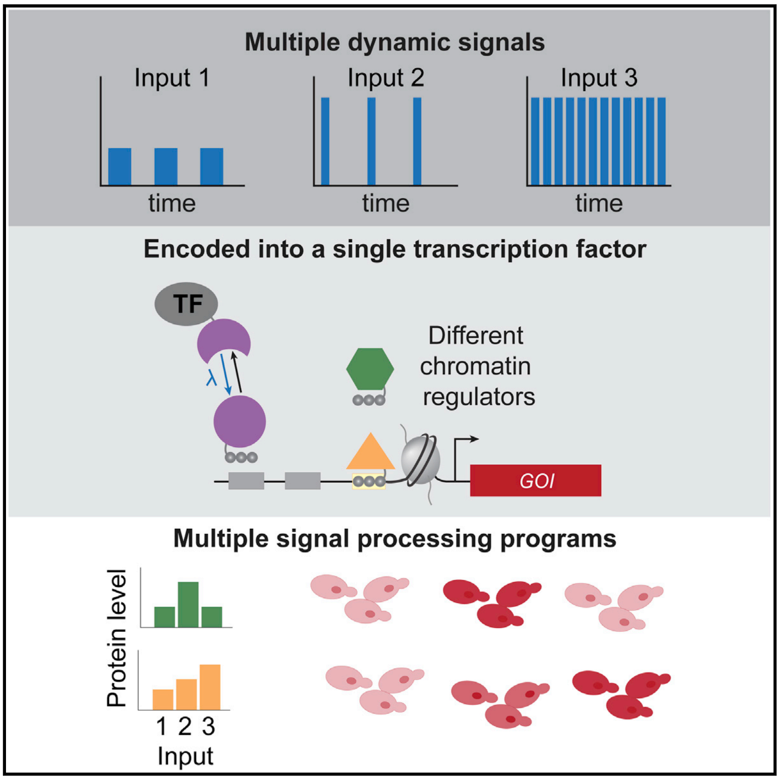
Cell Systems

Article

Mapping the dynamic transfer functions of eukaryotic gene regulation

Graphical abstract



Highlights

- Synthetic optogenetic system developed to study dynamic promoter inputs-outputs
- Mapped protein outputs for 119 dynamic pulsatile inputs
- Developed a mass-action model that explains complex filtering phenomena
- Transfer functions were modified by recruitment of chromatin-regulating proteins

Authors

Jessica B. Lee, Leandra M. Caywood, Jennifer Y. Lo, Nicholas Levering, Albert J. Keung

Correspondence

ajkeung@ncsu.edu

In brief

Biological information can be encoded in the dynamics of signaling components. Combining optogenetics and flow cytometry, this work describes a platform to map the protein expression response to a large parameter space of 119 dynamic inputs as well as co-recruitment of 101 chromatin-regulating proteins. It also provides quantitative synthetic biology, modeling, and information theory frameworks to understand and predict the complex and diverse promoter-to-protein transfer functions for a single gene.

Lee et al., 2021, Cell Systems *12*, 1–15
November 17, 2021 © 2021 Elsevier Inc.
https://doi.org/10.1016/j.cels.2021.08.003

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

Cell Systems

CellPress

Article

Mapping the dynamic transfer functions of eukaryotic gene regulation

Jessica B. Lee,[1] Leandra M. Caywood,[1] Jennifer Y. Lo,[1] Nicholas Levering,[1] and Albert J. Keung[1,2,*]
[1]Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27606, USA
[2]Lead contact
*Correspondence: ajkeung@ncsu.edu
https://doi.org/10.1016/j.cels.2021.08.003

SUMMARY

Biological information can be encoded within the dynamics of signaling components, which has been implicated in a broad range of physiological processes including stress response, oncogenesis, and stem cell differentiation. To study the complexity of information transfer across the eukaryotic promoter, we screened 119 dynamic conditions—modulating the pulse frequency, amplitude, and pulse width of light—regulating the binding of an epigenome editor to a fluorescent reporter. This system revealed tunable gene expression and filtering behaviors and provided a quantification of the limit to the amount of information that can be reliably transferred across a single promoter as ∼1.7 bits. Using a library of over 100 orthogonal chromatin regulators, we further determined that chromatin state could be used to tune mutual information and expression levels, as well as completely alter the input-output transfer function of the promoter. This system unlocks the information-rich content of eukaryotic gene regulation.

INTRODUCTION

There is ample evidence that biological information can be encoded in the dynamics of signaling components and not just in their biochemical identities (Behar and Hoffmann, 2010; Cai et al., 2008; Dalal et al., 2014; Hansen and O'Shea, 2013; Hao et al., 2013; Imayoshi et al., 2013; Inoue et al., 2016; Purvis et al., 2012). Cells, with a limited number of components, utilize dynamic signal processing to perform sophisticated functions in response to complex environments. Transcription factors (TFs) may be a particularly important archetype for this type of information transmission, as they are relatively low in diversity but must command many distinct and complex gene expression programs (Lee and Young, 2013). Indeed, through chemical and optogenetic approaches, the dynamics of TF nuclear-cytoplasmic translocation has been shown to control gene expression levels and population noise (An-adirekkun et al., 2020; Chen et al., 2020; Hansen and O'Shea, 2013; Rademacher et al., 2017). There is also evidence that different promoters can transduce dynamic TF input signals into distinct output responses (Chen et al., 2020; Hansen and O'Shea, 2016; Harton et al., 2019). Thus, developing a quantitative understanding of how dynamic TF signals are ultimately interpreted and processed by individual genes and promoters is clearly important.

A couple of compelling analogies can be drawn: (1) to information theory, with promoters analogous to information transfer channels and (2) to process control, with promoters acting as unit processes with dynamic input-output transfer functions. The nature of these channels or transfer functions might even be tunable by parameters such as promoter sequence, chromatin state, or three-dimensional chromatin topology. However, developing this type of robust quantitative framework poses considerable challenges. Mapping the transfer function of a single promoter seems ostensibly simple but faces the inherent technical difficulties of controlling dynamic properties of biological systems. The complex diversity of eukaryotic chromatin presents yet another formidable barrier. More specifically, there are three particularly pressing challenges. First, there is a broad range of dynamic input and output parameters that is technically challenging to access, control, and characterize. Second, as each individual promoter can be regulated by multiple distinct TFs and chromatin regulators (CRs), pleiotropic effects can confound global perturbations to nuclear TF levels or chromatin state. Finally, there are hundreds of distinct CRs that can alter how promoters interpret TF signals, resulting in a large experimental space to explore (Kouzarides, 2007; Li et al., 2007).

To address these challenges, we engineered both dynamic and static epigenome editors that bypass pleiotropic issues due to their locus specificity and thereby provide insight into the causal impacts of CRs and TFs on transcription (Bintu et al., 2016; Keung et al., 2014; Park et al., 2019; Polstein and Gersbach, 2015). To study the effects of TF signal dynamics on transcription, we employed an optogenetic system that dynamically recruited the transactivator VP16 to a genomically integrated fluorescent reporter. By pairing the optogenetic system with programmable Arduino-controlled LED arrays and single-cell fluorescence measurements by flow cytometry, we were able to capture and screen a large parameter space of dynamic inputs.

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

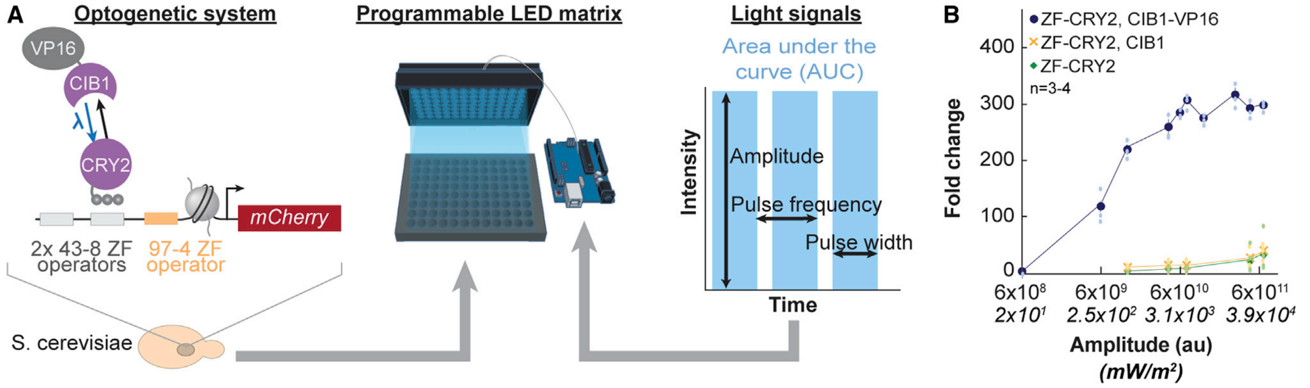


Figure 1. Optogenetics provides complete access to the dynamic parameter space

(A) Schematic of genetic and hardware systems. The optogenetic system was expressed in *S. cerevisiae* (left). ZF-CRY2-targeted operators were placed upstream of a minimal *CYC1* promoter driving the expression of mCherry. In the presence of blue light, CIB1-VP16 binds ZF-CRY2 and disassociates without blue light. Parameters of amplitude, frequency, and pulse width (right) were varied using a custom Arduino-controlled, individually addressable LED matrix (center). The area under the curve (AUC) is defined as the (amplitude) x (frequency) x (pulse width) x (duration of experiment).

(B) Fold change in fluorescence for various light intensity amplitudes for a constant, 14-h light pulse. Fold changes for control strains, ZF-CRY2 and ZF-CRY2+CIB1, are also shown. Error bars are SEM for 3 (ZF-CRY2 and CIB1-VP16) or 4 (ZF-CRY2, CIB1, and ZF-CRY2) biological replicates. Power density measurements were calculated using equation found in Figure S1G.

Using this experimental platform, we mapped protein outputs in response to 119 different optogenetic inputs that modulated the amplitude, frequency, and pulse width of VP16 recruitment. Input conditions with the same total signal but different dynamic parameters yielded outputs with over an order of magnitude difference and, therefore, acted as a filter. A kinetic model was developed to describe the complex transfer function captured by the experimental data, including the light pattern-filtering behavior. To further understand the reliability of the information transfer, we applied information theory to the single-cell distribution data and estimated the limits to the amount of information transmittable through each input mode—as well as with all input modes combined—with frequency modulation carrying the greatest amount of transmittable information and amplitude the least. Finally, we asked if co-recruitment of CRs to the promoter could alter its transfer function without any alteration to the promoter sequence. 101 CRs were constitutively recruited to the promoter. Many of them altered the gene expression response to dynamic VP16 inputs, including exhibiting complex types of transfer functions such as band-pass, low-pass, and high-pass frequency filtering. In addition, co-recruiting CRs with VP16 tuned the maximum possible amount of information that was transmittable through the single promoter. This study reveals the information-rich nature of eukaryotic gene expression even at just a single gene, implicates an interplay between dynamics and chromatin, and also provides quantitative synthetic biology, modeling, and information theory frameworks to understand and predict complex promoter-to-protein transfer functions.

RESULTS

Optogenetics provides complete access to the dynamic parameter space

To probe the transfer functions across a promoter, we developed an optogenetic system to recruit epigenome editors to a synthetic transcriptional reporter in arbitrary dynamic patterns (Figure 1A). A *CYC1* promoter drove expression of an mCherry reporter and was integrated into the *LEU2* locus of *Saccharomyces cerevisiae*. The *CYC1* promoter contained two identical binding sites (GAGTGAGGA) recognized by an engineered zinc finger (ZF) array “ZF43-8” and an orthogonal binding site recognized by ZF array “ZF97-4” (TTATGGGAG) (Keung et al., 2014; Khalil et al., 2012). The ZF97-4 binding site will be used later in this work. We fused ZF43-8 to cryptochrome 2 (*ZF-CRY2*) and cryptochrome-interacting basic helix-loop-helix to the transcriptional activator *VP16* (*CIB1-VP16*) and placed their expression under ATC and IPTG control (Keung et al., 2014), respectively. CRY2 binds CIB1 when exposed to blue light and dissociates upon light removal (Kennedy et al., 2010; Liu et al., 2008). This system has high temporal resolution with an association half-life of seconds and dissociation half-life of ~5 min (Rademacher et al., 2017). We also tested other optogenetic systems, different N- and C- terminal fusions, and several induction drug concentrations (Figures S1A–S1F). The final system was chosen for its robust activation with light and minimal activation without light. To deliver the light signals, an Arduino Due controlled individually addressable blue LEDs (wavelength = 455–465 nm) in a 96-well format.

To accurately map the effects of different dynamic input light patterns on eukaryotic gene expression, the system must operate at sub-saturation. Therefore, we first determined the dynamic range of the system and identified sub-saturation light amplitudes (i.e., intensities). We exposed the cells to lights that were constantly on for 14 h with a range of amplitudes. The resultant fold change in mCherry protein was measured by flow cytometry and was defined as the median fluorescence of cells incubated with a given light pattern divided by the median fluorescence of cells incubated without light (Figure 1B). The output signal saturated at intensities above 6×10^{10} au (~3,100 mW/m^{2}). Control cells expressing only ZF-CRY2 or ZF-CRY2 with CIB1 exhibited much lower levels of activation,

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

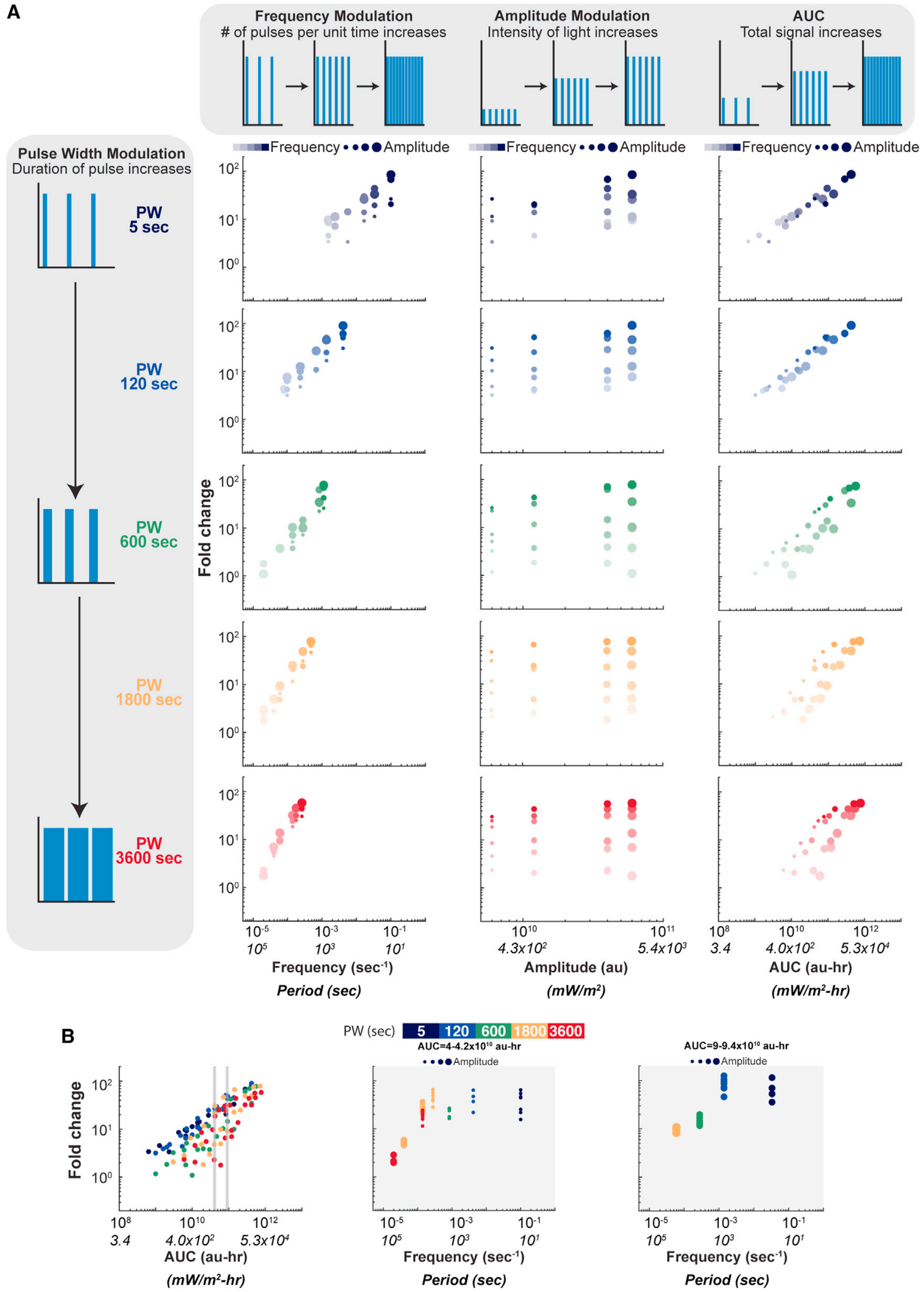


Figure 2. 119 dynamic signals provide comprehensive map of a eukaryotic transfer function

(A) Schematics illustrating frequency, amplitude, and pulse width modulation (FM, AM, and PWM) and AUC are shown along the top or left. Slope of fold change versus frequency increases with pulse width.

(legend continued on next page)

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

especially at sub-saturating levels. For further studies modulating all three dynamic parameters (amplitude, frequency, and pulse width), we chose sub-saturation amplitudes below 6×10^{10} au to ensure both comprehensive coverage of the dynamic parameter space and to minimize any activation due to ZF-CRY2 alone.

