nature genetics

Mapping the fine structure of a eukaryotic promoter input-output function

Arun S Rajkumar, Nicolas Dénervaud & Sebastian J Maerkl

The precise tuning of gene expression levels is essential for the optimal performance of transcriptional regulatory networks. We created 209 variants of the *Saccharomyces cerevisiae PHO5* promoter to quantify how different binding sites for the transcription factor Pho4 affect its output. We found that transcription-factor binding affinities determined *in vitro* could quantitatively predict the output of a complex yeast promoter. Promoter output was precisely tunable by subtle changes in binding-site affinity of less than 3 kcal mol^{-1}, which are accessible by modifying 1–2 bases. Our results provide insights into how transcription-factor binding sites regulate gene expression, their possible evolution and how they can be used to precisely tune gene expression. More generally, we show that *in vitro* binding-energy landscapes of transcription factors can precisely predict the output of a native yeast promoter, indicating that quantitative models of transcriptional regulatory networks are feasible.

DNA regulatory motifs are important regulators of gene expression[1], and considerable effort is currently being invested in deciphering the rules that govern promoter architecture. Transcription-factor binding-site affinities can be measured *in vitro*, and high-throughput methods permit the full characterization of binding preferences quantitatively or semiquantitatively[2]. *In vivo*, binding-site affinity has to be taken in the context of a promoter—its orientation, distance from a transcription start site, occlusion by nucleosomes, competing factors, cofactor requirements and genomic location. Thus, it is not clear to what extent *in vitro* binding-affinity measurements can quantitatively predict promoter output.

The effects of genomic context on a given binding site can be measured if its accessibility, orientation or location is varied across promoters and the promoter output is determined. Such promoter libraries can consist of synthetic promoters[3], mutated native promoters or enhancers[4,5] or native promoters[6]. Several recent studies constructed such libraries to study these factors in prokaryotic[7] and eukaryotic gene regulation. The latter studies focused on how binding-site context affects expression, and they established broad rules for promoter architecture[8,9] and determined the contribution of every base in a promoter to its regulatory function[10,11]. However, explicit knowledge of how binding-site affinity translates to *in vivo* promoter activity is still sparse. Studies addressing Pho4-regulated promoters were limited to 13 *PHO5* promoters[12] and 16 variants of a *VTC4* promoter, which had one of its two Pho4 binding sites ablated[13]. Other libraries designed to address the role of binding-site affinity consisted of randomly assembled promoters or ectopic binding sites in a chimeric context. Nonetheless, these studies revealed that weak binding sites can have substantial regulatory function in conjunction with strong sites[3] and demonstrated that *in vitro* affinity does reflect *in vivo* activity[14]. Given that extensive low-affinity binding sites in the genome have been identified and their potential role in transcription has been predicted or determined in individual cases[15–17], explicitly identifying the exact contribution of binding-site affinity to regulatory function is necessary for a quantitative understanding of gene regulation. Approaches to promoter tuning have used either multiples of strong transcription-factor binding sites or large affinity changes to adjust expression levels[14,18]. It was recently shown that nucleosome-disfavoring sequences can fine tune promoter output, and it has been suggested that a similar tuning precision may not be possible by varying transcription-factor binding-site affinity[19].

We mapped the fine structure of a eukaryotic promoter input-output function by synthesizing a library of defined promoters, which we fused to a reporter gene and integrated into the yeast genome. We chose to interrogate the *PHO5* promoter from budding yeast[20], as we recently determined the precise *in vitro* binding-energy landscape of Pho4, the master regulator of the *PHO5* promoter[21]. The output of these promoters (as determined by mCherry fluorescence) was measured at the single-cell level on a massively parallel, microfluidic array under continuous growth conditions (N.D. *et al.*, unpublished data). This platform provided precise temporal information on mCherry abundance changes and the cytoplasmic-to-nuclear translocation of Pho4. We determined promoter output and Pho4 translocation under different concentrations of inorganic phosphate, [P_i], which regulates the phosphate starvation response[22,23].

RESULTS

Construction of the synthetic promoter library

The *PHO5* promoter contains two Pho4 sites (**Fig. 1a**): a strong nucleosomal site that is occluded by a nucleosome under non-inducing

Institute of Bioengineering, School of Engineering, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland. Correspondence should be addressed to S.J.M. (sebastian.maerkl@epfl.ch).

Received 16 March; accepted 22 July; published online 18 August 2013; doi:10.1038/ng.2729

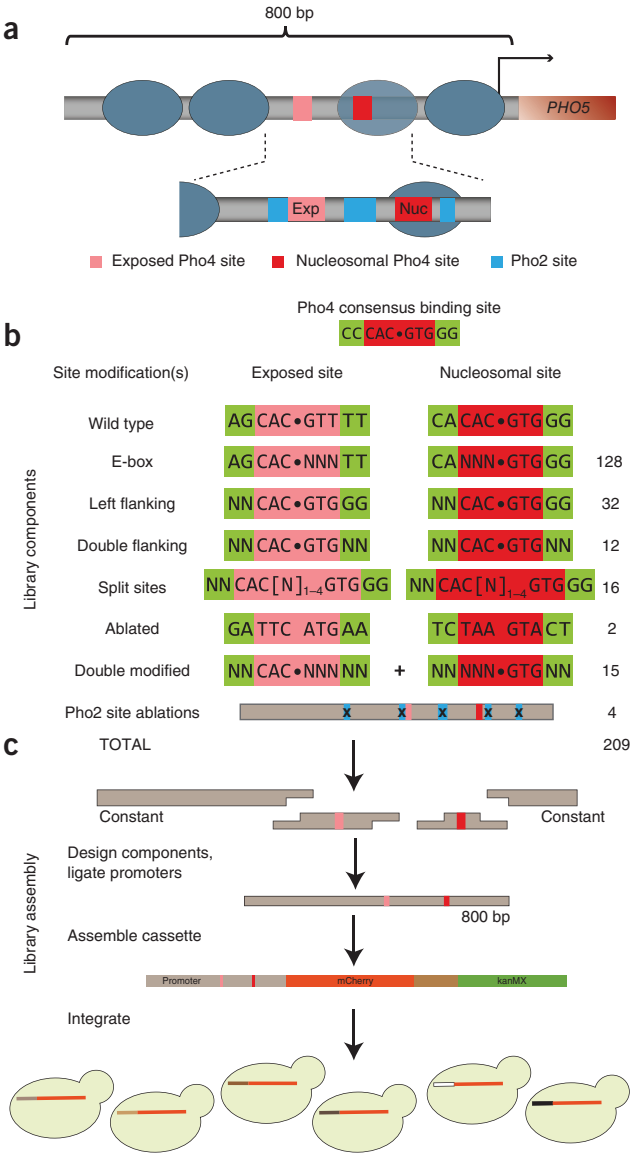

Figure 1 Design and assembly of the promoter library. (**a**) Schematic of the *PHO5* promoter illustrating nucleosome positioning, Pho4 and Pho2 sites. (**b**) Components of the synthetic promoter library. Each Pho4 site consists of an E-box (red or pink) and two flanking bases (green) on either side. The 209 promoter variants consist of Pho4 binding-site variants, ablations of Pho4 or Pho2 sites. (**c**) Assembly of synthetic promoters from component oligonucleotides. The promoter is broken into constant arms that flank sequences containing the modified sites, ligated, amplified and transformed into yeast at the *LYS2* locus. The synthetic promoters regulate expression of mCherry (orange). The components of the synthetic promoters are shown to scale with respect to the full-size *PHO5* promoter.

conditions and a weaker exposed site that is accessible under all conditions[24]. The two different accessibilities are thought to decouple the dynamic range of the promoter from the induction threshold, with the exposed site determining the promoter threshold and both sites contributing to the dynamic range[12]. Near the Pho4 sites are binding sites for Pho2, a co-regulator of *PHO5* (refs. 25,26). We designed and synthesized 209 promoters (**Fig. 1b**) to investigate how Pho4 binding sites contribute to the regulatory activity of the promoter. We constructed each promoter by high-temperature ligation from component oligonucleotides. We then transformed these constructs into yeast by homologous recombination into the *LYS2* locus (**Fig. 1c**)[27]. Although the promoters were integrated at the *LYS2* locus, their nucleosome positioning was identical to that of the native *PHO5* promoter and compared well to occupancy data from large-scale mapping studies (**Supplementary Fig. 1**)[28]. Therefore, the respective Pho4 sites remained nucleosomal and exposed despite the ectopic locus.

Single-cell analysis of promoter library induction

Our microfluidic device contains 1,152 growth chambers subdivided into three independent sections (**Fig. 2a**,**b**), which allowed us to study the entire library with three replicates per strain at either a single [P_i] or under three different [P_i]. The induction measurements for the same promoter across the subunits and between devices were repeatable to within 12% (**Supplementary Fig. 2**). Each experiment included strains with mCherry-tagged Pho4 and Pho2 (**Fig. 2c**). Whereas Pho4 localized from the cytoplasm to the nucleus upon induction, Pho2 remained in the nucleus regardless of [P_i] (data not shown). We observed that the fraction of cells in which Pho4 translocated to the nucleus decreased as [P_i] increased (**Fig. 2d**), but the kinetics of this translocation were independent of [P_i] (**Fig. 2e**).

mCherry expression was sigmoidal within the first 6 h of induction, after which mCherry abundance increased at a constant rate. This two-phase behavior was due to a decrease in the cell-division rate after 6 h of P_i starvation, and we thus modeled the transcription of mCherry up to this time point (**Fig. 2f**). We assumed that transcription activation coincided with Pho4 nuclear localization. We thus generated an ordinary differential equation (ODE) model in which the promoter switches from a basal to an activated level of transcription, with the normalized Pho4 nuclear localization determining the transition (**Fig. 2f**). Although simple, this model fit the observed mCherry abundance traces within the 6-h time period for all promoters (**Fig. 2g**). These transcription rates were our metric of promoter activity and correlated well with mCherry fluorescence abundance values (**Supplementary Fig. 3**). Promoter activity measured in bulk also correlated well with single-cell promoter activity (**Supplementary Fig. 4**), and basal transcription rates were independent of activated transcription rates (**Supplementary Fig. 5a**). Gene expression noise decreased with increasing transcription rate but increased with basal transcription rate (**Supplementary Fig. 6**).

Mapping the input-output function of the *PHO5* promoter

We measured our promoter library under fully inducing conditions (0 M P_i) to map the relation between *cis*-regulatory input (the promoter sequence) and promoter output. Unlike previous approaches, we generated a promoter library that contained binding-site variants for which a fully characterized *in vitro* binding energy landscape exists[21]. Our promoter variants gave rise to a wide range of expression levels, both above and below that of the wild-type promoter, with transcription rates varying over close to an order of magnitude (**Fig. 3**).

