter. Promoters with no native Pho4 binding sites and a low-affinity binding site cluster in the NFR (C-A) or NOR (A-C) lost function similar to promoters with the corresponding native Pho4 binding sites present (N-A and A-N, respectively). The wildtype C-N promoter showed significantly higher expression than wildtype C-A promoter (and A-N showed ablated levels of expression), suggesting that the low-affinity cluster in the NFR may contribute to nucleosome displacement to uncover the NOR. Our results show that low-affinity clusters can generate similar expression levels of the PHO5 promoter compared to what is conferred by the native cis-regulatory sequence of the NFR and the NOR. Partial recovery of function was expected based on the aforementioned work demonstrating that overexpressing Pho4 (presumably increasing its concentration and occupancy) will only partially compensate for Pho2 site ablation^{58,60}.

Indeed, when all Pho2 binding sites are deleted from the promoter, gene expression from the native promoter (delPho2 N-N) drops by around 6-fold, consistent with previous reports^{14,58}. Remarkably, in this Pho2-depleted scaffold, replacing the native Pho4 sites in either the NFR or NOR with clusters of low-affinity binding sites (delPho2 C-N, or delPho2 N-C, respectively) actually resulted in a similar or greater level of gene expression than what was achieved by the native binding

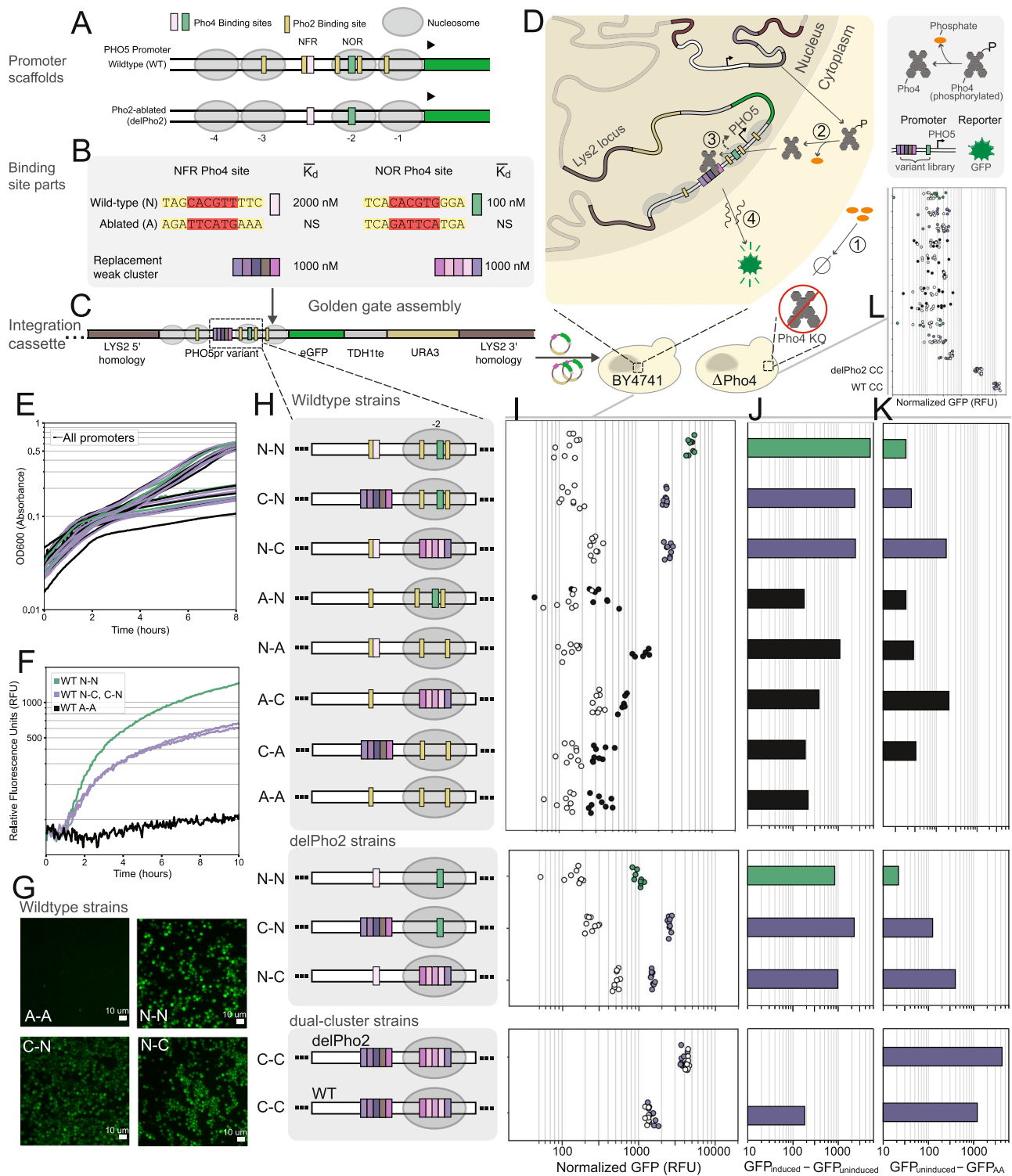


Fig. 7 | In vivo characterization in a native gene regulatory system. **A** PHO5 promoter scaffolds in which native Pho4 binding sites are replaced with low-affinity Pho4 clusters. **B** Legend of Pho4 binding site parts. Native binding regions containing Pho4 and Pho2 sites either have their Pho4 binding sites ablated, or are replaced with clusters of low-affinity Pho4 binding sites. **C** Promoters were assembled into a multi-cassette plasmid and integrated into the LYS2 locus of either BY4741 or Pho4 knockout (Δ Pho4) yeast strains. **D** Schematic of the response (induction) to phosphate starvation. **E** OD600 time series measurements for the full BY4741 library of promoters, both for uninduced and induced conditions. **F** Induction time series measurements for select cluster-containing strains.

