Molecular Systems Biology 8; Article number 576; doi:10.1038/msb.2012.7
Citation: ***Molecular Systems Biology*** **8:576**

www.molecularsystemsbiology.com

molecular systems biology

A regulatory role for repeated decoy transcription factor binding sites in target gene expression

Tek-Hyung Lee and Narendra Maheshri*

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
* Corresponding author. Department of Chemical Engineering, Massachusetts Institute of Technology, Room 66-558 25 Ames Street, Cambridge, MA 02139, USA. Tel.: +1 617 258 8986; Fax: +1 617 258 8224; E-mail: narendra@mit.edu

Received 10.1.12; accepted 27.2.12

Tandem repeats of DNA that contain transcription factor (TF) binding sites could serve as decoys, competitively binding to TFs and affecting target gene expression. Using a synthetic system in budding yeast, we demonstrate that repeated decoy sites inhibit gene expression by sequestering a transcriptional activator and converting the graded dose–response of target promoters to a sharper, sigmoidal-like response. On the basis of both modeling and chromatin immunoprecipitation measurements, we attribute the altered response to TF binding decoy sites more tightly than promoter binding sites. Tight TF binding to arrays of contiguous repeated decoy sites only occurs when the arrays are mostly unoccupied. Finally, we show that the altered sigmoidal-like response can convert the graded response of a transcriptional positive-feedback loop to a bimodal response. Together, these results show how changing numbers of repeated TF binding sites lead to qualitative changes in behavior and raise new questions about the stability of TF/promoter binding.
Molecular Systems Biology **8**: 576; published online 27 March 2012; doi:10.1038/msb.2012.7
Subject Categories: chromatin & transcription
Keywords: bimodality; tandem repeats; transcription factor decoys; transcriptional regulation

Introduction

The genomes of many organisms contain long tracts of repetitive nucleotide sequences known as tandem repeats (TRs) of DNA. Over 45% of the human genome is repeated sequence, mostly found in non-coding regions (Lander *et al*, 2001). While sometimes discounted as 'junk' DNA, TR length has been implicated in a number of different phenotypes and diseases. When TRs occur within the open reading frame of genes, their expansion/contraction directly affects protein structure or expression. For example, TRs within yeast adhesin genes can influence their adhesive and flocculent properties (Verstrepen *et al*, 2005), TRs within the *Runx-1* transcription factor (TF) gene in dogs dictate skull morphology (Fondon and Garner, 2004), and changes in TR number in contingency loci in many prokaryotes switch expression state by introducing frameshifts (Rando and Verstrepen, 2007). TRs within intergenic regions that are close to genes are also widely implicated in affecting gene expression. Expansion of trinucleotide repeats in untranslated regions or introns of genes has a causative role in triplet expansion diseases (Cummings and Zoghbi, 2000) often by silencing gene expression. Recent work in budding yeast demonstrates that TRs within promoters can influence gene expression by altering nucleosome structure or the number of TF binding sites (Vinces *et al*, 2009). Importantly, because variation in TR number is 100- to 1000-fold higher than single point mutation rates (Rando and Verstrepen, 2007), TRs represent an evolutionary reservoir of potential diversity. Indeed, the majority of spontaneous mutations in budding yeast are associated with repeated regions (Lynch *et al*, 2008).

Bioinformatic studies have found that many TRs in non-coding regions contain known TF binding sites (Horng *et al*, 2003); whether these sequences have functional roles remain unclear. One potential role for these TRs would be to serve as decoys, competitively binding the cognate TF and thereby influencing expression of target promoters. In mice, the major α-satellite TRs within pericentromeric heterochromatin contain binding sites for C/EBPα. These TRs sequester C/EBPα, leading to a reduction in gene expression at target genes of this activator (Liu *et al*, 2007). The ability of decoy binding sites in TRs to bind a TF could depend on chromatin-mediated accessibility. For example, in *Drosophila* the addition of drugs that increase accessibility to the heterochromatic GAGAA repeat within satellite V leads to increased sequestration of the GAGA factor and reduced expression of target genes (Janssen *et al*, 2000).

Simple kinetic models can clarify how the strength of protein/DNA interactions and protein stability impacts the function of repeated decoy TF binding sites on target gene expression. An intuitive notion is that decoy sites serve as competitive inhibitors, reducing the TF available to bind to target promoters. However, non-equilibrium models that include production and degradation of the TF demonstrate this is not always true. Previous theoretical work highlights the fact that if the degradation rate of the TF/decoy complex is

much lower than the unbound TF, the steady-state levels of unbound TF are independent of the presence of decoys, with no resulting effect on target gene expression (Burger *et al*, 2010). A second key parameter that influences the dose–response of target genes in the presence of decoys is the ratio of the affinity of TF/decoy and TF/promoter binding. If TF/decoy affinity is much higher, then as TF levels increase target gene expression is unchanged until all decoy sites are saturated. For a transcriptional activator, this leads to an increase in concavity of the dose–response curve between the TF and the target promoter and a sharper, sigmoidal-like threshold response (Buchler and Louis, 2008).

Here, we construct and model a synthetic system in budding yeast to quantitatively analyze the effect of TRs of decoy binding sites on target gene expression. We find that repeated decoy sites do decrease target gene expression. Furthermore, the dose–response is qualitatively altered from a graded to a sharper threshold response. Interpreted in the context of our model, these results indicate that TFs bind to repeated decoy binding sites more strongly than to the promoter. This surprising implication is supported by chromatin immunoprecipitation (ChIP) assays, which monitor TF occupancy in both regions. Moreover, we confirm the functional relevance of the altered dose–response by demonstrating the ability of decoys to change the graded dose–response of a positive-feedback loop to a bimodal response. Our results show how TRs of decoy sites have qualitative effects on gene expression and network behavior that depend on decoy site number. They also raise questions about the strength of activator/promoter interactions that lead to expression.

Results

Modeling the effects of decoy sites on target gene expression

To describe the effects of an array of repeated decoy sites on target gene expression, we consider the chemical transformations illustrated in Figure 1A and stated here:

$$\begin{aligned}
&\phi \xrightarrow{s} T\\
&T \xrightarrow{\delta} \phi\\
&TN \xrightarrow{\delta_N} \phi\\
&TP \xrightarrow{\delta_P} \phi\\
&T + N \underset{k^N_{off}}{\overset{k^N_{on}}{\rightleftharpoons}} TN\\
&T + P \underset{k^P_{off}}{\overset{k^P_{on}}{\rightleftharpoons}} TP
\end{aligned} \tag{1}$$

The TF (T) is a transcriptional activator which is produced constitutively and can potentially bind to either decoy sites (N) or the promoter (P). Species balances for free and bound forms are:

$$\begin{aligned}
T_0 &= T + TP + TN\\
N_0 &= N + TN\\
P_0 &= P + TP
\end{aligned} \tag{2}$$

Differential equations are formulated for T, TP and TN in the Supplementary information. While not explicitly shown here, the DNA corresponding to decoys (N) and the promoter (P) are also synthesized and diluted as cells grow. These processes and the synthesis and degradation of the unbound TF (T) are slow compared to fast binding and unbinding of the TF to the promoter or decoy sites (order 10's of minutes versus 10's of seconds—see Supplementary information for details). For now, we assume that decoy and promoter-bound TFs (TN and TP) degrade at rates identical to the unbound TF (T).