Pho4 is a basic helix-loop-helix (bHLH) transcription factor that binds to a palindromic decamer centered on a hexameric E-box sequence (CACGTG)[29]. We created promoters with all 128 possible exposed and nucleosomal E-box half-site variants. The measured transcription rates qualitatively resembled the *in vitro* binding energy landscape (**Fig. 4a**,**b**), peaking when the promoter contained a consensus E-box and falling off rapidly when any base was changed. In one instance where a modified E-box recreated a wild-type site, albeit shifted by two bases (sequence variant ACG (exposed)), the resulting transcription rate was considerably lower than that of the wild-type promoter, suggesting a position dependence that was possibly

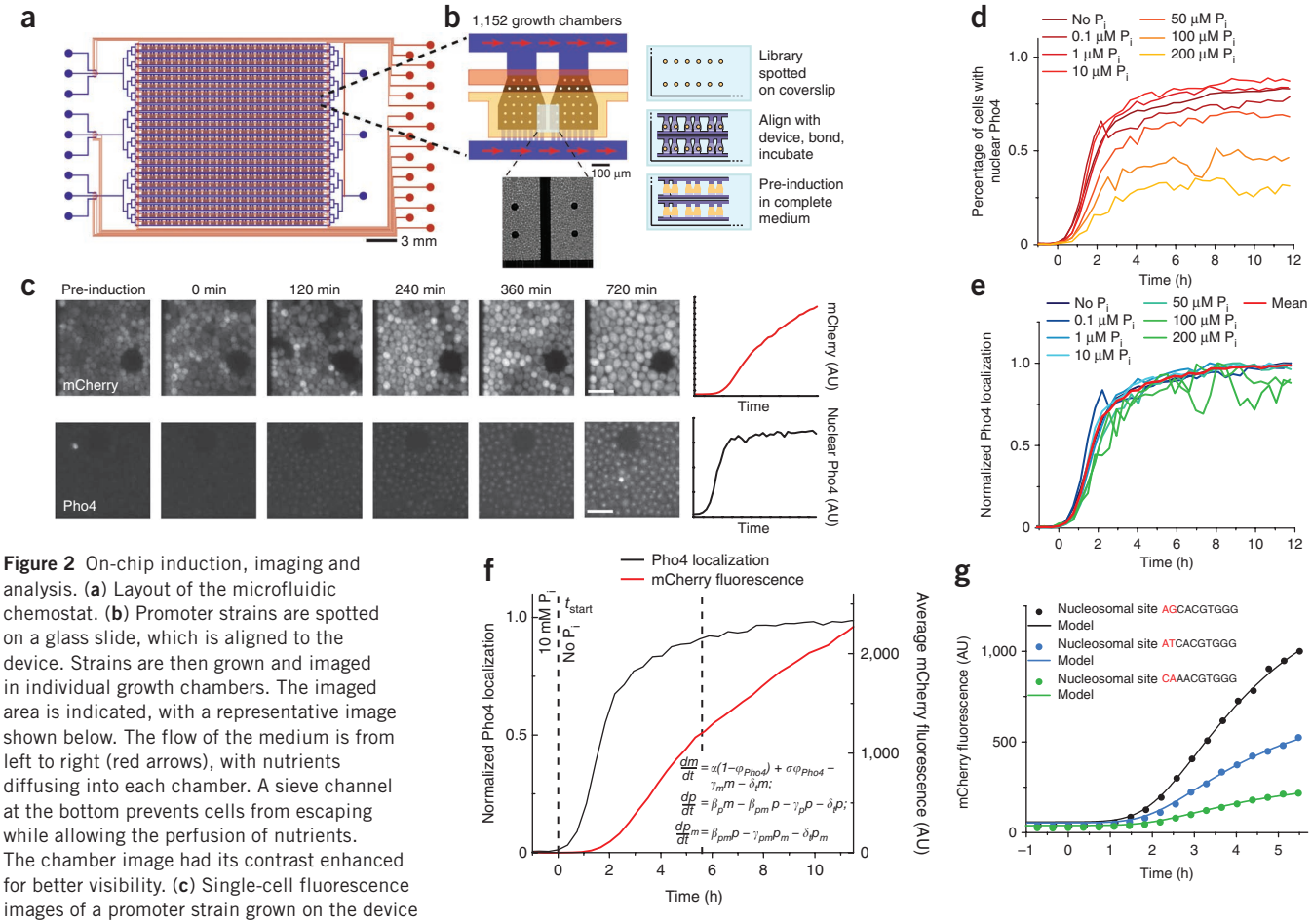

Figure 2 On-chip induction, imaging and analysis. (**a**) Layout of the microfluidic chemostat. (**b**) Promoter strains are spotted on a glass slide, which is aligned to the device. Strains are then grown and imaged in individual growth chambers. The imaged area is indicated, with a representative image shown below. The flow of the medium is from left to right (red arrows), with nutrients diffusing into each chamber. A sieve channel at the bottom prevents cells from escaping while allowing the perfusion of nutrients. The chamber image had its contrast enhanced for better visibility. (**c**) Single-cell fluorescence images of a promoter strain grown on the device (mCherry) and a strain with mCherry-tagged Pho4 (Pho4). Traces for the entire imaging period are shown on the right, with each curve showing the average fluorescence or fraction of nuclear mCherry of 200–300 individual cells. Scale bars, 15 µm. (**d**,**e**) Fewer cells relocate Pho4 to the nucleus as [P_i] increases (**d**), but localization kinetics are independent of [P_i] (**e**). (**f**) mCherry transcription and translation are modeled within the first 6 h of induction, indicated by the two dashed lines (α, basal transcription rate; σ, activated transcription rate; φ_{Pho4}, fraction of nuclear Pho4; m, mRNA; P, nascent protein; p_m, mature fluorescent protein; β, protein synthesis rate; γ, degradation rates; δ_t, dilution rate). (**g**) The model accurately fits mCherry expression for strong and weak promoters.

established by a nearby Pho2 site. The remaining affinity spikes in core E-box site variants could be explained by shifted E-boxes with suboptimal flanking bases (CGT (nucleosomal)) or split E-boxes (CCG (nucleosomal)). Thus, promoter output tuning is severely limited by modifications to the core E-box motif. Similar results have been obtained in recent attempts at promoter tuning with other transcription-factor binding-site variants[14,18,19].

In vitro measurements demonstrated that changes to the E-box–flanking bases affected affinity, extending the Pho4 binding site from a hexamer to a decamer[21,30]. Flanking-base changes led to modest affinity decreases *in vitro* of less than 1 kcal mol^{-1}, as compared to the affinity losses of more than 1 kcal mol^{-1} for core sequence variants. We constructed sets of promoters with modified E-box–flanking bases to determine whether flanking bases are physiologically relevant.

Changing 1–2 flanking bases resulted in different transcription rates *in vivo*, and these changes correlated broadly with *in vitro* binding energies. Furthermore, binding-site context affected preferences: whereas an exposed consensus site (CCCACGTGGG) showed the highest transcription rate (**Fig. 4c**), this was not the case for the nucleosomal site, where [A/T]GCACGTGGG resulted in the highest transcription

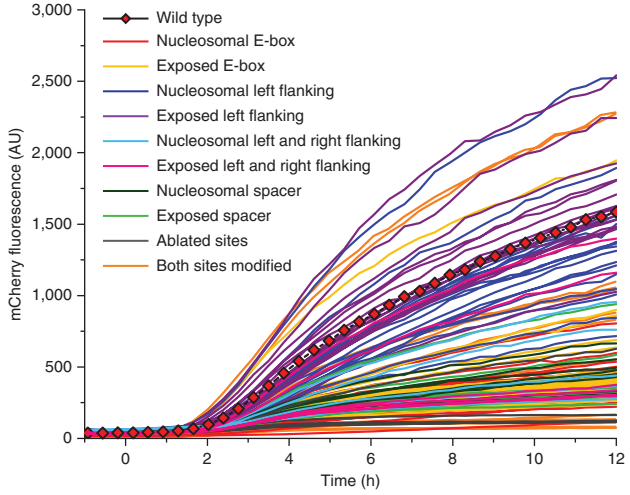

Figure 3 Induction traces for each promoter in the library under P_i starvation. $t = 0$ corresponds to the start of P_i starvation. The synthetic promoters allow expression well below and above that of the wild-type *PHO5* promoter (red diamonds). Each trace is the average of 2–6 measurements.