119 dynamic signals provide comprehensive map of a eukaryotic transfer function

Frequency, amplitude, and pulse width modulation (FM, AM, and PWM, respectively) present a large combinatorial space, which is challenging to capture experimentally; yet, it is crucial to do so in order to understand the transfer functions of eukaryotic promoters and to generate quantitative and predictive models. With the programmable LED array paired with flow cytometry, we delivered 119 distinct input signal patterns to yeast cultured in 96-well format and measured mCherry reporter endpoint fluorescence after 14 h (see Table S1 for a complete list of conditions). These patterns included 5 sets of conditions with pulse width held constant at 5, 120, 600, 1,800, or 3,600 s. We chose this range of pulse widths to include timescales similar to those found for several pulsatile TFs in *S. cerevisiae* (Dalal et al., 2014). For each pulse width, four amplitudes, 6×10^9, 1.2×10^{10}, 4×10^{10}, and 6×10^{10} au (2.5×10^2, 5.3×10^2, 2×10^3, and 3.1×10^3 mW/m^2), and 5–6 frequencies (between 2×10^{-5} and 1×10^{-1} s^{-1}, equivalent to periods between 50,000 and 10 s) were delivered to the cells. We defined the total input signal, or area under the curve (AUC), as the product of frequency, amplitude, pulse width, and the duration of the experiment. The throughput of the system allowed us to measure thousands of cells as well as four biological replicates per condition (see Table S1 for individual replicate data).

As expected, mCherry expression increased with amplitude, frequency, pulse width, and AUC (Figure 2A). However, whereas frequency and pulse width had strong effects on mCherry output (Figure 2A, left column), amplitude had much weaker effects (middle column). To quantify the relative effect of each light parameter on mCherry expression, we first standardized the amplitude, frequency, pulse width, and the resulting mCherry fold change using z-transformation. This allowed comparison between the coefficients of each input mode within a regression model (Schielzeth, 2010). We then fit the standardized data to a linear regression model, given in Figure S2A ($R^2 = 0.92$, p = 5.29e-57). This linear model confirmed that frequency had the largest coefficient and, therefore, greatest effect on fold change, whereas amplitude had the weakest effect.

The fact that the system did not respond equally to each mode of modulation suggested that the system might exhibit filtering behaviors, where input signal patterns that share identical input AUCs but—through different weightings of amplitude, frequency, and pulse width—could yield different output levels. Indeed, this signal-filtering property was observed over a wide range of AUCs (examples shown in Figure 2B). This indicates that this system has inherent signal-filtering capabilities because frequency, pulse width, and amplitude do not have equally proportional effects on mCherry expression. We found that filtering was not an artifact of measuring mCherry fluorescence at different time points after the last light pulse was delivered; the same filtering was observed even when the timing of the last light pulse was shifted relative to the end of the experiment (Figures S2B–S2E).

Model captures system behavior and filtering

With this set of dynamic input data, we asked if they could inform the development and architecture of mass-action models of eukaryotic gene expression, and if this model could subsequently provide additional insights into the filtering property of the system. We tested several previous models (Benzinger and Khammash, 2018; Chen et al., 2020; Hansen and O'Shea, 2013; Harton et al., 2019) along with some new architectures and found that a three-promoter-state model (probability of each promoter state is represented by $P_{unbound}$, P_{bound}, and P_{active}) best fit the experimental data with $R^2 = 0.865$ (Figures 3 and S3A). A quantitative comparison of model structures (Figure S3B) was performed by stratified K-fold cross validation (Figure S3C), which supported a three-promoter-state model over similar models with two or four promoter states. We also explored incorporating various numbers of Hill functions into the model as was done in other models of eukaryotic gene expression. We found that the inclusion of one or two Hill functions yielded the best fit for our system, but that inclusion of two Hill functions was not sufficiently better than one to warrant the extra structural complexity. Based on the cross validation, fit to the experimental data, and physical understanding of our specific system, we found the three-promoter-state model with inclusion of a single Hill function to best describe our data.

The filtering behavior observed in the experimental data was a notable feature of our optogenetic system. Therefore, we asked if our model of choice also captured this behavior. Indeed, our model reflected this filtering behavior (Figure 3C). The model suggested that filtering may arise from a combination of different mechanisms, as a variety of distinct input patterns with the same AUC gave rise to distinct outputs (Figures 3C and 3D). One contributor may be that the decay of promoter occupancy is not immediate; therefore, at high frequency there is an accumulation of "extra" promoter occupancy (Figure 3E). This could also explain why frequency has a large effect on the fold change.

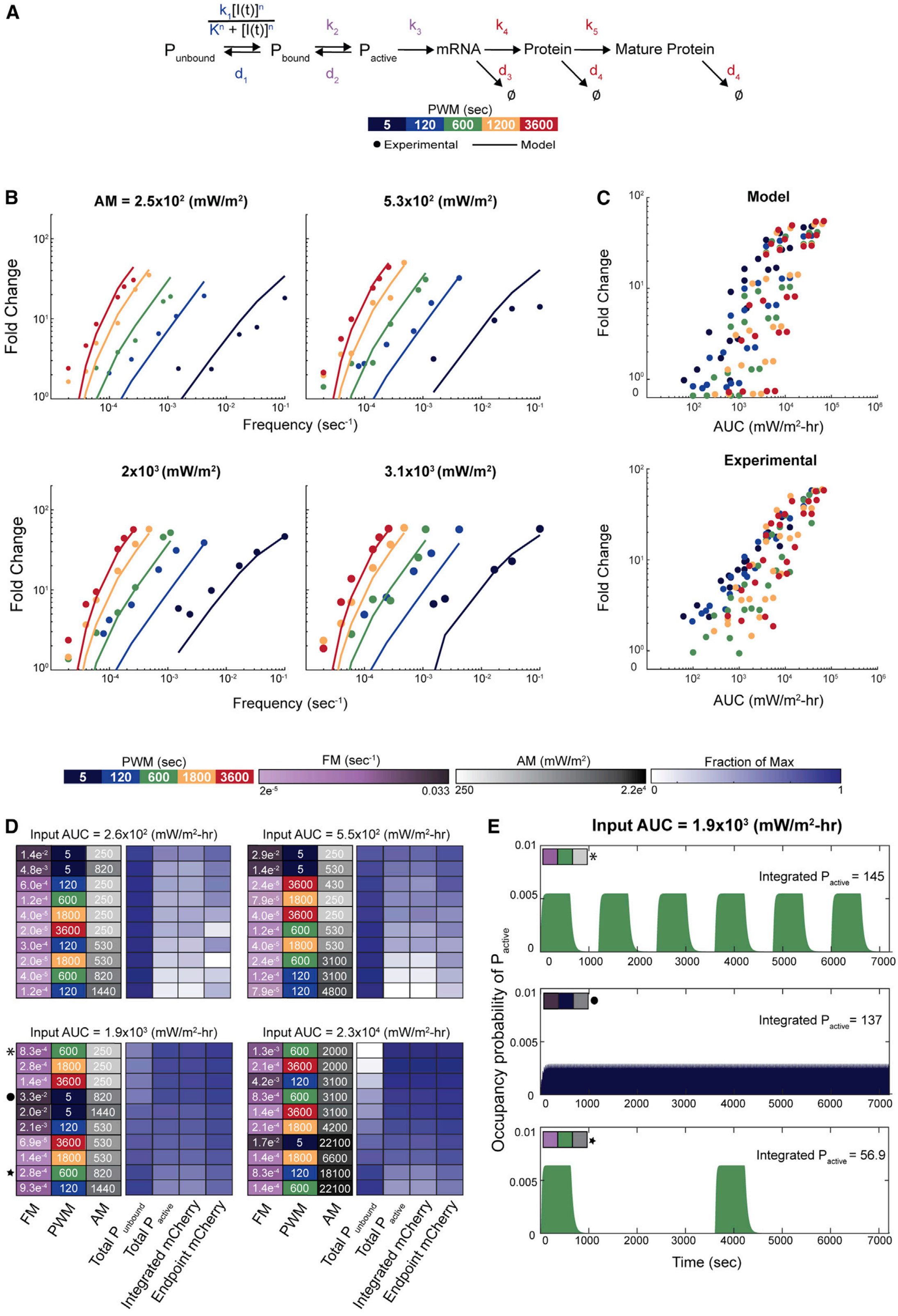
Single-cell measurements capture total population noise

Although understanding and mapping the transfer functions of a system is important, the reliability in achieving the same output repeatedly over time, or within a population of cells in response to the same input signal, is equally important. This reliability can be characterized by the noisiness of gene expression. Gene expression noise is an important and inherently stochastic process due to the low copy number of genes (Eldar and Elowitz, 2010; Elowitz and Leibler, 2000; Maheshri and O'Shea, 2007)

(B) All conditions are plotted together, further emphasizing the filtering behavior at constant AUCs (left). Two example AUCs are highlighted by vertical gray lines. The fold changes highlighted are graphed versus frequency (middle, right) to illustrate filtering behavior based on frequency.

(A and B) Fold change = (median fluorescence with blue light condition-autofluorescence)/(median fluorescence without blue light-autofluorescence). Each dot is the mean of 4–8 biological replicates. For fold change values for each replicate, see Table S1.

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003



Input AUC = 2.6x10^2 (mW/m^2-hr)

FM	PWM	AM
1.4e^{-2}	5	250
4.8e^{-3}	5	820
6.0e^{-4}	120	250
1.2e^{-4}	600	250
4.0e^{-5}	1800	250
2.0e^{-5}	3600	250
3.0e^{-4}	120	530
2.0e^{-5}	1800	530
4.0e^{-5}	600	820
1.2e^{-4}	120	1440

Input AUC = 5.5x10^2 (mW/m^2-hr)

FM	PWM	AM
2.9e^{-2}	5	250
1.4e^{-2}	5	530
2.4e^{-5}	3600	430
7.9e^{-5}	1800	250
4.0e^{-5}	3600	250
1.2e^{-4}	600	530
4.0e^{-5}	1800	530
2.4e^{-5}	600	3100
1.2e^{-4}	120	3100
7.9e^{-5}	120	4800

Input AUC = 1.9x10^3 (mW/m^2-hr)

	FM	PWM	AM
*	8.3e^{-4}	600	250
	2.8e^{-4}	1800	250
	1.4e^{-4}	3600	250
●	3.3e^{-2}	5	820
	2.0e^{-2}	5	1440
	2.1e^{-3}	120	530
	6.9e^{-5}	3600	530
	1.4e^{-4}	1800	530
★	2.8e^{-4}	600	820
	9.3e^{-4}	120	1440

Input AUC = 2.3x10^4 (mW/m^2-hr)

FM	PWM	AM
1.3e^{-3}	600	2000
2.1e^{-4}	3600	2000
4.2e^{-3}	120	3100
8.3e^{-4}	600	3100
1.4e^{-4}	3600	3100
2.1e^{-4}	1800	4200
1.7e^{-2}	5	22100
1.4e^{-4}	1800	6600
8.3e^{-4}	120	18100
1.4e^{-4}	600	22100

(legend on next page)

and, together with cell-to-cell variability in general cellular components, creates a distribution of single-cell outputs for each unique input (Elowitz and Leibler, 2000; Grabowski et al., 2019; Gregor et al., 2007; Rosenfeld et al., 2005; Tkacik et al., 2009). As noise plays an important role in determining the reliability of a system, we first quantified how noise in our system was affected by FM, AM, and PWM. We calculated the robust coefficient of variation (CV) of the population for all 119 input light conditions (Figure 4A).

When graphed against frequency, the CV exhibited less variation at low frequencies but a greater range of values for high frequencies (Figure 4A, left column). This was especially true for smaller pulse widths (5, 120, and 600 s). Amplitude did not have a large effect on the CV (Figure 4A, middle column). However, larger pulse widths generally had lower CVs (Figure 4A, left and middle columns). The fluorescence values were normalized by FSC-A before finding the CV in Figure 4A. We compared the CVs from non-normalized data (Figures 4B and S4A) with non-normalized data with a restricted cell size gating (Figures 4C and S4B) and to data normalized by FSC-A (Figures 4D and S4C). The CV was much lower and nearly constant for all AUCs and frequencies when mCherry fluorescence was normalized by size using FSC-A or restricted to similar size cells. Without normalization or size restriction, the CV displayed two regimes: high, nearly constant CV for low AUCs and descending CV for high AUC (AUC greater than 10^{11} au). This phenomenon agrees with prior work showing cell size as a major contributor to noise (Bar-Even et al., 2006; Newman et al., 2006), with removal of this noise due to size also providing an estimation of intrinsic noise (Newman *et al.*, 2006).

Quantifying the contribution of signaling dynamics to maximum mutual information

Although CVs of output distributions provide some qualitative understanding of how reliably information can be transmitted across a gene, a more quantitative understanding would be helpful in analyzing and comparing this and future systems. Borrowing concepts from information theory, the limits or reliability of an information transmission system can be quantified as the maximal mutual information (MI) (Shannon, 1948). The MI quantifies this limit by maximizing the difference in entropy of the input signal and entropy of the input given the output (see STAR Methods section for equation).

In previous studies, MI was used to quantify the reliability of information transmission through biological networks (Cheong et al., 2011; Hansen and O'Shea, 2015; Uda et al., 2013). Output distributions derived from different inputs will overlap to different extents. Any resulting fluorescence values within the region of overlap cannot be perfectly attributed to a single input signal pattern, resulting in some amount of error in the information transmitted across the promoter and expressed as protein. Systems with low error have lower overlap in their output distributions and have higher MI. Whereas, more overlap results in more error and lower MI (Figure 5A). Although an MI = 1 bit ostensibly implies the ability to distinguish two inputs without error, especially in biological systems, it is more generally a quantification of the ability to infer an input signal from an output signal with some error. Thus, although many biological systems have an MI < 1 (Cheong et al., 2011; Uda et al., 2013; Voliotis et al., 2014), these systems may still have functional biological relevance in resolving two or more inputs in the presence of error (Bowsher and Swain, 2014). Recently, Hansen and O'Shea used MI to characterize the gene expression response to Msn2p signaling using frequency or amplitude modulation. They observed that a single promoter could transmit (to protein expression) 1.58 bits using only amplitude modulation (Hansen and O'Shea, 2015). It was unclear if this is the maximum possible for a single promoter and whether more inputs and combining modes of modulation could provide a higher estimate of information limits.

As we have mapped many input-output responses for all three dynamic modes of modulation (FM, AM, and PWM), we asked what the information capacity limit of our single promoter was. To do this, we randomly selected subsets of the 119 different input signaling patterns and calculated the MI. The same overall dynamic range of mCherry expression was maintained for all subsets of signal input patterns. We repeated this process for increasing numbers of inputs per subset and found that the MI started around 1.45 bits and increased before plateauing near an MI of 1.7 bits (Figure 5B; Table S2). This indicated that the MI was dependent on the number of inputs and required a large screen of the parameter space to measure.

Biological signaling pathways can encode information through the amplitude, frequency, or pulse width of a shared signaling molecule (Batchelor et al., 2011; Hao and O'Shea, 2011; Purvis and Lahav, 2013). However, it is unclear which method is the most reliable. Hansen and O'Shea showed that promoters that bind Msn2p have higher information transduction capacities using amplitude rather than frequency modulation (Hansen and O'Shea, 2015). Here, we found that the MI for each mode of modulation depended on the constant values of the other two parameters (Figure 5C; Table S2). For example, the MI_{AM} increased at low frequency and leveled off at a frequency dependent upon the pulse width. There was only a small dependence of MI_{AM} on pulse width (Figure 5C, top row). There was a less pronounced increase in MI_{FM} as amplitude or pulse width increased (Figure 5C, middle row). MI_{PWM} showed little increase with amplitude but a large increase with frequency (Figure 5C, bottom row). Even when the MI was calculated without normalizing the fluorescence, the same trends were observed

Figure 3. Model captures system behavior and filtering

(A) Schematic of the three-promoter-state model used to fit the experimental data. Blue parameters (k_1, d_1, K, and n) were fit by the model and are unique to the blue light optogenetic system. Red parameters (d_3, k_4, d_4, and k_5) were adapted from literature and are unique to mCherry. Purple parameters (k_2, d_2, and k_3,) were fit by the model and are unique to VP16. $P_{unbound}$, P_{bound}, and P_{active} represent the probabilities of each promoter state.

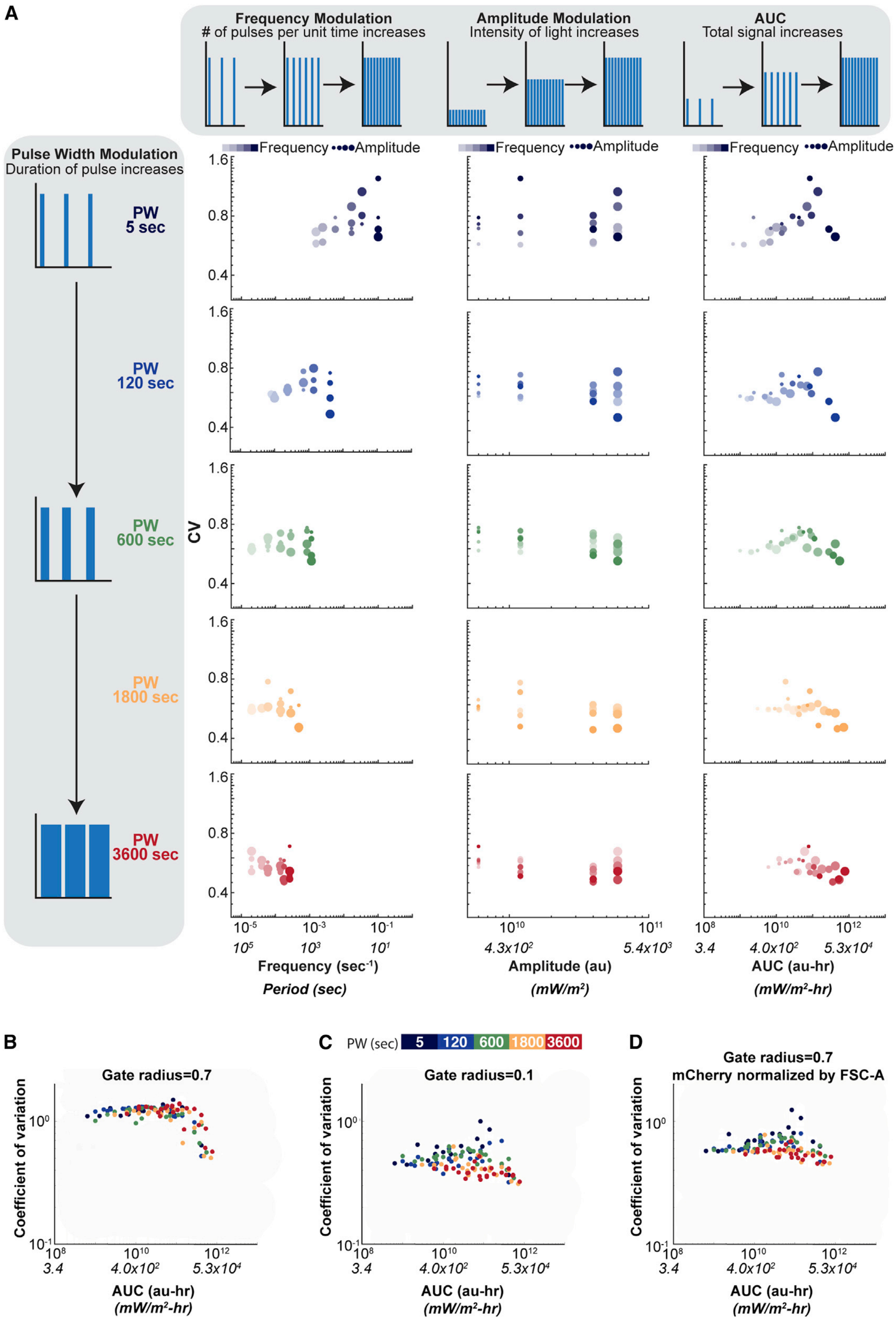
(B) The resulting fold changes for the model using the best-fitting parameter set (R^2 = 0.865) are shown as lines. Experimental data are dots.