Because there are few promoter sites compared with the number of free T and total decoy N_0 sites, we neglect the TP complex in the species balance for T_0, leading to the following expression for the free T sites, where $K_N = k^N_{off}/k^N_{on}$:

$$\frac{T}{T_0} = \frac{\left(1 - \frac{K_N}{T_0} - \frac{N_0}{T_0}\right) + \sqrt{\left(1 - \frac{K_N}{T_0} - \frac{N_0}{T_0}\right)^2 + 4\frac{K_N}{T_0}}}{2} \tag{3}$$

We assume gene expression is proportional to TP/P_0, the TF occupancy at the promoter (Bintu *et al*, 2005):

$$\frac{(TP)}{P_0} = \frac{T/T_0}{\frac{T}{T_0} + \frac{K_P}{T_0}} \tag{4}$$

where $K_P = k^P_{off}/k^P_{on}$. We will use concentration units of molecules per yeast nuclear volume (n.v.). Using Equations (3) and (4), we plot the dose–response of TF occupancy to $T_0 = T + TN$ when T_0 is varied by changing the synthesis rate, S. Because $T_0 = S/\delta$, choosing to plot TF occupancy versus T_0 or S are equivalent within a constant. When $K_N/K_P = 1$, increasing decoy number N_0 decreases target gene expression but does not change the shape of the dose–response curve (Figure 1B and Supplementary Figure 1A). However, the shape changes when decoy sites have much higher affinity ($K_N/K_P \ll 1$). As TF level increases, they bind and saturate decoy sites before leading to gene expression (Figure 1C and Supplementary Figure 1B). The result is a sharper, threshold dose–response that has previously been discussed (Buchler and Louis, 2008).

We can also consider the non-equilibrium effects of varying δ_N/δ, assuming $\delta = \delta_P$. In the first case, decoy-bound TF is protected from degradation and $\delta_N/\delta \ll 1$. When $K_N/K_P = 1$, faster turnover of unbound T decouples its level from the decoy-bound TN species, making it invariant to changes in decoy number N_0 (Burger *et al*, 2010). However, increasing N_0 does affect T_0 as the decoy-bound species TN increases at any given synthesis rate S. Therefore, the dose–response curve of TF occupancy versus T_0 is altered when decoys are added, whereas TF occupancy never changes with S (Supplementary Figure 1C). TF occupancy also never changes with S when $K_N/K_P \ll 1$, although now there is the sharper, threshold dose–response versus T_0 (Supplementary Figure 1D). In the second case, where $\delta_N/\delta \gg 1$, the dose–response is nearly identical to the case of $\delta_N = \delta_P$, but much larger changes in S are required to increase occupancy as the decoy-bound species degrades quickly (Supplementary Figure 1E and F).

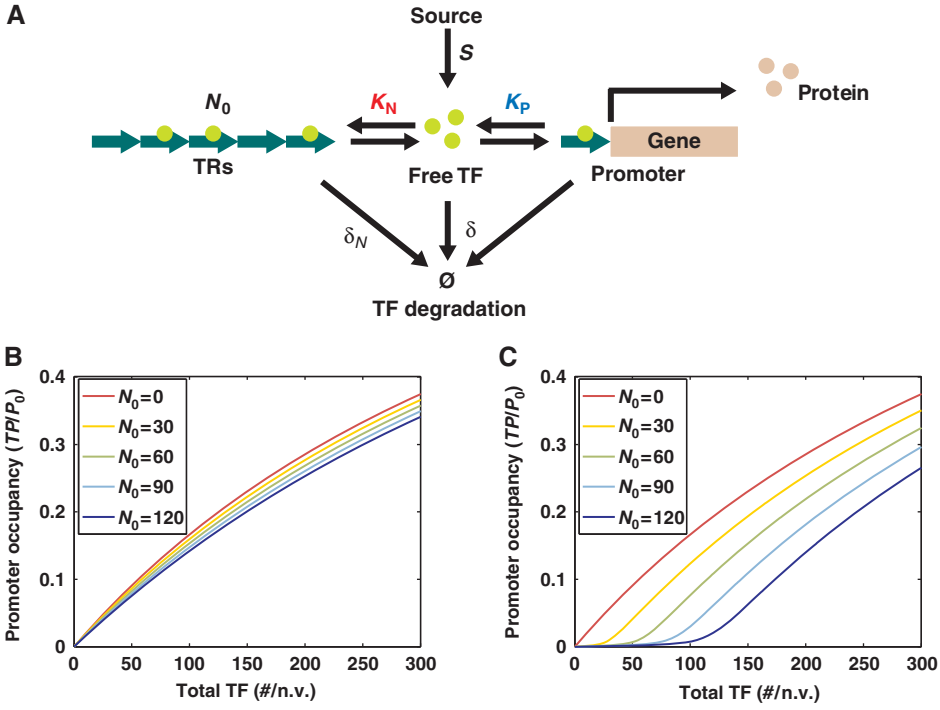

Figure 1 A simple model predicts that an array of decoy binding sites qualitatively alters the dose–response of a TF and target promoter depending on the strength of the TF/binding site interaction. (**A**) A simple model to describe the effects of TRs containing decoy binding sites on target gene expression. Important parameters include the number (N_0) and binding affinity ($1/K_N$) of decoy sites, the binding affinity ($1/K_P$) of promoter binding sites, and the production and degradation of each species. Details are described in the text. (**B**) Model predicted dose–response of expression versus total TF level, T_0, for various numbers, N_0, of decoy sites when the binding affinity of the TF for decoy and promoter sites are identical ($K_N = K_P = 500$). Decoy sites reduce expression but do not change the graded nature of the response. (**C**) As in B, but with promoter binding affinity set much lower than decoy binding affinity ($K_N = 1$, $K_P = 500$), which results in a more sigmoidal-like dose–response.

Contiguous arrays of tetO sites decrease target gene expression and lead to a sharper, sigmoidal-like response

We used the synthetic tet-OFF system, adapted for budding yeast (Gari *et al*, 1997), to measure the effect of TRs of decoy binding sites on target gene expression. Here, the tet-transcriptional-activator (tTA) binds specifically to the tet operator (tetO), and activates gene expression of a 7 × tetO promoter driving a downstream fluorescent reporter integrated at the *URA3* locus. We integrated a *MYO2* promoter driving tTA expression at the *ADE2* locus and introduced arrays containing various numbers of contiguous tetO sites on either single copy centromeric plasmids, high copy 2 μm plasmids, or by genomic integration at the *HIS3* locus (Figure 2A). The tetO arrays were derived from a 9-kb non-recombinogenic tetO array containing 240 × tetO binding sites spanned by 10 or 30 bp of random DNA sequence (a kind gift from D Sherrat; Lau *et al* (2003); Figure 2A). We verified array stability over the course of experiments (Supplementary information).

To measure the dose–response curve of the 7 × tetO promoter, we varied active tTA levels by abrogating the tTA–tetO interaction using doxycycline (dox) and monitored fluorescent reporter expression in exponentially growing cells. In Figure 2B, we compare the dose–response curve in the absence and presence of centromeric plasmid-borne tetO arrays. Adding decoy sites decreases expression at any given level of dox. On the basis of our understanding of the tTA–dox interaction (detailed below), varying dox is equivalent to changing the tTA synthesis rate. Therefore, the decreased expression implies that decoy-bound tTA is not protected from degradation and δ_N/δ cannot be much less than unity. We further verified that the decoy array reduces expression at a given tTA synthesis rate by placing tTA expression under the control of the methionine-inducible *MET3* promoter (Supplementary Figure 2). The simplest interpretation of these results is that decoy-bound and unbound tTA have the same degradation rate and we set $\delta_N = \delta$. We cannot exclude the possibility that $\delta_N \gg \delta$, but this does not change inferences about promoter and decoy binding strength (Supplementary Figure 1).