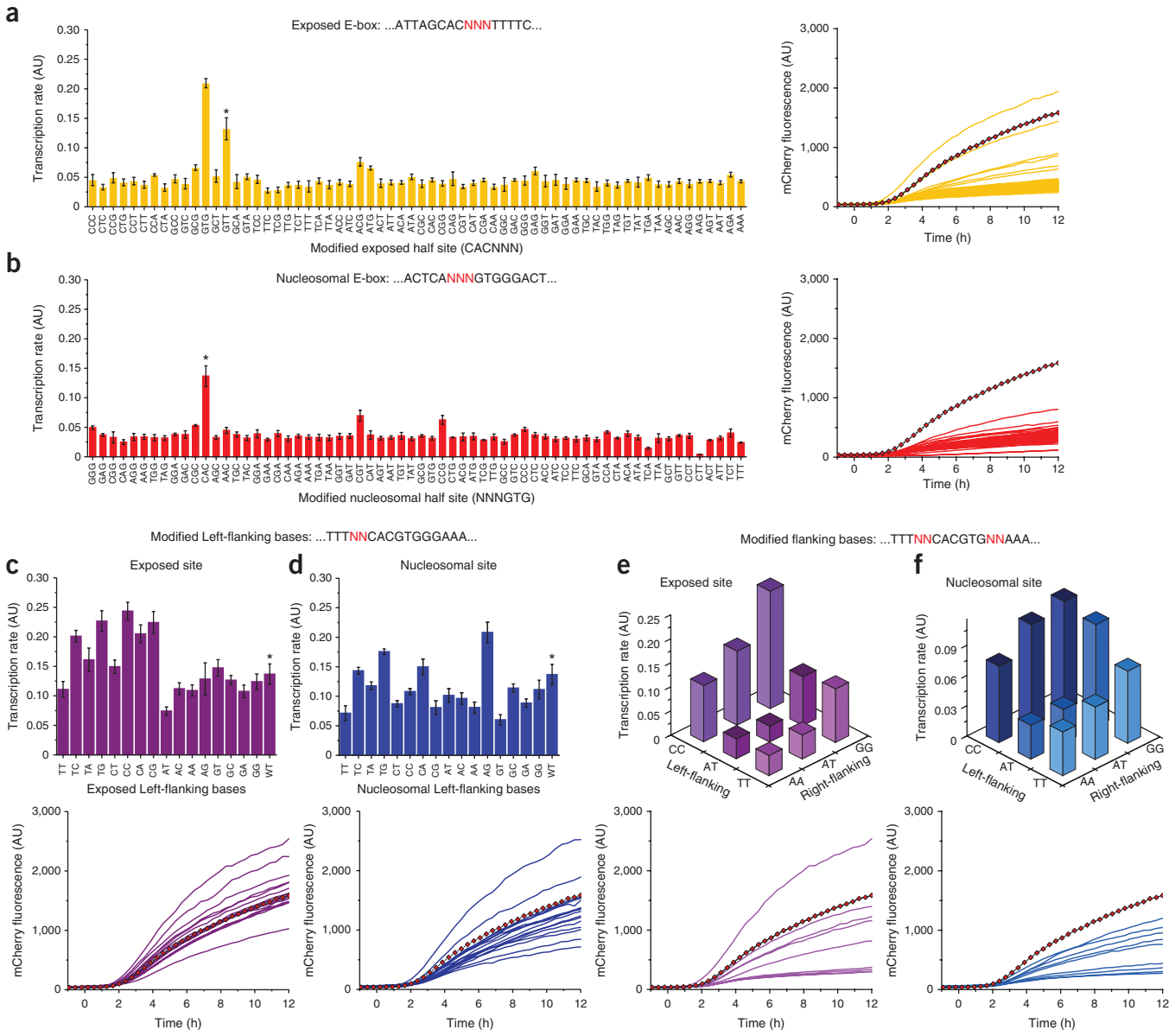

Figure 4 Single Pho4 site variants. Modifications to the exposed and nucleosomal Pho4 binding sites are highlighted in red, and the site-specific sequence context is shown. (**a**–**f**) The effects of modifying the E-box (**a**,**b**), left-flanking bases (**c**,**d**) and both pairs of flanking bases (**e**,**f**) on promoter activity are given as transcription rates and raw induction traces. Wild-type Pho4 sites are labeled with an asterisk. Transcription rate data are presented as the mean ± s.d. of 2–6 measurements per promoter, and traces for the wild-type promoter are highlighted as in **Figure 3**.

rate (**Fig. 4d**). In general, the effects of the flanking base seemed to be attenuated for the nucleosomal binding site, indicating that competition with nucleosomes may mask fine structures that are otherwise observed for binding sites in nucleosome-free regions. Promoter activity revealed that both sites showed a marked preference for C or G flanking the E-box as opposed to T, with the exposed site also disfavoring ANCACGTGGG and GNCACGTGGG (the former preferences having been observed *in vitro*[30]). Whereas modifying 1 or 2 bases led to decreased but appreciable transcription rates that were comparable to the values for promoters with relatively strong sites, flanking both sides of the E-box with unfavorable bases decreased transcription rates down to the level of promoters with low-affinity core E-boxes (**Fig. 4e**,**f**). These larger changes of 2–4 bases were recently measured in a simplified *VTC4* promoter in which all 16 palindromic flanking base sequences were investigated (CC/GG, CT/AG and so on)[13]. Unlike changes to the core E-box, 1- to 2-base changes in the flanking regions led to smaller changes in promoter output, and it is these small modifications in the E-box–flanking bases that allow fine tuning. Although most Pho4 site modifications resulted in transcription rates well below that of the wild-type promoter, we were also able to generate transcription rates above it, suggesting that the wild-type promoter output is near an optimum and is tuned to this specific level.

Split binding sites, dual sites, and binding-site ablations

When we measured the binding energy landscape of Pho4 *in vitro*[21,31], we found that several affinity spikes could be attributed to a split E-box motif. To investigate whether such sites could be bound *in vivo* and translate to promoter output, we created promoters with sites split by 1-, 2- or 4-base insertions. We found that Pho4 could still activate promoters with insertions of GC or GCGC (**Fig. 5a**).

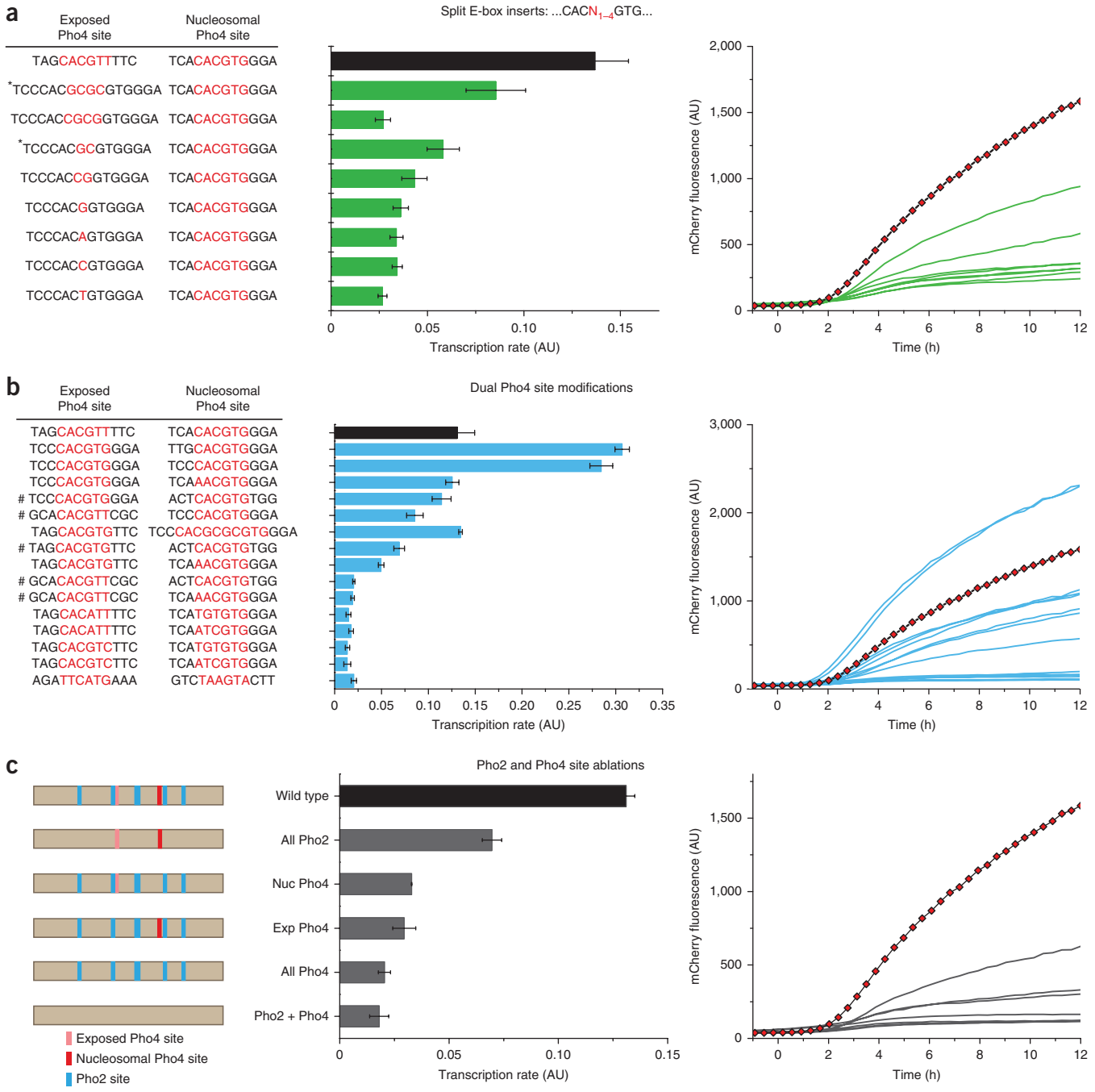

Figure 5 Effects of split sites, dual site modifications and ablations. (**a**) Split-exposed Pho4 binding sites. Promoters with secondary low-affinity binding sites are indicated with asterisks. The trace of the wild-type promoter is plotted as in **Figure 3**. (**b**) Fifteen promoters with pairs of simultaneously modified nucleosomal and exposed sites. Promoters marked with a hash contain reconstituted but shifted wild-type Pho4 sites. (**c**) The promoter library included a subset of promoters whose Pho4 (pink or red) and Pho2 (blue) sites were systematically ablated, and their induction was measured. Transcription rates (middle) and fluorescence (right) decreased considerably with the loss of all Pho2 and Pho4 sites. Nucl, nucleosomal; exp, exposed.

These fine structures were less pronounced in the nucleosomal position (**Supplementary Fig. 7**), which is similar to the attenuation observed with flanking-base substitutions. Calculating the occupancy of split E-boxes and secondary sites formed by the split led us to conclude that the output could be explained by the creation of two low-affinity binding sites: CCCACGCGTG/CACGCGTGGG (3.5 kcal mol^{-1}) and CCCACGCGCG/CGCGCGTGGG (2.9 kcal mol^{-1}).

Our library includes a number of promoters with both Pho4 sites modified (**Fig. 5b**). Increasing the affinity of either site resulted in a corresponding increase in promoter output, such that two high-affinity sites led to a more than twofold higher output than that of the wild-type promoter. The role of flanking bases in fine tuning promoter output was highlighted here as well, and we observed considerable output for a nucleosomal split site in conjunction with a high-affinity exposed site.

