Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast

Tali Raveh-Sadka[1,2,4], Michal Levo[1,2,4], Uri Shabi[3], Boaz Shany[1,2], Leeat Keren[1,2], Maya Lotan-Pompan[1,2], Danny Zeevi[1,2], Eilon Sharon[1,2], Adina Weinberger[1,2] & Eran Segal[1,2]

Understanding how precise control of gene expression is specified within regulatory DNA sequences is a key challenge with far-reaching implications. Many studies have focused on the regulatory role of transcription factor–binding sites. Here, we explore the transcriptional effects of different elements, nucleosome-disfavoring sequences and, specifically, poly(dA:dT) tracts that are highly prevalent in eukaryotic promoters. By measuring promoter activity for a large-scale promoter library, designed with systematic manipulations to the properties and spatial arrangement of poly(dA:dT) tracts, we show that these tracts significantly and causally affect transcription. We show that manipulating these elements offers a general genetic mechanism, applicable to promoters regulated by different transcription factors, for tuning expression in a predictable manner, with resolution that can be even finer than that attained by altering transcription factor sites. Overall, our results advance the understanding of the regulatory code and suggest a potential mechanism by which promoters yielding prespecified expression patterns can be designed.

Precise regulation of expression is vital to most biological functions. A key challenge is to unravel how this regulation is encoded within genomes and to be able to understand, predict and design expression patterns from regulatory sequences. Addressing this challenge requires knowledge of both the functional elements within regulatory sequences and the ways in which such elements combine to specify regulatory programs. Much of the research in the field has focused on the role of transcription factors and their binding sites[1]. Despite much progress, our ability to understand transcriptional regulation using only transcription factors remains limited. Recent studies suggested that nucleosome organization of regulatory regions is essential for bridging this gap. Because transcription factor–binding sites that are wrapped into nucleosomes are less accessible to binding by their cognate factor compared to sites located in nucleosome-free regions[2,3], we expect that the transcriptional output directed by the same site will depend on its surrounding nucleosome organization[4–7].

A strong determinant of nucleosome organization is the presence of homopolymeric stretches of deoxyadenosine nucleotides, referred to as poly(dA:dT) tracts. These sequence elements disfavor nucleosome formation[8,9] and are strongly associated with nucleosome depletion over the tract itself and its surrounding DNA in both *in vivo*[10,11] and *in vitro*[10,12] genome-wide studies. Notably, these tracts are highly abundant in eukaryotic genomes[13] and are particularly prevalent in promoters[14]. Consistent with these observations and with a role for nucleosome organization in determining transcriptional output, a study that altered the presence and length of a native poly(dA:dT) element in one yeast promoter showed that this tract can indeed stimulate expression, most likely by conferring increased accessibility to the nearby transcription factor–binding site[15]. Thus, sequences that strongly disfavor nucleosome formation may also serve as important promoter building blocks.

Despite this potentially important regulatory role of poly(dA:dT) tracts, very little is known about the extent and nature of their transcriptional effect. What is the magnitude of their effect compared to that of other regulatory elements? How does their effect depend on their own sequence properties and on properties of other regulatory elements such as the affinity of nearby transcription factor sites? And how does the overall arrangement of these tracts and other regulatory elements determine the transcriptional output?

Here, we systematically address the above questions by measuring the activities of 70 different promoter variants that we designed with poly(dA:dT) tracts that differ in their length, composition and distance from several distinct transcription factor sites. We further characterize the transcriptional effect of poly(dA:dT) tracts by nucleosome occupancy and single-cell expression measurements. Notably, we show that, by manipulating only poly(dA:dT) tracts, we can affect nucleosome organization and predictably alter the resulting transcriptional level to a significant extent, comparable to that attained by altering transcription factor sites. In fact, compared to binding site alterations, poly(dA:dT) manipulations can yield more gradual changes and may thus offer a genetic mechanism by which expression can be tuned with finer resolution.

[1]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. [2]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel. [3]Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel. [4]These authors contributed equally to this work. Correspondence should be addressed to E.S. (eran.segal@weizmann.ac.il) or A.W. (adina.weinberger@weizmann.ac.il).

Received 26 January; accepted 3 May; published online 27 May 2012; doi:10.1038/ng.2305

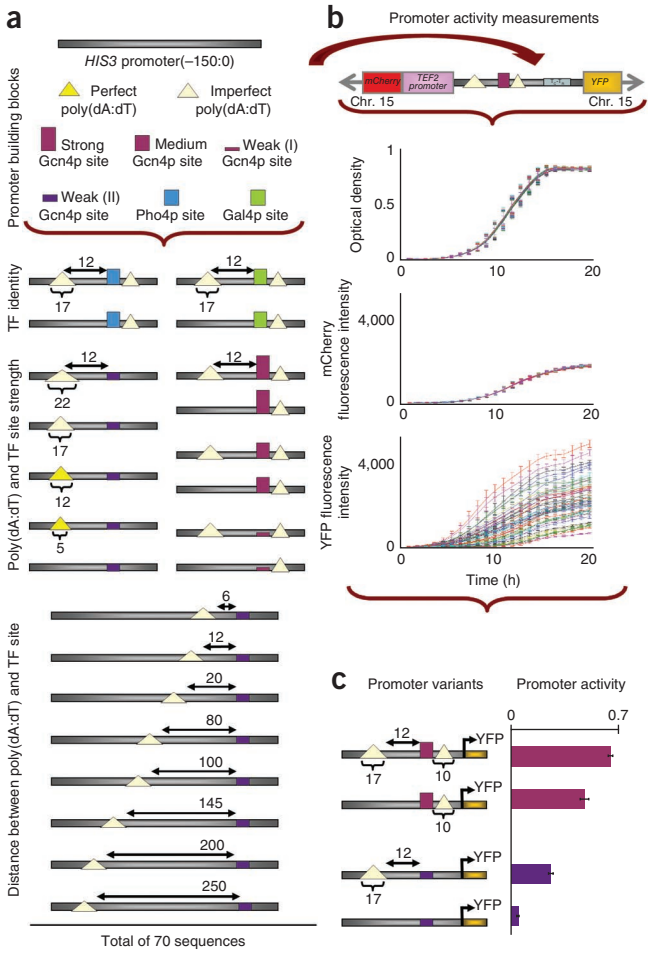

Figure 1 Schematic of library design, strain construction and promoter activity measurements. (**a**) Schematic of the promoter variants designed in this study. Shown are the transcription factor (TF)-binding sites and poly(dA:dT) tracts that were used as promoter building blocks (top). A total of 70 promoter variants were constructed in a sequence context based on the yeast *HIS3* promoter, in which a site for the transcriptional activator Gcn4p is flanked by two poly(dA:dT) tracts of 17 bp and 10 bp (Online Methods). Also shown are illustrations of the systematic changes that we made to these building blocks, which include changes to the properties of transcription factor–binding sites (identity and affinity), poly(dA:dT) tracts (length and composition) and to overall promoter architecture (distance between poly(dA:dT) tracts and transcription factor sites). Numbers by the arrows reflect distances in basepairs. (**b**) Promoter variants were inserted into a constant locus in the yeast genome immediately upstream of a YFP reporter (**Supplementary Fig. 1**). In addition, to control for experimental variability, all strains contained an mCherry reporter driven by a constant promoter. Measurements of optical density (OD), YFP and mCherry were taken over several hours using a robotically operated fluorometer. The low variation across strains in OD and mCherry measurements indicates that all strains grew similarly, that experimental variability was small and that the differences in YFP measurements are indeed attributable to the different promoter sequences. (**c**) Shown is the effect on promoter activity (computed from OD, YFP and mCherry measurements; **Supplementary Note**) of an upstream poly(dA:dT) tract in two sequence contexts (top, with a downstream poly(dA:dT) tract; bottom, without a downstream poly(dA:dT) tract and with a different Gcn4p site). For each promoter variant, the mean promoter activity ± two standard errors is shown from 6–12 independent experiments. Cells were grown and measured under conditions of amino-acid depletion.