For an accurate picture of the dose–response curve, one needs a model to translate an experimentally set external dox concentration to an active tTA level. To do so, we extended a previously reported and experimentally verified model of the dox–tTA interaction (Murphy *et al*, 2007; To and Maheshri, 2010). Key features of this model are (1) a constant flux of dox enters cells resulting in the intracellular dox concentration being linearly proportional to the external dox concentration, (2) two dox molecules bind to each tTA dimer in a non-cooperative manner to abrogate its DNA binding capability and (3) free, promoter-, decoy- and dox-bound tTA equilibrate on timescales faster than tTA degradation (~15 min half-life—To and Maheshri, 2010). Adding these interactions, the following two expressions can be used to find the fraction of unbound tTA

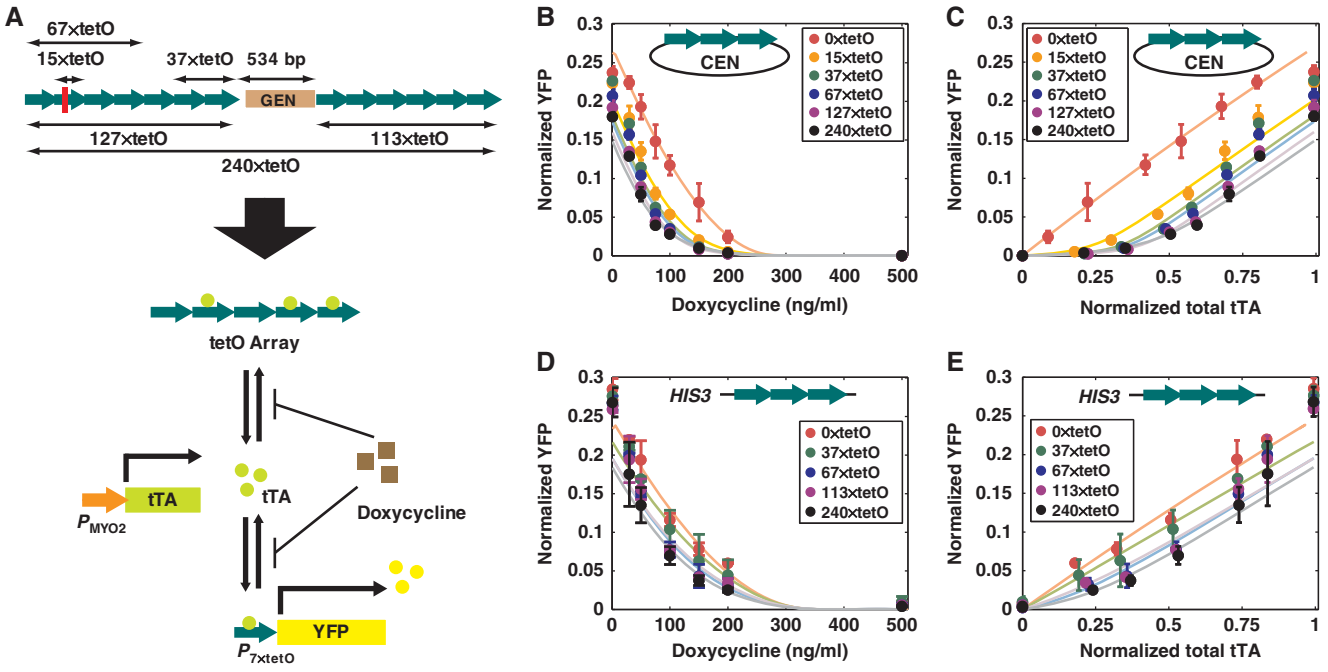

Figure 2 Arrays of tetO decoy sites reduce target gene expression and convert the graded dose–response between tTA and its target promoter to a sigmoidal-like response. (**A**) tTA is expressed constitutively from a chromosomally integrated *MYO2* promoter at the *ADE2* locus, and its activity is titrated by addition of dox. Activation of a tTA-responsive 7 × tetO promoter driving YFP integrated at the *URA3* locus was monitored by flow cytometry. Arrays of tetO binding sites of various sizes were created from a single 240 × tetO array. (**B**) Target gene expression at the 7 × tetO promoter is reduced at any dox concentration when tetO arrays are introduced on a centromeric plasmid. Dots represent experimental data, and solid lines correspond to fits of the kinetic model described in the main text to six data points at low tTA levels. (**C**) Using a model that accounts for dox–tTA interactions, the expression data in B can be plotted versus total tTA number (unbound + decoy-bound tTA). The arrays result in a qualitatively sharper change in the dose–response indicative of stronger binding of tTA to the array versus the promoter. (**D**, **E**) tetO arrays have a qualitatively similar, albeit weaker, effect when integrated at the *HIS3* locus in the chromosome. Error bars represent s.d. of 2 biological replicates. Source data is available for this figure in the Supplementary Information.

(T/T_0) and the free internal dox concentration (x):

$$\frac{T}{T_0} \approx \frac{x_0 / T_0}{2\frac{x}{K_s}\left(1+\frac{x}{K_s}\right)} \tag{5}$$

$$\frac{T}{T_0} = \left(1 + 2\frac{x}{K_s} + \left(\frac{x}{K_s}\right)^2\right)^{-1} \tag{6}$$

Here K_S is the thermodynamic affinity of tTA to dox, which we estimate as the previously reported affinity of tetR to dox of 0.21/n.v. (we assume a yeast n.v. of 5 μm^3; Degenkolb *et al*, 1991). The total dox concentration (x_0) is proportional to the external dox concentration, with a free parameter, K_M (a membrane partition coefficient for dox) that must be fit. We use the following expression, based on Equation (4), to relate T/T_0 to fluorescent reporter expression:

$$FP - FP_{\min} = (FP_{\max} - FP_{\min}) \times \frac{(TP)}{P_0} = k_{\max} \frac{T/T_0}{\frac{T}{T_0} + \frac{K_P}{T_0}} \tag{7}$$

FP is the measured fluorescent reporter expression, $FP_{\min}$ is the basal expression in the absence of TF, $FP_{\max}$ is the maximum expression, and $k_{\max} = FP_{\max} - FP_{\min}$. In previous work, we have established this model, with a 'Hill coefficient' of 1, for the 7 × tetO promoter (To and Maheshri, 2010). We can estimate the $FP_{\max}$ by measuring the expression of the promoter in positive feedback, and $FP_{\min}$ by measuring expression in strains without tTA or subject to very high dox levels (To and Maheshri, 2010).

The CFP/YFP fluorescent signals reported are normalized with respect to CFP/YFP signals measured in a yeast strain constitutively expressing the fluorescent protein from an *ADH1* promoter integrated at the *LEU2* locus. This allows direct comparison of fluorescent signals irrespective of fluorophore and method of measurement. Finally, when tTA expression is driven from the weak *MYO2* promoter, steady-state levels (T_0) are low enough that $T_0 < K_P$, and the dose–response is always in the linear range (To and Maheshri, 2010). This is confirmed when we use Equations (5)–(7) to fit the dose–response data in the absence of decoys, by varying two free parameters, K_P/T_0 and K_M/T_0.