G Strains were also characterized by single-cell fluorescence microscopy (in two independent replicates). **H–J** A Pho4-specific response was quantified as the difference in fluorescence intensity between induced and uninduced conditions. Note that in (**J**), bottom panel, the delPho2 C-C bar extends into the negative as seen in the raw data. **K** A non-specific response was quantified as the difference in fluorescence intensity in uninduced conditions between a given promoter strain and the ablated-ablated (AA) promoter strain. **L** The non-specific response can also be seen in a Pho4 knockout strain, with a similar relationship across promoter variants (Supplementary Fig. 7).

sites lacking the neighboring Pho2 sites. This implies that the Pho4-specific impact on gene expression of low-affinity Pho4 clusters is equal to or greater than for the native Pho4 binding sites and its potential co-operative interaction with Pho2. This then suggests that low-affinity clusters attain significant occupancy, which we previously established in vitro and in a minimal promoter in vivo, and can be extended to native and more complex promoter settings.

Expression in the delPho2 C-N strain does not decrease compared to the WT C-N strain (containing all Pho2 binding sites except for the one replaced by the cluster in the NFR), whereas compared to the WT N-C the delPho2 N-C strain shows lower expression, suggesting that the Pho2 binding site in the NFR is of greater importance to gene expression as has been suggested in the literature⁵⁸. Indeed, our results suggest that ablating the nucleosome-free Pho2 site hinders the efficacy of the NFR, likely in part by reducing cooperative binding to Pho4⁵⁸. However, our results suggest that the function of this regulatory region can be compensated for by strengthening the region's Pho4 mean occupancy through a low-affinity Pho4 binding cluster, leading to a higher level of gene expression.

Interestingly, similar to the Z₃EV promoter system, non-specific leaky expression correlated with the number of low-affinity binding sites (Fig. 7K). Dual-cluster replacement resulted in significant non-specific expression, which did not depend on Pho4. This is evidenced by the fact that expression did not depend on induction in the wildtype background. Furthermore, these strains were the only ones to exhibit a high level of expression at timepoint zero of induction, and their expression did not increase over time. This was the case for dual-cluster replacement both in the native promoter scaffold and in the scaffold depleted of Pho2 binding sites. In further agreement, these promoters showed significant expression in the ΔPho4 background (Fig. 7L). Replacement of the NOR alone by a weak-cluster was enough to drive a moderate level of non-specific expression, while this was not the case for cluster replacement in the NFR. Aside from the above discussed strains, all other strains showed little to no non-specific expression. They exhibited consistent, ablated levels of gene expression in the absence of induction, and in the ΔPho4 background.

Discussion

Our current understanding of how transcription factors bind to genomic regulatory regions to drive transcription has developed largely through experiments that characterize binding to individual binding sites, often with a focus on high-affinity consensus sites. The functional relevance of low-affinity binding sites, particularly clusters of low-affinity binding sites is less-well understood⁷. Methodological challenges associated with measuring binding of multiple transcription factors to low-affinity clusters has prevented precise measurements of how binding to these clusters compares with binding to individual consensus binding sites, creating uncertainty about the significance of low-affinity clusters. Indeed, through prior state-of-the-art biochemical methods used to measure collective binding in high-throughput, occupancy appeared driven primarily by strong binding sites¹⁹. Reports in the literature identified in vivo examples where low-affinity clusters are functionally important^{6–11}, or exhibit properties like binding synergy^{15,18}. But, the lack of a systematic analysis in vitro or in vivo limited us to inductive reasoning when trying to apply these findings to new contexts. By combining in vitro biochemical analysis with in vivo functional studies of gene regulation in the eukaryotic model system *S. cerevisiae* we were able to provide insights into how transcription factors interact with low-affinity clusters and were able to show that these clusters are functional in vivo.

We applied a high-throughput in vitro method to characterizing collective transcription factor binding to DNA sequences consisting of arrays of binding sites ranging over orders of magnitude in affinity. This permitted us to develop a quantitative biophysical understanding of how transcription factors interact with binding site clusters,

particularly to clusters consisting of low-affinity binding sites. Clusters of low-affinity binding sites achieved similar levels of occupancy as single high-affinity binding sites and we predicted that this level of occupancy can be achieved at physiologically relevant transcription factor concentrations. We verified this effect with two distinct transcription factors representing two of the largest transcription factor structural classes: zinc finger and bHLH. Small clusters of 2–6 binding sites with individual binding site affinities one order of magnitude lower than the consensus sequence were able to achieve this effect. We furthermore challenged the notion that overlapping binding sites are mutually exclusive due to steric effects, and thus would allow only one transcription factor to bind to overlapping sites at any given time. In fact, clusters of strong binding sites with a 1 bp overlap of a 9 bp consensus sequence were not mutually exclusive and were simultaneously bound by several transcription factors. Even a 3 bp overlap of a 9 bp consensus motif permitted some co-occurring binding, although here the steric factor strongly dominated. Similar trends were observed for weak binding site clusters with 1 and 3 bp overlapping sequences, indicating that strong binding sites permit and support partial binding of a transcription factor, allowing full occupancy of overlapping binding sites.

The high levels of occupancy achieved by low-affinity binding site clusters as determined by our in vitro characterization suggested that low-affinity binding site clusters may indeed be functional in vivo. We therefore first tested low-affinity binding site clusters in an engineered synthetic system in vivo. We placed low-affinity Zif268 binding site clusters in the minimal CYC1 promoter and used a β-estradiol inducible Zif268 transcription factor (Z₃EV) to test whether these promoters could give rise to expression. We found that binding site clusters of 3 to 6 weak binding sites with individual affinities one order of magnitude below the consensus sequence were able to activate transcription to similar levels as a promoter containing 3 consensus sequence binding sites. Therefore, in this synthetic system, small clusters of low-affinity binding sites are functional. We also observed a concurrent increase in promoter leakiness which we characterized and attributed to non-specific interactions of an endogenous transcription factor binding to these binding site clusters. Although clusters of low-affinity binding sites were functional in the synthetic Z₃EV-minCYC1 system, it was possible that this was due to high, super-physiological expression levels of the Z₃EV transcription factor. To control for this uncertainty we tested whether low-affinity binding site clusters could substitute single high-affinity binding sites in a native gene regulatory network. We substituted native binding sites in the nucleosome-free and nucleosome-occluded regions of the well-characterized PHO5 promoter, which is regulated by endogenous levels of the master regulator Pho4. In this context low-affinity binding site clusters were also able to functionally replace single high-affinity binding sites, recovering a large proportion of the native PHO5 gene expression level. This was particularly impressive considering that substitution of the native Pho4 binding sites in these two locations also ablated the neighboring Pho2 binding sites, which are known to further strengthen the effect of Pho4 binding to these single target sites and contribute to transcriptional activation of this promoter.