RESULTS

Promoter library design and experimental setup

To study the effect of various combinations of nucleosome-disfavoring sequences and of transcription factor sites on transcription, we constructed 70 variants derived from the native yeast *HIS3* promoter, which contains two poly(dA:dT) tracts flanking a single site for the transcriptional activator Gcn4p[16,17] (**Fig. 1a** and **Supplementary Fig. 1**). In our variants, we systematically manipulated the length and composition of poly(dA:dT) tracts, the affinity and identity of transcription factor sites and the overall arrangement of these elements (**Fig. 1a**). To accurately measure the promoter activity of these variants (**Fig. 1b**), we employed an experimental system in yeast based on fusing promoters to a fluorescent protein reporter[18]. We integrated different promoters into the same genomic location and upstream of the same yellow fluorescent protein (YFP). To the same genomic location and in all strains, we also integrated an mCherry fluorescent reporter downstream of a constant promoter, allowing us to control for the experimental variability of our system (Online Methods).

Using a robotically automated plate fluorometer, we accurately and robustly measured the expression of the YFP and mCherry reporters over time in living cell populations grown under the activating conditions of the regulating transcription factor in each designed promoter (Online Methods). Qualitatively similar results were obtained in synthetic complete medium (**Supplementary Figs. 2–6**). By also measuring the optical density of the same cell culture, indicative of population size, we could then calculate the average YFP production rate per cell per second, averaged over the entire exponential growth phase (Online Methods), which is hereafter referred to as the promoter activity (**Fig. 1b**). Because all variants were integrated into the same genomic location and the same mRNA transcript and protein were produced from every strain, differences in the measured YFP levels are attributable only to differences in the input promoter sequences, making our system ideal for unraveling the effects of various sequence elements on expression.

Poly(dA:dT) tracts significantly affect transcriptional outcome

To establish a significant and causal role for poly(dA:dT) tracts in determining transcriptional output, we first examined the effect of their deletion. Consistent with a previous study[15], we found that, in a sequence context derived from the native *HIS3* promoter, deletion of the 17-bp poly(dA:dT) tract upstream of the Gcn4p-binding site results in a significant reduction in promoter activity (**Fig. 1c**). To test this effect in a simpler promoter architecture and with a different site affinity, we generated additional variants in which we deleted the poly(dA:dT) tract immediately downstream of the Gcn4p-binding site and altered the site affinity[19–21] (**Supplementary Note**). Notably, in this context, deletion of the upstream poly(dA:dT) tract nearly abolished promoter activity, even though a competent site for Gcn4p is still present (**Fig. 1c**). These results show that, as predicted[7], poly(dA:dT) tracts have a causal effect on transcriptional output.

If the effect of poly(dA:dT) tracts on transcription is indeed mediated by their nucleosome-disfavoring nature, then we expect a stronger effect from tracts that disfavor nucleosome formation more strongly. Recent genome-wide measurements of nucleosome occupancy *in vivo* and *in vitro* showed that nucleosome depletion increases with both the length and perfection of poly(dA:dT) tracts[10,14,22]. To test whether these tracts' properties indeed affect promoter activity, we manipulated the length and composition of the poly(dA:dT) element upstream of the Gcn4p-binding site in both of the above sequence contexts (with and without the downstream tract), and measured the resulting promoter activities (**Fig. 2a**,**b**). In both contexts, we found similar promoter activities for a variant without an upstream poly(dA:dT)

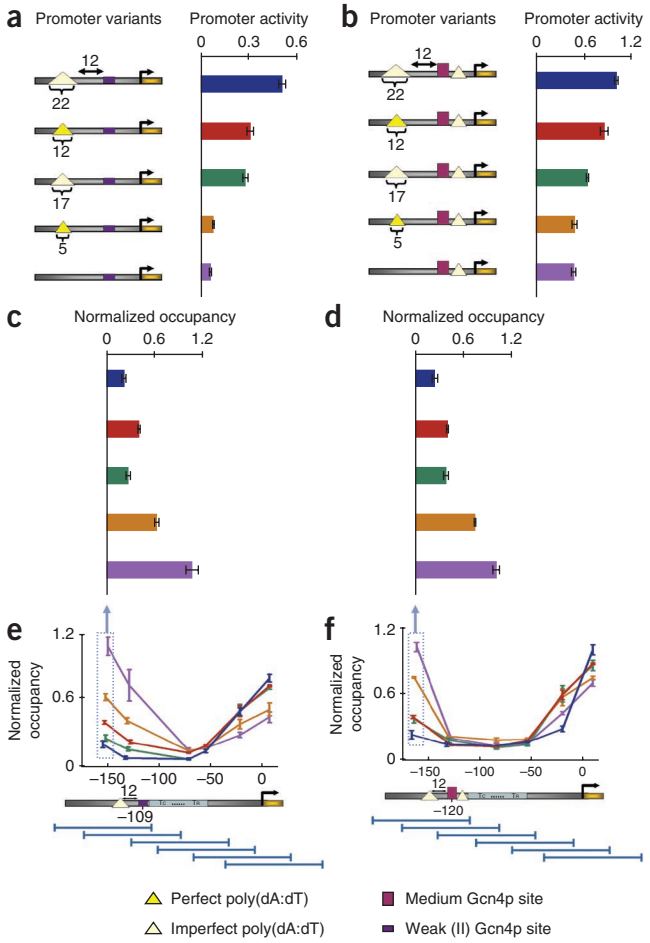

Figure 2 Poly(dA:dT) tracts significantly affect the transcriptional outcome, likely by altering nucleosome organization. (**a**) Schematics and promoter activity values of promoter variants that differ in the length and composition of the poly(dA:dT) element upstream of the Gcn4p site (12 bp upstream of the site) (**Supplementary Note**). Numbers by the arrows reflect the distances in basepairs between the poly(dA:dT) tract and the Gcn4p site (edge to edge). Shown are the mean promoter activity ± two standard errors obtained in 6–12 independent experiments. Cells were grown and measured under conditions of amino-acid depletion. (**b**) Schematics and promoter activity values are shown as in **a** for a different sequence context, with a different Gcn4p site and with a poly(dA:dT) element downstream of the site, as in the native *HIS3* promoter. (**c**) qPCR-based measurements of nucleosome occupancy for the promoters from **a** in a region surrounding the Gcn4p site and its upstream poly(dA:dT) tract (primer pair 1; **Supplementary Tables 3** and **4**). Measurements were performed on cells grown under conditions of amino-acid depletion. All values are normalized with respect to the measured nucleosome occupancy over the −1 nucleosome region in the *PHO5* promoter and are shown as mean ± two standard errors from 2–3 independent experiments. (**d**) qPCR-based measurements of nucleosome occupancy are shown as in **c** for the promoters from **b**. (**e**) qPCR-based measurements of nucleosome occupancy are shown as in **c** with several additional primer pairs that tile the promoter sequence. The *x* axis represents the distance from the translational start site, and the *x* coordinate of each plotted point represents the center of the primer pair used. The entire amplicon generated by each primer pair is represented by the blue segments below each plotted point (for primer pair sequences, see **Supplementary Tables 3** and **4**). Also shown (bottom) is a schematic of the promoter with the locations of the poly(dA:dT) elements, the Gcn4p site and the two known TATA boxes (T_c and T_R)[40]. (**f**) qPCR-based measurements of nucleosome occupancy and an accompanying schematic are shown as in **e** for the promoters from **b**.

element and for a variant containing a short and perfect 5-bp tract but significantly higher activity for a variant with a longer, perfect 12-bp tract. Notably, the activity of the variant with the perfect 12-bp tract was larger than (or equal to, in the context where the downstream tract was deleted) that of a variant containing a 17-bp tract with two mismatches but was significantly smaller than that of a variant containing a 22-bp tract with the same two mismatches. Thus, as expected from a nucleosome-mediated effect and consistent with the genome-wide nucleosome occupancy studies, we show that the length and sequence composition of poly(dA:dT) elements can be used interchangeably to tune their effect on transcription, with longer and more perfect tracts inducing higher transcriptional levels.