We measured the activity of promoters whose Pho2 and Pho4 sites were partially or completely ablated (**Fig. 5c**). Ablating all Pho2 and Pho4 sites led to a decrease in transcription rate down to a basal level that was approximately 15% of the wild-type activity. Ablating both Pho4 sites

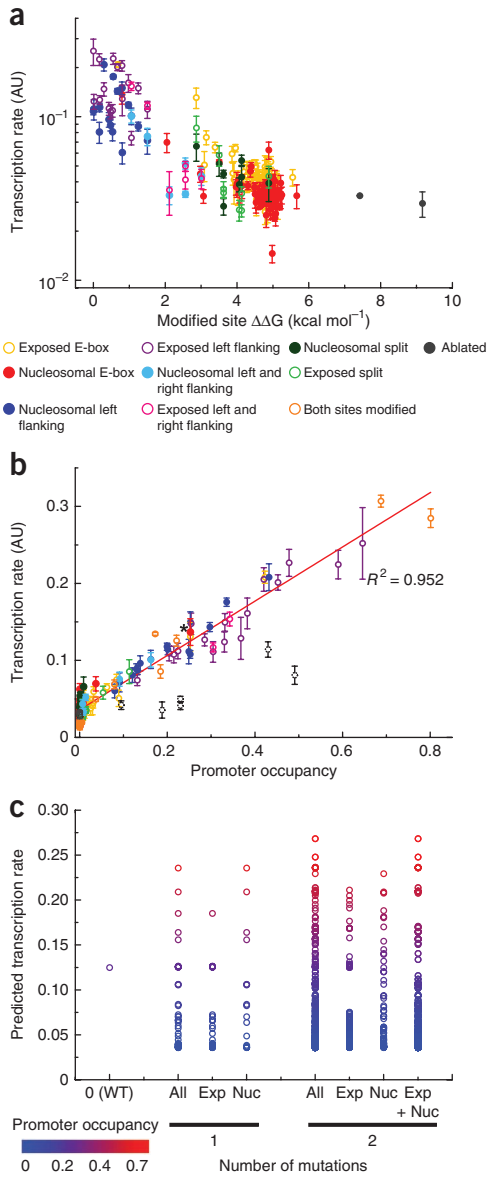

Figure 6 Correlation between *in vitro* binding-site affinity and *in vivo* promoter output. (**a**) Measured transcription rates correlate well with the affinity of the modified site calculated directly from the *in vitro* binding energy (R^2 = 0.826; **Supplementary Fig. 9**). (**b**) Calculating promoter occupancy (P_{bound}) using binding-site affinity, Cbf1 competition, nucleosome remodeling and transcription-factor competition quantitatively predicts *in vivo* promoter output. The wild-type promoter is marked with an asterisk. (**c**) One or two mutations to either site, or a single mutation to both sites, is sufficient to access the entire dynamic range of the promoter. Occupancies based on every single- and double-base mutation to the nucleosomal (nuc) and/or exposed (exp) sites were calculated, and the transcription rate was predicted. WT, wild type.

led to a similar decrease in transcription rate, indicating that Pho2 alone has no ability to activate transcription. Output was decreased by 30–40% when either Pho4 site was ablated. Notably, promoter output dropped to only 50% of wild type when all Pho2 sites were ablated compared to a near tenfold decrease that was reported in previous work[25]. We chose our Pho2 ablations to preserve nucleosome occupancy, which could explain this discrepancy (**Supplementary Fig. 8**).

In vitro affinity predicts *in vivo* promoter output

Using our quantitative *in vitro* binding-energy landscape of Pho4 (ref. 21), we compared the predicted affinities to the observed *in vivo* promoter output (**Fig. 6a**). Binding-site affinity alone accounted for 82% of the output variance for promoters with a single modified site (**Supplementary Fig. 9**). This compares favorably with results from previous studies, in which thermodynamic models were used to account for 59–75% of the expression of a synthetic promoter library[3,32].

Affinity losses of 4 kcal mol^{-1} resulted in baseline promoter output levels (**Fig. 5c**). Binding-site affinity changes of 4 kcal mol^{-1} or less were able to tune promoter output, with the entire dynamic range being accessible by affinity changes of 3 kcal mol^{-1} or less. A loss in affinity of 3 kcal mol^{-1} coincides with the dynamic range observed *in vitro*[21]. Corresponding to changes of 3 kcal mol^{-1} or less, affinity changes of up to ~0.3 bits represent the range in which promoter output was tunable (**Supplementary Fig. 10**). Changes of 2 or more bits resulted in mostly baseline promoter levels. In terms of information content, binding-site positions can be assigned to two broad classes: (i) positions with high information content, which are required for transcription-factor binding and promoter output, and (ii) positions with intermediate to low information content, which allow tuning of promoter output levels.

We implemented a statistical mechanics model that calculates the probability of Pho4 being bound to the promoter (P_{bound}) on the basis of both the exposed- and nucleosomal-site affinities[33,34]. A model based on these two parameters alone performed well, except for a set of promoters whose output was lower than expected (**Supplementary Fig. 11a**). Given recent suggestions that Cbf1 may act as a competitor for Pho4 (refs. 13,35), we included Cbf1 competition in our model, which only marginally affected the P_{bound} values (**Supplementary Fig. 11b**)[13]. This left two groups of promoters unexplained. The first group consisted of promoters with low-affinity exposed sites and a wild-type, strong nucleosomal site. The model considered both sites to be equally accessible and thus overestimated the contribution of the occluded site. We therefore included a nucleosome remodeling factor, which modulates the accessibility of the promoter as a function of the exposed or nucleosomal site affinity and accurately predicted this class of promoters (**Supplementary Fig. 11c**).

The final class of 22 promoters showed output that was lower than expected but that was nonetheless correlated with the P_{bound} values. These promoters fell within a class of promoters with modified flanking bases. In these promoters, we included stretches of TTT and AAA flanking the binding sites to prevent the formation of secondary Pho4 binding sites. Furthermore, only a subclass of these promoters showed lower-than-expected output. We thus hypothesized that the inclusion of the insulating stretches in combination with specific sequence motifs could give rise to a binding site that is recognized by one or more transcription factors other than Pho4 and Cbf1. This hypothesis was corroborated by the apparent sequence specificity of the promoter subset and the fact that the promoters in question had considerably higher basal fluorescence levels (**Supplementary Figs. 5b** and **11c**). We verified that the difference in output was not due to experimental variation by comparing a set of promoters that we analyzed across different devices, which correlated with R^2 = 0.97 and a slope of 1.045 (**Supplementary Fig. 12**). Searching STAMP[36] and ScerTF[37] for transcription factors with specificities comparable to that of the sequence motif present on these promoters identified Rpn4, Spt23, Mac1, Stb3 and Ypr015c. It is therefore possible that one or more of these transcription factors can compete with Pho4. Calculating P_{bound} values for promoters with abnormally high basal expression by including a specific competitor term with an abundance and affinity comparable to a generic transcription factor was able to accurately predict the

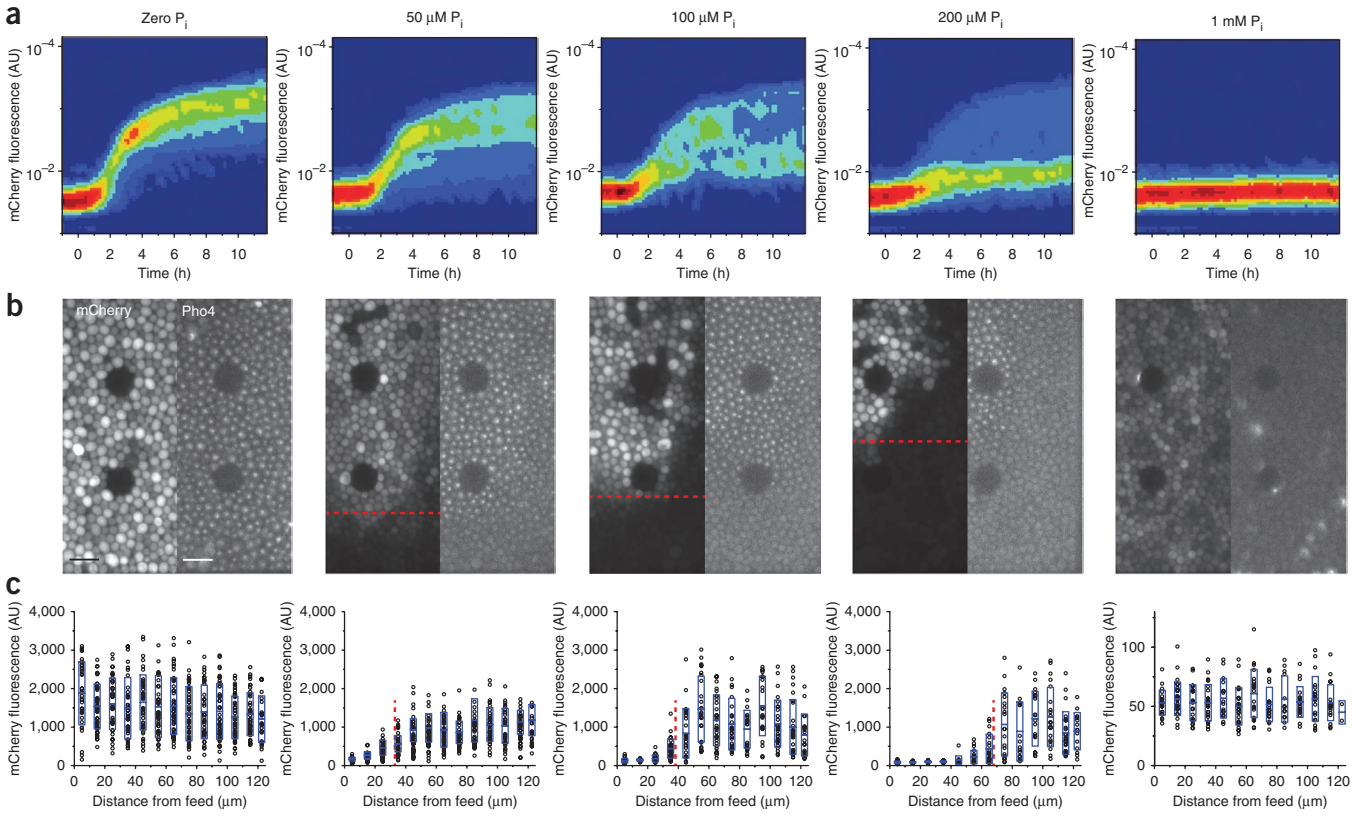

Figure 7 Phosphate-dependent activation of the *PHO5* promoter. (**a**) Histograms of single-cell fluorescence over time reveal its broadening and bifurcation as [P_i] is increased. (**b**) Side-by-side comparison of microscopic images of the corresponding chamber used in **a** and a representative Pho4 localization at the end of induction. Scale bars, 15 μm. The red dashed lines indicate the measured position of the fluorescent front. The lower edge of the images corresponds to the location of the sieve channels at the bottom of each chamber. (**c**) mCherry profiles across the chamber for the same promoter and [P_i] as in **a** and **b**. The boxes indicate the mean fluorescence for 5-μm intervals ± 1 s.d. The dashed lines indicate the position of the front, as in **b**.

output for these promoters (**Supplementary Fig. 11d**). A final group of six promoters remained refractory, four of which fell into a single class and two of which had basal expression values that were not high enough to be recalculated (**Fig. 6b**).

A model based on statistical mechanics, including Cbf1 competition, nucleosomal remodeling and a sequence-specific competitor ultimately explained 95% of the observed variation for 203 promoters (**Fig. 6b**). Promoters spanning all classes of modifications, including core E-box changes, flanking-base changes, split sites and dual site modifications, could be accurately predicted. It is thus possible to quantitatively model the output of complex native promoters with *in vitro*–determined binding affinities.