These findings have several important implications on our current understanding of transcriptional regulation, computational methods for predicting gene regulatory function, gene regulatory network evolution, and engineering of synthetic gene regulatory networks. We have shown that low-affinity binding site clusters are effective at activating transcription in vivo suggesting that for computational approaches to gene regulatory function prediction and transcription factor binding prediction based on ChIP-seq data, that low-affinity binding sites warrant closer consideration, and that clusters of low-affinity binding sites can be highly functional.

On a more fundamental level, it was not known whether transcription factor dwell time or occupancy is the critical parameter that

determines whether a binding region can give rise to functional gene expression. A single high-affinity binding site and clusters of low-affinity binding sites can both give rise to high occupancies at physiologically relevant transcription factor concentrations. But how this occupancy is achieved is fundamentally different between these two cases, with a single high-affinity binding site achieving high occupancy due to long transcription factor dwell times (low off-rates), whereas low-affinity binding site clusters achieve high occupancy as an ensemble average of many interactions of short dwell times (high off-rates). Our findings show that both systems can give rise to transcriptional activation and suggests that long transcription factor dwell times are not required for transcriptional activation. This may not apply to transcriptional repression where it is possible that long dwell times are critical for efficient repression. It will be interesting to assess whether low-affinity binding site clusters can function as effective repressive regulatory regions or not. Finally, whether there is an observable difference in regards to gene expression noise between low-affinity binding site clusters and single high affinity binding sites would also warrant further exploration⁶³.

From an evolutionary perspective, low-affinity binding site clusters are of considerable interest as they could serve as a functional intermediate between a non-regulated promoter and a highly evolved promoter containing one or more consensus sequence binding sites. It is interesting that clusters of low-affinity binding sites are functional, but that they also seem to come with an additional cost which is non-specific regulation or cross-talk. Small clusters of low-affinity binding sites thus might be easier to form from a non-specific sequence background, but are then under continued pressure to further evolve to more specific high-affinity binding sites. It is striking that as few as 3 low-affinity binding sites are sufficient to generate a highly functional regulatory region. Likewise, when approaching this notion from a transcription factor centric perspective, there exist many more possible low-affinity binding sites for any given transcription factor than specific sites, which also enables regulatory regions to evolve more readily but also is the reason for non-specific interactions and cross-talk. Transcriptional regulation has traditionally been considered as being encoded in strong, well-defined binding sites. Results from de Boer et al. supported a more finer-grained view of transcription by demonstrating that gene expression in promoters of *S. cerevisiae* is largely driven through the collective action of many weak regulatory contributions^{13,64}. Holding total occupancy constant, many weaker sites may allow for a greater number of binding sites and different transcription factors⁶⁴, increasing the potential of network connectedness, or the average degree of nodes found in a eukaryotic gene regulatory network.

Synthetic promoters to date have predominantly made use of consensus binding sites, high-affinity binding sites, clusters thereof⁶⁵, and most recently clusters of lower affinity binding sites in combination with transcription factor cooperativity⁶⁶. As engineering synthetic promoters and entire gene regulatory networks matures it is likely that increasingly precise gene expression levels will become necessary. This to some extent may be achievable with relatively few single binding sites of various affinity, but it may be worthwhile considering the use of low-affinity binding site clusters. Low-affinity binding site clusters may be more resilient to mutations for example. First, a single deleterious mutation in a cluster of several low-affinity binding sites will have a much smaller impact on the resulting gene expression level as a mutation in a single high affinity binding site. Second, a single base change in a single high-affinity binding site will result in a large decrease in binding affinity, whereas a single base change in a low-affinity site will lead to a much smaller relative decrease in affinity, or may even be silent in regards to affinity.

In summary we conducted a systematic characterization of low-affinity binding site clusters by conducting a quantitative biochemical analysis *in vitro* as well as functional studies *in vivo*. This work provided

insights into how transcription factors bind to such clusters and established that a small number of low-affinity binding sites in a local cluster can be highly functional *in vivo*. These insights improve our current understanding of gene regulatory networks, and our ability to engineer sophisticated gene regulatory networks *de novo*.

Methods

DNA target PCRs

Initial materials for producing iMITOMI DNA target libraries included tagged primers (5' Cy5-tagged: CTG/iCy5/TCGGCCGCTAACA, and 3' biotin-tagged: /5Bios/GTCATACCGCCGGA) ordered from IDT. Furthermore, DNA targets of interest were ordered as single-stranded 90-basepair long primers from IDT, which had 5' and 3' ends complementary to the tagged primers. For each library member, two 50 μ L PCR reactions were prepared according to the following recipe:

Two 50 μ L PCR products were mixed with 500 μ L of DNA binding buffer, combined into one purification column (Zymo DNA Clean & Concentrator 50), and PCR purified according to kit instructions. Samples were eluted into 80 μ L of Invitrogen UltraPure Distilled Water. DNA targets were run on an agarose gel for characterization, and only products with a single band of expected size were advanced to spotting.

Spotting plate preparation

75 μ L of each DNA target was introduced into independent wells of a 384 well plate. Plates were measured for fluorescence with an excitation wavelength of 570 ± 9 nm and an emission wavelength of 593 ± 9 nm, using a Biotek Synergy Mx Multi-mode reader. Based on these measurements, new fluorescence-equalized wells were prepared by diluting an appropriate amount of DNA with water to reach 72 μ L (no water was added for the well with the lowest fluorescence intensity, the others were equalized to this fluorescence intensity $\pm 5\%$, Supplementary Fig. SIC. 24 μ L of 2% Bovine Serum Albumin (in H₂O) was added to each well. The plate was sealed with aluminum and kept at -20 °C until time of use.