Equation (6) can be modified to fit the dose–response curves in the presence of decoy arrays:

$$\frac{T}{T_0} = \left(1 - (N_0/T_0)\frac{T/T_0}{T/T_0 + K_N/T_0}\right) \times \left(1 + 2\frac{x}{K_s} + \left(\frac{x}{K_s}\right)^2\right)^{-1} \tag{8}$$

By using (5), (7) and (8) with estimated values for K_M/T_0 and K_P/T_0, we fit two new parameters: N_0/T_0 and K_N/T_0. We report K_M/T_0, N_0/T_0 and K_N/K_P in Table I.

Two features of the fitting procedure deserve mention. First, we find estimates of K_M/T_0 across different data sets are similar, as expected for a property that is independent of decoy number. Second, because both changing N_0/T_0 and K_N/T_0 can

Table I Fit parameter estimates

tetO Array location	$\frac{K_M}{T_0}$ (ng/ml)		Number of tetO sites within array				
Centromeric plasmid	0.0071		15	37	67	127	240
		$\frac{K_N}{K_P}$[a]	0.002	0.001	0.001	0.001	0.002
		$\frac{N_0}{T_0}$	0.28 (1.1)[b]	0.35 (1.7)	0.38 (2.0)	0.43 (2.6)	0.49 (3.0)
		fold[c]	4.5	5.8	7.4	9	8.6
Integrated (*HIS3*)	0.0058		15	37	67	113	240
		$\frac{K_N}{K_P}$	—	0.08	0.01	0.02	0.007
		$\frac{N_0}{T_0}$	—	0.25 (0.29)	0.25 (0.71)	0.28 (0.65)	0.29 (0.97)
		fold	—	1	1.9	1.4	3.2
Multiple centromeric plasmids	0.0066		67	67/67	67/67/67		
		$\frac{K_N}{K_P}$	0.004	0.0004	0.0002		
		$\frac{N_0}{T_0}$	0.23 (0.76)	0.47 (3.5)	0.57 (7.5)		
		fold	1.9	6.6	5.2		
2-μm Plasmid	0.0087		Bin 2	Bin 3	Bin 4		
		$\frac{K_N}{K_P}$	0.008	0.002	0.0008		
		$\frac{N_0}{T_0}$	0.21 (0.73)	0.30 (1.6)	0.36 (2.5)		
		fold	1.6	2.6	3.5		

[a]The reported K_N/K_P was chosen such that its sum squared of residuals (SSRs) is only 1.25 times higher than the K_N/K_P found at the global minima, which is a good heuristic for the plateau region in Supplementary Figure 3.
[b]The parenthetical value for N_0/T_0 corresponds to the best estimate when $K_N/K_P = 0.1$.
[c]'Fold' represents the fold change of the SSR for $K_N/K_P = 0.1$ versus the SSR for the reported K_N/K_P.

reduce target gene expression, we analyzed the covariance between these two parameters. This is best seen in Supplementary Figure 3, where the sum of the squared residuals (SSRs) of the fit is given for various values of these two parameters. For any given K_N/T_0, there is a narrow range of N_0/T_0 that results in a good fit. In contrast, we found a range of K_N/T_0 spanning several orders of magnitude results in a good fit. The SSR in this range is plotted in Supplementary Figure 3 for various values of K_N/K_P. The minimum SSR value corresponds to a low $K_N/K_P \sim 10^{-6}$, but lies within a shallow plateau region. Because such a large change in affinity results in a physically nonsensical residence time for the TF, assuming a diffusion-controlled on-rate (see Supplementary information), we used the upper-bound of the plateau region as our estimate K_N/K_P. We defined this heuristically as when the SSR changes by less than 25%, leading to a reported K_N/K_P values that generally lie between 10^{-2} and 10^{-3} for all experiments (Table I). These values should be considered as order of magnitude estimates and clearly suggest a large difference in the strength of tTA binding to decoy sites versus productive binding events at the promoter exists. In Supplementary Figure 3, we compare these fits to a case where $K_N/K_P = 10^{-1}$, which does not describe the data well.

The model captures the experimental data, except at the two highest levels of tTA expression, where it systematically overpredicts the extent to which decoy sites decrease expression. This trend persists whether the tetO arrays are present on a centromeric plasmid (Figure 2B and C), integrated in the genome (Figure 2D and E) or present on a high copy plasmid (Supplementary Figure 4). At higher tTA levels, there is a decrease in the gap between target gene expression of strains with and without decoys. This feature cannot be explained for any choice of physical parameters by our model. The decreased gap implies either the decoys release bound tTA at higher tTA levels, increased array occupancy promotes gene expression at the promoter by an unknown mechanism or total tTA levels change in the presence of decoy sites at low dox levels (see Supplementary information for further discussion). The decreased gap is not dependent on our dox model; it remains even if the data are plotted as a function of dox rather than the total TF level.

Because of the shape of the dose–response curve, the model predicts $K_N/K_P \ll 1$. Describing the tTA–tetO interaction using a single thermodynamic affinity is likely a gross simplification. In reality, the residence time of tTA to either the promoter or decoy sites depends on a combination of (1) interactions with multiple tetO binding sites, (2) the conformation and chromatin state of the DNA, (3) interactions of tTA with other proteins that may also have affinity for DNA (general transcriptional machinery, chromatin remodeling factors and so on) and (4) other unknown factors. Some of these interactions are not at thermodynamic equilibrium. Nevertheless, the model serves as a useful framework for pinpointing what interactions must be different.

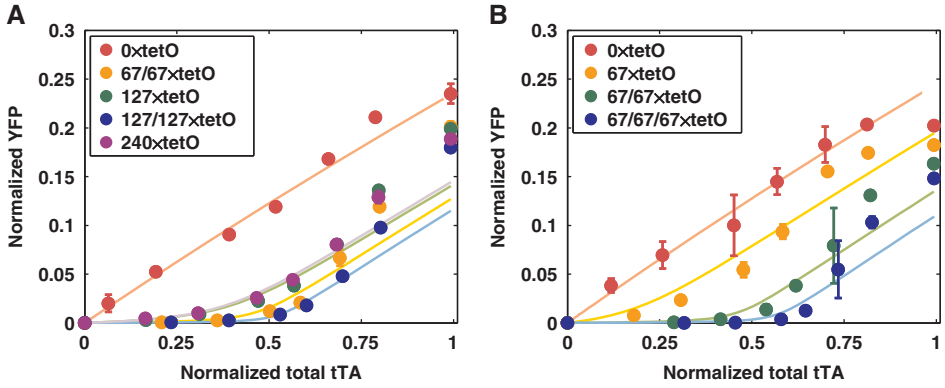

Figure 3 Non-contiguous tetO arrays sequester tTA more effectively than contiguous tetO arrays. (**A**) The dose–response of cells with a 7 × tetO promoter driving YFP in the presence of two separate centromeric plasmid-borne 67 × tetO arrays (67/67) versus one 127 × tetO array (127) or two centromeric plasmid-borne 127 × tetO arrays (127/127) versus one 240 × tetO array (240). In strains with a single contiguous array, a second empty plasmid is also present. Separating the array into two different locations increases its effectiveness in sequestering tTA. (**B**) A similar effect is seen when comparing expression in strains containing 1, 2 or 3 copies of 67 × centromeric plasmid-borne arrays. Error bars represent s.d. of 3 biological replicates. Source data is available for this figure in the Supplementary Information.