To further substantiate that these transcriptional changes were mediated by the effect of poly(dA:dT) tracts on nucleosome organization, we measured nucleosome occupancy along the promoter of all of the variants presented in **Figure 2a**,**b** using quantitative PCR (qPCR)[23]. Indeed, we found that alterations to poly(dA:dT) tracts affected nucleosome occupancy, with the most pronounced change being reduced occupancy near the tract (**Fig. 2c**–**f**). This reduction mainly affected the nucleosome occupancy of the nearby Gcn4p-binding site (**Fig. 2c**,**d**) and much less so the occupancy of more distant elements, such as the TATA box and the transcriptional start site (**Fig. 2e**,**f**). These measured occupancy differences are consistent with genome-wide correlations observed between poly(dA:dT) tracts and nucleosome occupancy[10,14], as variants containing longer and more perfect tracts indeed showed lower occupancy in their vicinity. Most notably, we found a general trend in which promoter variants with lower nucleosome occupancy over the nearby Gcn4p-binding site had higher promoter activity (**Fig. 2a**–**d**). We also tested the effect of poly(dA:dT) elements on nucleosome occupancy in an additional set of variants, again differing in the length and composition of the tract, but in which we mutated the Gcn4p-binding site (**Supplementary Fig. 7**). Here, too, we found a reduction in nucleosome occupancy upon addition of a nearby tract, suggesting that binding of Gcn4p is not required for this effect (**Supplementary Fig. 7c**). In terms of the effect on promoter activity, we observed a mild increase with longer and more perfect poly(dA:dT) tracts, even with a mutated Gcn4p-binding site, possibly indicating an effect of these tracts on the accessibility of other elements, such as the TATA box. However, the tract's effect on promoter activity was amplified in the presence of the intact site, in what can be referred to as a cooperative relationship[7,24] (**Supplementary Fig. 7a**,**b**).

Together, these results strongly suggest that the effect of poly(dA:dT) tracts on transcription is mediated, at least partially, by the reduced nucleosome occupancy and thus increased accessibility that these tracts confer on nearby promoter elements, such as transcription factor sites.

Poly(dA:dT) tracts offer general means to tune expression

Promoter elements that exert their transcriptional effect by modulating the binding of general regulators, such as histones, have the potential to regulate many target genes regardless of gene-specific regulators. To test whether such broad effects apply to poly(dA:dT) tracts, we replaced the Gcn4p-binding site with either a Gal4p or Pho4p site and measured the promoter activity of variants with and without a poly(dA:dT) element upstream of these sites in the activating conditions of these transcription factors (Online Methods). Indeed, in all cases, we found that the promoter activity of variants that contain a poly(dA:dT) element upstream of the regulatory factor site was significantly higher than that of promoters lacking this element (**Fig. 3a**), suggesting that poly(dA:dT) elements can serve as general modulators of expression.

Although the poly(dA:dT) tract had a qualitatively similar stimulatory effect in all of the above cases, the magnitude of the effect differed

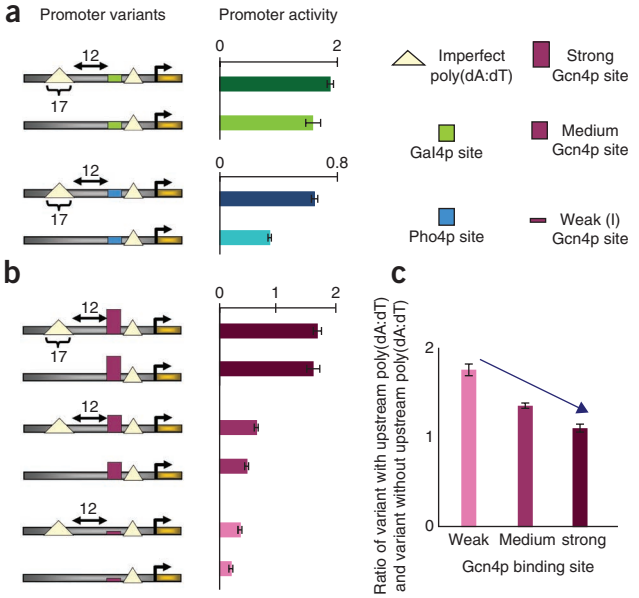

Figure 3 The transcriptional effects of poly(dA:dT) tracts are evident in promoters regulated by different transcription factors, and its magnitude is inversely proportional to the affinity of the transcription factor site. (**a**) Schematics and activity measurements for promoter variants with or without a poly(dA:dT) tract upstream of either a Gal4p (top) or a Pho4p (bottom) site (mean ± two standard errors from 4 and 5 independent experiments, respectively). Numbers above the arrows reflect the distance in basepairs between the poly(dA:dT) tract and the Gcn4p-binding site (edge to edge). Cells were grown and measured under conditions known to activate each regulator. (**b**) Schematics and activity measurements for promoter variants with or without a poly(dA:dT) tract upstream of a strong (top), medium (middle) or weak (bottom) binding site for Gcn4p. Measurements are shown as the mean promoter activity ± two standard errors from 6–12 independent experiments when cells were grown and measured under conditions of amino-acid depletion. (**c**) For each pair of promoter variants with the same Gcn4p site from **b**, the ratio of promoter activity values between the variant containing the upstream poly(dA:dT) element and the variant lacking this element is shown. The blue arrow indicates that the magnitude of the transcriptional effect of the poly(dA:dT) tract is inversely proportional to the affinity of the Gcn4p site. Error bars show ± two standard errors of the ratio as calculated across 6–12 independent experiments.

for each factor. These differences likely depend on the effective concentration of the regulating factor in the measured condition, on the affinity of the site and possibly also on the exact mechanism by which different factors activate transcription. Intuitively, as the transcription factor concentration or site affinity is higher, it can better compete with nucleosomes for access to the DNA[7] and, thus, the added benefit from a nearby poly(dA:dT) element is likely smaller. Indeed, when examining the effect of poly(dA:dT) tracts in the vicinity of Gcn4p-binding sites with different affinities, we found that, although their effect was always stimulatory, it was stronger near lower-affinity sites (**Fig. 3b**,**c**). This result complements a related observation, according to which the effect of poly(dA:dT) elements varies inversely with the concentration of Gcn4p[15].

Taken together, our results show that poly(dA:dT) elements can serve as general modulators of expression, where the magnitude of their effect depends on properties of their surrounding promoter architecture.

Effect of poly(dA:dT) tracts on transcription is context specific

In addition to manipulating the properties of poly(dA:dT) tracts and transcription factor sites, altering the relative arrangement of these elements is another means by which the range of attainable transcriptional outputs may be increased. This expectation stems from the nonlocal, distance-dependent effect that nucleosome-disfavoring elements likely confer on the surrounding nucleosome organization: whereas the probability of being wrapped within a nucleosome is low for basepairs in the immediate vicinity of poly(dA:dT) tracts, basepairs further away are also expected to be influenced by the presence of such tracts, although to a smaller extent[7,14,25]. Thus, the location of poly(dA:dT) tracts relative to other elements such as transcription factor sites may differentially affect the accessibility of these elements and, consequently, the transcriptional output.