His-tag protein purification

For each transcription factor, bacteria containing the corresponding plasmid were overnight cultured with ampicillin (Amp), and inoculated into two 1 L baffled flasks each containing 500 mL LB and Amp (100 μ g/ml). Cultures were grown at 37 °C at 260 rpm for 2 h, induced with IPTG (1 mM) and grown for an additional 3 h. Cells were centrifuged at 3220 g at 4 °C for 10 min in 50 mL falcon flasks. Cell pellets were resuspended in buffer A (sequential transfers, using total volume of 7.5 mL), and sonicated (Vibra cell 75186 sonicator, probe tip diameter: 6 mm) with 20 s ON 20 s OFF for four cycles using 70% amplitude. Lysate was distributed to 2 mL tubes and centrifuged for 20 min at 4 °C at 20,000 g. Supernatant was removed and dispensed into a regenerated Ni Sulfate column. The column was washed with 25 mL buffer A, then protein was eluted in 5 mL buffer B. The sample was dialysed overnight in 1 L of HT buffer at 4 °C with magnetic stirring, followed by 3 h of dialysis with HT stock buffer under similar conditions. The sample was run on an SDS-PAGE gel shown in Supplementary Fig. 2 for characterization.

Control and flow wafer fabrication

Mold-fabrication using silicon wafers was performed through a photolithography process in the CMi cleanrooms at EPFL, as described previously^{22,67}. In brief, for the control layer, a silicon wafer was primed in a Tepla 300, and SU-8 photoresist (GM 1070) was spin coated onto the wafer using a Sawatec spin coater to reach a height of 30 μ m. After a soft bake the wafer was exposed (365 nm illumination, 20 mW/cm² light intensity) using a chrome mask for 10 s on a MABA6 mask aligner. The wafer was post exposure baked and then developed using propylene glycol methyl ether acetate (PGMEA), followed by a hard bake.

For the flow layer the silicon wafer was treated with hexamethyldisilazane (HMDS) vapor using a YesIII oven. AZ 9260 photoresist was spin coated onto the wafer using an EVG150 modular cluster tool to reach a height of around 14 μm . After the wafer was baked it was left for a 1 h relaxation period, followed by UV exposure using an MABA6 mask aligner. The total dose was 660 mJ/cm split into two exposures of 18s with a 10s wait period in between (20 mW/cm² light intensity). The wafer was developed again using the EVG150 (AZ 400K developer), and then baked at 160 °C for 2 h.

Designs of microfluidic devices are available on our group's website: lbcn.epfl.ch.

PDMS chip fabrication

Wafers were first treated with chlorotrimethylsilane (TMCS) vapor to facilitate removal of cured polydimethylsiloxane (PDMS) from the wafers. PDMS curing agent and elastomer (amounts of 1 g and 20 g for the flow layer, 10 g and 50 g for the control layer, respectively) were mixed (2000 rpm for 1 min) and defoamed (2200 rpm for 2 min) using a Thinky ARE-250 centrifugal mixer. PDMS from the control layer was poured on the control wafer and left to degas in a vacuum chamber, followed by baking for 20 min. For the flow layer, 4 mL of PDMS was spin-coated onto the flow-wafer using a SCS G3P-8 spin coater (1800 rpm for 35 s), followed by baking for 20 min. Control inlets were punched using a Schmidt Press manual hole puncher and 21-gauge (OD 0.04") pins (Technical Innovations, Inc.), and the control PDMS blocks were aligned to the flow layer under a stereoscope using coaxial illumination. Aligned chip layers were bonded at 80 °C for 90 min. Next, PDMS chips were cut off the flow layer, and flow inlets and outlets were punched using a 900 mm pin.

Epoxy slide coating and microarraying

VWR slides with cut edges and plain ends (631–1550) were cleaned in a solution of 600 mL milliQ water, 120 mL 25% Ammonia solution, and 150 mL H₂O₂ (on a hotplate at 80 °C). Slides were then washed with milliQ water and dried. Next, slides were left to incubate for 20 min in a bath of 891 mL toluene and 5 mL of (3-Glycidyloxypropyl)trimethoxysilane (GPS, Sigmaaldrich Cat. 440167). Slides were rinsed with toluene and dried, followed by baking at 120 °C for 120 min.

Epoxy-coated glass slides were spotted with the purified 90 bp DNA targets using a Genetix Qarray² robot with an MP2.5 pin (Arrayit). Spotting chambers were aligned to DNA spots on the glass slide under a microscope, and the device was baked for 8 h at 80 °C.

IMITOMI experiments

Control lines were filled with tap water and pressurized at 172 kPa. The neck valve was pressurized to prevent premature target resolubilization. Flow lines were pressurized at 48 kPa. Surface chemistry was conducted in the remainder of the chip²², by sequentially patterning the surface using 15 min flow steps with biotin-BSA, then neutravidin, followed by pressurization of the button valve and a second biotin-BSA flow. PBS washes were conducted between each flow step for 5 min. This resulted in available neutravidin binding sites under the button valve in the detection chamber, whereas the remainder of neutravidin binding sites in the chip's flow channels were passivated by biotin-BSA. DNA targets in spotting chambers were resolubilized by pressurizing the PBS flow line, closing the outlet and opening the neck valve to force air out of the PDMS and PBS into the spotting chambers. After a PBS wash with the neck valve closed to prevent any contamination, the sandwich valves were pressurized to isolate chambers from one another, and the neck and button valves were opened to allow the biotinylated Cy5-tagged DNA targets to diffuse to and bind under the button valves in the detection chambers for 90 min. The button valve was then closed, and the chip was washed with PBS. The chip surface was blocked using Promega

wheat-germ extract to reduce non-specific binding of concentrated fluorescent protein solutions. The chip was multiplexed into 8 sections using the multiplexing valves. Different concentrations of purified mScarlet-tagged transcription factor were flowed into different sections of the chip, followed by incubation for 60 min with the button valves open to allow binding to reach equilibrium. The chip was scanned to quantify the free protein concentration at equilibrium (compared against a background scan after the subsequent wash step). The button valve was closed to isolate the bound protein at equilibrium, and the chip was washed, and scanned again to quantify the bound protein and amount of DNA present (compared against a background scan before fluorescent protein or fluorescent DNA was introduced into the detection chamber, respectively).