To present the accessible range of promoter output in terms of mutational distance, we predicted transcription rates for all possible 1- or 2-base mutations in the Pho4 binding sites (**Fig. 6c**). Promoters with a single mutation were tunable over a transcription range of up to ~0.23 arbitrary units (AU) and were reasonably fine grained. Two mutations, either within the exposed or nucleosomal site or with one mutation in each, resulted in finely graded promoter output values with transcription rates of up to 0.27 AU. The entire dynamic range is thus readily accessible through a small number of mutational changes in either one or both Pho4 binding sites. We obtained the same result when considering experimentally measured promoters (**Supplementary Fig. 13**).

PHO5 promoter activity at intermediate [P_i]

We measured our promoter library under seven [P_i]: 0 μM, 0.1 μM, 10 μM, 50 μM, 100 μM, 200 μM and 1 mM. In addition to determining promoter output levels, we concurrently measured Pho4 nuclear translocation (**Fig. 7**). From 0 to 10 μM [P_i], all cells translocated Pho4 into the nucleus, and each cell fully induced mCherry expression. At intermediate [P_i] of 50–200 μM, we observed that cells near the bottom of the growth chamber (proximal to the sieve channels) did not translocate Pho4 to detectable levels and only induced mCherry to low levels, whereas cells further into the microchemostat showed the same translocation and induction behavior as cells cultured without P_i (**Fig. 7a**,**b**). The front between Pho4 translocating and nontranslocating cells moved deeper into the microchemostat as we increased [P_i] (**Fig. 7b**,**c**) until the entire visible section of the microchemostat failed to translocate Pho4 at 1 mM P_i. This suggested that a P_i gradient was established as a result of rapid consumption of P_i by cells near the nutrient supply inlet at low to medium [P_i]. We modeled P_i diffusion and cellular uptake across the chamber and found that the presence of the front and its P_i-dependent position could be explained using an experimentally determined P_i uptake rate of 0.6865 fmol min^{-1} per cell[38], which indicated that the critical [P_i] for full Pho4 translocation seemed to be between 12 and 15 μM (**Supplementary Fig. 14**).

Chromatin has been shown to decouple dynamic range from induction threshold for the *PHO5* promoter[12]. The dynamic range can be defined as the difference of expression in high [P_i] (no induction) and in 0 M P_i (full induction). The induction threshold is defined as the minimal [P_i] that gives rise to promoter output. Decoupling of dynamic range from induction threshold suggests that these two parameters can be tuned independently, meaning that two promoters can in principle have the same dynamic range (the same output under

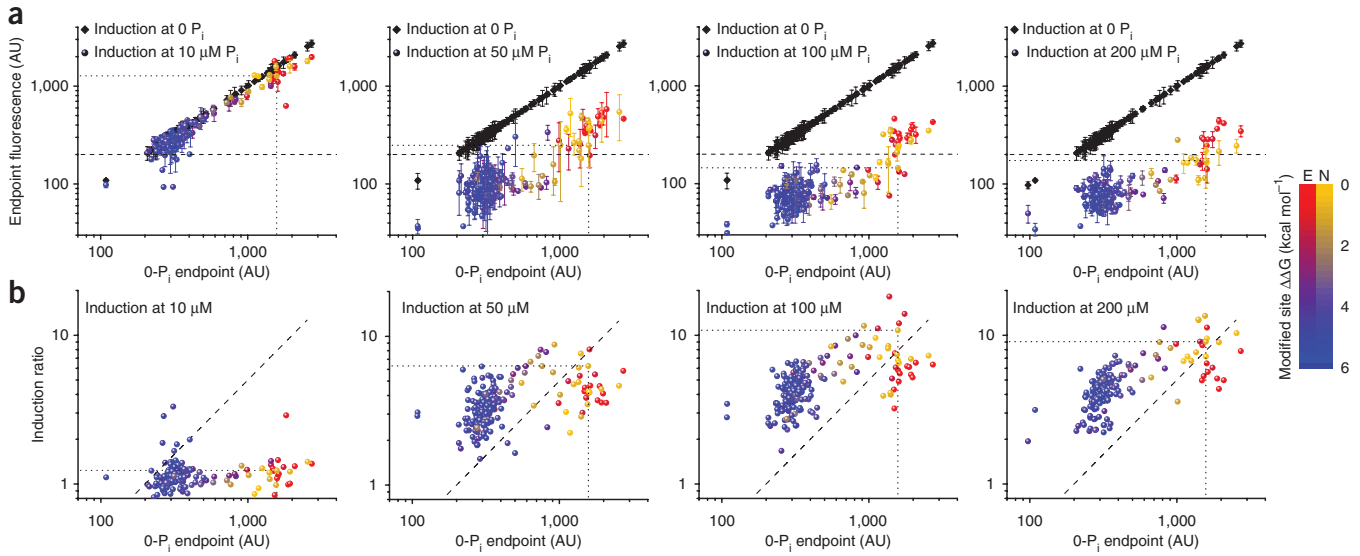

Figure 8 Promoter dynamic range and induction threshold. (**a**) Plotting promoter output for a subset of the library at various [P_i] shows that the synthetic promoters decouple activation threshold from dynamic range for a certain range of output levels. No output is observed at intermediate [P_i] for promoters with a sufficiently low output under full induction, whereas promoters with high-affinity binding sites and correspondingly high output under full induction induced appreciably at intermediate [P_i]. The dotted lines indicate the wild-type promoter, and the dashed line is set at 200 AU. (**b**) Induction ratio (full induction output/output at intermediate [P_i]) as a function of full induction. Induction ratio plots are colored by exposed (E) or nucleosomal (N) binding-site affinity. The dotted and dashed lines are the same as those in **a**.

fully inducing conditions) but different induction thresholds. Dynamic range is a function of all Pho4 binding sites, whereas induction threshold has been suggested to depend solely on the exposed site. In 0 M P_i, promoters with either a low- or high-affinity exposed binding site are remodeled, leading to full induction, whereas at intermediate [P_i], an exposed low-affinity binding site cannot remodel and induce the promoter. This model stipulates that the affinity of the nucleosomal Pho4 site has no influence on induction threshold, meaning that promoters with a high-affinity nucleosomal site but a low-affinity exposed site are not expected to induce at intermediate [P_i].

To determine whether our synthetic promoters are decoupled and how decoupling may depend on binding-site affinity, we plotted promoter outputs rank ordered by their respective output values at 0 M P_i (full induction) (**Fig. 8** and **Supplementary Figs. 15** and **16**). At 10 μM P_i, all promoters fully induced to similar levels as they did at 0 M P_i (**Fig. 8a**). At 50 μM P_i, promoters with full-induction output below 1,000 AU did not induce and were therefore completely decoupled. Most promoters with a higher dynamic range (above 1,000 AU under full induction) induced to appreciable levels under this condition. Notably, we observed that not only did promoters with high-affinity exposed sites induce at intermediate [P_i], but promoters with high-affinity nucleosomal sites did as well. These nucleosomal promoter variants all contained a low-affinity exposed site and were thus not expected to induce under these conditions. Therefore, high-affinity nucleosomal sites also contribute to setting the induction threshold.

Because the dynamic range is a function of all Pho4 binding sites and because both the exposed and nucleosomal sites define the induction threshold, it may be difficult to achieve a high dynamic range and full decoupling at intermediate [P_i]. To visualize this relationship, we defined the promoter induction ratio as the promoter output under fully inducing conditions divided by the output at a given intermediate [P_i] (**Fig. 8b**). At 50 μM P_i, essentially all promoters with a higher or equal dynamic range to that of the wild-type *PHO5* promoter had a lower induction ratio. The wild-type promoter also only marginally induced at 50 μM P_i. At 100 μM and 200 μM P_i, the wild-type promoter was completely decoupled (did not induce). Only one other synthetic promoter had better performance than the wild-type promoter at 50 μM P_i, and a few promoters performed equally well at 100 μM and 200 μM P_i. This suggests that the wild-type promoter is optimized to achieve the highest possible output under fully inducing conditions while preventing induction at intermediate [P_i] (a high dynamic range and a high induction ratio).

DISCUSSION

We synthesized a defined promoter library to explicitly study to what extent transcription-factor binding-site affinity contributes to the activity of an inducible promoter with multiple binding sites. The lack of systematic data on this relationship has already been noted in broader work[14].

Recent work on tuning promoter output has focused on relatively large binding-site affinity changes and/or generating binding-site duplications. Here we show that it is possible to fine tune promoter output by directly modifying the affinity of a single Pho4 binding site in the *PHO5* promoter. Pho4 has been known to bind to a hexameric E-box motif, and we previously showed that the flanking bases of Pho4 contribute to affinity *in vitro* and distinguish the closely related transcription factors Pho4 and Cbf1 (ref. 35). Here we show that modification of 1–2 flanking bases allows for fine tuning of promoter output. Furthermore, subtle affinity changes of less than 3 kcal mol^{-1} or 0.3 bits are capable of fine tuning promoter output over the entire dynamic range of the promoter. Changes to the core E-box motif resulted in a marked reduction in output in almost all cases. These subtle changes in binding-site sequence also showed that the tuning of promoter output through transcription-factor binding-site variations should be readily accessible by evolution. Our results also confirm that transcription-factor binding-energy landscapes can be translated directly to modulating and fine tuning promoter output *in vivo*. In fact, a relatively simple binding model based on *in vitro* binding affinities was ultimately able to explain 95% of the observed variance, indicating that precise quantitative predictions of *in vivo* promoter output based on *in vitro* biophysical measurements are possible.

Studying the P_i-dependent induction behavior of our promoter library recapitulated previous findings[12,39] but showed that it is both the exposed and the nucleosomal Pho4 sites in the *PHO5* promoter that define induction threshold as opposed to solely the exposed site, as has been previously stipulated[12]. Furthermore, the wild-type *PHO5* promoter seems to be optimized to achieve a high dynamic range while at the same time preventing induction at intermediate [P_i]. Only a few of our synthetic promoters achieved a similarly high dynamic range and fold-induction ratio performance.

Although numerous studies have made considerable progress in enumerating the rules governing promoter input-output relations, predicting promoter output remains a challenge. A native yeast promoter such as *PHO5* (ref. 12) contains many degrees of freedom. *PHO5* is regulated by at least two transcription factors, Pho4 and Pho2, whose binding sites are in close proximity to one another and whose regulation of *PHO5* is cooperative[25,26,35]. The binding-site location with respect to the transcription start site and nucleosome positioning, as well as the nucleosome positioning itself, is also of importance[14,19,40,41]. In addition, computational models need to be developed[42–44], and supported by precise quantitative data describing many if not all components of a given system. Despite these complexities, progress in the field is promising and may soon permit the quantitative prediction of promoter output levels for transcriptional regulatory networks in *S. cerevisiae*.