To test this hypothesis, we constructed several sets of promoter variants, on the basis of the two above sequence contexts, in which we placed the same poly(dA:dT) tract at varying distances, ranging from 6 bp to 250 bp, upstream of the Gcn4p-binding site (**Fig. 4** and **Supplementary Fig. 8**). Compared to promoters that lack the upstream tract altogether, we found that in all sets, poly(dA:dT) tracts significantly affected promoter activity, even when they were placed 200 bp from the Gcn4p-binding site, with the magnitude of this effect varying greatly across the different distances (**Fig. 4a**,**b**). As a general trend, we found that promoter activity decreases as the distance between the poly(dA:dT) tract and the Gcn4p-binding site increases. This decrease is not entirely monotonic, as several variants deviated from this trend. Some of these deviations may result from creation or obstruction of transcription factor sites that may have occurred due to specific changes in the location of the poly(dA:dT) tract.

According to theory, placing an element strongly disfavoring nucleosomes at increasing distances from a binding site in a promoter with an otherwise equal probability for forming nucleosomes along its length is predicted to have a periodically decaying effect on nucleosome occupancy over the site and, consequently, an anticorrelated periodic effect on the transcriptional output[7,25]. Thus, the observed trend of decay in promoter activity does not fully agree with the naïve theoretical expectation. Possible explanations for this gap may be related to non-uniformity in the probability of nucleosome formation along the promoter, cooperative interactions between nearby nucleosomes or the effects of active chromatin remodeling, all of which are not accounted for in the simple theoretical predictions described.

To gain further insights into the sources of the observed trend, we measured nucleosome occupancy over the Gcn4p-binding site for a subset of the variants in **Figure 4a**. We found that varying the location of poly(dA:dT) tracts resulted in differences in nucleosome occupancy over the Gcn4p-binding site (measurements of additional promoter regions depicted in **Fig. 2e** show that more downstream regions were less variable among the tested variants). Moreover, whereas promoter activity decreased as the distance between the tract and the Gcn4p-binding site increased, nucleosome occupancy over the site generally increased (**Fig. 4e**).

Overall, our results show that altering the location of poly(dA:dT) tracts relative to a transcription factor site can affect the site's nucleosome occupancy and, correspondingly, promoter activity, yielding a multitude of transcriptional outputs, as demonstrated by the gradual trend of decay in promoter activity as the distance between these elements increases.

Poly(dA:dT) tracts can tune expression levels with fine resolution

Our variants show that manipulations of poly(dA:dT) tracts can affect transcription with magnitudes that are on par with those attained by modifying transcription factor sites. For instance, adding

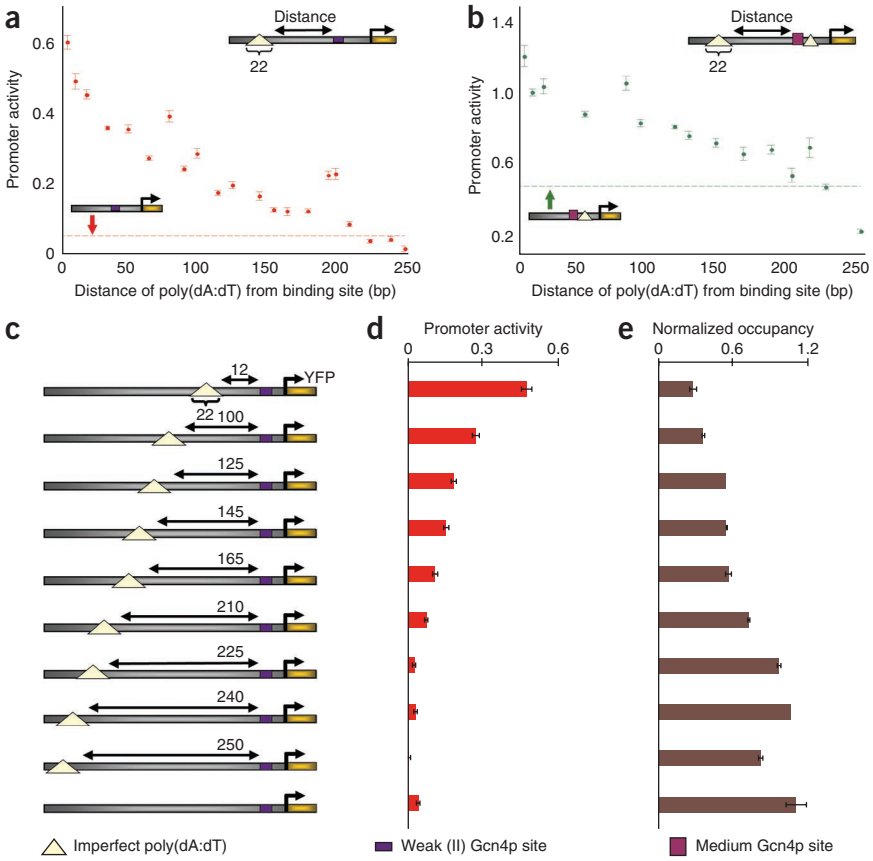

Figure 4 The transcriptional effects of poly(dA:dT) tracts depend on their distance from other promoter elements. (**a**) Promoter activity measurements for a set of promoter variants differing in the location of an upstream 22-bp poly(dA:dT) tract relative to the Gcn4p site (schematic defined in inset). Shown are the mean promoter activity ± two standard errors from 6–12 independent experiments when cells were grown under conditions of amino-acid depletion. The horizontal line represents the promoter activity of a variant that lacks the upstream poly(dA:dT) tract. (**b**) Promoter activity measurements as in **a** for promoter variants constructed in another promoter context with a different Gcn4p site and with a poly(dA:dT) tract downstream of the site (schematic defined in inset). (**c**) Schematics of a subset of the promoters in **a**. Numbers above the arrows reflect the distance in basepairs between the poly(dA:dT) tract and the Gcn4p site (edge to edge). (**d**) Promoter activity values for the promoters in **c**. Shown are the mean promoter activity ± two standard errors from 6–12 independent experiments. Cells were grown and measured under conditions of amino-acid depletion. (**e**) qPCR-based measurements of nucleosome occupancy for the promoters in **c** in a region surrounding the Gcn4p site and in the immediately upstream region (primer pair 1; **Supplementary Tables 3** and **4**). Measurements were performed on cells grown under conditions of amino-acid depletion. All values are normalized with respect to the measured nucleosome occupancy over the –1 nucleosome region in the *PHO5* promoter and are shown as the mean ± two standard errors from 2 independent experiments (for distances of 125 bp and 240 bp, only one measurement was performed).

a tract upstream of a Gcn4p-binding site increased promoter activity to a comparable extent to that attained by strengthening the site (**Supplementary Fig. 6**). Notably, it seems that beyond offering a complementary mechanism by which significant transcriptional changes can be achieved, poly(dA:dT) tracts may also offer additional functionality. The gradual changes in transcription that we observed when sampling only a small fraction of the possible alterations that can be made to properties of poly(dA:dT) tracts suggests that these tracts may provide a mechanism by which expression levels can be tuned with very fine resolution, perhaps with greater resolution than can be achieved by changes to transcription factor sites.

Examination of published Gcn4p binding data[20] demonstrates this concept. Of 57 mutations to the native Gcn4p-binding site in the *HIS3* promoter, which include all but two of the possible point mutations to the site, only one mutation (a single-basepair change that converts the site into a consensus site[20]) increases Gcn4p binding and *HIS3* expression, and most other mutations abolish Gcn4p binding[20]. When introducing the single-basepair mutation that changes the native Gcn4p-binding site into the consensus site, we observed a substantial, ~3-fold increase in promoter activity (**Fig. 5**), which, together with data from the above study[20], suggests that intermediate activity levels cannot be achieved by any of the tested mutations to the Gcn4p-binding site. Notably, we found that we could produce such intermediate promoter activity levels by altering properties of the poly(dA:dT) tract that is upstream of the Gcn4p-binding site, including changes to its length, composition and distance from the Gcn4p-binding site (**Fig. 5**).

As understanding of the transcriptional effect of various properties of poly(dA:dT) tracts improves, additional manipulations to these tracts, allowing for an even finer resolution, can be designed, thus providing a promising means to obtain promoters with prespecified transcriptional outputs. Aside from this beneficial use, such changes to poly(dA:dT) tracts may have a significant role in the evolution of gene expression, by offering an efficient genetic mechanism by which genomes can evolve and fine tune transcriptional levels.