Larger contiguous tetO arrays are less effective than multiple non-contiguous tetO arrays

We were surprised to find that at any given dox concentration, increasing array size beyond 67 × tetO sites had little additional effect on decreasing target gene expression. In fact, the actual number of tetO sites in the array is not input into the model; the effective number, N_0/T_0, estimated is remarkably similar for 67 ×, 113 ×, 127 × and 240 × array sizes (Table I). However, when decoy site numbers are increased by using high copy plasmids N_0/T_0 does increase. To determine whether the contiguous nature of the additional decoy sites was responsible for the decrease in their effectiveness, we constructed yeast strains that had multiple centromeric plasmid-borne decoy arrays. We found that two copies of a plasmid-borne 67 × array was more effective versus both one copy of a plasmid-borne 127 × array and one copy of a plasmid-borne 240 × array (Figure 3A and Supplementary Figure 5). Further increases in the copy number of plasmid-borne arrays continue to lower gene expression (Figure 3B and Supplementary Figure 5) confirming a split, non-contiguous array of tetO sites is more effective in sequestering tTA than a contiguous array.

The effects of decoys on target gene expression can be explained in terms of tTA binding

Our proposed mechanism, whereby repeated decoy sites decrease target gene expression, is through competitive tTA binding. This mechanism requires the fraction of decoy-bound tTA molecules to be anticorrelated with target gene expression and similar for the 67 ×, 113 ×, and 240 × tetO arrays. We tested this using quantitative ChIP experiments to measure occupancy of a 3X–HA-tagged tTA at both the promoter and the array. We first measured tTA occupancy at the promoter for cells with 0 ×, 15 ×, 67 × and 240 × tetO sites at various dox levels. The '% INPUT' ChIP signal used here reflects the percentage of input chromatin that is immunoprecipitated by an HA-specific antibody. In Figure 4A, we observe that promoter occupancy is roughly proportional to expression at the 7 × tetO promoter.

Using the same chromatin samples, we next measured tTA occupancy at a particular region shared among the different tetO arrays (red region in Figure 2A). Array occupancy clearly increases with increasing tTA level and not just at low tTA levels (Figure 4B and Supplementary Figure 6). At any given tTA level, the ChIP signal is weaker for the 240 × array compared with the 67 × array. This is expected as the 67 × and 240 × arrays are equally potent in reducing target gene expression. Therefore, these data provide direct support that the 'per site' occupancy of tTA on the 67 × array is several-fold higher than the 240 × array. The extremely high ChIP signals we observe at the array when high levels of tTA are expressed may be because any sheared fragment encompassing the probed region also has additional tTA molecules present on tetO sites adjacent to the probed region. Therefore, not only is there a strong interaction between the 3X-HA tag and the anti-HA antibody, but also many antibodies will be bound to each probe, resulting in highly efficient precipitation of tTA.

To determine whether tTA preferentially binds to the tetO array as compared to the promoter at low tTA levels, we plotted the ChIP signal for the array versus the promoter (Figure 4B). At lower tTA levels, the slope of this plot is smaller and array occupancy changes more significantly than promoter occupancy. At higher tTA levels this slope increases, but there is still significant binding of tTA to the array, which clearly does not saturate. This change in slope could be attributed to anti-cooperative binding of tTA to the tetO array. Higher occupancy of the array could bend the DNA and/or alter chromatin and affect subsequent binding. However, the difference in slopes that marks a transition from tight tTA–tetO interactions on the array to weaker interactions is more subtle than anticipated; especially given the sharp change in concavity we observe in the expression data. We hypothesized that perhaps not all the binding sites in the 7 × tetO promoter are weaker binding as compared with the array. ChIP measures an 'average occupancy' across multiple binding sites (as the chromatin was sheared to an average size of 300 bp that encompasses

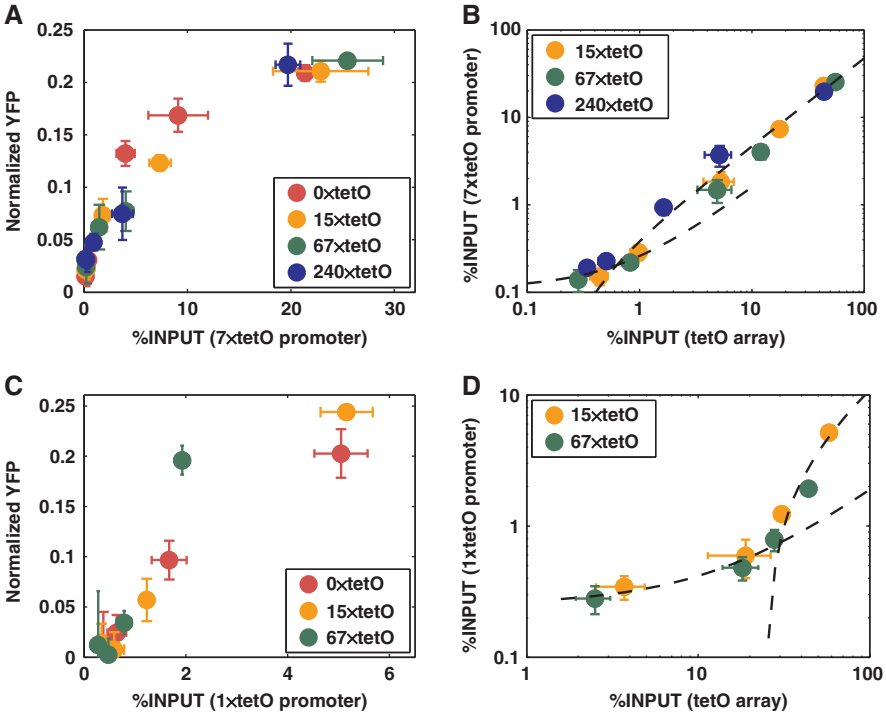

Figure 4 tTA binds to the tetO array and the tetO promoter with different strengths. (**A**) Occupancy of a C-terminal 3X–HA-tagged tTA is monitored at the 7 × tetO promoter by ChIP in the presence and absence of centromeric-borne tetO arrays at 0, 100, 200, 500 and 1000 ng/ml dox. A region including only the 7 × tetO binding sites is probed. ChIP signal (% of INPUT DNA measured in the immunoprecipiated (IP) sample) of tTA varies nearly linearly with target gene expression. (**B**) tTA occupancy was also monitored at a particular region in the tetO arrays encompassing 6 × tetO sites (denoted in Figure 2A). Promoter occupancy is plotted versus tetO array occupancy on a log–log scale. The relative change in promoter versus tetO occupancy is slightly different at low tTA levels versus higher tTA levels. Dotted lines are straight lines (that appear curved on a log–log plot) that could represent two binding regimes. (**C**) and (**D**) are identical to A and B, respectively, but with a 1 × tetO promoter. Expression varies nearly linearly with tTA promoter occupancy, but there is a dramatic difference in the relative binding of tTA to the promoter versus the array as the array occupancy increases. Error bars represent s.d. of triplicate chromatin samples prior to IP. Source data is available for this figure in the Supplementary Information.

multiple sites on average). To address this, we repeated the experiment using a 1 × tetO promoter. We see a similar linear relationship between promoter occupancy and expression (Figure 4C). Strikingly, when comparing the ChIP signal for the decoy arrays versus the 1 × tetO promoter, the transition between the binding regimes is much more dramatic (Figure 4D).