Fluorescence microscopy

Fluorescent scanning was conducted using a Nikon Ti Eclipse and the NIS-Elements software. Chips were placed on the microscope stage and a grid of all chambers on the chip was built interactively using NIS-elements, by specifying the top left and bottom right chambers. The chip was taped down and ensured to be in a flat plane, otherwise a focus surface can also be registered in the software. The lasers were never left on, to avoid photobleaching, and instead the chip was imaged at 20X magnification, generally with 500 ms exposure, in the Cy5 channel (DNA) or mCherry channel (transcription factor). Transcription factor was flowed into the chip in increasing concentrations to minimize non-specific background signal.

Media and growth conditions

Cells were incubated for yeast transformation at 30 °C shaking at 250 rpm in yeast extract peptone dextrose (YPD) medium (Sigma-Aldrich) (10 g/L yeast extract, 20 g/L peptone and 20 g/L glucose). Plates contains 20 g/L of agar (10752-36, Alfa).

Promoter library yeast strains were cultured in synthetic complete (SC) medium lacking uracil (1.92 g of yeast synthetic dropout medium supplement without uracil, Y1501-20G, Sigma-Aldrich), 6.7 g Yeast Nitrogen Base (YNB) without amino acids (Y0626-1KG, Sigma-Aldrich), were dissolved in 960 mL of deionized water and autoclaved. After sterilization we added 40 mL 50% filter sterilized glucose solution to a final volume of 1 L. For plate reader experiments of PHO5 promoter library strains we use a different recipe of SC, SC phosphate-free (PF) and SC phosphate-rich (PR) medium, a modified recipe from that described in ref. 57. In this case we add per 1 L, 20 g of glucose, 5.6 g YNB with ammonium sulfate, without phosphates, without sodium chloride (MP 4027-812), 0.79 g Complete Supplement Mixture (CSM) (MPB-114500022) and 0.1 g sodium chloride. The amount of adenine and tryptophan was supplemented over the CSM to final concentrations of 0.13 g/L adenine and 0.1 g/L tryptophan to suppress auto-fluorescence. SC PF medium contains 0.55 g/L of potassium chloride instead of 1 g/L of monobasic potassium phosphate that contains SC PR medium.

Promoter library and yeast strain construction

Promoter variants were ordered as gene fragments from TWIST Bioscience. The PHO5 promoter library integration cassette consisted of two LYS2 homology arms (each 500 bp), PHO5 promoters variants, yeast codon optimized enhanced green fluorescent protein (yoEGFP), TDH1 terminator sequence (177 bp), and URA3 auxotrophic marker. The minimal CYC1 promoter library integration cassette consisted of two LYS2 homology arms (each 500 bp), CYC1 minimal promoter variants, yoEGFP ENO2 terminator and URA3 auxotrophic marker. All integration cassettes were constructed hierarchically by using modular part plasmids following the YeastToolkit (YTK) workflow⁵⁵.

Briefly, multiple YTK part plasmids were assembled into cassette plasmids via Bsal goldengate assembly and then multiple cassette

plasmids were assembled to the multi-cassette plasmids via BsmBI goldengate assembly. Goldengate reactions are prepared as following: add the volume necessary to have 100 ng of each plasmid, except for the backbone that was added 25 ng, 2 μ L of 10x T4 ligase buffer (NEB), 1 μ L of T4 ligase (NEB), 1 μ L of BsaI or BsmBI (NEB), and adding water to final volume of 20 μ L. Thermocycler setup: (BsmBI assembly: 45 °C for 2 min/BsaI assembly: 42 °C for 2 min, 16 °C for 5 min) \times 25 cycles, followed by a final digestion step at 60 °C for 10 min and a heat inactivation at 80 °C for 20 min. To assemble PHO5 promoter library cassette plasmids (from pSC189 to pSC201), assembly connector ConLS part plasmid (pYTK002), PHO5 promoters variant part plasmids (from pSC170 to pSC182), yoEGFP part plasmid (pMC068), TDH1 terminator part plasmid (pYTK056), assembly connector ConRE part plasmid (pYTK072) were assembled into the backbone plasmid (pYTK095) via a BsaI goldengate assembly. PHO5 promoter library integration cassette plasmids (from pML001 to pML013) were constructed by BsmBI assembly of cassettes plasmids (from pSC189 to pSC201) into the backbone plasmid that contains two LYS2 homology arms and URA3 auxotrophic marker (pSC109).

To assemble the minimal CYC1 promoter library cassette plasmids (from pAS009 to pAS016 and from pML022 to pML033), assembly connector ConLS part plasmid (pYTK002), CYC1 minimal promoters variant part plasmids (from pAS001 to pAS008 and from pAS058 to pAS069), yoEGFP part plasmid (pMC068), ENO2 terminator part plasmid (pYTK055), assembly connector ConRE part plasmid (pYTK072) were assembled into the backbone plasmid (pYTK095) via a BsaI goldengate assembly. CYC1 minimal promoter library integration cassette plasmids (from pAS017 to pAS024 and from pML034 to pML045) were constructed by BsmBI assembly of cassettes plasmids (from pAS009 to pAS016 and from pML022 to pML033), into the backbone plasmid that contains two LYS2 homology arms and URA3 auxotrophic marker (pSC109).

For all plasmid cloning we used NEB 10-beta Competent *E. coli* (High Efficiency) (New England Biolabs, Cat C3019H). We followed the transformation protocol from the manufacturer. We conducted bacterial selection and growth in Lysogeny Broth (LB) plates or LB medium at 37 °C with supplement of appropriate antibiotics (chloramphenicol 34 μ g/mL, ampicillin 100 μ g/mL or kanamycin 50 μ g/mL). All plasmids were sequence-verified by sanger sequencing (Microsynth Ecoli NightSeq).