METHODS

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

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J.L. Garcia-Cordero for help with modeling phosphate uptake and the Shore lab at the University of Geneva for their assistance with the library construction. This work was supported by a SystemsX.ch grant DynamiX-RTD (2008/005) to S.J.M. and the École Polytechnique Fédérale de Lausanne.

AUTHOR CONTRIBUTIONS

S.J.M. and A.S.R. designed the library constructs. A.S.R. synthesized the promoter libraries, performed nucleosome occupancy measurements and characterized the promoter library in bulk. N.D. carried out the on-chip experiments. N.D. and S.J.M. designed and implemented the transcription model. A.S.R. and S.J.M. analyzed the data. A.S.R. and S.J.M. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Segal, E. & Widom, J. From DNA sequence to transcriptional behaviour: a quantitative approach. *Nat. Rev. Genet.* **10**, 443–456 (2009).
2. Geertz, M. & Maerkl, S.J. Experimental strategies for studying transcription factor–DNA binding specificities. *Brief Funct. Genomics* **9**, 362–373 (2010).
3. Gertz, J., Siggia, E.D. & Cohen, B.A. Analysis of combinatorial *cis*-regulation in synthetic and genomic promoters. *Nature* **457**, 215–218 (2009).
4. Patwardhan, R.P. *et al.* High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.* **27**, 1173–1175 (2009).
5. Kwasnieski, J.C., Mogno, I., Myers, C.A., Corbo, J.C. & Cohen, B.A. Complex effects of nucleotide variants in a mammalian *cis*-regulatory element. *Proc. Natl. Acad. Sci. USA* **109**, 19498–19503 (2012).
6. 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).
7. Kinney, J.B., Murugan, A., Callan, C.G. Jr. & Cox, E.C. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc. Natl. Acad. Sci. USA* **107**, 9158–9163 (2010).
8. Cox, R.S. III, Surette, M.G. & Elowitz, M.B. Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.* **3**, 145 (2007).
9. Gertz, J. & Cohen, B.A. Environment-specific combinatorial *cis*-regulation in synthetic promoters. *Mol. Syst. Biol.* **5**, 9 (2009).
10. Patwardhan, R.P. *et al.* Massively parallel functional dissection of mammalian enhancers *in vivo*. *Nat. Biotechnol.* **30**, 265–270 (2012).
11. Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277 (2012).
12. Lam, F.H., Steger, D.J. & O'Shea, E.K. Chromatin decouples promoter threshold from dynamic range. *Nature* **458**, 6 (2008).
13. Aow, J.S. *et al.* Differential binding of the related transcription factors Pho4 and Cbf1 can tune the sensitivity of promoters to different levels of an induction signal. *Nucleic Acids Res.* **41**, 4877–4887 (2013).
14. Sharon, E. *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* **30**, 521–530 (2012).
15. Tanay, A. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Res.* **16**, 962–972 (2006).
16. Kahana, S. *et al.* Functional dissection of IME1 transcription using quantitative promoter-reporter screening. *Genetics* **186**, 829–841 (2010).
17. Su, T.C., Tamarkina, E. & Sadowski, I. Organizational constraints on Ste12 *cis*-elements for a pheromone response in *Saccharomyces cerevisiae*. *FEBS J.* **277**, 3235–3248 (2010).
18. Khalil, A.S. *et al.* A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647–658 (2012).
19. Raveh-Sadka, T. *et al.* Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nat. Genet.* **44**, 743–750 (2012).
20. Svaren, J. & Hörz, W. Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast. *Trends Biochem. Sci.* **22**, 5 (1997).
21. Maerkl, S.J. & Quake, S.R. A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **315**, 233–237 (2007).
22. Oshima, Y. The phosphatase system in *Saccharomyces cerevisae*. *Genes Genet. Syst.* **72**, 323–334 (1997).
23. Ogawa, N., DeRisi, J.L. & Brown, P.O. New components of a system for phosphate accumulation and polyphosphate metabolism in *Saccharomyces cerevisiae* revealed by genomic expression analysis. *Mol. Biol. Cell* **11**, 13–22 (2000).
24. Venter, U., Svaren, J., Schmitz, J., Schmid, A. & Hörz, W. A nucleosome precludes binding of the transcription factor Pho4 *in vivo* to a critical target site in the PHO5 promoter. *EMBO J.* **13**, 8 (1994).
25. Barbarić, S., Münsterkötter, M., Goding, C.R. & Hörz, W. Cooperative Pho2-Pho4 interactions at the PHO5 promoter are critical for binding of Pho4 to UASp1 and for efficient transactivation by Pho4 at UASp2. *Mol. Cell Biol.* **18**, 11 (1998).
26. Barbarić, 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).
27. Rajkumar, A.S. & Maerkl, S.J. Rapid synthesis of defined eukaryotic promoter libraries. *ACS Synth. Biol.* **1**, 483–490 (2012).
28. Lee, W. *et al.* A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.* **39**, 1235–1244 (2007).
29. Shimizu, T. *et al.* Crystal structure of PHO4 bHLH domain–DNA complex: flanking base recognition. *EMBO J.* **16**, 4689–4697 (1997).
30. Fisher, F. & Goding, C.R. Single amino acid substitutions alter helix-loop-helix protein specificity for bases flanking the core CANNTG motif. *EMBO J.* **11**, 7 (1992).
31. Maerkl, S.J. & Quake, S.R. Experimental determination of the evolvability of a transcription factor. *Proc. Natl. Acad. Sci. USA* **106**, 18650–18655 (2009).
32. Mogno, I., Vallania, F., Mitra, R.D. & Cohen, B.A. TATA is a modular component of synthetic promoters. *Genome Res.* **20**, 1391–1397 (2010).
33. Granek, J.A. & Clarke, N.D. Explicit equilibrium modeling of transcription-factor binding and gene regulation. *Genome Biol.* **6**, R87 (2005).
34. Bintu, L. *et al.* Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.* **15**, 116–124 (2005).
35. Zhou, X. & O'Shea, E.K. Integrated approaches reveal determinants of genome-wide binding and function of the transcription factor Pho4. *Mol. Cell* **42**, 826–836 (2011).
36. Mahony, S. & Benos, P.V. STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic Acids Res.* **35**, W253–W258 (2007).
37. Spivak, A.T. & Stormo, G.D. ScerTF: a comprehensive database of benchmarked position weight matrices for *Saccharomyces* species. *Nucleic Acids Res.* **40**, D162–D168 (2012).
38. Wykoff, D.D. & O'Shea, E.K. Phosphate transport and sensing in *Saccharomyces cerevisiae*. *Genetics* **159**, 10 (2001).
39. Springer, M., Wykoff, D.D., Miller, N. & O'Shea, E.K. Partially phosphorylated Pho4 activates transcription of a subset of phosphate-responsive genes. *PLoS Biol.* **1**, E28 (2003).
40. Raveh-Sadka, T., Levo, M. & Segal, E. Incorporating nucleosomes into thermodynamic models of transcription regulation. *Genome Res.* **19**, 1480–1496 (2009).
41. Chiang, D.Y., Nix, D.A., Shultzaberger, R.K., Gasch, A.P. & Eisen, M.B. Flexible promoter architecture requirements for coactivator recruitment. *BMC Mol. Biol.* **7**, 16 (2006).
42. Janssens, H. *et al.* Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster* even skipped gene. *Nat. Genet.* **38**, 1159–1165 (2006).
43. Kim, H.D. & O'Shea, E.K. A quantitative model of transcription factor-activated gene expression. *Nat. Struct. Mol. Biol.* **15**, 1192–1198 (2008).
44. Mao, C. *et al.* Quantitative analysis of the transcription control mechanism. *Mol. Syst. Biol.* **6**, 431 (2010).

ONLINE METHODS

Promoter library and strain construction. Each component of the promoter library is a synthetic derivative of the *PHO5* promoter—the 800 bp upstream of the *PHO5* gene in *S. cerevisae*—with its Pho4 or Pho2 sites modified or ablated. We recently described the construction of these libraries[27]. In brief, the promoter was broken into nine 60- to 90-bp oligonucleotide pairs with 30-bp overlaps for synthesis. Promoter oligonucleotide pairs were phosphorylated and then ligated at 60 °C using 9°N ligase (New England Biolabs), followed by amplification and cloning into plasmid pBS34 (obtained from the Yeast Resource Center) carrying mCherry and a G418 resistance marker. Sequence-verified promoter constructs were amplified and transformed into the *LYS2* locus of yeast strain BY4741 by homologous recombination using the lithium acetate/polyethylene glycol (PEG) method[45]. We screened transformants for correct integration by semi-nested colony PCR. PCR-positive clones were retained for experiments after promoter sequence reconfirmation.

Pho4 site modifications were as described in the text. Pho4 ablations were designed from *in vitro* binding data and contained the lowest possible binding energies for each segment of a Pho4 binding site. We also avoided reconstituting weak Pho4 sites considering the local sequence. Pho2 ablations targeted known[25] and predicted binding sites[46,47]. As the Pho2 binding site is AT rich, we did not use ablations from the literature[25] to preserve nucleosome positioning. We disrupted each Pho2 site[47] and checked nucleosome positioning with the online prediction tool at the website of the Segal lab (http://genie.weizmann.ac.il/software/nucleo_prediction.html). We only retained ablations that did not substantially perturb nucleosome occupancy.

pho5Δ strains were constructed by replacing *PHO5* and its promoter with the *SpHIS5* gene from plasmid pKT210 (ref. 48) (obtained from EUROSCARF) or a selected promoter construct, as required. Deletion was verified by colony PCR of an amplicon spanning the 3′ end of either insert and a region 145 bp downstream of the *PHO5* locus.

Bulk library induction. Promoter library strains were grown for 26 h in 96-well plates in yeast extract peptone dextrose (YPD) (Sigma-Aldrich) containing 200 µg/ml G418 (Sigma-Aldrich) at 30 °C and were then diluted 30-fold and allowed to re-enter log phase in synthetic complete medium containing 10 mM P_i. Cells were washed twice in P_i-free synthetic medium[12] and diluted to a starting optical density (OD) of 0.1–0.2 in P_i-free medium in black 96-well plates with transparent bottoms (Nunc). mCherry fluorescence was measured every 6 min for 12–16 h on a plate reader (BioTek SynergyMx) and was normalized to cell number by dividing by the OD at each time point. Hill functions were fit to the normalized data, with the parameters of interest being the maximal output, F_{max}, and the time to half-maximal output, $t_{1/2}$.