(C) Fold change generated by the model shows close similarity to the filtering observed in the experiment in Figure 2B.

(D) Heatmaps show total integrated occupancy of $P_{unbound}$, P_{active}, total integrated mCherry, and endpoint mCherry values with the input light pattern indicated on the left, for four different AUCs. Values were log transformed and normalized to the max value within all heatmaps.

(E) Holding amplitude and AUC constant or pulse width and AUC constant, the integrated occupancy of P_{active} is higher for higher frequency conditions.

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003



(legend on next page)

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

(Figure S5). This indicates that MI measured from AM and PWM depends on what the frequency is and that a large parameter space is necessary to determine the limits of information transmittance for each mode of modulation.

When comparing the maximum MI for each mode of modulation, AM was less reliable (1.12 bits) than PWM (1.23 bits) and FM (1.48 bits). This agrees with our previous assessment that frequency had the greatest effect on fold change. The histograms of the outputs provide some insight into why AM had relatively low MI (Figure 5D). For AM, the amplitudes of 1.2×10^{10}, 4×10^{10}, and 6×10^{10} au (5.3×10^2, 2×10^3, and 3.1×10^3 mW/m^2) had a high degree of overlap and were near saturation when the pulse width and frequency were at high values. Additionally, the 6×10^9 au (2.5×10^2 mW/m^2) histogram had a very broad peak (CV = 0.57 ± 0.02, SEM), which decreased the MI. In fact, low amplitude conditions exhibited broad distributions in general. In contrast, both PWM and FM outputs were more distinguishable, even when the other parameters were at their maximum values. They also exhibited tighter distributions and, therefore, had higher MIs. The reasons our system had higher reliability using FM over AM, which was converse to what was found by Hansen and O'Shea (Hansen and O'Shea, 2015), could be attributed to several factors, including: (1) different mechanisms of activation of Msn2p and VP16; (2) different promoter sequence structure, e.g., location of binding sites relative to transcription binding site; (3) differences in the binding kinetics; and (4) different genomic location of the reporter and, therefore, different initial chromatin states of the promoter. All of these factors could contribute to differences in MI.

Chromatin regulators tune maximum information content

We were able to map the transfer function and quantify the reliability of information transfer for our single synthetic promoter. However, prior work has shown that different promoters can exhibit varying regulatory behaviors, including distinct dynamic ranges of expression, activation kinetics, and noise (Hansen and O'Shea, 2013, 2015; Hao et al., 2013), with a likely explanation for these differences being distinct local chromatin states. Yet, an individual promoter can also exist in diverse chromatin states that might alter the way it responds to input signals (Hansen and O'Shea, 2013; Li et al., 2007). We hypothesized that chromatin state, defined by a complex combination of features, including nucleosome positioning, nucleosome modifications, three-dimensional topology, and the presence of diverse chromatin-regulating proteins could alter both the maximum information transmittable by a single promoter and the nature of its transfer function without any change to its DNA sequence.

To determine if the chromatin landscape could change the MI without altering the promoter sequence, we created a library comprising ZF97-4 fused to each of 101 CRs chosen for a diversity of putative activities and membership in a variety of protein complexes, e.g., SAGA, TFIID, and SWI/SNF (see Table S3 for a list of CRs). These CRs were constitutively recruited to the *CYC1* promoter (Figure 6A). We have previously characterized and confirmed the activity of CRs in this library that have histone modifying domains at this locus via ChIP-seq (Keung et al., 2014). VP16 was then dynamically recruited using the optogenetic system. Given the large number of yeast strains in this library, we focused on varying frequency while keeping amplitude and pulse width constant, as frequency had yielded the greatest MI when VP16 was recruited alone. However, we did additionally map a subset of the CRs using AM and PWM (Figures S6C, S6D, S7C, and S7D), which showed that CRs can tune MI through all three modes of modulation. Many CRs maintained similar overall low (Hda3p, Chz1p, and Hir2p) or high (Rxt3p and Swc3p) MIs across all three modes of modulation.

Four input frequencies were measured: 0 (i.e., dark), 6.7×10^{-4}, 3.3×10^{-2}, and 1×10^{-1} s^{-1} (periods of 0, 1,500, 30, and 10 s). Amplitude and pulse width were held constant at 6×10^{10} au (3.1×10^3 mW/m^2) and 5 s, respectively. With these experimental conditions, we obtained an $\mathrm{MI}_{\mathrm{FM}}$ value for the promoter with each distinct CR recruited (Figure 6B). The $\mathrm{MI}_{\mathrm{FM}}$ ranged from 0.064 ± 0.02 (SEM) for Caf40p to 1.34 ± 0.04 for Arp8p. The values of $\mathrm{MI}_{\mathrm{FM}}$ in Figure 6 are lower than those shown in Figure 5 because only four distinct input light conditions were tested for each CR instead of the 6–7 used in Figure 5. To provide a means for mutual comparison, each $\mathrm{MI}_{\mathrm{FM}}$ was normalized to the $\mathrm{MI}_{\mathrm{FM}}$ of the yeast strain that recruited only VP16 and no CR. Of note, there were strains with large $\mathrm{MI}_{\mathrm{FM}}$ variability among the biological replicates (for example, Hir2p and Nap1p). This is expected, because the VP16-only strain exhibited relatively large variability when fewer inputs were used to calculate MI. As more inputs were included, the MI became more consistent between replicates (Figure 5B). However, even with only four distinct inputs, it was apparent that CRs affected the $\mathrm{MI}_{\mathrm{FM}}$ ($p = 1.18 \times 10^{-7}$, ANOVA). Rxt3p had one of the top two MIs for all three modes of modulation. Additionally, Hir2p, Hda3p, Chz1p, and Caf4p consistently had low MI (less than 0.2 bits) for each mode. This suggests that chromatin may regulate how "fine-tunable" a gene is and how much information can be transmitted reliably via transcription. The changes in MI and output fluorescence were not simply due to overexpression of the CRs but required locus-specific targeting to the reporter (Figures S6A and S6B).

To gain further insight, we also clustered CRs that had low (less than 0.5) and high (above 1.15) $\mathrm{MI}_{\mathrm{FM}}$ (Figures S6E–S6G). Through gene ontology (Cherry et al., 2012), we found significant enrichment within the low-$\mathrm{MI}_{\mathrm{FM}}$ cluster of CRs involved in RNA catabolic process, mitochondrion organization, organelle fission, peroxisome organization, and regulation of translation (Fisher exact test with Bonferroni-Holm correction, $p < 0.05$).

Figure 4. Single-cell measurements capture total population noise

(A) The robust coefficient of variation (CV) calculated for all light conditions. Schematics illustrating each mode of modulation and AUC are shown along the top or left side. Dot color, size, and shade correspond to pulse width, amplitude, and frequency, respectively. Fluorescence values are normalized by FSC-A prior to calculating the CV.

(B–D) Three different strategies for determining the CV were explored, with the resulting CV graphed versus AUC. Gating strategies used are: (B) large radius, (C) small radius, and (D) large radius and FSC-A normalization (see Figure S4 for details). Each dot is the mean of 4–8 biological replicates. For CV values for each replicate, see Table S1.

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

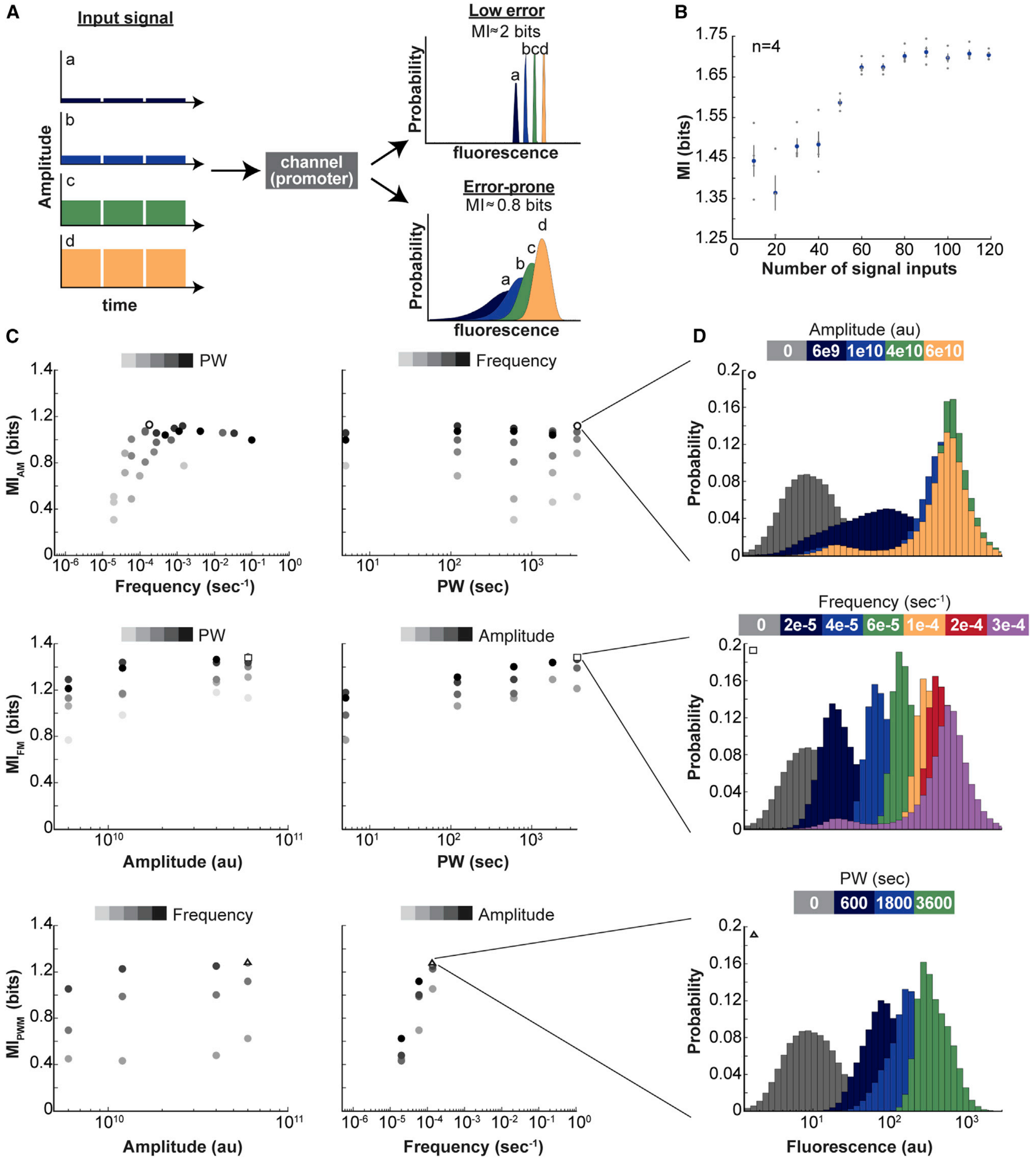


Figure 5. Quantifying the contribution of signaling dynamics to maximum mutual information

(A) Schematic showing potential signal inputs (left) for a promoter that is modeled as a noisy channel and two extreme cases of possible outcomes (right). Outputs with low error and 4 inputs have a maximal mutual information (MI) of approximately 2 bits. The information is transmitted without error. Outputs with a large amount of overlap, or error, and 4 inputs have an MI of approximately 0.8 bits. There is a low amount of information transmitted.

(B) Plot of MI for all three modes of modulation as the number of signal inputs increases. The total fold change range was constant for all combinations of inputs. Error bars are SEM, n = 4 random sets of light conditions.

(C) MI for the different modes of modulation as a function of the other two parameters (held constant for each dot). Each dot is the mean of 4 biological replicates.

(legend continued on next page)

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

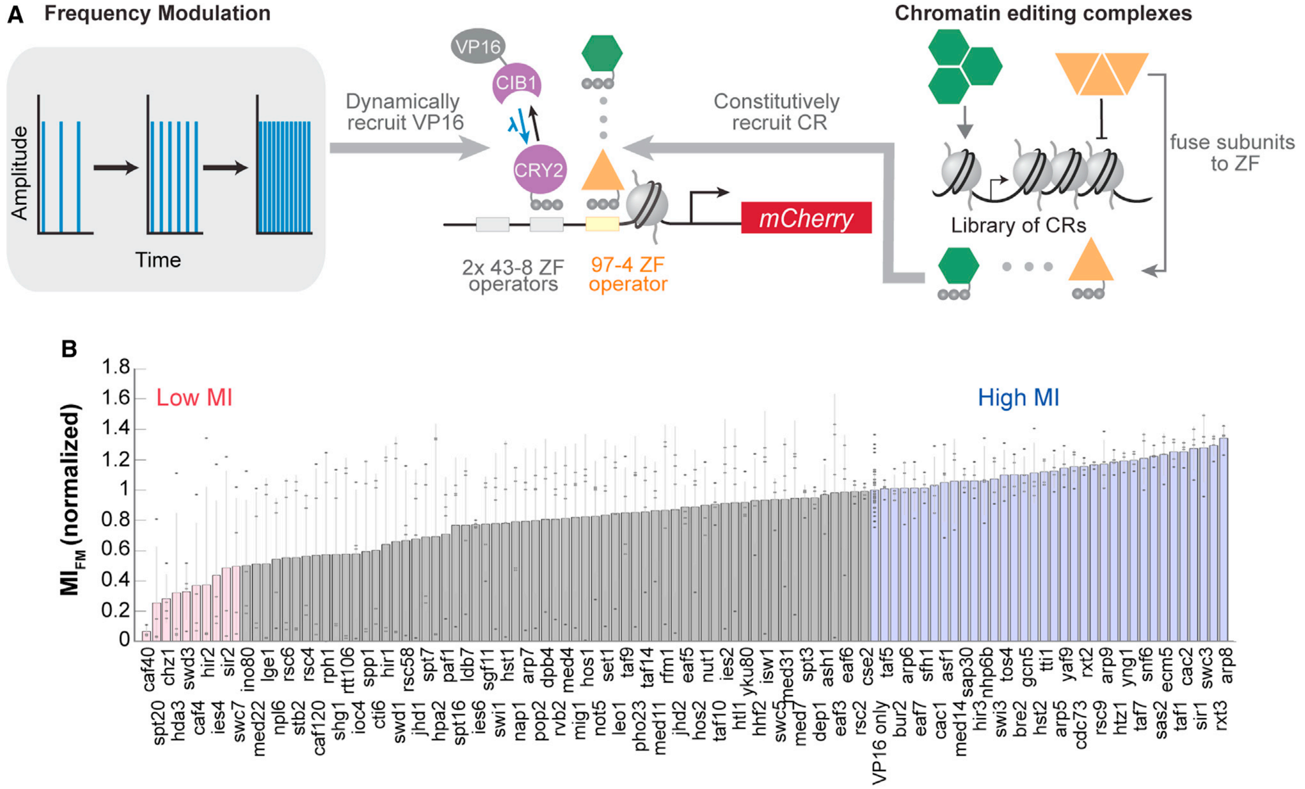


Figure 6. Chromatin regulators tune maximum information content

(A) Eukaryotes utilize a diverse set of protein complexes capable of editing the epigenome. 101 subunits of these complexes were each fused to a 97–4 zinc finger (A, right). This allowed recruitment of the chromatin regulators to the same promoter (center) as the optogenetically controlled VP16 (left).

(B) Maximal mutual information for frequency modulation using frequencies 0, 6.7 × 10^{-4}, 3.3 × 10^{-2}, and 1 × 10^{-1} s^{-1} for the library of ZF97-4-CRs. Error bars are SEM. Gray dots are biological replicates. n = 3–4. VP16 only strain is 48 biological replicates.

When clustered by gene ontology based on function, helicase and nuclease activity had the highest MI$_{FM}$ values, whereas enzyme binding and structural molecule activity had the lowest (Figure S6G). Additionally, CRs involved in DNA recombination and response to DNA damage had the highest MI$_{FM}$ values, suggesting that these CRs may be recruited in natural situations to enhance the reliability of signals to induce DNA repair.

Chromatin regulators diversify the transfer functions achievable by a single promoter

Previous studies have shown that different promoters can exhibit discrete transfer functions (Hansen and O'Shea, 2013, 2016; Harton et al., 2019), not just alterations in MI. We asked whether constitutive recruitment of CRs could alter the transfer function of a single promoter without changes to the DNA sequence. We asked whether CRs alter the qualitative signal-filtering properties of the promoter. For example, can CRs allow the promoter to respond preferentially to low- or high-frequency input signals and not just shift the dynamic range of the output response?