The PHO5 promoter library yeast strains were made by transforming *S. cerevisiae* strains BY4741 and Δ Pho4 (Yeast Knockout Library, Horizon Discovery Ltd.), with the integration cassettes and we obtained 26 variant strains (from sML034 to sML059). The min-CYC1 promoter library integration cassettes were used to transform the parental strains BY4741 and sSCO51 to obtain 40 variant strains (from sAS001 to sAS008 and from sML073 to sML104). In brief, sSCO51 has integrated into HO locus the yeast codon optimized mScarlet-i (yomScarlet-i) under the control of a modified GAL1 promoter with six Zif268 binding sites as described⁶⁸. Upstream, in the same HO locus, this strain has integrated the artificial transcription factor Z3EV (a fusion of the mouse transcriptional factor Zif268 DNA binding domain, the ligand binding domain of the human estrogen receptor and viral protein 16) that is selectively relocated from the cytoplasm to the nucleus in the presence of β -estradiol. All promoter library yeast strains were made by homologous recombination of the integration cassettes, into LYS2 locus, using the lithium acetate/polyethylene glycol (PEG) method⁶⁹. For yeast integration all integration cassette plasmids were digested with NotI and after that the DNA was purified (ZYMO DNA clean and concentrator-25 kit). The transformants were selected in SC medium lacking uracil plates at 30 °C. We screened transformants for correct integration by colony PCR followed by sanger sequencing of the colony PCR product (Microsynth).

Yeast plater reader characterization

Promoter library strains were grown overnight in 3 ml in SC medium lacking uracil at 30 °C shaking at 250 rpm. Cultures were diluted to an optical density (O.D) of 0.175 in fresh SC medium lacking uracil. Cells grown until log phase, OD ~ 0.8, then PHO5 promoter library strains were washed twice in SC phosphate-free (PF) medium and diluted to starting OD of 0.1–0.2 in SC PF medium and SC phosphate-rich (PR) medium. Minimal CYC1 promoter strains were diluted to starting OD of 0.1–0.2 in SC medium lacking uracil and SC medium lacking uracil with β -estradiol at 200 nM. Cells were cultured in 96-well plates with clear flat bottom (Nunc) and covered with a gas-permeable membrane (Breathe-Easy membrane, Sigma-Aldrich). OD, GFP and mScarlet was measured every 10 min for 20–24 h on a plate reader (BioTek SynergyMx). 96-well plates were incubated at 30 °C with continuous agitation while the plate reading is not measuring. Background of the media was subtracted, then data was normalized dividing by the OD at each time point.

The same growth and inducing conditions were used to induce promoter library for characterization by fluorescence microscopy. The only difference being that 96-well plates with clear conical bottoms (Nunc) were used for growing the yeast for microscope analysis. Strains were imaged at log phase, at an OD of around 0.8, using a Nikon Eclipse Ti fluorescent microscope. Sample visualization was conducted under a 60x objective through the NIS-Elements software (Nikon Instruments). Cells were imaged in bright-field and fluorescence modes. Images were processed and data analyzed using Image J.

Yeast flow cytometry characterization

CYC1 minimal promoter strains were grown overnight in SC medium lacking uracil at 30 °C shaking at 250 rpm. Cultures were diluted at O.D. of 0.175 in fresh SC medium lacking uracil. Cells were grown until log phase, OD 0.8 and then were diluted and induced to starting OD of 0.1–0.2 in SC medium lacking uracil with β -estradiol (Sigma-Aldrich) at 200 nM. Cells were induced for 12 h. Cells were diluted to OD 0.5 in phosphate buffered saline (PBS) to assay fluorescent expression by flow cytometry. GFP and mScarlet measurements were made with a BD LSRFortessa. For all data we acquired 10,000 events. Events were gated by forward (FSC-A) and side scatter (SSC-A) to select the population of interest. Then doublets were excluded gating by FSC-W (Pulse Width (W)) and FSC-H (Pulse Height (H)). The median of the fluorescent distribution were calculated. Positive and negative GFP and mScarlet controls were used to ensure that the instrument maintained proper calibration. The level of gene expression from each promoter was measured from three independent cultures.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support this study are available from the corresponding author upon reasonable request. All data is available in the following GitHub repository (https://github.com/eukaryoting/systematic_analysis_of_low-affinity_clusters). Source data are provided with this paper.

Code availability

All code for data processing, analysis, and visualization is available in the following GitHub repository (https://github.com/eukaryoting/systematic_analysis_of_low-affinity_clusters).