Together with our model, these data support the notion that the residence time of tTA on both the 1 × tetO promoter and a subset of binding sites in the 7 × tetO promoter is shorter than on decoy tetO sites, but only when the tetO array region is relatively free of bound tTA. At higher tTA levels, tTA continues to bind the remaining vacant tetO sites in the array, albeit weakly. If some tetO sites within the 7 × tetO promoter do bind tTA as tightly as those in the array, then an additional copy of the promoter should affect gene expression in a manner similar to the array. To test this, we constructed a centromeric plasmid containing the 7 × tetO promoter driving CFP and introduced it into the usual yeast strain expressing tTA and containing an integrated 7 × tetO promoter driving YFP. Addition of this plasmid reduces target YFP expression and changes the concavity of the dose–response (Supplementary Figure 7). Introducing a centromeric plasmid containing 6 × tetO sites from the 7 × tetO promoter, but without the minimal *CYC1* promoter or the CFP open reading frame, has similar effects. Therefore, a subset of the tetO sites in the 7 × tetO promoter binds to tTA as strongly as the tetO sites present in the array region. This would explain why 7 × tetO promoter binding is relatively strong even at low tTA levels (Figure 4B). It also implies that the location of tetO sites or the regional chromatin environment does not contribute significantly to stronger binding of tTA versus the productive promoter binding.

The altered dose–response induced by TRs converts the behavior of a positive-feedback loop from a graded to a bimodal response

The dox titration data in Figures 2 and 3 demonstrate that repeated decoy arrays are effective at decreasing target gene expression and do so in a manner that converts the linear dose–response to one with an inflection point. As an additional, more stringent test of this conversion occurring, we added tetO arrays to a strain containing a 1 × tetO promoter driving tTA expression in a transcriptional positive feedback (Figure 5A). The tTA levels were indirectly assayed by expression from a 1 × tetO promoter driving YFP expression. When we titrate the feedback strength of a 1 × tetO promoter in positive feedback using dox, we observe a graded response (Figure 5B and D), as has been previously shown as the 1 × tetO promoter response in the absence of feedback is gradual and nearly linear (To and Maheshri, 2010; Supplementary Figure 8). If the decoy sites generate an inflection in the dose–

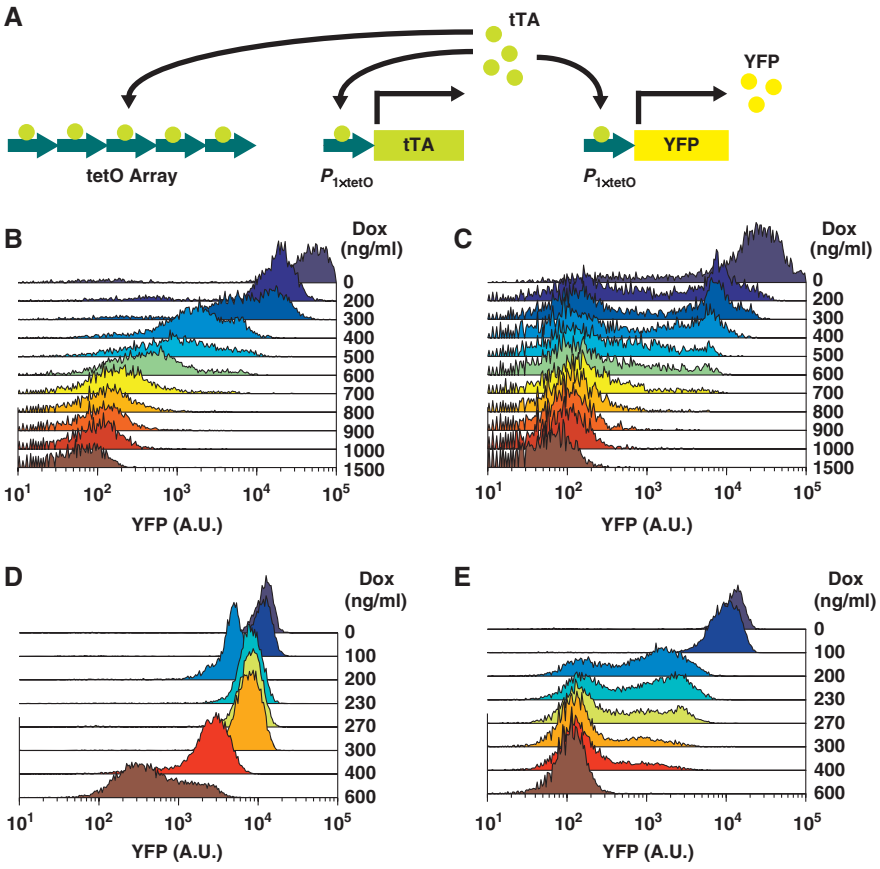

Figure 5 Adding tetO decoy sites converts a positive-feedback loop from a graded to switch-like bimodal response. (**A**) Two 1 × tetO promoters driving tTA and YFP expression are chromosomally integrated at the *HIS3* and *LEU2* locus, respectively. (**B**) Analytical flow cytometry of single-cell fluorescence distributions at various dox concentrations (ng/ml). In the absence of the tetO array, the behavior response of the positive feedback where the promoter is embedded is graded. The feedback strength was changed by controlling dox concentration. (**C**) In the presence of a 2-μm plasmid-borne 127 × tetO array, a graded response was changed to a bimodal response. Chromosomal integration of two 67 × tetO arrays at the *TRP1* and *URA3* locus is less potent, but also converted the (**D**) graded positive-feedback response into (**E**) the bimodal response over a smaller range of dox concentrations. Source data is available for this figure in the Supplementary Information.

response, then their addition could lead to bistable gene expression when in positive feedback. Indeed, we see bimodal expression when we introduce a 2-μm plasmid containing a 127 × tetO array, integrate two copies of a 67 × tetO array in the genome (Figure 5C and E) or add a centromeric plasmid containing a 240 × tetO array (data not shown). Importantly, addition of decoy sites has no effect on the noise in gene expression (Supplementary Figure 9), and therefore this expression is not due to the noise-induced bimodality (To and Maheshri, 2010). The bimodal expression could be due to bistability; however, other explanations are also possible. For example, the sharper threshold response created by the addition of the tetO array may read out slow fluctuations in an upstream factor as a bimodal response. Regardless, addition of the decoy array results in a qualitative change in the response.

Discussion

Previous evidence suggests that TRs of decoy binding sites can sequester a transcriptional activator and inhibit its target gene expression (Janssen *et al*, 2000; Liu *et al*, 2007). However, these studies were qualitative in nature, limited to one or two levels of the activator and a fixed number of TRs. To better understand the consequences of such decoy sites, we used the synthetic tet–OFF system in budding yeast to study how repeated arrays of tetO decoy sites influenced expression of tTA-inducible tetO promoters. We find that decoy sites reduce expression from a tTA-inducible promoter and alter its dose–response curve, converting it from graded to more sigmoidal-like response. Using a simple mathematical model, we show that the observed dose–response can occur if decoy sites bind to tTA with high affinity as compared with the promoter. We confirmed this idea by using ChIP experiments to monitor tTA binding at both promoter and decoy tetO sites. These results are surprising given the tetO sites in both regions are identical in sequence.