To address this question, we clustered all sets of biological replicates by their pattern of fluorescence output responses to low-, medium-, and high-frequency input signals (6.7 × 10^{-4}, 3.3 × 10^{-2}, and 1 × 10^{-1} s^{-1}) (Figures 7 and S7). This was performed in an unsupervised manner using k-means clustering, with behaviors defined as: low pass, linear, band pass, saturation, and band stop (Figure 7A, top row). This demonstrates that the same promoter can exhibit multiple types of filtering, tunable by CRs. Example CRs for each cluster are shown in the bottom row of Figure 7A. A list of all CRs and their clusters is given in Table S3. Histograms of the output for each example is shown in Figure 7B. It should be noted that, although Hda3p had all of its biological replicates grouped into the band-stop cluster using FM, the histograms indicate (Figure 7B) that the band-stop behavior of Hda3p is not very well defined (hence we displayed these data as partially faded out). Indeed, band stop may be a relatively rare filtering property as Hda3p was the only CR to have all four biological replicates in the band-stop cluster. However, with PWM and AM, Hda3p and Gcn5p were also clustered into the band-stop cluster (Figures S7C

(D) Fluorescence histograms of single-cell distributions for different modes of modulation for the parameter set resulting in the highest MI for the specified mode, indicated by O, □, or Δ in (C) and (D). All values of individual replicates are in Table S2. For our calculation of MI, we used the fluorescence normalized to FSC-A to make our values more comparable with microscopic data that use the average pixel intensity per cell (Hansen and O'Shea, 2015). However, we have included the unnormalized MI values in Table S2; the same trends are seen with the unnormalized values (Figure S5).

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

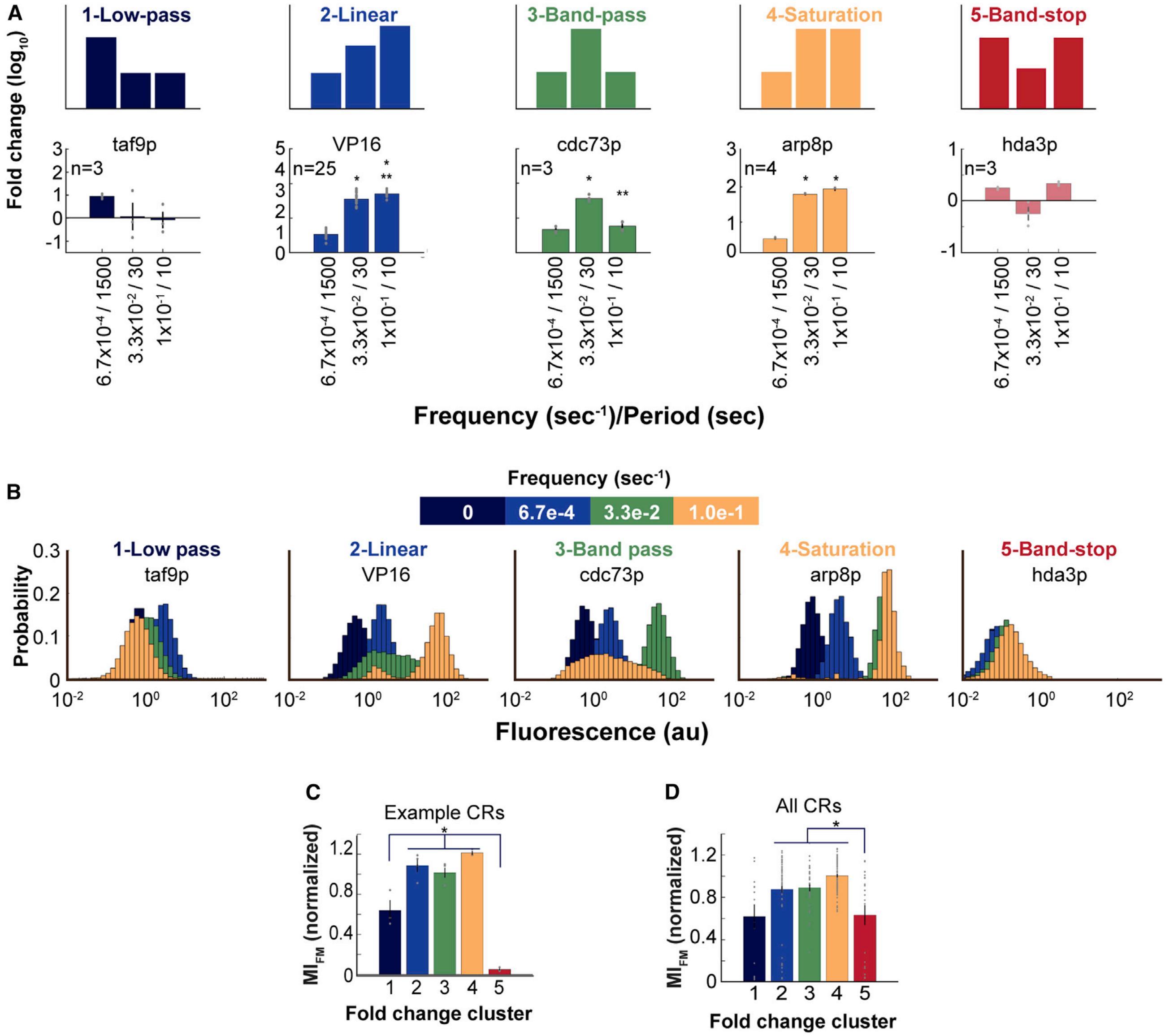


Figure 7. Chromatin regulators diversify the transfer functions achievable by a single promoter

(A) (A, top row) The strains containing ZF97-4 CRs were grouped into five clusters (A, top), using k-means clustering algorithm. Each cluster exhibited a different type of signal filtering. (A, bottom row) Example CRs for each cluster. $^{*}p < 0.05$ compared with frequency = 6.7×10^{-4} s^{-1}, $^{**}p < 0.05$ compared with frequency = 3.3×10^{-2} s^{-1}, Tukey-Kramer post-hoc. Numbers of biological replicates (n) are listed in the figure.

(B) Histograms for a single replicate of the CRs shown in (A). Note that the histograms for hda3p did not demonstrate band-stop behavior.

(C) Average MI_{FM}, for example, CRs shown in (A). $^{*}p < 0.05$, Tukey-Kramer post-hoc.

(D) Average MI_{FM} of all strains within each cluster. $^{*}p < 0.05$, Tukey-Kramer post-hoc. Error bars are SEM. Number of strains for each cluster is in Table S3.

and S7D). Additionally, band-stop transfer functions (e.g., Hda3p) had significantly lower MI_{FM} than the other clusters (Figures 7C and 7D). This is most likely due to narrower dynamic ranges and higher noise, as this type of filtering may be harder to achieve mechanistically. When assessing the MI for each cluster, the trend suggested that CRs may need to sacrifice MI and information transmission capacity to achieve signal-filtering properties such as low-pass and band-stop filtering (Figures 7C and 7D). We found that there were few significantly enriched gene ontology terms for each FM cluster (Figure S7B). For the band-pass cluster, CRs with RNA binding and oxidoreductase activity were significantly enriched using gene ontology based on function. Finally, as expected, the noise of each strain was generally inversely proportional to the fold change (Figures S7A and S7E).

DISCUSSION

Many TFs exhibit pulsatile behavior in response to stress. We addressed the question of how an individual gene interprets this type of dynamic input signal. Using optogenetics to induce 119 distinct dynamic input signals, we mapped the transfer function

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

of an individual promoter as well as the associated noise and reliability of information transmission as a function of dynamic parameters. A three-promoter-state kinetic model was able to capture the transfer function and signal filtering across a broad range of total input AUCs. We further showed that both the qualitative nature of the transfer function and the quantitative maximum information content of the gene could be tuned by constitutive recruitment of CRs to the promoter. This work directly demonstrates the signal processing potential of a single individual gene and develops molecular and computational tools that can be used to harness it.

Epigenome editors, CRs fused to DNA-binding domains, are an increasingly important tool in both biological research and therapeutic development (Adamson et al., 2016; Keung et al., 2014; Liu et al., 2018; Park et al., 2019; Thakore et al., 2016). Their functions have been largely viewed as inducing static changes in state, for example, in which the alteration of histone modifications or recruitment of a transactivator/repressor might lead to up or downregulation of transcription. However, it is now clear that both the dynamic recruitment of editors themselves as well as their impact on the interpretation and processing of other dynamic signals can have profound regulatory effects, including the filtering of different types of dynamic signals well beyond just monotonic on or off control. Such properties have previously been shown to be tunable through mutations in proteins or alterations of protein scaffolds (Bashor et al., 2019; Hao et al., 2013). It is now evident that altering the epigenome can also regulate filtering properties without changing gene or protein sequences. This could be used to confer useful functions such as expressing therapeutic proteins only within a specific range of input signals.

It is also clear that, although the expression strength of an output signal can be tuned by altering the concentration of an input epigenome editor or TF using conventional inducible systems (i.e., LacI or TetR), this type of amplitude-based control may not always be ideal. For example, we found frequency modulation was able to confer a similar output dynamic range as amplitude modulation but with tighter population distributions and therefore greater MI and reliability. Furthermore, when combining all three dynamic parameters, MI was further increased, enhancing the amount of information that could reliably be transmitted by the gene. By achieving more possible output states with lower error for a limited number of inputs, tighter control over output responses is possible and may be particularly important in applications that are sensitive to expression levels such as regulating immune responses.

In addition to informing the design of synthetic biological tools such as epigenome editors, this work suggests consideration of how both the fidelity and inherent transfer functions of natural signaling systems may exhibit considerable differences between cell types and/or over time. The transfer functions and the MI of the same individual genes may switch how they interpret dynamic signals in distinct cell types or in distinct cell states—or during the progression of cancer, aging, or normal development. Many natural systems shown to interpret dynamic signals may also alter their interpretations or transfer functions depending on time and space, including neural cell fate decision making (Imayoshi et al., 2013; Marshall, 1995) and cancer proliferation (Bugaj et al., 2018). Many other biological processes have been linked to dynamic pulsing, such as B-cell activation (Inoue et al., 2016) and responses to radiation (Purvis et al., 2012).

The exploration of dynamic signaling provides opportunities to continue shifting biological engineering to quantitative frameworks borrowed from disciplines in the physical sciences and engineering, but it also contributes to those frameworks due to the distinctive properties of biological systems. For example, this work presents analogies to the concept of dynamic transfer functions common in process control theory, which formalizes the description and prediction of how outputs are controlled by input signals. Yet, as we showed, a gene regulated by chromatin is a highly complex "unit process" that can morph its transfer function to have distinct filtering properties, without a change in gene sequence. Changing the filtering behavior of a transfer function in a conventional unit process such as a chemical reactor might, in contrast to a biological system, require drastic actions such as altering the reactor's material properties or shape.

Information theory also provides a theoretical basis to move from phenomenological frameworks of dose-dependent gene responses that assumes continuous and graded control over gene expression levels, to thinking about true information transmission more rigorously. Notably, we, as well as others (Billing et al., 2019; Cheong et al., 2011; Dubuis et al., 2013; Grabowski et al., 2019; Hansen and O'Shea, 2015; Harton and Batchelor, 2017; Jetka et al., 2019; Selimkhanov et al., 2014; Tkacik et al., 2009; Tudelska et al., 2017; Uda et al., 2013), have shown that these biological unit processes from signal transduction cascades to gene expression have seemingly low information content of less than 1.5 bits. Although we found this limit can be further expanded to ~1.7 bits, this may still seem low and initially presents a conundrum for how biological systems can exert such high-level functions within highly variable and complex environments. However, biological systems function and make decisions with error. In other words, cellular inferences of an input based on an output measurement are often imperfect; therefore, the maximum MI of 1.7 found in this work does not mean eukaryotic genes can interpret only $2^{1.7}$ states. Rather many more states could be inferred, with the MI simply a quantification of how well one can infer the input signal from an output given some amount of uncertainty. Furthermore, there are several strategies that could increase the amount of information transmitted. For example, each gene can respond to multiple TFs and other factors including nucleosome remodelers and three-dimensional topology. Many promoters especially in mammalian systems can be quite large, promoting the ability to sense additional inputs. The diversity of multiple inputs could further increase the MI of genes. Furthermore, linking multiple components into higher order circuits can yield overall greater information transmission as well as lend precision or robustness to input-output responses (Barkai and Leibler, 1997). For all of these reasons, the MI of biological systems may be even higher than currently measured. As a case in point, the simple addition of just one additional input factor (i.e., recruiting CRs such as Arp8p or Rxt3p) was able to increase the MI of the reporter in our system (Figure 6). The ability to increase MI could lead to more complex biological sensors, whereas reducing MI could provide expression

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

systems that are more robust to environmental stressors (Billing et al., 2019). Finally, it is also likely that biological systems have evolved to respond to specific types of naturally occurring input signal; therefore, it would also be interesting to determine how synthetically measured MIs map to natural information transmission and if experimental systems may either be limited or augmented in information transmission capacity.

There are many avenues to expand into and explore. In our work, we relied on endpoint measurements that could be rapidly measured by flow cytometry. However, information can also be stored in the dynamics of the output signal, e.g., the production rate, time delay of repression/activation, or oscillatory behavior. High throughput approaches that can track the output dynamics of thousands of cultures would unlock this potential space for investigation. While we also investigated a single promoter, different promoter structures would likely confer distinct transfer functions (Hansen and O'Shea, 2016). Additional factors that could be explored include the effect of gene duplications, tuning the binding kinetics and/or cooperativity of TFs, assessing species differences, and exploring the contribution of multiple inputs, which would already have nice quantitative frameworks to build upon from process control theory (i.e., multiple input multiple output or "MIMO" control). Continued advances in experimental and computational systems that can handle the large parameter space of dynamic signals will unlock our ability to measure, quantify, and understand information transmission in biological systems and reveal the underpinnings of how limited numbers of components can give rise to the rich complexity of biological functions.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - ○ Cell culture
- METHOD DETAILS
  - ○ Cloning and plasmid construction
  - ○ Oligonucleotides table
  - ○ Cell strain generation
  - ○ Yeast strains table
  - ○ LED matrix construction and calibration
  - ○ LED intensity measurement with power meter
  - ○ Flow cytometry
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ Fold change and noise calculation
  - ○ Maximal mutual information calculation
  - ○ Determining signal filtering clusters
  - ○ Statistical analyses
  - ○ Model selection
  - ○ Deterministic model construction
  - ○ Parameter screen and model fitting

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cels.2021.08.003.

ACKNOWLEDGMENTS

This work was supported by the NSF Emerging Frontiers in Research and Innovation program (1830910), the NCSU Provost's Professional Experience Program (to N.L.), the NCSU Office of Undergraduate Research (to J.Y.L.), post 9/11 GI Bill (to J.B.L.), and an NIH T32 Molecular Biotechnology Training Program Fellowship (T32GM133366, supporting J.B.L. and L.C.). We thank Mike Mantini and members of the Keung Lab for their help in designing and building light matrices. We acknowledge the computing resources provided on Henry2, an NCSU HPC cluster.

AUTHOR CONTRIBUTIONS

J.B.L. and A.J.K. designed research. J.B.L., J.Y.L., N.L., and L.M.C. performed experiments. J.B.L. analyzed data. L.M.C. fitted kinetic model. J.B.L., L.M.C, and A.J.K. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 31, 2020
Revised: May 26, 2021
Accepted: August 4, 2021
Published: August 31, 2021

REFERENCES

Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y.W., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. Cell *167*, 1867–1882.e21. https://doi.org/10.1016/j.cell.2016.11.048.

An-Adirekkun, J.M., Stewart, C.J., Geller, S.H., Patel, M.T., Melendez, J., Oakes, B.L., Noyes, M.B., and McClean, M.N. (2020). A yeast optogenetic toolkit (yOTK) for gene expression control in Saccharomyces cerevisiae. Biotechnol. Bioeng. *117*, 886–893. https://doi.org/10.1002/bit.27234.

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y., and Barkai, N. (2006). Noise in protein expression scales with natural protein abundance. Nat. Genet. *38*, 636–643. https://doi.org/10.1038/ng1807.

Barkai, N., and Leibler, S. (1997). Robustness in simple biochemical networks. Nature *387*, 913–917. https://doi.org/10.1038/43199.

Bashor, C.J., Patel, N., Choubey, S., Beyzavi, A., Kondev, J., Collins, J.J., and Khalil, A.S. (2019). Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. Science *364*, 593–597. https://doi.org/10.1126/science.aau8287.

Batchelor, E., Loewer, A., Mock, C., and Lahav, G. (2011). Stimulus-dependent dynamics of p53 in single cells. Mol. Syst. Biol. *7*, 488. https://doi.org/10.1038/msb.2011.20.

Behar, M., and Hoffmann, A. (2010). Understanding the temporal codes of intra-cellular signals. Curr. Opin. Genet. Dev. *20*, 684–693. https://doi.org/10.1016/j.gde.2010.09.007.

Benzinger, D., and Khammash, M. (2018). Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation. Nat. Commun. *9*, 3521. https://doi.org/10.1038/s41467-018-05882-2.

Billing, U., Jetka, T., Nortmann, L., Wundrack, N., Komorowski, M., Waldherr, S., Schaper, F., and Dittrich, A. (2019). Robustness and information transfer within IL-6-induced JAK/STAT signalling. Commun. Biol. *2*, 27. https://doi.org/10.1038/s42003-018-0259-4.

Bintu, L., Yong, J., Antebi, Y.E., McCue, K., Kazuki, Y., Uno, N., Oshimura, M., and Elowitz, M.B. (2016). Dynamics of epigenetic regulation at the single-cell level. Science *351*, 720–724. https://doi.org/10.1126/science.aab2956.

Bowsher, C.G., and Swain, P.S. (2014). Environmental sensing, information transfer, and cellular decision-making. Curr. Opin. Biotechnol. *28*, 149–155. https://doi.org/10.1016/j.copbio.2014.04.010.

Bugaj, L.J., Sabnis, A.J., Mitchell, A., Garbarino, J.E., Toettcher, J.E., Bivona, T.G., and Lim, W.A. (2018). Cancer mutations and targeted drugs can disrupt dynamic signal encoding by the Ras-Erk pathway. Science *361*, 892. https://doi.org/10.1126/science.aao3048.

Cai, L., Dalal, C.K., and Elowitz, M.B. (2008). Frequency-modulated nuclear localization bursts coordinate gene regulation. Nature *455*, 485–490. https://doi.org/10.1038/nature07292.