On-chip culture, induction and analysis. Promoter library strains, along with mCherry-labeled Pho2 and Pho4 strains and a control, were grown overnight at 30 °C in YPD with 200 µg/ml G418 in 96-well plates with V-shaped wells (Nunc) sealed with a Breathe-Easy membrane (Sigma-Aldrich). The plates were spun down at 2,400 r.p.m. for 3 min before spotting. Strains were spotted on an epoxy-coated coverslip using a DNA microarrayer (Qarray2, Genetix) as a 24 × 48 array. The design and manufacture of the microchemostat chip and the imaging platform are described elsewhere (N.D. *et al.*, unpublished data). A polydimethylsiloxane (PDMS) chip was aligned to the coverslip so that each chemostat chamber was aligned with a spot corresponding to a library strain. After chip priming, chemostats were perfused with synthetic complete medium, and the cells were allowed to recover for 16–20 h. During this time, the library was imaged every 30 min with a low-magnification objective using the NIS-Elements viewer (Nikon Instruments) to monitor growth and identify problems (for example, abnormal growth, empty chambers or cross contamination) with individual chambers. Chambers with problems were discarded from the analysis. During the induction experiment, cells were imaged in bright-field and fluorescence modes. Images were taken every 20 min (the shortest interval possible for whole-chip imaging) for 1 h while growing in P_i-rich medium and during a 12-h induction period while in P_i-free medium. mCherry expression was imaged using a Texas Red filter.

Image data were stored and analyzed on a five-node IBM server with 50 TB of storage. An automated image analysis pipeline was used to extract single-cell fluorescence data (N.D. *et al.*, unpublished data). MATLAB scripts were used to separate each chamber per image. Single cells were segmented, parameterized and filtered on the basis of their size, eccentricity and proximity to the image boundaries. Outliers in each chamber's background-subtracted, single cell–fluorescence distribution were discarded. We processed data using the OpenPBS queue manager to speed up analysis, allowing us to analyze up to 40 data sets in parallel. Data for all promoters are provided in **Supplementary Data**.

Calculation of Pho4p binding-site energies and information content. Synthetic promoter binding-site energies were represented as the change in binding energy from the consensus, $\Delta\Delta G_i$. Assuming symmetrical binding, we calculated $\Delta\Delta G_i$ for the strongest 10-mer in a 24-base sequence centered on the modified binding site. $\Delta\Delta G_i$ was the sum of the *in vitro* binding energies of Pho4 to E-boxes and flanking bases identical to the site of interest[21].

Changes in the information content of each binding site were calculated using a position-weight matrix derived from the associated changes in binding energy for each single-base modification in a 10-base Pho4 binding site[49]. The information content, I_i, associated with each sequence position in the matrix is:

$$I_{\text{seq}}(i) = \sum_i f(b,i)\log_2\frac{f(b,i)}{P(b)}$$

where $f(b,i)$ is the frequency of base b at position i in the site, and $P(b)$ is the genomic frequency of base b in *S. cerevisiae*. $f(b,i)$ is in turn calculated from the binding energy $\Delta\Delta G_{b,i}$ of a site having base b at position i normalized over all bases:

$$f(b,i) = \frac{P(b,i)}{\sum_b P(b,i)}$$

where

$$P(b,i) = \frac{1}{1+e^{(\Delta\Delta G_{b,i}/RT)}}$$

where R is the gas constant and T is temperature. The total information content of a promoter's modified Pho4 site is calculated as:

$$\sum_i f(b,i)I_{\text{seq}}(i)$$

The quantity being summed is equivalent to the height of each base in a sequence logo representing the Pho4 binding site[50].

Modeling of transcription. Each promoter was assumed to have a basal transcription rate α under high $[P_i]$ and a transcription rate σ after induction with φ_{Pho4} that determine the relative contribution of each. φ_{Pho4} is the fraction of cells with Pho4 translocated into the nucleus at time t. Pho4 stays in the nucleus once translocated and does not oscillate (**Supplementary Figs. 17** and **18**). Assuming that promoter activation occurred after Pho4 translocation into the nucleus, we defined rate equations for mRNA (m), nascent mCherry (p) and mature mCherry (p_m) synthesis as follows:

$$\frac{dm}{dt} = \alpha(1-\varphi_{\text{Pho4}}) + \sigma\varphi_{\text{Pho4}} - \gamma_m - \delta_t m$$

$$\frac{dp}{dt} = \beta_p m - \beta_{pm} p - \gamma_p p - \delta_t p$$

$$\frac{dp_m}{dt} - \beta_{pm} p - \gamma_{pm} p_m - \delta_t p_m.$$

Nascent mCherry protein is synthesized with a rate β_p and matures to a fluorescent state with a rate β_{pm}. All products decrease by degradation (γ) and dilution by cell growth (δ_t) and, in the case of nascent protein, maturation.

doi:10.1038/ng.2729

Rate estimates were taken from the literature where possible or approximated as listed below.

mRNA degradation. $\gamma_m = 1.4 \times 10^{-2}$ min^{-1}. The reported half-life of PGK mRNA is $t_{1/2} = 70$ min, with $\gamma = 1 \times 10^{-2}$ min^{-1}, and the range of mRNA half-lives are $t_{1/2} \approx 3$–90 min, with $\gamma = 7.7 \times 10^{-3}$ to 2.3×10^{-1} min^{-1} (ref. 51). Similar ranges were reported in another study[52].

Protein degradation. $\gamma_p = \gamma_{pm} = 1.7 \times 10^{-3}$ min^{-1}. The reported range of protein half-lives is $t_{1/2} \approx 10^{-5}$ to 10^{-1} min^{-1} with a median value of 1.69×10^{-2} min^{-1} (ref. 53).

Translation rate. $\beta_p = 10$ amino acids (aa) s^{-1} per ribosome × 60 s per min × 1/237 proteins per aa = 2.53 proteins min^{-1}. We assumed that one mRNA molecule is occupied by only one ribosome at a time, which translates to 10 aa s^{-1} per mRNA (the reported range is 2.8–10 aa s^{-1} per ribosome[54]). A ribosome density of one ribosome per mRNA is slightly smaller than the expected average density of about one ribosome per 156 nt or 4.6 ribosomes per mCherry mRNA[55]. Using the higher translational rate of 10 aa s^{-1} per ribosome roughly offsets this.

Maturation rate. $\beta_{pm} = 6.7 \times 10^{-2}$ min^{-1} derived from literature values for mCherry (4.6×10^{-2} min)[56] and Venus (3.7×10^{-2} min^{-1})[57].

Dilution rate. $\delta_t = 1 \times 10^{-2}$ min^{-1}, based on 7×10^{-3} min^{-1} from the literature[58]. As the doubling time of 1×10^{-2} min^{-1} is 1.4 times faster than the observed doubling time, we repeated the model fits with a doubling time of $\delta_t = 7 \times 10^{-3}$ min^{-1}, which also gave rise to robust fits, the same relative transcription rates and slightly higher absolute transcription rates.

α and σ were estimated for each promoter using the 'fminsearch' function in MATLAB. α was determined independently of σ by first fitting to mCherry expression levels before induction using the same set of ODEs as described above but lacking σ and φ_{Pho4}. The value determined for α was then used in a second round of fits to determine σ. The traces in **Figure 2f** were generated using the 'ode15' function in MATLAB from the model parameters.

Promoter occupancy model. For the initial analysis, we used the following equations to calculate the binding and occupancy probabilities[33]:

$$P_i = \frac{1}{1 + e^{\beta \Delta\Delta G_i}}$$

$$P_{\text{occ}} = 1 - \prod_{i-1}^{\text{windows}} (1 - P_i)$$

$$\Delta\Delta G_i = (\Delta G_i - \Delta G_c)$$

where P_i is the probability that Pho4 is bound to site i, P_{occ} is the promoter occupancy, ΔG_i is the affinity to site i, and ΔG_c is the affinity to the consensus sequence. The product is calculated for all possible binding sites on the promoter (windows). All affinities were taken from the previous report by Maerkl and Quake[21]. β in all equations is $k_b T$, were k_b is the Boltzmann constant and T is temperature, or 0.592 kcal mol^{-1}.

Because the sequences of our synthetic promoters varied only at the exposed and nucleosomal Pho4 binding sites, we scanned 24 bp around these sites to determine the highest-affinity binding site in these regions, which simplified P_{occ} to the following:

$$P_{\text{occ}} = 1 - [(1 - P_e) \times (1 - P_n)]$$

where P_e and P_n are the binding probabilities for the exposed and nucleosomal sites of highest affinity, respectively. This simplified approach to determining P_{occ} was equivalent to calculating P_{occ} over all possible binding sites.

To include competition with other factors, we calculated binding probabilities on the basis of a statistical mechanics approach[34]. We calculated Pho4 promoter occupancy (P_{bound}) as follows:

$$P_{\text{bound}} = \frac{P_e^p + P_n^p + P_e^p P_n^p}{1 + P_e^p + P_n^p + P_e^p P_n^p}$$

for Pho4 binding alone, and

$$P_{\text{bound}} = \frac{P_e^p + P_n^p + P_e^p P_n^p}{1 + P_e^p + P_n^p + P_e^p P_n^p + P_e^c + P_n^c + P_e^c P_n^c}$$

for Pho4 binding with Cbf1 competition. We omitted terms such as $P_e^p P_n^c$ and so on for simplicity. This simplification is not expected to affect the computed values or decrease the prediction accuracy, as the Cbf1 contribution is low and the contribution of these higher terms is also low.

Thus Pho4 binding with Cbf1 competition and a second specific competitor is calculated as:

$$P_{\text{bound}} = \frac{P_e^p + P_n^p + P_e^p P_n^p}{1 + P_e^p + P_n^p + P_e^p P_n^p + P_e^c + P_n^c + P_e^c P_n^c + P_{\text{comp}}}$$

With the individual weights given by:

$$P_i^p = \left(\frac{P}{NS}\right) e^{\beta(\Delta G_{ns}^p - \Delta G_i^p)}$$

$$P_i^c = \left(\frac{C}{NS}\right) e^{\beta(\Delta G_{ns}^c - \Delta G_i^c)}$$

$$P_{\text{comp}} = \left(\frac{S}{NS}\right) e^{\beta \Delta G_s}$$

where ΔG_{ns}^p and ΔG_{ns}^c are the nonspecific (ns) binding affinities of Pho4 (p) and Cbf1 (c), respectively, of −6.82 kcal mol^{-1}. ΔG_i^p and ΔG_i^c are the specific binding affinities of Pho4 and Cbf1 to sequence i, respectively, where i is the exposed (e) or nucleosomal (n) binding site. The numbers of Pho4 (P) and Cbf1 (C) molecules per cell were estimated from single-cell fluorescence values of GFP-tagged Pho4 and Cbf1 strains and were 622 and 1,422 proteins per cell, respectively. NS is the number of nonspecific binding sites available for binding in the genome and was set to 405,000. This value is one order of magnitude lower than a rough estimate based on nucleosome-free DNA in the yeast genome[59] but gave rise to identical values when compared to the simplified P_i equation described above. As the specific competitor is not known, we estimated S to be 1,215 proteins per cell and used a ΔG_s of 4.1 kcal mol^{-1}.