Our ChIP data suggest the presence of stronger and weaker binding regimes for the tTA–tetO interaction that depend on tTA binding at nearby sites, as well as whether the binding event leads to gene expression. At low levels, tTA must bind to decoy sites (and a subset of binding sites in the 7 × tetO promoter) with over 10-fold higher 'effective' affinity, compared with tTA binding to tetO sites in the promoter that result in productive gene expression. At higher tTA levels, either tTA binding to the promoter becomes stronger or tTA binding to the decoy sites becomes weaker. We favor the latter interpretation. If the former was true, the dose–response of promoters when

tTA is titrated would not be linear, under the assumption that expression is proportional to tTA occupancy. How might anti-cooperative binding within the decoy array come about? While 10–30 bp spacing between tetO sites insures that tTA binding to one tetO site cannot sterically hinder binding to adjacent sites, binding may indirectly affect adjacent sites through bending or twisting DNA, shifting nucleosomes or recruitment of other factors. In the Supplementary information, we provide an order of magnitude estimate of the total steady-state nuclear tTA level when expressed from the *MYO2* promoter of 10^2. Using this estimate, based on our data and model, the binding transition occurs when roughly 10–30 tTAs are bound to 67×, 113× and 240× integrated tetO arrays, and approximately 40–60 tTAs are bound to equivalent centromeric repeats. These results are similar to *in vitro* studies of dimeric lacI binding to a 256× tandem lacO array embedded within the λ-phage genome. The authors find only 2.5% (~13) of the available lacO sites are bound at concentrations of lacI that should saturate the array, and lacI–lacO binding affinity appears to be inversely dependent on the occupancy of lacI on the array (Wang *et al*, 2005). Finally, tetR binding to multiple tetO sites present between a synthetic enhancer and promoter in *E. coli* has been found to affect gene expression in a manner consistent with anticooperative binding (Amit *et al*, 2005). While we have explained binding at decoy tetO sites using strong and weak regimes, the ChIP data are certainly consistent with a continuous decrease in affinity with tTA occupancy.

The ChIP experiments can also help to understand a puzzling aspect of the expression data: adding additional contiguous tetO sites to the 67×tetO array has little further effects on target gene expression—the 'effective' number of strong binding sites is nearly equivalent. However, the potency of these arrays increases when separated by placement on different plasmids or portions of the DNA. These observations can be explained if two features of tTA–tetO binding are true. First, every tetO site should bind with similar strength in the low-occupancy strong binding regime. This is consistent with the ChIP signal at the 240×tetO array being lower than the 67× array at low tTA levels (Figure 4, Supplementary Figure 6)—tTA samples fourfold more sites with 240×repeats, hence binding at any particular region is lower. Second, the transition to the weak binding regime should be dictated by the absolute number of tTA bound to the array, probably through a long-range interaction that reduces the affinity of neighboring vacant tetO sites. Because the 240×tetO array consists of a 113× and 127× array separated by 534 bp (Figure 2B), this long-range interaction occurs over at least 100's of bps of DNA. When the array is split apart by placement on different plasmids, strong tTA binding at one array does not affect another array—hence the number of strong binding sites (or effective binding sites) scale with the number of split arrays as observed (Figure 3, Table I).

Perhaps even more unexpected than the anti-cooperative binding of tTA to the decoy sites is the large difference in tTA binding to the promoter versus decoy sites. The molecular origin of this difference remains unresolved. Some possibilities that specifically increase the tTA–tetO residence time within decoy sites could include: a unique chromosomal location and/or the chromatin environment of the array, the multivalent nature of the tTA–tetO array interaction and/or active recruitment of transcriptional machinery by tTA. However, these possibilities are less likely in light of the fact that tetO sites within the 7×tetO promoter (whether alone or in a context of the promoter) can modify the dose–response in a manner consistent with strong tTA binding (Supplementary Figure 7).

An alternative idea is that the affinity of TFs to binding sites within an active promoter is significantly altered, perhaps because of active processes that destabilize the TF binding during a productive initiation cycle. FRAP measurements of TF occupancy on large gene arrays suggest TFs are highly dynamic (Darzacq *et al*, 2007; Darzacq *et al*, 2009; Karpova *et al*, 2008), and only a small fraction of binding events actually result in expression. At this promoter, at the expression levels in this paper, initiation events are infrequent, with a frequency between approximately 0.0015 and 0.015 per min over the range of expression (To and Maheshri, 2010). The model only distinguishes 'promoter binding' from 'decoy binding' by requiring gene expression to be proportional to transcriptionally competent 'promoter binding' events only. Probably at the 7×tetO promoter, not all binding events are 'promoter binding' (Figure 4). Previous FRAP studies are unable to distinguish between these types of events.

As has been put forward in Buchler and Louis (2008), molecular titration provides a simple mechanism to generate ultrasensitivity that is a crucial ingredient for the rich dynamical behavior of biological networks, including multistable and oscillatory behavior. This mechanism has been elegantly demonstrated in the context of protein–protein interactions (Buchler and Cross, 2009). Tight sigma factor/anti-sigma factor interactions have been suggested to introduce bistability in prokaryotic gene networks (Tiwari *et al*, 2010). It is likely that this mechanism operates in RNA–RNA and RNA–protein interactions as well, particularly in the context of regulatory microRNA's, whose affinity for targets is easily tuned (Bartel, 2009; Mukherji *et al*, 2011). Our work extends this paradigm to DNA–protein interactions, where it may be generally true if 'promoter binding' events competent for transcription are necessarily weaker than other TF/DNA interactions. This conversion qualitatively changed the behavior of a transcriptional positive feedback involving tTA, converting its response from a graded to switch-like. Because greater numbers of decoy sites can have more potent effects, our work points to yet another mechanism whereby the microevolution of TR number can lead to qualitative phenotypic changes. It will be important to confirm the generality of this response and the importance of non-contiguous decoy sites by studying native TFs. Of particular interest might be the behavior of single input modules (Alon, 2007)—a gene regulatory network motif where one TF controls the expression of many genes. If clustered binding sites present in promoters can function as high-affinity decoys, the dose–response of a lower affinity class of promoters within a single input module may be ultrasensitive because of the presence of a higher affinity class of clustered sites in other promoters within the motif.

Materials and methods

Strain and plasmid construction

All yeast strains were derived from the W303 background (Thomas and Rothstein, 1989). Strain construction was performed using

standard methods of yeast molecular biology (Guthrie and Fink, 2004). Details of the tTA and tetO promoters are given in the study by To and Maheshri, 2010. Strains and plasmids used are listed in Supplementary Tables 1 and 2.

tTA titration by doxycycline

Yeast cells were grown in synthetic medium with 2% glucose overnight at 30 °C, then diluted (OD_{600} ~0.01 to 0.05) and grown in the same medium with various concentrations of dox (Sigma) in a 96 deep–well plate for at least 8 h, maintaining exponential phase. Then, cells were diluted again (OD_{600} ~0.01 to 0.05) and grown for at least 8 h to insure reporter expression reached the steady state. After incubation, cells were placed on ice or at 4 °C and fluorescence intensities were measured by flow cytometry.

Models

The basic model for gene expression in the presence of decoys, and a comprehensive model encompassing the basic model as well as the tTA/dox interaction have been deposited in BioModels under accession nos. 1202270000 and 1202270001.