Chen, S.Y., Osimiri, L.C., Chevalier, M., Bugaj, L.J., Nguyen, T.H., Greenstein, R.A., Ng, A.H., Stewart-Ornstein, J., Neves, L.T., and El-Samad, H. (2020). Optogenetic control reveals differential promoter interpretation of transcription factor nuclear translocation dynamics. Cell Syst *11*, 336–353.e24. https://doi.org/10.1016/j.cels.2020.08.009.

Cheong, R., Rhee, A., Wang, C.J., Nemenman, I., and Levchenko, A. (2011). Information transduction capacity of noisy biochemical signaling networks. Science *334*, 354–358. https://doi.org/10.1126/science.1204553.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res *40*, D700–D705.

Christiano, R., Nagaraj, N., Fröhlich, F., and Walther, T.C. (2014). Global proteome turnover analyses of the yeasts S. cerevisiae and S. pombe. Cell Rep *9*, 1959–1965. https://doi.org/10.1016/j.celrep.2014.10.065.

Dalal, C.K., Cai, L., Lin, Y.H., Rahbar, K., and Elowitz, M.B. (2014). Pulsatile dynamics in the yeast proteome. Curr. Biol. *24*, 2189–2194. https://doi.org/10.1016/j.cub.2014.07.076.

Dubuis, J.O., Tkacik, G., Wieschaus, E.F., Gregor, T., and Bialek, W. (2013). Positional information, in bits. Proc. Natl. Acad. Sci. USA *110*, 16301–16308. https://doi.org/10.1073/pnas.1315642110.

Eldar, A., and Elowitz, M.B. (2010). Functional roles for noise in genetic circuits. Nature *467*, 167–173. https://doi.org/10.1038/nature09326.

Elowitz, M.B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. Nature *403*, 335–338. https://doi.org/10.1038/35002125.

Gonze, D., and Abou-Jaoudé, W. (2013). The Goodwin model: behind the hill function. PLoS One *8*, e69573. https://doi.org/10.1371/journal.pone.0069573.

Grabowski, F., Czyż, P., Kochańczyk, M., and Lipniacki, T. (2019). Limits to the rate of information transmission through the MAPK pathway. J. R. Soc. Interface *16*, 20180792. https://doi.org/10.1098/rsif.2018.0792.

Gregor, T., Tank, D.W., Wieschaus, E.F., and Bialek, W. (2007). Probing the limits to positional information. Cell *130*, 153–164. https://doi.org/10.1016/j.cell.2007.05.025.

Hansen, A.S., and O'Shea, E.K. (2013). Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression. Mol. Syst. Biol. *9*, 704. https://doi.org/10.1038/msb.2013.56.

Hansen, A.S., and O'Shea, E.K. (2015). Limits on information transduction through amplitude and frequency regulation of transcription factor activity. eLife *4*, e06559. https://doi.org/10.7554/eLife.06559.

Hansen, A.S., and O'Shea, E.K. (2016). Encoding four gene expression programs in the activation dynamics of a single transcription factor. Curr. Biol. *26*, R269–R271. https://doi.org/10.1016/j.cub.2016.02.058.

Hao, N., Budnik, B.A., Gunawardena, J., and O'Shea, E.K. (2013). Tunable signal processing Through modular control of transcription factor translocation. Science *339*, 460–464. https://doi.org/10.1126/science.1227299.

Hao, N., and O'Shea, E.K. (2011). Signal-dependent dynamics of transcription factor translocation controls gene expression. Nat. Struct. Mol. Biol. *19*, 31–39. https://doi.org/10.1038/nsmb.2192.

Harton, M.D., and Batchelor, E. (2017). Determining the limitations and benefits of noise in gene regulation and signal transduction through single cell, microscopy-based analysis. J. Mol. Biol. *429*, 1143–1154. https://doi.org/10.1016/j.jmb.2017.03.007.

Harton, M.D., Koh, W.S., Bunker, A.D., Singh, A., and Batchelor, E. (2019). p53 pulse modulation differentially regulates target gene promoters to regulate cell fate decisions. Mol. Syst. Biol. *15*, e8685. https://doi.org/10.15252/msb.20188685.

Imayoshi, I., Isomura, A., Harima, Y., Kawaguchi, K., Kori, H., Miyachi, H., Fujiwara, T., Ishidate, F., and Kageyama, R. (2013). Oscillatory control of factors determining multipotency and fate in mouse neural progenitors. Science *342*, 1203–1208. https://doi.org/10.1126/science.1242366.

Inoue, K., Shinohara, H., Behar, M., Yumoto, N., Tanaka, G., Hoffmann, A., Aihara, K., and Okada-Hatakeyama, M. (2016). Oscillation dynamics underlie functional switching of NF-κB for B-cell activation. NPJ Syst. Biol. Appl. *2*, 16024. https://doi.org/10.1038/npjsba.2016.24.

Jetka, T., Nienałtowski, K., Winarski, T., Błoński, S., and Komorowski, M. (2019). Information-theoretic analysis of multivariate single-cell signaling responses. PLoS Comput. Biol. *15*, e1007132. https://doi.org/10.1371/journal.pcbi.1007132.

Kennedy, M.J., Hughes, R.M., Peteya, L.A., Schwartz, J.W., Ehlers, M.D., and Tucker, C.L. (2010). Rapid blue-light-mediated induction of protein interactions in living cells. Nat. Methods *7*, 973–975. https://doi.org/10.1038/nmeth.1524.

Keung, A.J., Bashor, C.J., Kiriakov, S., Collins, J.J., and Khalil, A.S. (2014). Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. Cell *158*, 110–120. https://doi.org/10.1016/j.cell.2014.04.047.

Khalil, A.S., Lu, T.K., Bashor, C.J., Ramirez, C.L., Pyenson, N.C., Joung, J.K., and Collins, J.J. (2012). A synthetic biology framework for programming eukaryotic transcription functions. Cell *150*, 647–658. https://doi.org/10.1016/j.cell.2012.05.045.

Khmelinskii, A., Keller, P.J., Bartosik, A., Meurer, M., Barry, J.D., Mardin, B.R., Kaufmann, A., Trautmann, S., Wachsmuth, M., Pereira, G., et al. (2012). Tandem fluorescent protein timers for in vivo analysis of protein dynamics. Nat. Biotechnol. *30*, 708–714. https://doi.org/10.1038/nbt.2281.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705. https://doi.org/10.1016/j.cell.2007.02.005.

Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237–1251. https://doi.org/10.1016/j.cell.2013.02.014.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707–719. https://doi.org/10.1016/j.cell.2007.01.015.

Liu, H.T., Yu, X.H., Li, K.W., Klejnot, J., Yang, H.Y., Lisiero, D., and Lin, C.T. (2008). Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in Arabidopsis. Science *322*, 1535–1539. https://doi.org/10.1126/science.1163927.

Liu, Y.X., Yu, C., Daley, T.P., Wang, F.Y., Cao, W.S., Bhate, S., Lin, X.Q., Still, C., Liu, H.L., Zhao, D.H., et al. (2018). CRISPR activation screens systematically identify factors that drive neuronal fate and reprogramming. Cell Stem Cell *23*, 758–771.e8. https://doi.org/10.1016/j.stem.2018.09.003.

Louvion, J.F., Havaux-Copf, B., and Picard, D. (1993). Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast. Gene *131*, 129–134. https://doi.org/10.1016/0378-1119(93)90681-r.

Maheshri, N., and O'Shea, E.K. (2007). Living with noisy genes: how cells function reliably with inherent variability in gene expression. Annu. Rev. Biophys. Biomol. Struct. *36*, 413–434. https://doi.org/10.1146/annurev.biophys.36.040306.132705.

Marshall, C.J. (1995). Specificity of receptor tyrosine kinase signaling: transient VERSUS sustained extracellular signal-regulated kinase activation. Cell *80*, 179–185. https://doi.org/10.1016/0092-8674(95)90401-8.

Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature *441*, 840–846. https://doi.org/10.1038/nature04785.

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

Park, M., Patel, N., Keung, A.J., and Khalil, A.S. (2019). Engineering epigenetic regulation using synthetic read-write modules. Cell *176*, 227–238.e20. https://doi.org/10.1016/j.cell.2018.11.002.

Polstein, L.R., and Gersbach, C.A. (2015). A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. Nat. Chem. Biol. *11*, 198–200. https://doi.org/10.1038/nchembio.1753.

Purvis, J.E., Karhohs, K.W., Mock, C., Batchelor, E., Loewer, A., and Lahav, G. (2012). p53 dynamics control cell fate. Science *336*, 1440–1444. https://doi.org/10.1126/science.1218351.

Purvis, J.E., and Lahav, G. (2013). Encoding and decoding cellular information through signaling dynamics. Cell *152*, 945–956. https://doi.org/10.1016/j.cell.2013.02.005.

Rademacher, A., Erdel, F., Trojanowski, J., Schumacher, S., and Rippe, K. (2017). Real-time observation of light-controlled transcription in living cells. J. Cell Sci. *130*, 4213–4224. https://doi.org/10.1242/jcs.205534.

Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S., and Elowitz, M.B. (2005). Gene regulation at the single-cell level. Science *307*, 1962–1965. https://doi.org/10.1126/science.1106914.

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. Methods Ecol. Evol. *1*, 103–113. https://doi.org/10.1111/j.2041-210X.2010.00012.x.

Selimkhanov, J., Taylor, B., Yao, J., Pilko, A., Albeck, J., Hoffmann, A., Tsimring, L., and Wollman, R. (2014). Systems biology. Accurate information transmission through dynamic biochemical signaling networks. Science *346*, 1370–1373. https://doi.org/10.1126/science.1254933.

Shannon, C.E. (1948). A mathematical theory of communication. Bell Syst. Tech. J. *27*, 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x.

Singh, P. (2015). BlahutArimoto.m (GitHub).

Slonim, N., Atwal, G.S., Tkacik, G., and Bialek, W. (2005). Information-based clustering. Proc. Natl. Acad. Sci. USA *102*, 18297–18302. https://doi.org/10.1073/pnas.0507432102.

Thakore, P.I., Black, J.B., Hilton, I.B., and Gersbach, C.A. (2016). Editing the epigenome: technologies for programmable transcription and epigenetic modulation. Nat. Methods *13*, 127–137. https://doi.org/10.1038/nmeth.3733.

Tkacik, G., Walczak, A.M., and Bialek, W. (2009). Optimizing information flow in small genetic networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *80*, 031920. https://doi.org/10.1103/PhysRevE.80.031920.

Tudelska, K., Markiewicz, J., Kochańczyk, M., Czerkies, M., Prus, W., Korwek, Z., Abdi, A., Błoński, S., Kaźmierczak, B., and Lipniacki, T. (2017). Information processing in the NF-kappa B pathway. Sci. Rep. *7*, 15926. https://doi.org/10.1038/s41598-017-16166-y.

Uda, S., Saito, T.H., Kudo, T., Kokaji, T., Tsuchiya, T., Kubota, H., Komori, Y., Ozaki, Y.-i., and Kuroda, S. (2013). Robustness and compensation of information transmission of signaling pathways. Science *341*, 558–561. https://doi.org/10.1126/science.1234511.

Voliotis, M., Perrett, R.M., McWilliams, C., McArdle, C.A., and Bowsher, C.G. (2014). Information transfer by leaky, heterogeneous, protein kinase signaling systems. Proc. Natl. Acad. Sci. USA *111*, E326–E333. https://doi.org/10.1073/pnas.1314446111.

Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., and Brown, P.O. (2002). Precision and functional specificity in mRNA decay. Proc. Natl. Acad. Sci. USA *99*, 5860–5865. https://doi.org/10.1073/pnas.092538799.

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
NEB Turbo Competent *E. coli*	NEB	C2984I
NEB Stable Competent *E. coli*	NEB	C3040I
Chemicals, Peptides, and Recombinant Proteins		
Anhydrotetracycline	Fisher Scientific	Cat# 50-595-757
IPTG	Fisher Scientific	Cat# BP175510
Beta-estradiol	Fisher Scientific	Cat#AAL0380103
Cycloheximide	Fisher Scientific	Cat#AAJ6690103
YNB+nitrogen	Sigma	Cat# Y0626
CSM	Sunrise Scientific	Cat#1001-100
CSM-LEU	Sunrise Scientific	Cat#1005-010
CSM-LEU-URA	Sunrise Scientific	Cat#1038-010
CSM-LEU-HIS	Sunrise Scientific	Cat#1011-010
CSM-LEU-HIS-URA	Sunrise Scientific	Cat#1015-010
CSM-LEU-HIS-URA-TRP	Sunrise Scientific	Cat#1002-010
Adenine hemisulfate	Sunrise Scientific	Cat#1905-010
Deposited Data		
Flow cytometry data	This paper	GitHub-Zenodo:10.5281/zenodo.5116822
Experimental Models: Organisms/Strains		
Saccharomyces cerevisiae	Stratagene	YPH500
Oligonucleotides		
See Method details (Oligonucleotides table)	N/A	N/A
Recombinant DNA		
See Table S4 (Plasmids Table)	N/A	N/A
Software and Algorithms		
MATLAB 2018b	MathWorks	https://www.mathworks.com/downloads/
Arduino IDE	Arduino	https://www.arduino.cc/en/Main/Software
TinkerCAD	Autodesk	https://www.tinkercad.com/
Dynamic-Transfer-Functions	This paper	GitHub-Zenodo:10.5281/zenodo.5116822
MATLAB 2020b	MathWorks	https://www.mathworks.com/downloads/
Jupyter Notebook (anaconda3) 6.1.4	Jupyter	https://jupyter.org/install
Python 3.7.7	Python	https://www.python.org/downloads/release/python-377/
Deposited Model Data	This paper	GitHub-Zenodo:10.5281/zenodo.5116822
Deposited Model Code	This paper	GitHub-Zenodo:10.5281/zenodo.5116822

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Albert J. Keung (ajkeung@ncsu.edu).

Materials availability

Plasmids generated in this study have been deposited at Addgene as plasmids 174063–174070.

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

Data and code availability

- **Source data statement**: Flow cytometry source data have been deposited at GitHub and are publicly available as of the date of publication. DOI is listed in the key resources table.
- **Code statement**: All original code has been deposited at GitHub and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

The background cell line for all experiments in this study was YPH500 *(α, ura3-52, lys2-801, ade2-101, trp1Δ63, his3Δ200, leu2Δ1)* (Stratagene). Cells were cultured in synthetic drop-out media or complete media made (Sunrise Scientific) with YN-B from Sigma and 2% w/v glucose. Our host strain was generated by genomically integrating an expression cassette that constitutively expresses TetR, LacI, and GEV (Louvion et al., 1993) (cloned into single-integrating plasmid pNH607[*HO*]). Constitutive expression of the repressors in glucose-containing media ensured low basal levels of expression of ZF-CRY2 and CIB1-VP16 from the engineered *GAL1* promoters, which was relieved by the respective addition of the chemical inputs, ATC and IPTG, along with β-estradiol to the medium.

METHOD DETAILS

Cloning and plasmid construction

All plasmid constructs were created using standard molecular biology techniques and Gibson isothermal assembly. Plasmids were grown and prepared from either NEB Turbo or Stable competent cells. The CR plasmid library was synthesized as previously described (Keung *et al.*, 2014). In short, primer sequences were obtained from the *Saccharomyces* Genome Database (SGD). These primers (synthesized by Integrated DNA Technologies) were used to amplify full length CR ORFs from wild-type yeast (BY4742). SbfI and NotI flanking restriction sites were used to ligate the PCR products to the C-terminus of (3xFLAG)-(nuclear localization sequence)-(97-4 zinc finger array)-(17 amino acid glycine-serine linker) using plasmid pJL50.

Oligonucleotides table

Oligo ID	Sequence
JLp161	ggatcactagtGGTACCGAAGTACGGATTAGA
JLp162	gatcacgcgtCCACACAATTATAAGCAAAGGG
JLp46	ggactcctgcaggAAGATGGACAAAAAGACTATAGTTTG
JLp47	agcggccgcTCATTTGCAACCATTTTTTCCCA
JLp251	ggataaaatgtgataactaatcagcggtacAGAGTGAGGACTCGAAAATATTAAT
JYLp4	accagtgaataattcttcacctttagacatTTTAATATCTAGATTAGTGTGTGTATTTGT
JYLp5	acacaaacacaaatacacacactaatctagatattaaaATGTCTAAAGGTGAAGAATTATTCAC
JYLp6	ctagccgcggtaccaagcttactcgagTTACACCTGCCTTGAGGGA
JLp163	gctcgagctgcagatgaatcgtagatacGGAGGTTCTGGCGGTGGAAGT
JLp169	cttcggtaccactagtggatccgaattcgcCAAAGCCGAATCCACCACGG

(*Continued on next page*)

Continued

Oligo ID	Sequence
JLp321	ggaggtagcggtggtggcagcggttCCTGCAGGtaagcggccgccaccgcggtggagctctaagcaaata
JLp322	tatttgcttagagctccaccgcggtggcggccgcttaCCTGCAGGaaccgctgccaccaccgctacctcc
JLp342	ggactggtaccATAGCTTCAAAATGTTTCTACTC
JLp343	agatcgggcccATTAAAACTTAGATTAGATTGCTAT
JLp46	ggactcctgcaggAAGATGGACAAAAAGACTATAGTTTG
JLp47	agcggccgcTCATTTGCAACCATTTTTTCCCA
JYLp3	ggataaaatgtgataactaatcagcACTAGTcagatccgccaggc
JYLp7	caaaggtagttccctcaaggcaggtgtaaCTCGAGTAAGCTTGGTACCG
JYLp8	cttagagctccaccgcggtggcggccgcCTTCGAGCGTCCCAAAACCT
JLp1	gaattcccgggGCGAGCGCCGAAGCTAG
JLp2	ggatcggatccTCAGTATCTACGATTCATCTGCAGC

Cell strain generation

Strains were constructed by sequential plasmid transformations using standard lithium acetate-based transformation techniques. Plasmids were first linearized using PmeI or SbfI. Following transformation, cells were grown on selective auxotrophic minimal media (Sunrise). Strains are listed in the Yeast strains Table, while plasmids are listed in Table S4. *TRP4* auxotrophic marker constructs were integrated into AmpR of the *LEU2* construct.