A simple model accounts for many effects of poly(dA:dT) tracts

Given that the transcriptional effect of many of the above changes to poly(dA:dT) tracts could be qualitatively explained by their nucleosome-disfavoring nature[7,14,25], we next asked whether these changes can also be explained quantitatively using current models of transcriptional control. Notably, most existing models for predicting expression from regulatory sequence are based solely on the sequence preferences of transcription factors[26–29] and, as such, would predict the same transcriptional output for promoter variants that differ only in properties of poly(dA:dT) tracts. Thus, these models cannot explain the significant differences in promoter activity that we observed for such variants, underscoring the need to incorporate the effect of nucleosomes and their differing sequence affinities into models of transcriptional regulation.

A framework that models the binding competition between transcription factors and histones was previously proposed[7,30]. Here, we applied this framework (Online Methods) to compute the probability that Gcn4p is bound to its site in each promoter variant, as a proxy

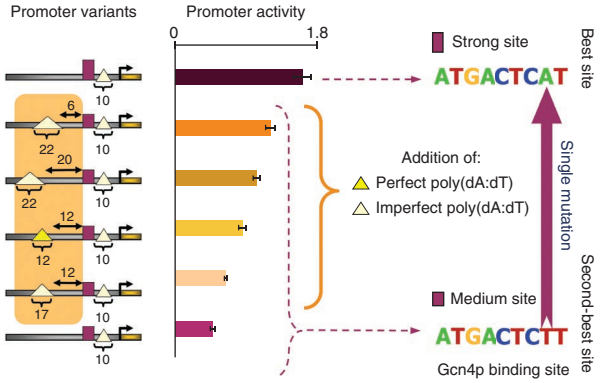

Figure 5 Changes to poly(dA:dT) tracts may allow tuning of expression levels with finer resolution than that allowed by changes to the transcription factor site. Schematic and activity measurements for promoter variants that contain either the consensus (strong) or the native (medium) Gcn4p site. Measurements are shown as the mean promoter activity ± two standard errors from 6–12 independent experiments when cells were grown under conditions of amino-acid depletion. Addition of various poly(dA:dT) tracts (four middle promoter variants with orange highlight) to the promoter with the native Gcn4p site resulted in intermediate promoter activity levels that spanned the large change in activity caused by the single-basepair mutation (T→A) that converts the native Gcn4p site to a consensus site.

measure of the transcriptional output[7] (**Fig. 6a**). As the binding preferences of both Gcn4p[19,20,31,32] and histones[10] are relatively well characterized, only two free parameters, for the concentrations of Gcn4p and histones, need to be estimated (**Supplementary Note**).

Applying this framework to the variants in which manipulations to poly(dA:dT) tracts resulted in qualitatively predicted effects on promoter activity shows that we can also understand these effects quantitatively. Notably, it is the incorporation of nucleosomes and, specifically, poly(dA:dT) tracts (**Supplementary Note**) into the model that accounts for the effects of manipulations to the presence, length and composition of poly(dA:dT) tracts (explaining 46% of the remaining variability in the data compared to a model lacking nucleosomes and achieving a Pearson's R^2 of 0.94; **Fig. 6b**).

However, some gaps still remain in our mechanistic understanding of the effect of poly(dA:dT) tracts. Most notably, variants with manipulations to the location of the upstream poly(dA:dT) element show a clear trend both in measured nucleosome occupancy over the site and in promoter activity values, and in this sense are predictable, yet this trend only partially matches the model predictions. Nevertheless, if we assess our overall ability to explain the entire data set with current models based on the binding of both transcription factors and histones, we obtain fairly good results (employing the same parameters as those chosen for the smaller set provides a lower bound of $R^2 = 0.75$) (**Fig. 6c** and **Supplementary Note**).This result is significantly better than the performance of models learned when permuting the association between promoter variants and promoter activity values (in 1,000 permutations, average $R^2 = 7\% \pm 5\%$; maximum $R^2 = 28\%$; $P < 1 \times 10^{-3}$).

Taken together, although some aspects of the transcriptional effects of poly(dA:dT) tracts remain unexplained, our results show that a relatively simple mechanistic model can account for much of the variability in the data and show the usefulness of incorporating nucleosomes and their sequence preferences into models of transcriptional control.

Poly(dA:dT) tracts affect cell-to-cell expression variability

Next, we wished to go beyond the effects that poly(dA:dT) tracts have on the mean promoter activity of a cell population and gain some insight into the effects that these tracts have at the single-cell level. Although several studies analyzed single-cell expression measurements of native promoters[33,34], the studied promoters differed by many properties, making it difficult to attribute the different measured behaviors to specific promoter elements. Our designed variants thus provide a unique opportunity to isolate the effect of poly(dA:dT) tracts, which is particularly intriguing, given that these tracts likely affect the promoter's nucleosome organization, which in turn has been linked to cell-to-cell expression variability[14,35,36].

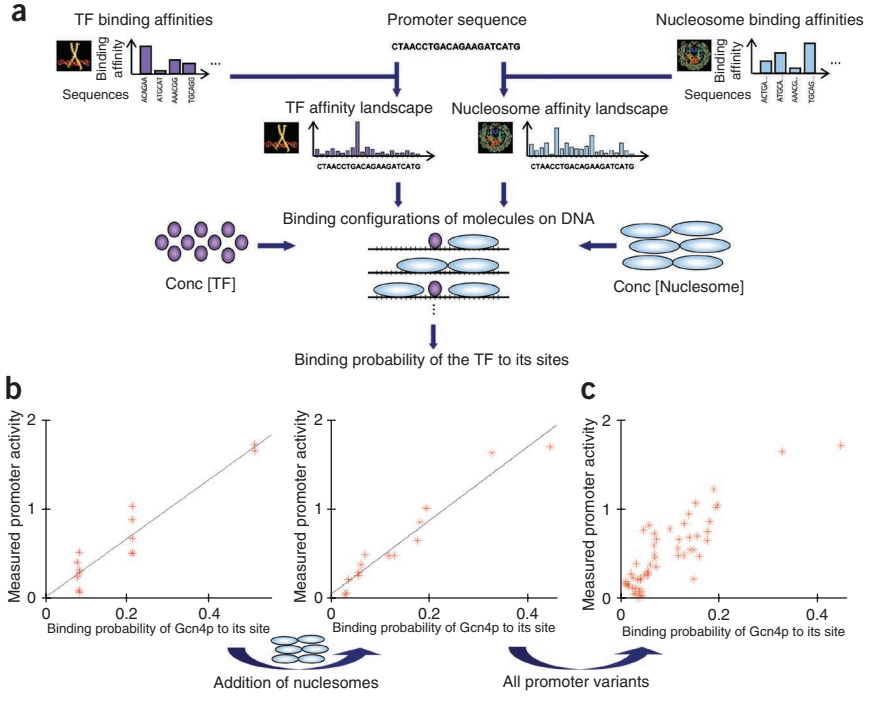

Figure 6 A mechanistic model for transcription accounts for many of the transcriptional effects of poly(dA:dT) tracts. (**a**) Schematic of the modeling framework that we employed, allowing us to compute the probability that Gcn4p is bound to its known site in each promoter variant, given as input the binding preferences and concentrations of Gcn4p and histones. (**b**) For a subset of the promoter variants (Online Methods), we show scatter plots of measured promoter activity and the binding probability of Gcn4p to its site predicted by a model that does not incorporate nucleosomes (left; Pearson's $R^2 = 0.89$, Spearman correlation = 0.81, normalized mutual information = 0.39; **Supplementary Note**) versus a model that does (right; Pearson's $R^2 = 0.94$, Spearman correlation = 0.99, normalized mutual information = 0.83). (**c**) Scatter plot of measured promoter activity and predicted binding probability of Gcn4p to its site for all Gcn4p-regulated promoter variants (Pearson's $R^2 = 0.75$, Spearman correlation = 0.82, normalized mutual information = 0.53) using the same parameter setting as in the right panel of **b**.