Flow cytometry

Analytical flow cytometry on yeast cells were performed using a Beckton-Dickinson LSRII HTS equipped with a 405-nm laser and 450/50-nm filter (CFP), a 488-nm laser and 530/30-nm filter (YFP) and a 561-nm laser and 610/20-nm (RFP) filter. For each sample, at least 30 000 cells were measured. Yeast cells without fluorescent reporters or a strain constitutively expressing YFP or CFP from an *ADH1* promoter were always used as negative and positive controls, respectively. This enabled normalization and comparison of the YFP or CFP intensity from measurements performed on different days. Reported data include the densest region of a forward versus side scatter plot of analyzed cells, representing 15% of population.

Quantitative ChIP

ChIP was performed as in Aparicio *et al* (2004) with slight modifications. Briefly, yeast strains grown overnight were diluted to OD_{600} of ~0.001 in 200 ml synthetic medium with 2% glucose and then grown to mid-exponential phase (a final OD_{600} between 0.7 and 1.0). Crosslinking was performed by resuspending cell pellets in 5.6 ml of 37% formaldehyde and incubating for 20 min at room temperature, followed by addition of 10 ml of 2.5 M glycine to quench the reaction. Fixed cells were vortexed with glass beads for 1 h at 4 °C for lysis. Chromatin was sheared using a Microson Ultrasonic Cell Disruptor, with 6×10 s cycles at a power setting of 8. Chromatin was immunoprecipitated with Dynabead (Invitrogen)—Anti-HA High Affinity rat monoclonal antibody (Roche) complex as previously described (Lee *et al*, 2006). qPCR was performed on an Applied Biosystems 7300 real-time PCR machine. PCR efficiency of primers targeting the tetO promoter and array were confirmed to be >1.85, using serial dilutions of either sheared chromosomal DNA or a highly concentrated IP DNA containing the tetO promoter and array. This also determined the threshold cycle (C_t) range for linear amplification, and all C_t values for INPUT and IP DNA were within this range.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

We thank K Verstrepen, KD Wittrup and members of the Maheshri Laboratory for useful discussions and comments. This work was funded by the Human Frontiers Science Program RGY2007 (to NM), NIH Award R01GM95733 (to NM), NSF BBBE 1033316 (to NM) and Massachusetts Institute of Technology startup funds (to NM).

Author contributions: NM conceived the research. NM and T-HL designed the experiments, analyzed the data and wrote the paper. T-HL performed the experiments.

References

Alon U (2007) Network motifs: theory and experimental approaches. *Nat Rev Genet* **8:** 450–461

Amit R, Garcia HG, Phillips R, Fraser SE (2011) Building enhancers from the ground up: a synthetic biology approach. *Cell* **146:** 105–118

Aparicio O, Geisberg JV, Struhl K (2004) Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences *in vivo*. *Curr Protoc Cell Biol* Chapter 17: Unit 17.7 (doi:10.1002/0471143030.cb1707s23)

Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* **136:** 215–233

Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Phillips R (2005) Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev* **15:** 116–124

Buchler NE, Cross FR (2009) Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol Syst Biol* **5:** 272

Buchler NE, Louis M (2008) Molecular titration and ultrasensitivity in regulatory networks. *J Mol Biol* **384:** 1106–1119

Burger A, Walczak AM, Wolynes PG (2010) Abduction and asylum in the lives of transcription factors. *Proc Natl Acad Sci USA* **107:** 4016–4021

Cummings CJ, Zoghbi HY (2000) Trinucleotide repeats: mechanisms and pathophysiology. *Annu Rev Genomics Hum Genet* **1:** 281–328

Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD, Singer RH (2007) *In vivo* dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol* **14:** 796–806

Darzacq X, Yao J, Larson DR, Causse SZ, Bosanac L, de Turris V, Ruda VM, Lionnet T, Zenklusen D, Guglielmi B, Tjian R, Singer RH (2009) Imaging transcription in living cells. *Annu Rev Biophys* **38:** 173–196

Degenkolb J, Takahashi M, Ellestad GA, Hillen W (1991) Structural requirements of tetracycline-Tet repressor interaction: determination of equilibrium binding constants for tetracycline analogs with the Tet repressor. *Antimicrob Agents Chemother* **35:** 1591–1595

Fondon JW3rd, Garner HR (2004) Molecular origins of rapid and continuous morphological evolution. *Proc Natl Acad Sci USA* **101:** 18058–18063

Gari E, Piedrafita L, Aldea M, Herrero E (1997) A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in *Saccharomyces cerevisiae*. *Yeast* **13:** 837–848

Guthrie C, Fink GR (2004) *Guide to yeast genetics and molecular and cell biology. Part A*. Amsterdam: Elsevier Academic Press

Horng JT, Lin FM, Lin JH, Huang HD, Liu BJ (2003) Database of repetitive elements in complete genomes and data mining using transcription factor binding sites. *IEEE Trans Inf Technol Biomed* **7:** 93–100

Janssen S, Cuvier O, Muller M, Laemmli UK (2000) Specific gain- and loss-of-function phenotypes induced by satellite-specific DNA-binding drugs fed to *Drosophila melanogaster*. *Mol Cell* **6:** 1013–1024

Karpova TS, Kim MJ, Spriet C, Nalley K, Stasevich TJ, Kherrouche Z, Heliot L, McNally JG (2008) Concurrent fast and slow cycling of a transcriptional activator at an endogenous promoter. *Science* **319:** 466–469

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K *et al* (2001) Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921

Lau IF, Filipe SR, Soballe B, Okstad OA, Barre FX, Sherratt DJ (2003) Spatial and temporal organization of replicating Escherichia coli chromosomes. *Mol Microbiol* **49:** 731–743

Lee TI, Johnstone SE, Young RA (2006) Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc* **1:** 729–748

Liu X, Wu B, Szary J, Kofoed EM, Schaufele F (2007) Functional sequestration of transcription factor activity by repetitive DNA. *J Biol Chem* **282:** 20868–20876

Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, Thomas WK (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci* **105:** 9272–9277

Mukherji S, Ebert MS, Zheng GX, Tsang JS, Sharp PA, van Oudenaarden A (2011) MicroRNAs can generate thresholds in target gene expression. *Nat Genet* **43:** 854–859

Murphy KF, Balazsi G, Collins JJ (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc Natl Acad Sci* **104:** 12726–12731

Rando OJ, Verstrepen KJ (2007) Timescales of genetic and epigenetic inheritance. *Cell* **128:** 655–668

Thomas BJ, Rothstein R (1989) Elevated recombination rates in transcriptionally active DNA. *Cell* **56:** 619–630

Tiwari A, Balazsi G, Gennaro ML, Igoshin OA (2010) The interplay of multiple feedback loops with post-translational kinetics results in bistability of mycobacterial stress response. *Phys Biol* **7:** 036005

To TL, Maheshri N (2010) Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science* **327:** 1142–1145

Verstrepen KJ, Jansen A, Lewitter F, Fink GR (2005) Intragenic tandem repeats generate functional variability. *Nat Genet* **37:** 986–990

Vinces MD, Legendre M, Caldara M, Hagihara M, Verstrepen KJ (2009) Unstable tandem repeats in promoters confer transcriptional evolvability. *Science* **324:** 1213–1216

Wang YM, Tegenfeldt JO, Reisner W, Riehn R, Guan X, Guo L, Golding I, Cox EC, Sturm J, Austin RH (2005) Single-molecule studies of repressor–DNA interactions show long-range interactions. *Proc Natl Acad Sci USA* **102:** 9796–9801