Yeast strains table

	Marker loci					
Strain ID	***HO***	***URA3***	***TRP4* ***	***LEU2***	***HIS3***	**Figure**
Y11	pNH607					1B, 2, 3,4, 5, 6, 7
JY28	pNH607			pJL29		
JY29	pNH607	pJL30		pJL29		1B
JY138	pNH607	pJL30		pJL29	pJL45	1B
JY145	pNH607	pJL30	pJL38	pJL29	pJL32	1B, 2, 3,4, 5, 6, 7
CR library	pNH607	pJL30	pJL50-EE	pJL29	pJL32	5, 6, 7
JY30	pNH607	pJL30		pJL29	pJL32	

LED matrix construction and calibration

Three LED matrices were made. Each had a LED housing unit 3D printed using black polylactic acid plastic. Each housing unit was designed to fit a standard 96-well plate with a single, programmable LED for each well. The plans for the housing unit were created in TinkerCad and are available upon request. Female socket pins were glued to the housing unit to connect to each LED. 60 or 92 blue LEDs (Chanzon, 100F5T-YT-WH-BL) were connected to 220 Ω resistors before being connected to 16-channel servo driver breakout boards (PCA9685, Adafruit). Three or five breakout boards were used for each 60 LED or 92 LED matrix, respectively. In addition, 12 LEDs were controlled directly from the PWM pins (0-11) on the Arduino Due. Each matrix was controlled by an Arduino Due, using I2C. Arduino code was written using the Arduino IDE to control the pulse width, intensity, and frequency of light pulses.

Calibration of the LEDs was done by attaching each LED matrix to a black 96-well plate with a flat, clear bottom (Corning, 3788) and taking 59 images across each well using a microscope (Nikon Ti-Eclipse, 20x SP objective, z=4486 μm) through a DAPI filter cube (Chroma Technology, 96360) with exposure time set to 100ms. The pixel intensity was extracted using a custom Matlab code. For a single well, pixel intensities for each image were read using the *imread* function. The total pixel intensities for each image were summed and then divided by the number of images. The average intensity was also determined for a well without an LED. This value was subtracted from all wells' intensities to get the working LED intensity. This was done for three Arduino inputs and fit to a line for each well. The calculated values were used as initial inputs for the intensities used for the experiments. The intensities were then checked and adjusted before each experiment to be within 20 percent of the desired intensity.

LED intensity measurement with power meter

The LED intensities can be converted to mW/m^2 using Figure S1 panel G. The power meter measurements were taken using a PM100D power meter (ThorLabs) with a S140C probe. A M134L01 fiber patch cable (0600 μm core, 0.5 NA, FC/PC to SMA, 1 m length) was connected to the probe via the FC/PC connector. For each well, the SMA connector was held against the bottom of a clear, flat-bottom plate (Corning, 3788) connected to the LED matrix. Multiple readings were taken at various locations for each well, and the mean was plotted and fitted to a line as shown in Figure S1 panel G.

Flow cytometry

Yeast colonies were picked from plates and cultured 24-48 h in the appropriate auxotrophic SD media. Cultures were diluted to ~0.1 OD600 with auxotrophic dropout media that contained 0.4 μg/mL ATC, 10mM IPTG, 5μM of beta-estradiol, and 0.02 mg/mL adenine. Cells were incubated at 30°C and 900 RPM, in the dark, for 8-9 h to allow for expression of ZF-CRY2 and CIB1-VP16. Cells were then diluted 1:30 with 200 μL SD-complete media, containing the same chemicals as above, into U-bottom, black 96-well plates (Costar, 3792). Samples were prepared as much as possible in a red light environment to reduce premature binding of CIB1 and CRY2. Plates were attached to the LED matrices and incubated at 30°C for 14 h at 500 RPM. Replicate plates were grown in the dark. The shaking speed was reduced to prevent damage of the LED matrices.

Prior to flow cytometry, 100 μL of 0.03 mg/mL of cycloheximide was added to each sample. Samples were then incubated in the dark at room temperature for 1 h to allow for mCherry maturation. Fluorescent measurements were taken using a MACSQuant VYB (Miltenyi Biotec). A maximum of 20,000 events were collected per sample. Plates were stored at 4°C while waiting for other plates to be run on flow cytometer. All samples were run within 8 h of adding cycloheximide.

QUANTIFICATION AND STATISTICAL ANALYSIS

Fold change and noise calculation

All samples were gated using SSC-A and FSC-A, using a custom MATLAB code based on methods described previously (Newman *et al.*, 2006). To summarize, the FSC-A and SSC-A were natural log transformed. Cells outside a circle of radius 0.7 around the median FSC-A and SSC-A were excluded from further analyses. Any gated samples with less than 250 events were also excluded from further analyses.

The population medians of the fluorescence distributions were calculated for the gated populations. For the most figures (except Figures 4B and 4C), the fluorescence was normalized by the FSC-A (size) before continuing. This was done to be consistent with microscopic measurements from a previous study that uses the mean pixel value per cell (Hansen and O'Shea, 2015). The autofluorescence value of *S. cerevisiae* YPH500 cells harboring no genomic integrations was subtracted from these values. "Fold change" values were calculated as the ratio of fluorescence values from cells exposed to a given blue light pattern to those from cells grown without blue light. Four isogenic strains were grown for each light condition. The "coefficient of variation", or CV is the robust CV calculated using the equation: 0.5 * [intensity(at 84.13 percentile) - intensity(at 15.87 percentile)] / median. Outliers were identified using MATLAB's *isoutlier* function, which classifies values as an outlier if it is more than three scaled median absolute deviations away from the median fold change or CV. Any outliers were excluded from the means graphed in Figures 2 and 3. Any samples with less than 2000 cells were also excluded.

To minimize the variability due to the large number of plates in the CR screens (Figure 6), each plate with blue light was normalized to the strain with VP16 only (JY145) with light always on and light intensity at 6x10^{10} au, which was grown in the same plate. Each plate without blue light was normalized to VP16 only (JY145) with no light, grown in the same plate. Population medians were used to calculate the fold change.

Maximal mutual information calculation

The maximal mutual information was found as previously described in (Cheong *et al.*, 2011; Hansen and O'Shea, 2015; Shannon, 1948). For each sample, events were gated as described in the *Flow Cytometry* section. Then the mCherry measurements were normalized to the FSC-A measurements. Any samples with less than 200 events were excluded. The responses were discretized using logarithmically sized bins. The mutual information I(R;S), measured in bits, was calculated by

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

$$I(R;S) = H(R) - H(R,S) \qquad \text{(Equation 1)}$$

$$I(R;S) = \sum_{i,j} p\left(R_i, S_j\right) log_2\left(\frac{p(R_i, S_j)}{p(R_i)p(s_j)}\right) \qquad \text{(Equation 2)}$$

where S is the signal input and R is the observed response output. H(R) is the entropy of the response and H(R,S) is the entropy of the response given the signal. The response given the signal, p(Ri,Sj) is the mCherry fluorescent measurements. The response distributions, given by p(R) and p(s), are unknowns. The maximal mutual information was found by solving the optimization in Equation 2.

$$MI(R;S) = I(R;S) \text{ for } \sum_i p(S_i) = 1;\ p(S_i) \geq 0. \qquad \text{(Equation 3)}$$

The above optimization was solved using the Blahut-Arimoto algorithm from code written by Piyush Singh (Singh, 2015). The MI was corrected for bias due to the number of bins by varying the number of bins from 5 to 50. The unbiased MI was calculated as the mean of MIs calculated using 21-41 bins, which is within the plateau region of MI versus number of bins. The MI was then corrected for under-sampling using jackknife sampling as described previously (Cheong *et al.*, 2011; Hansen and O'Shea, 2015; Slonim et al., 2005). The means shown in Figures 5 and 6 are of the unbiased MIs from 3-4 isogenic strains.

Determining signal filtering clusters

Clustering was completed on individual replicates for all the epigenome editors. To discover the clusters depicted in Figure 6, strains that had low fold change and low variability of fold change among frequencies were removed. The fold changes were then logarithmically transformed. The remaining strains were grouped into 5 clusters using the *kmeans* function in MATLAB with correlation as the distance metric. The centroids from the resulting clusters were slightly modified to fit the behaviors in Figure 6C. The centroids are as follows: Cluster 1: 0.756341421875678, -0.3, -0.338099818683020; Cluster 2: -0.736510820962514, 0.102405013200597, 0.65; Cluster 3: -0.351484685217366, 0.769986528990965, -0.418501843773599; Cluster 4: -0.775512363928304, 0.550835601762814,0.3; Cluster 5: 0.0178768566235528, -0.685625005390339, 0.667748148766786. The fold changes were then reclustered using these centroids with the *pdist2* function, again with correlation as the distance metric.

Statistical analyses

Details of statistical tests can be found in the figure captions. One-way and n-way ANOVA tests were performed using either the *anova1* or nanova functions, respectfully. For the comparison among multiple conditions, a Tukey's honest significant difference criterion (T-K analysis) was used via the *multcompare* function in MATLAB with a 95 percent confidence interval. The analysis of covariance (ANCOVA) was performed using the *aoctool* function. The Fisher-exact tests were performed using MATLAB function, *fishertest*. Data distributions were directly measured by flow cytometry. For all analyses, see GitHub.

Model selection

Twenty different two-, three-, and four-promoter-state models were tested with a variety of architectures and between 3 and 9 fitted parameters. The models were screened using two metrics: first, the residual sum of the squares on the model outputted endpoints and experimental endpoints; and second, comparison to the expected time course curve shape based on literature (Hansen and O'Shea, 2013; Harton et al., 2019; Benzinger and Khammash, 2018). Six models with various number of promoter states and number and placement of Hill functions were selected for more thorough analysis (Figure S3B). These models were compared using stratified K-fold cross validation with the scikit-learn Python module, with four folds created based on preserving the percentage of samples for each pulse width. For each training set and each model, 1000 random kinetic parameter sets generated by latin hypercube sampling were run and the parameter set that resulted in the highest R^2 value was used to run the associated testing set. The overall score for each model was calculated as the average R^2 across the four testing sets (Figure S3C). The three-promoter-state, one Hill and three-promoter-state, two Hill models had the highest cross validation scores, but given the similarity of the scores the three-promoter-state, one Hill model was chosen to avoid unnecessary complexity in the model.

Deterministic model construction

To better understand the relationship between dynamic inputs and gene expression outputs in our system, a deterministic kinetic model was created, which is described by the following set of ODEs:

$$\frac{dP_{unbound}}{dt} = d_1 P_{bound} - \frac{k_1 I(t)^n}{K^n + I(t)^n} P_{unbound} \qquad \text{(Equation 1)}$$

$$\frac{dP_{bound}}{dt} = \frac{k_1 I(t)^n}{K^n + I(t)^n} P_{unbound} + d_2 P_{active} - (d_1 + k_2) P_{bound} \qquad \text{(Equation 2)}$$

Please cite this article in press as: Lee et al., Mapping the dynamic transfer functions of eukaryotic gene regulation, Cell Systems (2021), https://doi.org/10.1016/j.cels.2021.08.003

$$\frac{dP_{active}}{dt} = k_2 P_{bound} - d_2 P_{active} \qquad \text{(Equation 3)}$$

$$\frac{d[mRNA]}{dt} = k_3 P_{active} - d_3 [mRNA] \qquad \text{(Equation 4)}$$

$$\frac{d[mCherry_1]}{dt} = k_4 [mRNA] - (d_4 + k_5)[mCherry_1] \qquad \text{(Equation 5)}$$

$$\frac{d[mCherry_2]}{dt} = k_5 [mRNA] - d_4 [mCherry_2] \qquad \text{(Equation 6)}$$

Here, d_1 and k_1 are the transition rates between the unbound and bound promoter states and d_2 and k_2 are the transition rates between the bound and active promoter states. The transcription, translation, and maturation rates are k_3, k_4, and k_5, respectively. The mRNA and mCherry degradation rates are d_3 and d_4. Four parameters were experimental constants d_3, k_4, d_4, and k_5, and seven were model fitted parameters, d_1, k_1, n, K, d_2, k_2, d_3, k_3. The fit of the model was assessed using the coefficient of determination. $P_{unbound}$, P_{bound}, P_{active}, represent the probability of the promoter being in a given state and are each between 0 and 1 and must sum to 1 at any point in time. A Hill function was used to describe the transition between $P_{unbound}$ and P_{bound}. The input function is I(t), and is based on the PWM, AM, and FM of the light condition, and the input amplitudes were 6 x 10^1 to 6 x $10e^2$ au rather than 6 x 10^9 to 6 x 10^{10} au to prevent overflow error.

Parameter screen and model fitting

To fit the model to the data, sets of parameters fit by the model (d_1, k_1, n, K, d_2, k_2, d_3, and k_3) were stochastically generated using Latin Hypercube Sampling (LHS) using the Surrogate Modeling Toolbox in Python. The ODEs were solved numerically using *odeint* in Python, and the model outputs at 14 h were compared to the experimental values using the residual sum of the squares. The fitting was performed in two steps: initially, 1500 randomly generated sets of parameters, each sampled over an appropriate range, were run through the model. Then, the parameter set that resulted in the highest R^2 value was used to "fine-tune" the LHS sampling range, and new sets were generated and run through the model. The fine-tuning process was iterated through a total of over 5500 parameter sets, resulting in a R^2 of 0.865 (Figure 3B). The model fitted parameters are shown in Figure S3C.

The experimental parameters unique to mCherry (d_3, k_4, d_4, and k_5), were treated as constants determined by literature values. The degradation rates were found by randomly sampling within ranges provided by literature and fit to the entire dataset while all other parameters were held constant. 3000 parameter sets were tested, with the mRNA degradation rate (d_3) ranging from 1.08 x 10^{-4} – 1.12 x 10^{-4} s^{-1} (Chen et al., 2020; Wang et al., 2002) and mCherry degradation rate (d_4) ranging from 1.83e-5 – 2.56e-5 s^{-1} (Harton et al., 2019; Christiano et al., 2014). Both degradation rates include an additional 2.5 h to account for dilution, and the best parameters were chosen by comparing model to experimental endpoints. For the rest of the fitting, experimental parameters were held constant at d_3 = 1.08 x 10^{-4} s^{-1}, d_4 = 1.85 x 10^{-5} s^{-1}, k_4 = 0.25 s^{-1},and k_5 = 0.000288 s^{-1} (Hasen and O'Shea, 2013; Khemlinskii et al., 2012). Model fitted parameters unique to the blue light optogenetic system (d_1, k_1, n, and K) were fit based around the literature ranges of the dissociation and association rates of the system. d_1 was found by sampling within 0.003 - 0.004 s^{-1}, k_1 between 2 x 10^{-5} – 2 s^{-1}, K between 0.1 – 10,000 mM, and n between 0.5 – 4 (Rademacher et al., 2017; Hansen and O'Shea, 2013; Gonze and Abou-Jaoude, 2013). The total forward ON rate for the first step was then filtered to be between 0.1 – 30 s for both I(t) = 60 au and I(t) = 600 au, mimicking the estimated binding rate of CRY2 to CIB1 (Rademacher et al., 2017). All other model fitted parameters were sampled between 2 x 10^{-5} – 2 s^{-1}. All model fitted parameters used in Figure 3, as well as the experimental constants, are shown in Figure S3C.

Cell Systems, Volume *12*

Supplemental information

Mapping the dynamic transfer functions of eukaryotic gene regulation

Jessica B. Lee, Leandra M. Caywood, Jennifer Y. Lo, Nicholas Levering, and Albert J. Keung

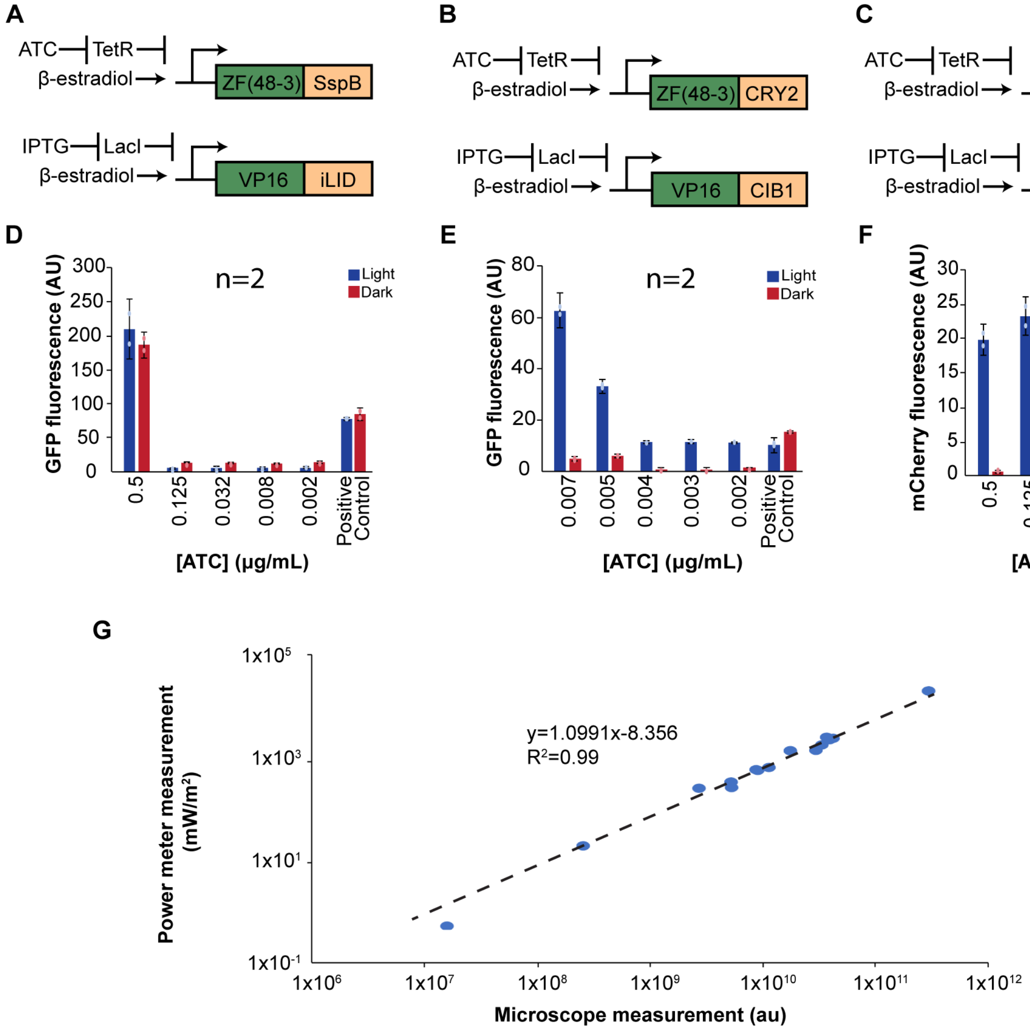


Figure S1. Related to Figure 1. Development of optogenetic system. (A-C) Schematics of DNA constructs tested. In addition to CRY2/CIB1, improved light-induced dimer (iLID) and its binding partner (SspB) were tested. (D) The iLID/SspB system showed no light-specific inducible activation. (E) Fluorescent output for ZF43-8-CRY2/VP16-CIB1 showed light-inducible activation. (F) Switching the fusion of CIB1 and VP16 produced robust light-inducible activation with minimal activation in the dark. The light condition was a single pulse (~6e10 au) for 6.5-7 hours. IPTG concentration was 20 mM for all plots. Error bars are 95% confidence intervals. n=2. Positive control was ZF43-8-VP16. (G) Light intensity relationship between microscope and power meter measurements provides a relationship between au and power. Equation on graph is a linear fit with y=$\log_{10}$(Power meter measurement) and x=$\log_{10}$(Microscope measurement).