To include nucleosome remodeling in our model, we calculated the probability of the nucleosome being remodeled as follows:

$$P_{\text{rem}} = \frac{e^{\beta \Delta G_{\text{nuc}}}}{1 + e^{\beta \Delta G_{\text{nuc}}} + \left(\frac{P}{NS}\right) e^{\beta(\Delta G_n^p + \Delta G_i^p)} + \left(\frac{C}{NS}\right) e^{\beta(\Delta G_n^c - \Delta G_i^c)}}$$

Remodeling depends primarily on the exposed Pho4 site affinity. But as we observed that high-affinity nucleosomal Pho4 sites can also lead to remodeling (or at least destabilize the nucleosome) we calculated P_{rem} using either the exposed or nucleosomal Pho4 binding site, depending on the affinity of the exposed Pho4 binding site. As the wild-type Pho4 binding site (affinity: $\Delta\Delta G$ = 2.86473 kcal mol^{-1}) is capable of full promoter induction and, thus, remodeling, we calculated P_{rem} on the basis of the exposed site affinity if its value was above 2.86474 kcal mol^{-1} or the nucleosomal site affinity if the exposed site affinity was below this threshold. ΔG_{nuc} was set to 4.3 kcal mol^{-1}. Inclusion of P_{rem} in the P_{bound} calculation leads to the following:

$$P_{\text{bound + rem}} = P_{\text{bound}} P_{\text{rem}}$$

Adding Cbf1 as a specific competitor had only a minor effect on the predicted promoter occupancy (**Supplementary Fig. 11b**), which was an expected result, as Pho4 dominates under full induction conditions and Cbf1 competition seems to affect GTCACGTGAC binding sites only. Adding nucleosome remodeling affected primarily promoters with a low-affinity exposed site and a wild-type nucleosomal site (**Supplementary Fig. 11c**). These promoters gave rise to aberrantly high occupancy estimates when using a standard P_{occ} or P_{bound} model, as the standard models assume that all sites are equally accessible.

A model combining Cbf1 competition and nucleosomal remodeling predicted most of the promoters accurately, except for a group of 22 promoters that showed lower *in vivo* output than predicted, and one promoter showed higher-than-predicted output (B7W-NNNN1) (**Supplementary Fig. 11c**). The 22 promoters whose outputs were too low fell within a class of promoters in which either the left-flanking or right-flanking bases were modified. In these promoters, we included upstream and downstream stretches of TTT and AAA, respectively, to insulate the binding site and prevent the formation of secondary high-affinity Pho4 binding sites. Furthermore, only a subclass of these promoters in which the first flanking base was predominantly either A or G (19/22 promoters) showed lower-than-expected output values, whereas the remaining promoters of the same class behaved as expected. We thus hypothesized that the inclusion of the insulating stretches in combination with specific sequence motifs could give rise to a binding site that is recognized by one or more yeast transcription factors other than Pho4 and Cbf1. This hypothesis was corroborated by the apparent sequence specificity of the promoter subset and the fact that the promoters in question had higher basal fluorescence levels (**Supplementary Figs. 5b** and **11**).

We ascertained the correctness of these promoters by resequencing. Additionally, we verified that the difference in output was not due to experimental variation between experiments by comparing a set of promoters analyzed across experiments, which correlated with $R^2 = 0.97$ and a slope of 1.045 (**Supplementary Fig. 12**). Experimental variation of ~5% (even with a conservative margin of 10–20%) does not explain the observed difference in output of ~200%. Finally, because not all promoters containing TTT and AAA insulating stretches were affected and because both nucleosomal and exposed sites were equally affected, we also excluded the possibility that failure of Pho2 binding was the reason for the decreased output. We thus recalculated P_{stat} for promoters with a basal fluorescence value above 46 AU and included a specific competitor that could bind to the single modified site. By including this specific competitor, we were able to accurately predict output for most of the promoters for which the basal rate was higher than for the rest of the synthetic promoter library (**Supplementary Fig. 11d**,**e**).

To test whether any yeast transcription factor with known sequence motifs could bind to our hypothesized motif in the 22 promoters with low output, we searched the STAMP[36] database using the consensus upstream motif of RATTTRNCAC as the input (where R is either A or G). STAMP returned three possible matches: Rpn4 (**WTTTGCCAC**CN), Spt23 (Y**TTTANN**T) and Mac1 (**WTTTGC**T). A search of ScerTF[37] with a motif of AATTTACCAC produced Stb3 (AATT**TTTCAC**T) and Ypr015c (GG**ATTTAC**G) with an input string of ATTTA. Bases colored red are complementary to the promoter binding site. It is thus feasible that one or more of these transcription factors binds to these specific binding sites and competes with Pho4 binding, leading to lower-than-expected output.

Only six promoters could not be accurately predicted: LF8, LF6w-G4, RF11w, RF12w, RF15w and RF16w. LF8 and LF6w-G4 had higher-than-normal basal output, but not high enough for these promoters to be recalculated with a model incorporating the specific competitor (in which case they would have been correctly predicted as well). The RF promoters have either a TAT or TTT left-flanking base sequence. We removed these six refractory promoters from our final linear regression fit.

The one promoter that showed higher-than-predicted output, B7w-NNNN1, had a split binding-site motif in the nucleosomal position. We improved our prediction of this promoter by taking into account that the split binding motif generates two low-affinity binding sites.

Modeling phosphate uptake. Two-dimensional numerical simulations for phosphate consumption were carried out using finite element analysis software (COMSOL Multiphysics 4.2). The transport of diluted species interface (*chds*) was used to calculate the phosphate concentration along the length of the microfluidic channel. We considered a rectangle 250-µm wide by 1-µm high to account for the cell packing density inside the microfluidic channels. The rectangle was discretized in triangular meshes with a maximum element size of 0.5 µm. The diffusion coefficient and affinity constant (K_m) for the phosphate rate were set to 10^{-9} m^2 s^{-1} and 0.058 mol m^{-3}, respectively[38,60]. The reaction rate of phosphate uptake was modeled as a first-order reaction:

$$\frac{V_{max}[P_i]}{[P_i] + K_m}$$

where V_{max} is the maximum rate and $[P_i]$ is the concentration of phosphate in the channel. The boundary conditions consisted of no flux through the length of the rectangle and the same constant concentration at the sides. A parametric sweep study and a time-dependent study were carried out to obtain the phosphate concentration along the channel.

Measuring P_i activation concentrations. We chose to analyze chambers closest to the feed inlet for a given promoter to minimize any decreases in $[P_i]$ across a chip. We binned and averaged single-cell endpoint fluorescence for every 10 µm from the feed inlet. We fit Hill functions to the position-dependent data, taking the distance to half-maximum fluorescence to be an estimate of the location of the fluorescent front. We then calculated the correlation coefficient and residual sum of squares between the fronts at various $[P_i]$ and the estimated front positions at different P_i uptake rates assuming activation $[P_i]$ values between 10 and 40 µM and $y = x$. The $[P_i]$ whose fronts correlated best with the measured fronts was chosen to be the activation concentration.

Mapping of nucleosome positions. Nucleosome positioning was measured by quantitative PCR (qPCR) on a synthetic wild-type promoter integrated at the *PHO5* or *LYS2* locus using modification of an existing method[12]. The latter strains lacked *PHO5* and its promoter to prevent target contamination with native promoter DNA. In brief, a 125 ml culture of the strain of interest at the mid-log phase was harvested, spheroplasted using zymolyase (Amsbio) and digested with 56–600 U micrococcal nuclease (ThermoFisher). The reaction was carried out for 20 min at 37 °C and stopped with 5% sodium dodecyl sulfate (SDS) in 250 mM ethylenediaminetetraacetic acid (EDTA). Nucleosome-protected DNA was purified after treatment with 0.5 mg/ml proteinase K (Sigma-Aldrich; 3 h at 37 °C) by phenol-chloroform extraction and treatment with 0.1 mg/ml RNase (Qiagen; 2 h at 37 °C) before analysis by agarose gel electrophoresis. Samples showing maximal mononucleosome DNA enrichment at ~150 bp were gel purified using the QIAquick Gel Extraction kit.

Nucleosome positions were mapped by qPCR on purified mononucleosome DNA using 22 pairs of primers tiled ~25 bp on average to cover 600 bp of the promoter and the first 50 bases of mCherry or *PHO5*. qPCR was carried out on an ABI 7500 Fast Cycler with Power SyBR Green PCR Mix (Life Technologies) as per the manufacturer's recommendations. qPCR values were normalized to a positive control targeting the *GAL4* locus.

45. Gietz, R.D. & Schiestl, R.H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 31–34 (2007).
46. MacIsaac, K.D. *et al.* An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* **7**, 113 (2006).
47. Zhu, C. *et al.* High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res.* **19**, 556–566 (2009).
48. Sheff, M.A. & Thorn, K.S. Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. *Yeast* **21**, 661–670 (2004).
49. Schneider, T.D. Information content of individual genetic sequences. *J. Theor. Biol.* **189**, 427–441 (1997).
50. Schneider, T.D. & Stephens, R.M. Sequence logos:a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990).
51. Wang, Y. *et al.* Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. USA* **99**, 5860–5865 (2002).
52. Miller, C. *et al.* Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol. Syst. Biol.* **7**, 458 (2011).
53. Belle, A., Tanay, A., Bitincka, L., Shamir, R. & O'Shea, E.K. Quantification of protein half-lives in the budding yeast proteome. *Proc. Natl. Acad. Sci. USA* **103**, 13004–13009 (2006).
54. Karpinets, T.V., Greenwood, D.J., Sams, C.E. & Ammons, J.T. RNA:protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis. *BMC Biol.* **4**, 30 (2006).
55. Arava, Y. *et al.* Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **100**, 3889–3894 (2003).
56. Shaner, N.C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572 (2004).
57. Sinha, S., van Nimwegen, E. & Siggia, E.D. A probabilistic method to detect regulatory modules. *Bioinformatics* **19**, i292–i301 (2003).
58. Di Talia, S., Skotheim, J.M., Bean, J.M., Siggia, E.D. & Cross, F.R. The effects of molecular noise and size control on variability in the budding yeast cell cycle. *Nature* **448**, 947–951 (2007).
59. Mavrich, T.N. *et al.* A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* **18**, 1073–1083 (2008).
60. Crank, J. *The Mathematics of Diffusion* (Clarendon Press, Oxford, 1986).

doi:10.1038/ng.2729