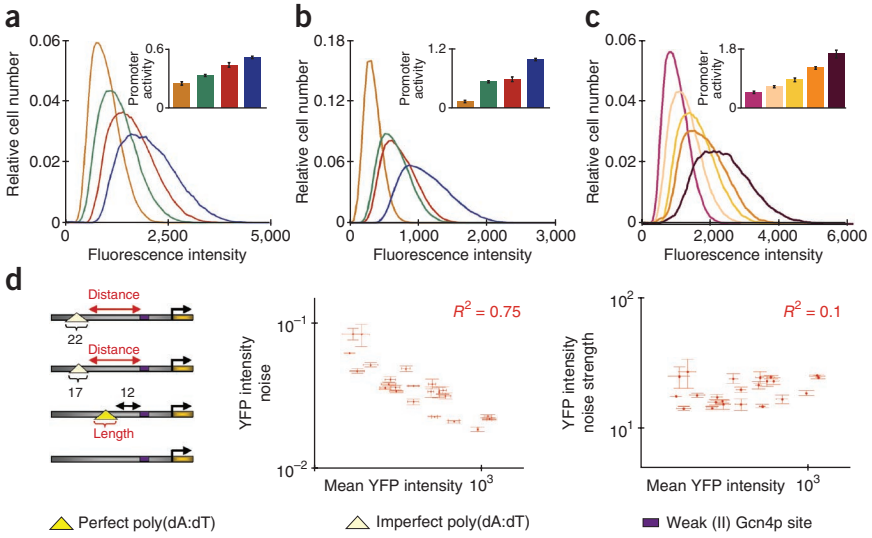

Figure 7 Poly(dA:dT) tracts affect cell-to-cell expression variability. (**a**) For a subset of the promoter variants in **Figure 2a** (one variant omitted for visual clarity) we show (in corresponding colors) the distributions of measured YFP intensities from a representative flow cytometry experiment. Inset, promoter activity measurements presented in **Figure 2a** for comparison. (**b**) The distribution of YFP intensities are shown as in **a** for a subset of the promoter variants in **Figure 2b**. (**c**) The distribution of YFP intensities are shown as in **a** for a subset of the promoter variants in **Figure 5**. (**d**) For promoter variants differing in the length, composition and location of the upstream poly(dA:dT) tract (V3–V6, V17–V19, V21, V22, V24–V31, V33, V34,V52–V55, **Supplementary Table 1**), we calculated the mean (μ) and variance (σ^2) of YFP intensities measured in mid-log phase across the population. Shown are plots of noise (defined as σ^2/μ^2) or noise strength (defined as σ^2/μ) versus the mean YFP intensity (μ) for each promoter. The value presented for each promoter is the mean of the statistic computed over three independent flow cytometry experiments ± two standard errors (Online Methods). The Pearson's R^2 value was computed on the basis of log values. We observed similar results for variants containing the downstream poly(dA:dT) (data not shown).

To study the transcriptional effect of poly(dA:dT) tracts at the single-cell level, we conducted flow cytometry measurements for a subset of strains that differed in the properties of their poly(dA:dT) tracts. Reassuringly, the average YFP levels in the flow cytometer were in excellent agreement with the promoter activity measurements (Pearson's $R^2 = 0.97$). Examining the single-cell distribution of YFP intensities, we found that all variants showed a unimodal distribution of YFP intensities, such that, for variants with a higher mean promoter activity, the entire distribution and not a specific subpopulation of cells shifted accordingly (**Fig. 7a–c**). As a further characterization, we computed both the transcriptional noise (σ^2/μ^2, variance divided by the square mean of YFP intensity) and noise strength (σ^2/μ, variance divided by the mean of YFP intensity) of all promoter variants (similar to a previous approach[37]), and found a strong anticorrelation between the mean and noise of the measured variants (Pearson's $R^2 = 0.75$). Notably, in contrast to this strong anticorrelation, the correlation between mean expression and noise strength was very weak (Pearson's $R^2 = 0.1$) (**Fig. 7d**). Under certain assumptions, noise corresponds to the frequency of protein expression bursts, whereas noise strength corresponds to the number of proteins produced at each burst[37,38]. Notably, our results thus suggest that manipulations to poly(dA:dT) tracts mainly affect the frequency of protein bursts and, to a much lesser extent, the size of each burst.

Taken together, our results show that poly(dA:dT) tracts also have predictable effects on expression at the single-cell level and provide initial insights into the dynamics by which these effects are exerted.

DISCUSSION

In summary, our study advances the understanding of how transcriptional regulation is encoded within genomic sequences by showing the importance of poly(dA:dT) tracts and providing a comprehensive characterization of their transcriptional effect. Our results strongly suggest that poly(dA:dT) tracts exert their transcriptional effects by altering nucleosome organization and, thus, the accessibility of regulatory promoter elements. Thus, poly(dA:dT) tracts may offer a general means to tune expression, applicable to promoters regulated by different transcription factors. Moreover, due to the many ways by which these tracts can be manipulated and combined with other promoter elements, they enable a wide range of transcriptional outputs and may allow fine tuning of expression with a higher resolution than that allowed by solely relying on transcription factor sites. Notably, we show various ways in which the transcriptional effects of poly(dA:dT) tracts are predictable. The transcriptional effects of most manipulations to poly(dA:dT) tracts are quantitatively explained by a simple mechanistic model. Even variants with manipulations to the location of poly(dA:dT) tracts, which still pose a challenge to our theoretical understanding, show a clear trend in both promoter activity and nucleosome occupancy measurements.

Despite these advances, several interesting research avenues await further exploration. First, as stated above, current mechanistic models are still unable to explain all of the variability generated by poly(dA:dT) tracts. Accumulation of promoter activity measurements for additional variants, specifically designed to bridge these gaps, will likely facilitate the formulation and estimation of more complex models incorporating, for instance, components such as TBP, chromatin remodelers and more complex interactions between the molecules involved. Second, although our results are compatible with the possibility that the transcriptional effect of poly(dA:dT) is mediated by nucleosome occupancy changes over adjacent transcription factor sites, they do not exclude the involvement of additional promoter elements, such as the transcriptional start site and TATA box. Decoupling such effects, if they exist, and uncovering the causal chain of underlying events is an unresolved and challenging task.

Given our results, the known prevalence of poly(dA:dT) tracts in eukaryotic promoters and their association with nucleosome depletion genome-wide, it would also be interesting to characterize the effect of these elements in additional promoter contexts (for several such examples, see ref. 18 and **Supplementary Fig. 9**). Third, although we showed initial results relating poly(dA:dT) manipulations to changes in cell-to-cell variability, a more comprehensive characterization of these effects on single-cell transcriptional dynamics, measured across time and compared to the effects of other promoter elements, will surely provide additional important insights. Finally, as histones are highly conserved across eukaryotes, so are nucleosome sequence preferences[14,39], and, thus, it would be interesting to study the transcriptional effect of poly(dA:dT) tracts in higher eukaryotes.

Taken together, our results show an important role for sequences that disfavor nucleosome formation in transcriptional regulation and suggest that, by combining these elements with other promoter building blocks, directed design of promoter sequences with prespecified expression patterns may be within reach.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We wish to dedicate this paper to Jon Widom who inspired and assisted us greatly throughout this project. This work was supported by grants from the European Research Council (ERC) and the US National Institutes of Health (NIH) to E. Segal. E. Segal is the incumbent of the Soretta and Henry Shapiro career development chair. T.R.-S. and M.L. thank the Azrieli Foundation for the award of an Azrieli Fellowship. We thank G. Hornung for his help with the analysis of flow cytometry measurements and E. Mochly for his help with the preparation of **Supplementary Figure 9**.