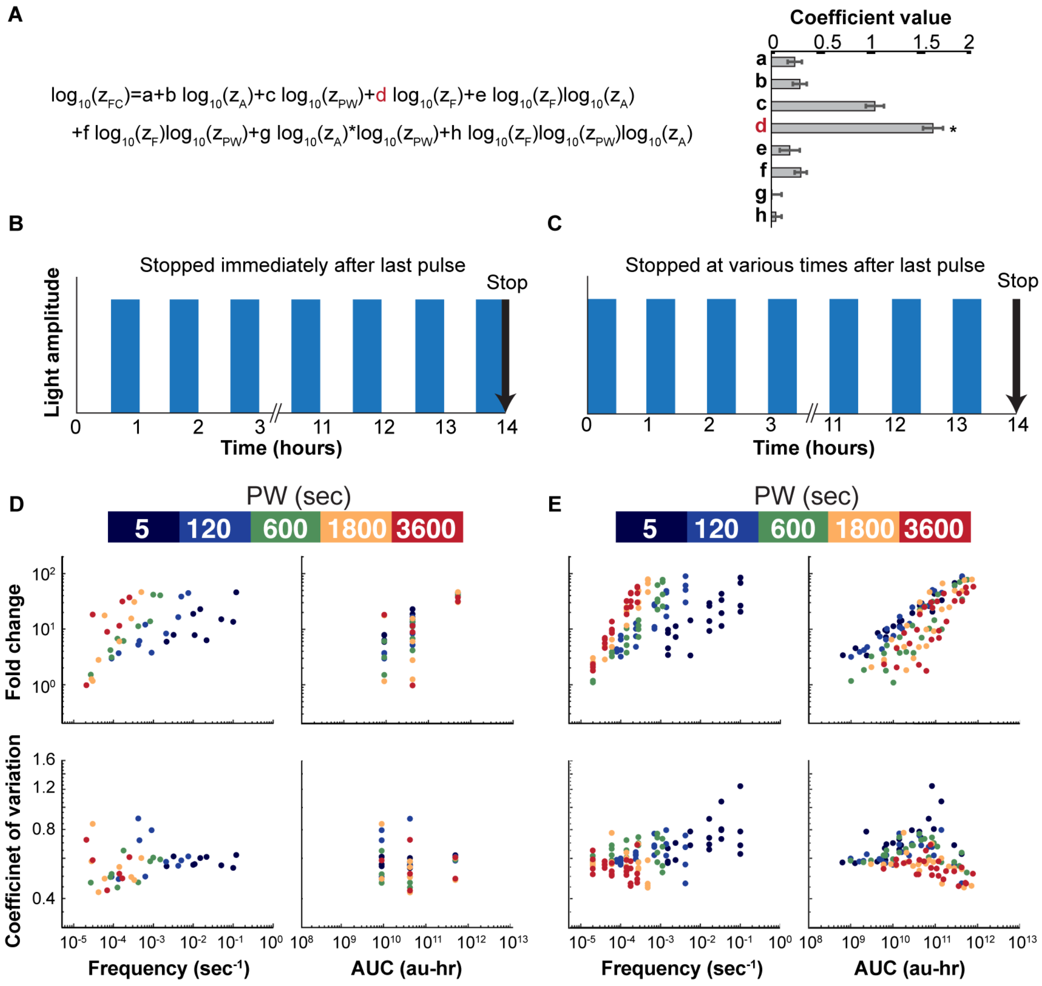
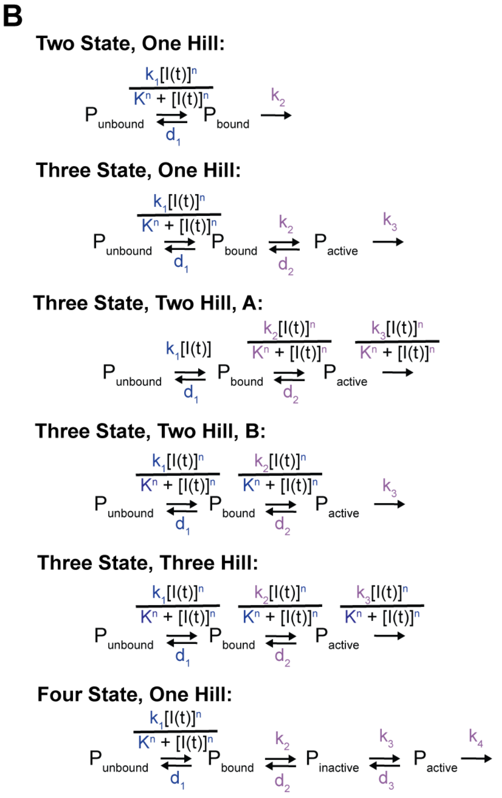


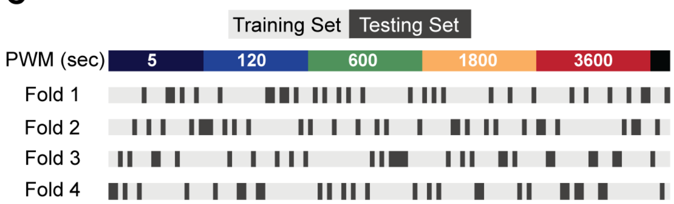
Figure S2. Related to Figure 2 and Figure 4. Sampling time does not account for filtering. (A) Fitted linear regression equation using the z-transformed parameters (z_A, z_{PW}, z_F, where A=amplitude, PW=pulse width, and F=frequency). The coefficient for z_F is significantly higher than the others (*p=0.05), meaning that frequency has the greatest effect on the resulting output. To determine whether the time after the last light pulse affects the fold change, we tested two scenarios (B-C). (B) The light pulses began at different times in order for them all to synchronously end at the same time. CHX was added to each sample immediately after the last pulse ended. (C) The light pulses began at the same time for all conditions, but the time after the last light pulse—and addition of cycloheximide (CHX)—varied according to equation: 1/F-PW. (D) Results of the scenario shown in panel B, chosen for three AUCs. (E) Results from scenario shown in panel C. This is also shown in Figures 2B and 4D. Both scenarios exhibited linear trends of fold change versus frequency with higher pulse widths having higher slopes. The filtering behavior was still observed in scenario A, where distinct mCherry outputs were achieved at the same AUCs. Noise for scenario shown in panel B.

A

Constant	Meaning	Value
d_1	CRY2-CIB1 dissociation rate	3.98×10^{-3} sec^{-1}
k_1	CRY2-CIB1 binding rate constant	1.66×10^{-2} sec^{-1}
K	Activation Coefficient	90.41
n	Hill Coefficient	0.964
d_2	Inactive State to Bound State	2.34×10^{-2} sec^{-1}
k_2	Bound State to Inactive State	2.05×10^{-4} sec^{-1}
k_3	Transcription rate	8.64×10^{-5} sec^{-1}
d_3	mRNA degradation rate	1.09×10^{-4} sec^{-1}
k_4	Translation rate	2.50×10^{-1} sec^{-1}
d_4	mCherry degradation rate	1.85×10^{-5} sec^{-1}
k_5	mCherry maturation rate	2.88×10^{-4} sec^{-1}



C



	Three State					
	One Hill	Two Hill A	Two Hill B	Three Hill	Two State	Four State
Fold 1	0.916	0.665	0.767	0.783	-1234	0.845
Fold 2	0.441	0.607	0.779	0.581	-1822	0.492
Fold 3	0.900	0.810	0.732	0.808	-1480	0.855
Fold 4	0.817	0.809	0.853	0.804	-1662	0.607
Average	0.768	0.723	0.783	0.744	-1550	0.700

D

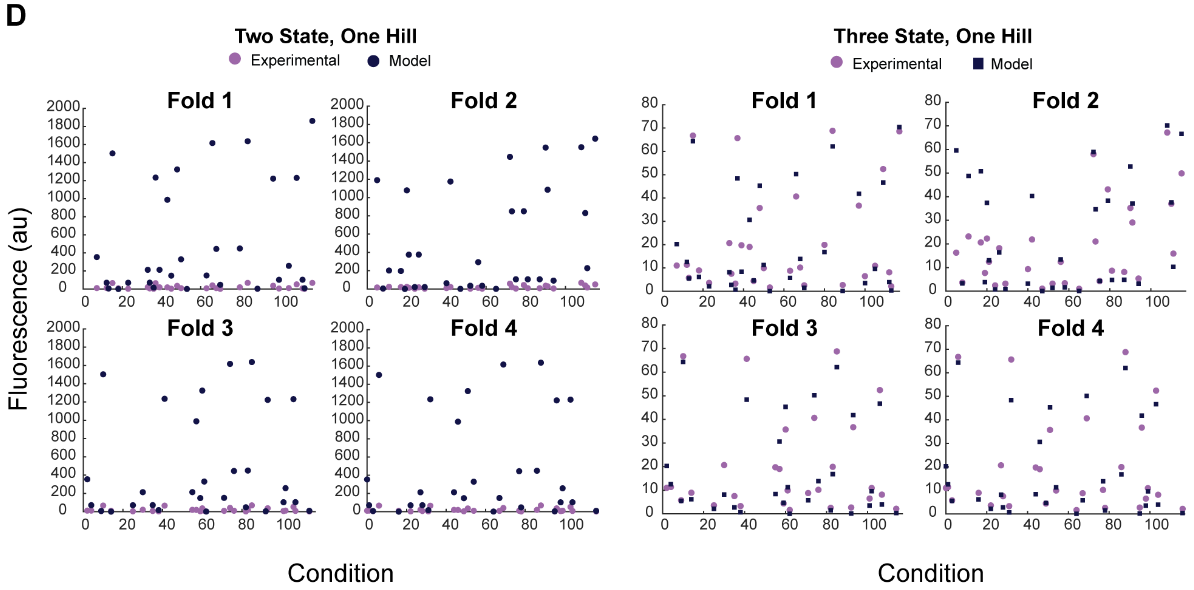


Figure S3. Related to Figure 3. Comparison of different models. (A) Values of fitted parameters for the three-promoter-state, one Hill function model. (B) Model structures tested in stratified K-fold cross validation, including a two-, four-, and multiple three-promoter-state architectures. (C) Visualization of training and testing sets for each fold in stratified K-fold cross validation. The percentage of samples in each pulse width are maintained across the folds. The R^2 values are shown for each model and fold. The average is the average R^2 for the model over each fold. The R^2 for the two-state model was very negative for the bounds we set for the parameters (see Methods section). Indeed, we see that the two-state model cannot produce a good fit (Panel D, left) for the bounds used for fitting. If k_1 is much smaller, then the two-promoter-state might have a better fit like that achieved by the three-promoter state (Panel D, right).

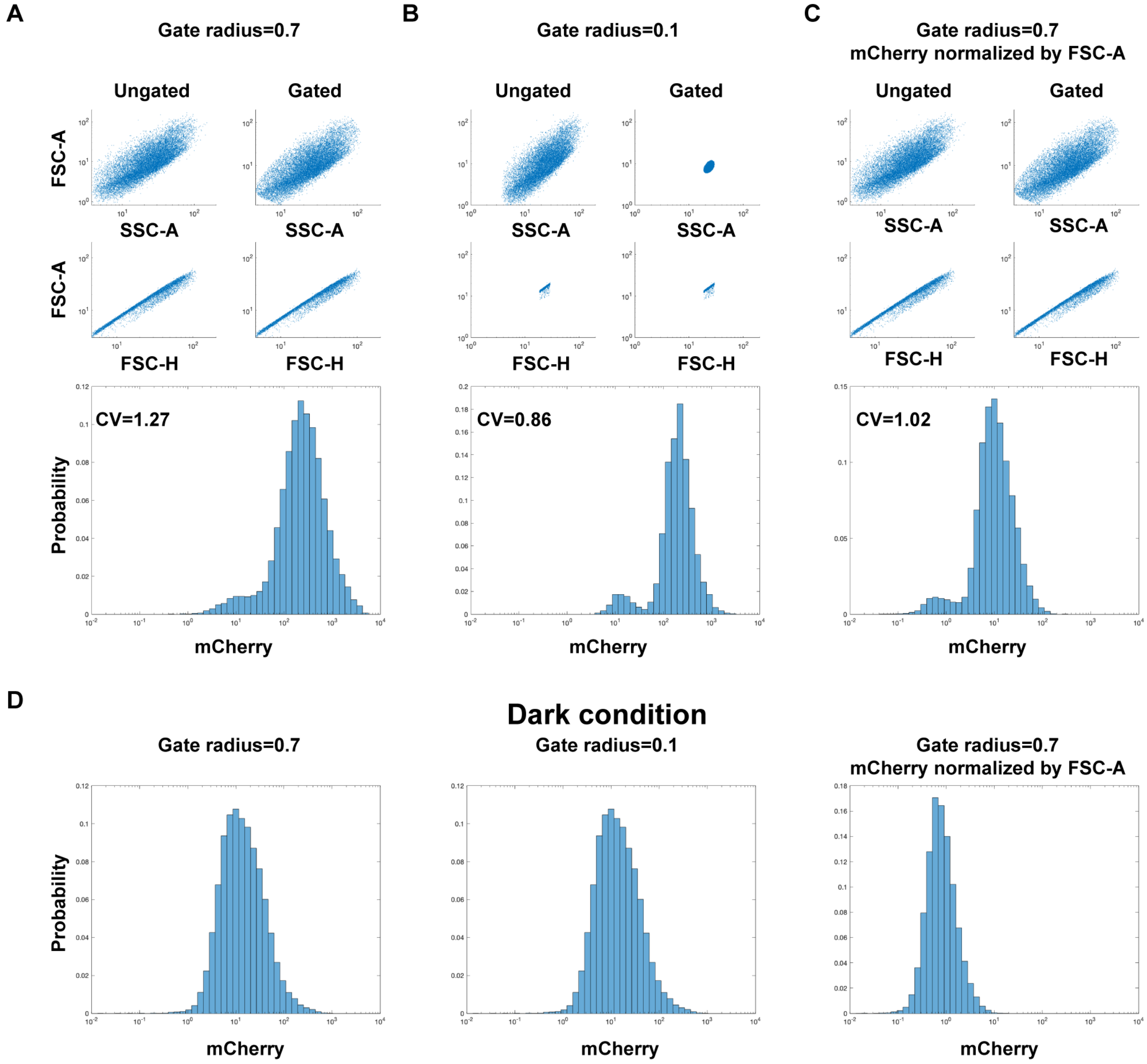


Figure S4. Related to Figure 4. Effects of size gating and normalization on noise. (A-C) Flow cytometry gating strategy for ZF-CRY2/CIB1-VP16 strain with blue light (amplitude=6e9 au, frequency=0.1 sec^{-1}, and pulse width=5 sec). (A) FSC-A vs SSC-A was gated with a large radius (0.7 with logarithmically transformed data). First, the data is gated by removing events outside of a radius of 0.7 with logarithmically-transformed data (top). Then doublets were removed by keeping events within gate=median(log(FSC-H)/log(FSC-A))+.1>log(FSC-H)./log(FSC-A)>median(log(FSC-H)/log(FSC-A)-.1 (Second row, note not many events were excluded in this example). Resulting mCherry histogram is shown at the bottom with a CV=1.27. (B) Gating strategy for small radius (0.3 with logarithmically transformed data) for FSC-A vs SSC-A. Doublet gating was done the same as in A. The resulting histogram shows the reduction in noise with CV=0.86. (C) Gating strategy with large gating radius (same as A) and fluorescence divided by FSC-A. Histogram shows a CV between the other two gating strategies, CV=1.02. (D) Histograms of mCherry fluorescence for the ZF-CRY2/CIB1-VP16 strain without the addition of blue light (dark).