References

1. Gotea, V. et al. Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Res.* **20**, 565–577 (2010).

2. Ezer, D., Zabet, N. R. & Adryan, B. Homotypic clusters of transcription factor binding sites: A model system for understanding the physical mechanics of gene expression. *Computational Struct. Biotechnol. J.* **10**, 63–69 (2014).
3. Lifanov, A. P., Makeev, V. J., Nazina, A. G. & Papatsenko, D. A. Homotypic regulatory clusters in drosophila. *Genome Res.* **13**, 579–588 (2003).
4. Wunderlich, Z. & Mirny, L. A. Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet.* **25**, 434–440 (2009).
5. Estrada, J., Wong, F., DePace, A. & Gunawardena, J. Information integration and energy expenditure in gene regulation. *Cell* **166**, 234–244 (2016).
6. Crocker, J. et al. Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* **160**, 191–203 (2015).
7. Kribelbauer, J. F., Rastogi, C., Bussemaker, H. J. & Mann, R. S. Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes. *Annu. Rev. Cell Developmental Biol.* **35**, 357–379 (2019).
8. Crocker, J., Noon, E. P.-B., & Stern, D. L. The soft touch: low-affinity transcription factor binding sites in development and evolution. In *Current Topics in Developmental Biology*, volume 117, pages 455–469. Elsevier, 2016.
9. Gaudet, J. & Mango, S. E. Regulation of organogenesis by the caenorhabditis elegans foxa protein pha-4. *Science* **295**, 821–825 (2002).
10. Farley, E. K. et al. Suboptimization of developmental enhancers. *Science* **350**, 325–328 (2015).
11. Jiang, J. & Levine, M. Binding affinities and cooperative interactions with bhlh activators delimit threshold responses to the dorsal gradient morphogen. *Cell* **72**, 741–752 (1993).
12. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
13. de Boer, C. G. et al. Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat. Biotechnol.* **38**, 56–65 (2020).
14. Rajkumar, A. S., Dénervaud, N. & Maerkl, S. J. Mapping the fine structure of a eukaryotic promoter input-output function. *Nat. Genet.* **45**, 1207–1215 (2013).
15. Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U. & Gaul, U. Predicting expression patterns from regulatory sequence in drosophila segmentation. *Nature* **451**, 535–540 (2008).
16. Mogno, I., Kwasnieski, J. C., & Cohen, B. A. Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. *Genome Res.* **23**, 1908–1915, (2013).
17. Weingarten-Gabbay, S. & Segal, E. The grammar of transcriptional regulation. *Hum. Genet.* **133**, 701–711 (2014).
18. Castellanos, M., Mothi, N. & Muñoz, V. Eukaryotic transcription factors can track and control their target genes using DNA antennas. *Nat. Commun.* **11**, 1–13 (2020).
19. Levo, M. et al. Unraveling determinants of transcription factor binding outside the core binding site. *Genome Res.* **25**, 1018–1029, (2015).
20. Zhu, F. et al. The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76–81 (2018).
21. Kim, S. et al. Probing allostery through DNA. *Science* **339**, 816–819 (2013).
22. Maerkl, S. J. & Quake, S. R. A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **315**, 233–237 (2007).
23. Maerkl, S. J. & Quake, S. R. Experimental determination of the evolvability of a transcription factor. *Proc. Natl Acad. Sci. USA* **106**, 18650–18655 (2009).
24. Rockel, S., Geertz, M., Hens, K., Deplancke, B. & Maerkl, S. J. iSLIM: a comprehensive approach to mapping and characterizing gene regulatory networks. *Nucleic Acids Res.* **41**, e52 (2013).
25. Blackburn, M. C., Petrova, E., Correia, B. E. & Maerkl, S. J. Integrating gene synthesis and microfluidic protein analysis for rapid protein engineering. *Nucleic Acids Res.* **44**, e68–e68 (2015).
26. Le, D. D. et al. Comprehensive, high-resolution binding energy landscapes reveal context dependencies of transcription factor binding. *Proc. Natl Acad. Sci.* **115**, E3702–E3711 (2018).
27. Fordyce, P. M. et al. De novo identification and biophysical characterization of transcription-factor binding sites with microfluidic affinity analysis. *Nat. Biotechnol.* **28**, 970–975 (2010).
28. Aditham, A. K., Markin, C. J., Mokhtari, D. A., DelRosso, N. & Fordyce, P. M. High-throughput affinity measurements of transcription factor and DNA mutations reveal affinity and specificity determinants. *Cell Syst.* **12**, 112–127 (2021).
29. Martin, L. L. et al. Systematic reconstruction of RNA functional motifs with high-throughput microfluidics. *Nat. Methods* **9**, 1192–1194 (2012).
30. Gerber, D., Maerkl, S. J. & Quake, S. R. An in vitro microfluidic approach to generating protein-interaction networks. *Nat. Methods* **6**, 71–74 (2009).
31. Einav, S. et al. Discovery of a hepatitis C target and its pharmacological inhibitors by microfluidic affinity analysis. *Nat. Biotechnol.* **26**, 1019–1027 (2008).
32. Geertz, M., Shore, D. & Maerkl, S. J. Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proc. Natl Acad. Sci.* **109**, 16540–16545 (2012).
33. Bates, S. R. & Quake, S. R. Highly parallel measurements of interaction kinetic constants with a microfabricated optomechanical device. *Appl. Phys. Lett.* **95**, 73705–73705 (2009).
34. Popp, A. P., Hettich, J. & Gebhardt, J. C. M. Altering transcription factor binding reveals comprehensive transcriptional kinetics of a basic gene. *Nucleic Acids Res.* **49**, 6249–6266 (2021).
35. Clauß, K. et al. DNA residence time is a regulatory factor of transcription repression. *Nucleic Acids Res.* **45**, 11121–11130 (2017).
36. Callegari, A. et al. Single-molecule dynamics and genome-wide transcriptomics reveal that nf-kb (p65)-DNA binding times can be decoupled from transcriptional activation. *PLoS Genet.* **15**, e1007891 (2019).
37. Lickwar, C. R., Mueller, F., Hanlon, S. E., McNally, J. G. & Lieb, J. D. Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484**, 251–255 (2012).
38. Loffreda, A. et al. Live-cell p53 single-molecule binding is modulated by c-terminal acetylation and correlates with transcriptional activity. *Nat. Commun.* **8**, 1–12 (2017).
39. Biggin, M. D. Animal Transcription Networks as Highly Connected, Quantitative Continua. *Developmental Cell* **21**, 611–626 (2011).
40. Alberts, B. et al. Molecular biology of the cell. In *Molecular Biology of the Cell*. 2002.
41. Hermsen, R., Tans, S., & Wolde, P. R. T. Transcriptional regulation by competing transcription factor modules. *PLoS Comput. Biol.* **2**, e164 (2006).
42. Panne, D. The enhanceosome. *Curr. Opin. Struct. Biol.* **18**, 236–242 (2008).
43. Ezer, D., Zabet, N. R. & Adryan, B. Physical constraints determine the logic of bacterial promoter architectures. *Nucleic Acids Res.* **42**, 4196–4207 (2014).
44. Hoch, M., Gerwin, N., Taubert, H. & Jackle, H. Competition for overlapping sites in the regulatory region of the drosophila gene kruppel. *Science* **256**, 94–97 (1992).
45. Guturu, H., Doxey, A. C., Wenger, A. M. & Bejerano, G. Structure-aided prediction of mammalian transcription factor complexes in