AUTHOR CONTRIBUTIONS

T.R.-S., M.L. and E. Segal conceived the project, designed promoter variants and analyzed the data. T.R.-S., M.L., A.W. and E. Segal planned all experiments. D.Z. and A.W. developed protocols for robotic strain assembly and activity measurements. D.Z., U.S. and M.L.-P. constructed the master strain. U.S. participated in the design of the variants, constructed the majority of variants and, together with M.L.-P., constructed strains. T.R.-S. and M.L. performed expression measurements. B.S. performed nucleosome occupancy measurements. L.K. participated in expression measurements and in their initial analysis. E. Sharon, together with T.R.-S., M.L. and L.K., developed the analysis pipeline. T.R.-S., M.L. and E. Segal wrote the manuscript. A.W. and E. Segal supervised and guided the research.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2305.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Ptashne, M. & Gann, A. *Genes and Signals* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2002).
2. Kornberg, R.D. & Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285–294 (1999).
3. Polach, K.J. & Widom, J. Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J. Mol. Biol.* **254**, 130–149 (1995).
4. Kim, H.D. & O'Shea, E.K. A quantitative model of transcription factor–activated gene expression. *Nat. Struct. Mol. Biol.* **15**, 1192–1198 (2008).
5. Lam, F.H., Steger, D.J. & O'Shea, E.K. Chromatin decouples promoter threshold from dynamic range. *Nature* **453**, 246–250 (2008).
6. Liu, X., Lee, C.K., Granek, J.A., Clarke, N.D. & Lieb, J.D. Whole-genome comparison of Leu3 binding *in vitro* and *in vivo* reveals the importance of nucleosome occupancy in target site selection. *Genome Res.* **16**, 1517–1528 (2006).
7. Raveh-Sadka, T., Levo, M. & Segal, E. Incorporating nucleosomes into thermodynamic models of transcription regulation. *Genome Res.* **19**, 1480–1496 (2009).
8. Anderson, J.D. & Widom, J. Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Mol. Cell. Biol.* **21**, 3830–3839 (2001).
9. Segal, E. & Widom, J. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* **19**, 65–71 (2009).
10. Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362–366 (2009).
11. Yuan, G.C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626–630 (2005).
12. Zhang, Y. *et al.* Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions *in vivo*. *Nat. Struct. Mol. Biol.* **16**, 847–852 (2009).
13. Dechering, K.J., Cuelenaere, K., Konings, R.N. & Leunissen, J.A. Distinct frequency-distributions of homopolymeric DNA tracts in different genomes. *Nucleic Acids Res.* **26**, 4056–4062 (1998).
14. Field, Y. *et al.* Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLOS Comput. Biol.* **4**, e1000216 (2008).
15. Iyer, V. & Struhl, K. Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J.* **14**, 2570–2579 (1995).
16. Knutson, B.A. & Hahn, S. Domains of Tra1 important for activator recruitment and transcription coactivator functions of SAGA and NuA4 complexes. *Mol. Cell. Biol.* **31**, 818–831 (2011).
17. Kuo, M.H., vom Baur, E., Struhl, K. & Allis, C.D. Gcn4 activator targets Gcn5 histone acetyltransferase to specific promoters independently of transcription. *Mol. Cell* **6**, 1309–1320 (2000).
18. Zeevi, D. *et al.* Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. *Genome Res.* **21**, 2114–2128 (2011).
19. 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).
20. Hill, D.E., Hope, I.A., Macke, J.P. & Struhl, K. Saturation mutagenesis of the yeast his3 regulatory site: requirements for transcriptional induction and for binding by GCN4 activator protein. *Science* **234**, 451–457 (1986).
21. Zhu, C. *et al.* High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res.* **19**, 556–566 (2009).
22. Yuan, G.C. & Liu, J.S. Genomic sequence is highly predictive of local nucleosome depletion. *PLOS Comput. Biol.* **4**, e13 (2008).
23. Sekinger, E.A., Moqtaderi, Z. & Struhl, K. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell* **18**, 735–748 (2005).
24. Miller, J.A. & Widom, J. Collaborative competition mechanism for gene activation *in vivo*. *Mol. Cell. Biol.* **23**, 1623–1632 (2003).
25. Kornberg, R.D. & Stryer, L. Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* **16**, 6677–6690 (1988).
26. Beer, M.A. & Tavazoie, S. Predicting gene expression from sequence. *Cell* **117**, 185–198 (2004).
27. Bintu, L. *et al.* Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.* **15**, 116–124 (2005).
28. Gertz, J., Siggia, E.D. & Cohen, B.A. Analysis of combinatorial *cis*-regulation in synthetic and genomic promoters. *Nature* **457**, 215–218 (2009).
29. 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).
30. Wasson, T. & Hartemink, A.J. An ensemble model of competitive multi-factor binding of the genome. *Genome Res.* **19**, 2101–2112 (2009).
31. MacIsaac, K.D. *et al.* An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* **7**, 113 (2006).
32. Zhu, C. *et al.* High-resolution DNA binding specificity analysis of yeast transcription factors. *Genome Res.* **19**, 556–566 (2009).
33. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nat. Genet.* **38**, 636–643 (2006).
34. Taniguchi, Y. *et al.* Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
35. Bai, L., Charvin, G., Siggia, E.D. & Cross, F.R. Nucleosome-depleted regions in cell-cycle-regulated promoters ensure reliable gene expression in every cell cycle. *Dev. Cell* **18**, 544–555 (2010).
36. Raser, J.M. & O'Shea, E.K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
37. Cai, L., Friedman, N. & Xie, X.S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).
38. Friedman, N., Cai, L. & Xie, X.S. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* **97**, 168302 (2006).
39. Ercan, S., Lubling, Y., Segal, E. & Lieb, J.D. High nucleosome occupancy is encoded at X-linked gene promoters in *C. elegans*. *Genome Res.* **21**, 237–244 (2011).
40. Chen, W. & Struhl, K. Saturation mutagenesis of a yeast his3 "TATA element": genetic evidence for a specific TATA-binding protein. *Proc. Natl. Acad. Sci. USA* **85**, 2691–2695 (1988).

ONLINE METHODS

Promoter library construction. *Construction of yeast master strain.* A single master strain containing all elements that are shared by all strains in our library was constructed by genomically integrating a construct of *ADH1* terminator -mCherry reporter gene -*TEF2* promoter -*HIS3* proximal promoter (100 bp upstream of the transcriptional start site) -YFP reporter gene -*ADH1* terminator -*NAT1* (see the detailed sequence of master strain in the **Supplementary Note**) into the Y8205 yeast strain (courtesy of C. Boone (University of Toronto)) at the deleted *HIS3* locus. This strain served as the basis for all of our strains.

Construction of promoter variants. The variable part of each promoter was constructed separately and inserted into the master strain by genomic integration, using *URA3* as a selection marker. Promoter variants were constructed either by Biomatik or as described in refs. 41–43, using the 150 bp upstream of the transcriptional start site in the native *HIS3* promoter and a plasmid sequence (F1 phage origin) of 361 bp in length taken from pRS426 (**Supplementary Fig. 1**; see full sequences of variants in **Supplementary Table 1**).

Each variant was amplified and linked to a *URA3* selection marker (sequences for amplification primers are given in **Supplementary Table 2**) using assembly PCR[41]. The resulting PCR product was designed to have 50 bp on its 5′ end and 92 bp on its 3′ end that are homologous to the genomic integration site. Integration into the genome was performed by homologous recombination, as described[44]. All steps were performed in 96-well plates, using a robotic procedure for yeast transformation[44]. The promoter sequence inserted into each strain was validated by sequencing.

Promoter building blocks. In the variable parts of the construct in each strain, we used various transcription factor–binding sites (for Gcn4p, Gal4p and Pho4p) and poly(dA:dT) elements as promoter building blocks (**Supplementary Note**).