- conserved non-coding elements. *Philos. Trans. R. Soc. B: Biol. Sci.* **368**, 20130029 (2013).
46. Fry, C. J. & Farnham, P. J. Context-dependent transcriptional regulation. *J. Biol. Chem.* **274**, 29583–29586 (1999).
47. Stanojevic, D., Small, S. & Levine, M. Regulation of a segmentation stripe by overlapping activators and repressors in the drosophila embryo. *Science* **254**, 1385–1387 (1991).
48. Ptashne, M. Principles of a switch. *Nat. Chem. Biol.* **7**, 484–487 (2011).
49. Dhiman, A. & Schleif, R. Recognition of overlapping nucleotides by *araC* and the σ subunit of RNA polymerase. *J. Bacteriol.* **182**, 5076–5081 (2000).
50. Humphrey, G. W., Englander, E. W. & Howard, B. H. Specific binding sites for a pol iii transcriptional repressor and pol ii transcription factor *yy1* within the internucleosomal spacer region in primate *alu* repetitive elements. *Gene Expr. J. Liver Res.* **6**, 151–168 (1996).
51. Lee, N. L., Gielow, W. O. & Wallace, R. G. Mechanism of *araC* auto-regulation and the domains of two overlapping promoters, *pc* and *pbad*, in the l-arabinose regulatory region of *escherichia coli*. *Proc. Natl Acad. Sci.* **78**, 752–756 (1981).
52. Lee, J. Y. et al. Multiple *sgrnas* with overlapping sequences enhance *crispr/cas9*-mediated knock-in efficiency. *Exp. amp; Mol. Med.* **50**, 1–9 (2018).
53. Voss, T. C. et al. Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell* **146**, 544–554 (2011).
54. Casini, A. et al. R2odna designer: computational design of biologically neutral synthetic dna sequences. *ACS Synth. Biol.* **3**, 525–528 (2014).
55. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).
56. Mclsaac, R. S. et al. Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res.* **41**, e57–e57 (2013).
57. Lam, F. H., Steger, D. J. & O’Shea, E. K. Chromatin decouples promoter threshold from dynamic range. *Nature* **453**, 246–250 (2008).
58. Barbaric, S., Munsterotter, M., Goding, C. & Horz, W. Cooperative *pho2-pho4* interactions at the *pho5* promoter are critical for binding of *pho4* to *uaspl1* and for efficient transactivation by *pho4* at *uaspl2*. *Mol. Cell. Biol.* **18**, 2629–2639 (1998).
59. Barbaric, S., Münsterkötter, M., Svaren, J. & Hörz, W. The homeodomain protein *pho2* and the basic-helix-loop-helix protein *pho4* bind DNA cooperatively at the yeast *pho5* promoter. *Nucleic Acids Res.* **24**, 4479–4486 (1996).
60. Fascher, K.-D., Schmitz, J. & Hörz, W. Role of trans-activating proteins in the generation of active chromatin at the *pho5* promoter in *s. cerevisiae*. *EMBO J.* **9**, 2523–2528 (1990).
61. Nourani, A., Utley, R. T., Allard, S. & Côté, J. Recruitment of the *nua4* complex poises the *pho5* promoter for chromatin remodeling and activation. *EMBO J.* **23**, 2597–2607 (2004).
62. Korber, P. & Barbaric, S. The yeast *pho5* promoter: from single locus to systems biology of a paradigm for gene regulation through chromatin. *Nucleic Acids Res.* **42**, 10888–10902 (2014).
63. Azpeitia, E. & Wagner, A. Short residence times of DNA-bound transcription factors can reduce gene expression noise and increase the transmission of information in a gene regulation system. *Front. Mol. Biosci.* **7**, 67 (2020).
64. Tanay, A. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Res.* **16**, 962–972 (2006).
65. Khalil, A. S. et al. A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647–658 (2012).
66. Bashor, C. J. et al. Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science* **364**, 593–597 (2019).
67. Laohakunakorn, N., Lavickova, B., Swank, Z., Laurent, J. & Maerkl, S. J. Steady-state cell-free gene expression with microfluidic chemostats. *Methods Mol. Biol.* **2229**, 189–203 (2021).
68. Mclsaac, R. S. et al. Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res.* **41**, e57–e57 (2012).
69. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the *liac/ss* carrier dna/peg method. *Nat. Protoc.* **2**, 31–34 (2007).

Acknowledgements

S.J.M. and I.I. thank Prof. David Shore (University of Geneva) for many insightful discussions. This work was supported by the following grants to S.J.M.: a Swiss National Science Foundation SystemsX.ch IPhD grant (SNF:51PHPO 157292 / SysX:2014/242), a Swiss National Science Foundation Sinergia grant (189910), and by the European Research Council under the European Union’s Horizon 2020 research and innovation program grant 723106.

Author contributions

A.S. performed all in vitro experiments, data analysis and modeling used in this manuscript. M.L. performed all in vivo experiments, M.L. and A.S. performed data analysis. I.I. and S.J.M. designed the initial iMITOMI method which was further refined by A.S. I.I. and S.J.M. designed the *Pho4* target library. A.S. and S.J.M. designed the *Zif268* target library. I.I. designed the recombinant proteins and cloned the expression plasmids. I.I. and A.S. generated the recombinant proteins. I.I. collected and analyzed preliminary *Pho4* data. E.O. supported I.I. and A.S. with development of the iMITOMI method, modeling, and data analysis. S.C., M.L., and A.S. generated the yeast strains. A.S. and S.J.M. wrote the manuscript with input from all authors. S.J.M. conceived and directed the study.

Competing interests

The following authors have financial interest in Adaptiv Biosystems and declare no non-financial competing interest: A.S., S.J.M. The following authors declare no competing financial or non-financial interests: M.L.M, I.I., E.J.O., and S.C.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-32971-0>.

Correspondence and requests for materials should be addressed to Sebastian J. Maerkl.

Peer review information *Nature Communications* thanks the other anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022