Plate reader measurements. Strains containing the different promoters were grown on synthetic complete medium in 96-well plates at 30 °C for a period of 2 d, allowing them to reach stationary phase. We then transferred 5 µl of cells into 96-well plates containing 150 µl of fresh medium. The following types of medium were used: (i) synthetic complete medium (SCD), containing 2% glucose as a carbon source; (ii) amino-acid depletion medium, which is based on yeast nitrogen base without amino acid supplementation, except for leucine and histidine, and with 2% glucose as a carbon source; (iii) gal medium, which is synthetic complete medium containing 2% galactose as a carbon source; and (iv) pho medium, which is synthetic complete medium lacking only phosphate.

The type of medium chosen for the measurements of each variant was the one known to induce its expression (amino-acid depletion, gal and pho media for variants containing Gcn4p-, Gal4p- and Pho4p-binding sites, respectively). All variants were additionally measured in SCD (**Supplementary Figs. 2–6** and **Supplementary Note**). Absorbance (OD_{600}) and fluorescence (YFP and mCherry) were then measured periodically (every 20 min) by a plate fluorometer (Tecan F-500) using a robotic system (**Fig. 1**).

Nucleosome occupancy assay. *Spheroplasts preparation and MNase digestion.* Strains containing different promoter variants (**Fig. 2**) were cultivated at 30 °C in 300 ml of amino-acid depletion medium. Exponentially growing cells were harvested, centrifuged (4,000 RPM for 5 min) and washed with double-distilled water (ddH_2O) twice. The pellet was re-suspended in 3 ml of spheroplasting buffer (1 M sorbitol, 1 mM β-mercaptoethanol and 9 mg of lyticase (Sigma, L5263-200KU)) and incubated for 15 min at 30 °C. Spheroplasts were centrifuged (3,000 RPM for 5 min) and washed twice with 1 M sorbitol. The pellet was resuspended with 2.25 ml of MNase digestion buffer (1 M sorbitol, 50 mM NaCl, 10 mM Tris, pH 7.5, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM β-mercaptoethanol and 0.5 mM spermidine (Sigma, S0266-1G) with 0.075% Nonidet P–40 (Igepal; Sigma, CA-630)). Spheroplasts were divided into two samples. One sample was further divided and digested with MNase by the addition of 200 U (Worthington)/ml of MNase (Sigma, N5386-500UN) for several minutes (5 min, 10 min, 12 min and 15 min were tested) at 30 °C, and digestions were terminated with stop solution (1/10 sample volume of 250 mM EDTA and 5% SDS). The other sample was used for extracting genomic DNA and was transferred directly to the stop solution.

DNA purification. Both MNase-treated, and genomic DNA samples were incubated with 0.25 mg/ml RNase (Sigma,R5500-100MG) for 1 h at 37 °C, and DNA was purified with a MinElute PCR purification kit (Qiagen). MNase-treated samples were electrophoretically separated on a 2% agarose gel, and mononucleosome-sized (~147-bp) fragments were excised from the gel. The fragments were inserted into dialysis tubes (Gene Bio Application, D012-100) and electroeluted for 45 min at 120 V. DNA was ethanol precipitated and suspended in ddH_2O.

Quantitative PCR analysis. Primer pairs tiling the promoter region (~210 bp upstream and ~60 bp downstream of the transcriptional start site) were designed to yield an amplicon of 95–107 bp. Sequences of primers are provided (**Supplementary Tables 3** and **4**). Quantitative PCR analysis was performed by RT-PCR (StepOnePlus, Applied Biosystems) using a ready-mix kit (KAPA, KK4605). Reactions were performed in 10 µl with final primer concentrations of 200 µM using the following program: 3 min at 95 °C and 40 cycles of 3 s at 95 °C, 20 s at 59 °C and 10 s at 72 °C. To account for as much experimental noise as possible, the preparation of genomic DNA was coupled to the preparation of mononucleosomes, and a dilution series of this genomic DNA was used to establish standard curves specific for each primer, strain and mononucleosome preparation. In general, R^2 values from linear regression exceeded 0.995, and the range was sufficiently large to avoid extrapolation. All primer pairs used did not show heterogeneous PCR products (according to melting curve analysis). Each PCR plate contained a primer pair amplifying the −1 nucleosome at the *PHO5* promoter. The nucleosome occupancy values obtained with this primer pair were used as a normalization factor to control for plate-to-plate variation and to allow the comparison of nucleosome occupancy across different promoters.

Flow cytometry measurements. Strains containing the different promoters were grown on synthetic complete medium in 96-well plates at 30 °C for a period of 2 d, allowing them to reach stationary phase. We then transferred 5 µl of cells into 96-well plates containing 150 µl of fresh amino-acid depletion medium and grew them to mid-log phase. YFP fluorescence intensity was measured by flow cytometry on the BD LSRII system (BD Biosciences). The excitation wavelength was 350 nm for YFP and 740 nm for mCherry. Approximately 150,000 cells were collected from each well at a flow rate of 1 µl/s.

Analysis pipelines. *Plate reader measurements.* The raw data for each experiment consist of OD (indicative of cell population size), YFP and mCherry reads for each of the strains in the 96-well plate measured every 20 min. Analysis of data is almost fully automated and consists of outlier removal, subtraction of background levels, detection of exponential growth phase and calculation of promoter activity (YFP production rate per cell per second averaged over the entire exponential phase). Additional details on the data analysis pipeline and calculation of promoter activity are provided (**Supplementary Note** and **Supplementary Fig. 10**).

Flow cytometry measurements. FCS files were imported to Matlab using an available script, outlier wells and cells were discarded and cells were gated on the basis of forward scatter (FSC) and side scatter (SSC) values. For the derivation of statistical measures, YFP values were corrected to further account for changes in cell size within the gated population using robust multiple linear regression of YFP versus FSC-A and SSC-A. Additional details are provided (**Supplementary Note**).

Model description. To quantitatively assess the ability to explain measured promoter activity values, we applied a thermodynamic framework that models the binding competition between transcription factors and histones for access to DNA sequences[7]. Gcn4p and histone affinity models were constructed on the basis of published data, and various values for their concentrations were scanned (**Supplementary Fig. 11**; detailed model description in the

Supplementary Note). Promoter sequences were extended by 2,000 bp in each direction, on the basis of the genomic context in which they appear in our strains, to avoid any boundary effects on nucleosome binding. The promoter sequences of variants in which a Gcn4p-binding site is located at a fixed location from the upstream poly(dA:dT) tract (V1–V8, V13–V16, V18 and V38; **Supplementary Table 1**) were given as input for the model application described in **Figure 6b**. All variants with a Gcn4p-binding site (V1–V8 and V13–V60; **Supplementary Table 1**) were given as input to the model application described in **Figure 6c**. In all model applications, we computed (given the promoter sequences, the affinity models and concentration values for Gcn4p and histones) the binding probability of Gcn4p to its site. This probability served as a proxy for the measured promoter activity. We assessed the prediction quality of various model parameterizations using several statistical measures, namely Pearson's R^2, Spearman's rank correlation coefficient and a normalized mutual information score (details on the estimation of prediction quality are given in the **Supplementary Note**). The specific contribution of the poly(dA:dT) elements to the improvement in predictions is discussed in the **Supplementary Note** and summarized in **Supplementary Table 5**.

41. Ben Yehezkel, T. *et al.* De novo DNA synthesis using single molecule PCR. *Nucleic Acids Res.* **36**, e107 (2008).
42. Linshiz, G. *et al.* Recursive construction of perfect DNA molecules from imperfect oligonucleotides. *Mol. Syst. Biol.* **4**, 191 (2008).
43. Shabi, U. *et al.* Processing DNA molecules as text. *Syst. Synth. Biol.* **4**, 227–236 (2010).
44. Gietz, R.D. & Schiestl, R.H. Microtiter plate transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 5–8 (2007).