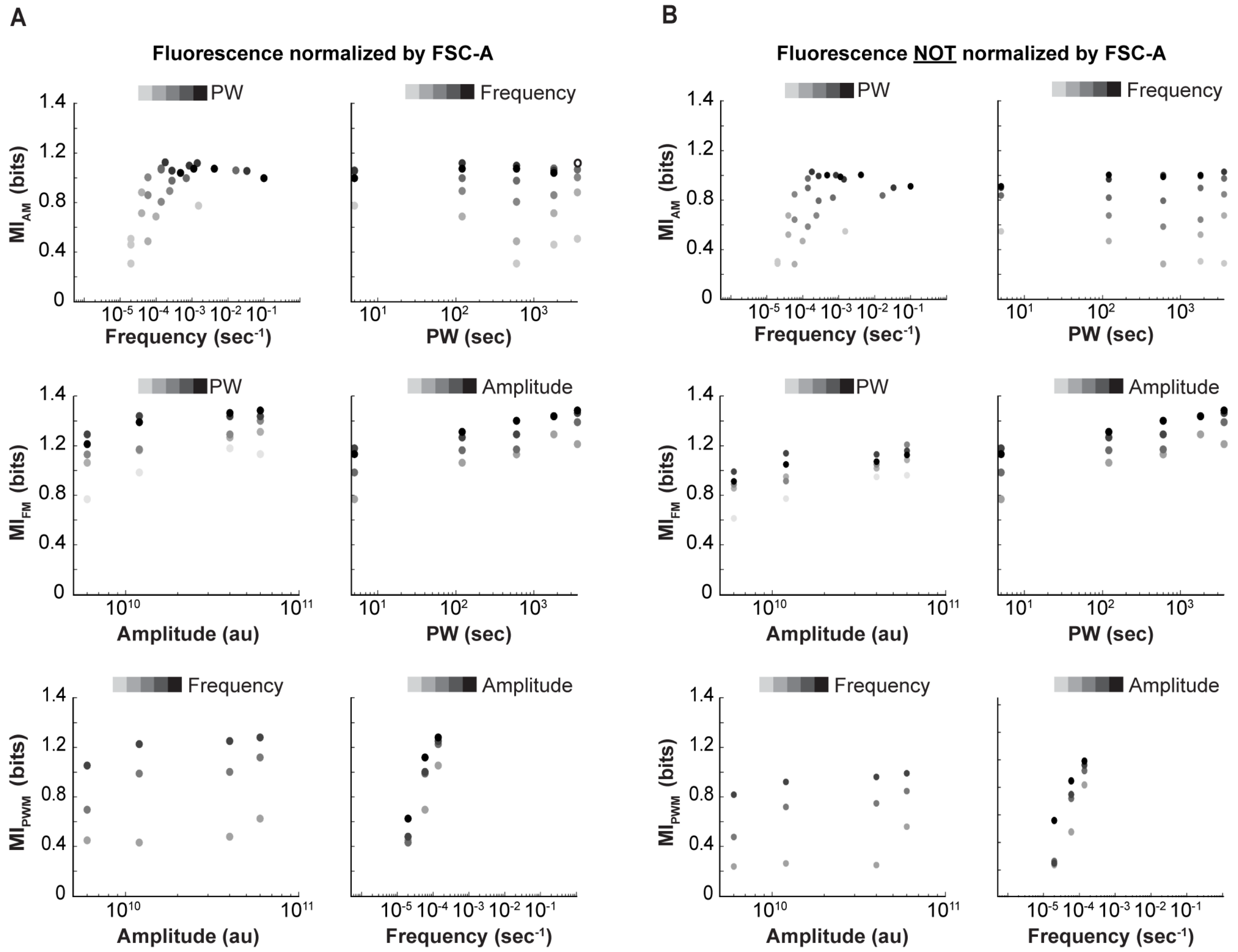


Figure S5. Related to Figure 5. MI calculated without FSC-A normalization shows same trends. (A) MI plots with fluorescence normalized by FSC-A (same as Figure 5). (B) MI plots with fluorescence NOT normalized by FSC-A. Both methods exhibit the same trends. However, the non-normalized values are lower. Each dot is the mean of 4 biological replicates. For a list of all values graphed, see Table S2.

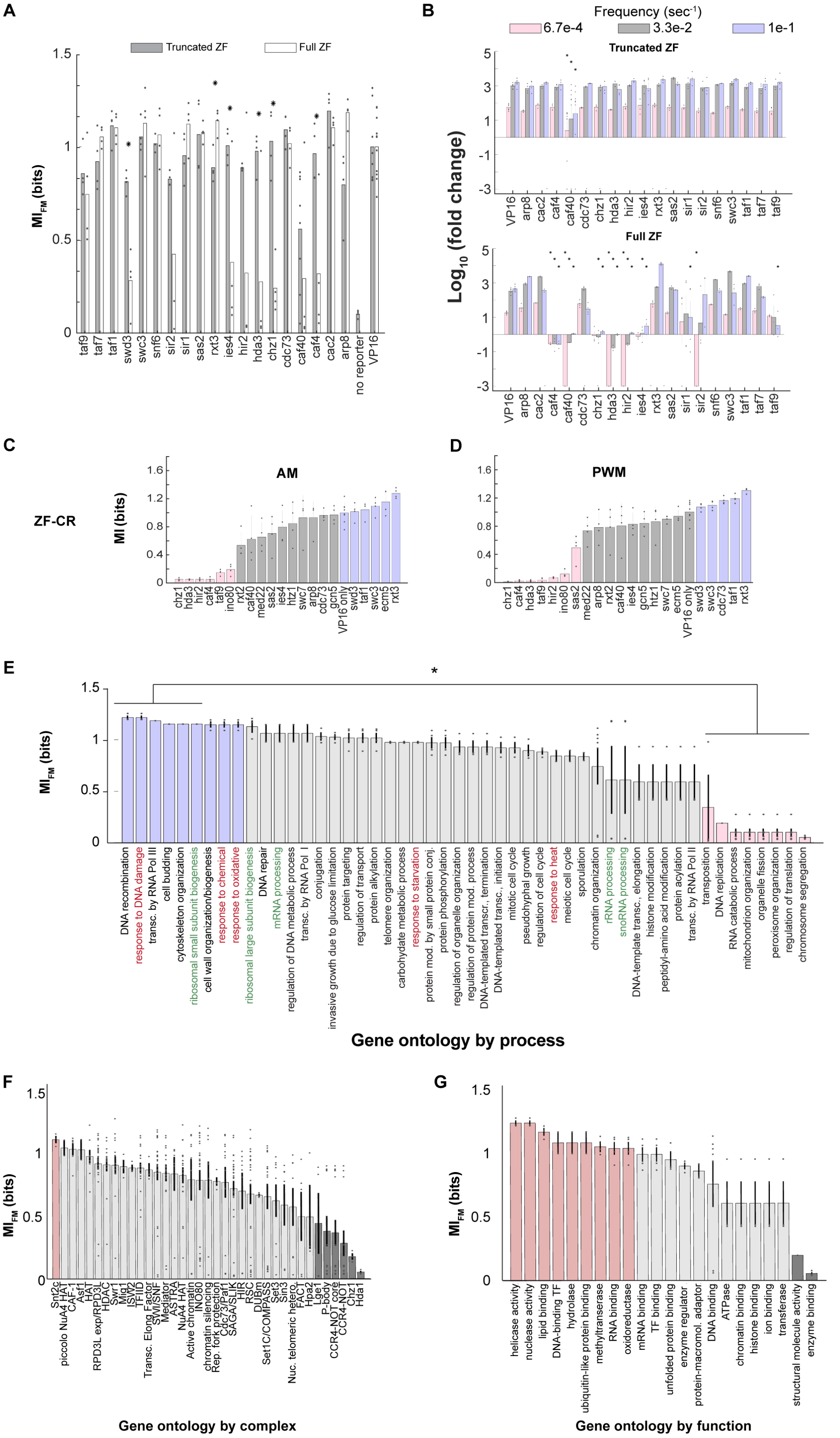

A
Truncated ZF
Full ZF
MI$_{FM}$ (bits)
1.5
1
0.5
0
taf9
taf7
taf1
swd3
swc3
snf6
sir2
sir1
sas2
rxt3
ies4
hir2
hda3
chz1
cdc73
caf40
caf4
cac2
arp8
no reporter
VP16
B
Frequency (sec^{-1})
6.7e-4
3.3e-2
1e-1
Truncated ZF
Full ZF
Log$_{10}$ (fold change)
5
3
1
-1
-3
VP16
arp8
cac2
caf4
caf40
cdc73
chz1
hda3
hir2
ies4
rxt3
sas2
sir1
sir2
snf6
swc3
taf1
taf7
taf9
C
ZF-CR
AM
MI (bits)
1.6
1.2
0.8
0.4
0
chz1
hda3
hir2
caf4
taf9
ino80
rxt2
caf40
med22
sas2
ies4
htz1
swc7
arp8
cdc73
gcn5
VP16 only
swd3
taf1
swc3
ecm5
rxt3
D
PWM
1.6
1.2
0.8
0.4
0
chz1
caf4
hda3
taf9
hir2
ino80
sas2
med22
arp8
rxt2
caf40
ies4
gcn5
htz1
swc7
ecm5
VP16 only
swd3
swc3
cdc73
taf1
rxt3
E
*
MI$_{FM}$ (bits)
1.5
1
0.5
0
DNA recombination
response to DNA damage
transc. by RNA Pol III
cell budding
cytoskeleton organization
ribosomal small subunit biogenesis
cell wall organization/biogenesis
response to chemical
response to oxidative
ribosomal large subunit biogenesis
DNA repair
mRNA processing
regulation of DNA metabolic process
transc. by RNA Pol I
conjugation
invasive growth due to glucose limitation
protein targeting
regulation of transport
protein alkylation
telomere organization
carbohydrate metabolic process
response to starvation
protein mod. by small protein conj.
protein phosphorylation
regulation of organelle organization
regulation of protein mod. process
DNA-templated transcr., termination
DNA-templated transc., initiation
mitotic cell cycle
pseudohyphal growth
regulation of cell cycle
response to heat
meiotic cell cycle
sporulation
chromatin organization
rRNA processing
snoRNA processing
DNA-template transc., elongation
histone modification
peptidyl-amino acid modification
protein acylation
transc. by RNA Pol II
transposition
DNA replication
RNA catabolic process
mitochondrion organization
organelle fission
peroxisome organization
regulation of translation
chromosome segregation
Gene ontology by process
F
MI$_{FM}$ (bits)
1.5
1
0.5
0
Snt2c
piccolo NuA4 HAT
CAF-1
Asf1
HAT
RPD3L exp/RPD3L
HDAC
Swr1
Mlh1
ISW2
TFIID
Transc. Elong Factor
SWI/SNF
Mediator
ASTRA
NuA4 HAT
Active chromatin
INO80
chromatin silencing
Rep. fork protection
Cdc73/Paf1
SAGA/SLIK
HIR
RSC
DUBm
Set1C/COMPASS
Set3
Sin3
Nuc. telomeric hetero.
FACT
Hpa2
Lge1
P-body
CCR4-NOT core
CCR4-NOT
Chz1
Hda1
Gene ontology by complex
G
MI$_{FM}$ (bits)
1.5
1
0.5
0
helicase activity
nuclease activity
lipid binding
DNA-binding TF
hydrolase
ubiquitin-like protein binding
methyltranserase
RNA binding
oxidoreductase
mRNA binding
TF binding
unfolded protein binding
enzyme regulator
protein-macromol. adaptor
DNA binding
ATPase
chromatin binding
histone binding
ion binding
transferase
structural molecule activity
enzyme binding
Gene ontology by function

Figure S6. Related to Figure 6. Additional modes of modulation and control experiments. (A-B) Replacing the 97-4ZF with a truncated, non-targeting 42-10ZF (CCCCGCCACCTAAAAACCCACCTGAGA), we measured the MI_{FM} for a subset of CRs (Panel A). The frequencies used were the same as the main text: 6.7e-4, 3.3e-2, and 1e-1 sec^{-1} (Periods of 1500, 30, and 10 sec). Several (6 out of19) of the CRs had significantly different MIs with the full ZF compared to the truncated ZF (* $p<0.05$, Welch t-test). n=2-4, except VP16 and caf40p (truncated ZF) have 8 biological replicates. (B) Additionally, we saw similar transfer functions (fold changes) for the truncated ZF subset and VP16, with only caf40p having significantly different fold changes for each frequency (top). The same CRs fused to the full, targeting ZF had 9 out 19 CRs with at least one fold change significantly different from the VP16 only strain with the same light frequency (bottom).*$p<0.05$, T-K analysis compared to VP16 only for each frequency condition). This implies that the changes in MI were due to recruitment of the 43-8ZF-CR to the locus and not global effects of over-expression of the CR. (C) MI for a subset of 97-4ZF-CRs using amplitude modulation (0, 6e9, 1.2e10, and 6e10 au). n=4 biological replicates. (D) MI for a subset of 97-4ZF-CRs using pulse width modulation (0, 5, 120, and 600 sec). (E-G) Average MI_{FM} for each gene ontology based on process, complex, and function. Chromatin regulators for each gene ontology were determined using the genes in the Yeast Genome Database (Cherry et al., 2012). All MIs are normalized to the MI of VP16 for the same set of conditions. All MI values are given in Table S3.

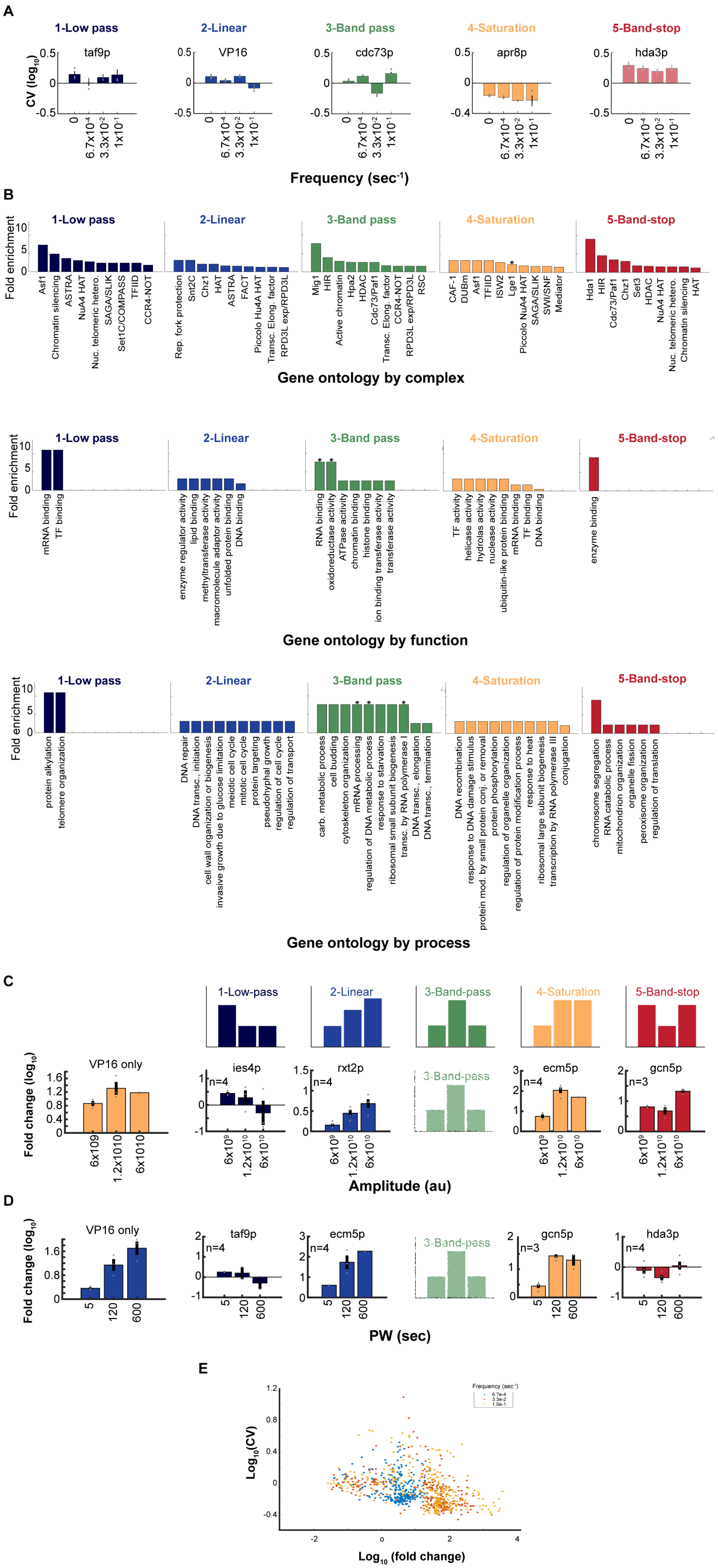

A
1-Low pass
2-Linear
3-Band pass
4-Saturation
5-Band-stop
taf9p
VP16
cdc73p
apr8p
hda3p
CV ($\log_{10}$)
0.5
0
-0.5
-0.4
0
6.7x10-4
3.3x10-2
1x10-1
Frequency (sec-1)
B
1-Low pass
2-Linear
3-Band pass
4-Saturation
5-Band-stop
Fold enrichment
10
5
0
Asf1
Chromatin silencing
ASTRA
NuA4 HAT
Nuc. telomeric hetero.
SAGA/SLIK
Set1C/COMPASS
TFIID
CCR4-NOT
Rep. fork protection
Snt2C
Chz1
HAT
ASTRA
FACT
Piccolo Hu4A HAT
Transc. Elong. factor
RPD3L exp/RPD3L
Mig1
HIR
Active chromatin
Hpa2
HDAC
Cdc73/Paf1
Transc. Elong. factor
CCR4-NOT
RPD3L exp/RPD3L
RSC
CAF-1
DUBm
Asf1
TFIID
ISW2
Lge1
Piccolo NuA4 HAT
SAGA/SLIK
SWI/SNF
Mediator
Hda1
HIR
Cdc73/Paf1
Chz1
Set3
HDAC
NuA4 HAT
Nuc. telomeric hetero.
Chromatin silencing
HAT
Gene ontology by complex
1-Low pass
2-Linear
3-Band pass
4-Saturation
5-Band-stop
Fold enrichment
10
5
0
mRNA binding
TF binding
enzyme regulator activity
lipid binding
methyltransferase activity
macromolecule adaptor activity
unfolded protein binding
DNA binding
RNA binding
oxidoreductase activity
ATPase activity
chromatin binding
histone binding
ion binding transferase activity
transferase activity
TF activity
helicase activity
hydrolase activity
nuclease activity
ubiquitin-like protein binding
mRNA binding
TF binding
DNA binding
enzyme binding
Gene ontology by function
1-Low pass
2-Linear
3-Band pass
4-Saturation
5-Band-stop
Fold enrichment
10
5
0
protein alkylation
telomere organization
DNA repair
DNA transc., initiation
cell wall organization or biogenesis
invasive growth due to glucose limitation
meiotic cell cycle
mitotic cell cycle
protein targeting
pseudohyphal growth
regulation of cell cycle
regulation of transport
carb. metabolic process
cell budding
cytoskeleton organization
mRNA processing
regulation of DNA metabolic process
response to starvation
ribosomal small subunit biogenesis
transc. by RNA polymerase I
DNA transc., elongation
DNA transc., termination
DNA recombination
response to DNA damage stimulus
protein mod. by small protein conj. or removal
protein phosphorylation
regulation of organelle organization
regulation of protein modification process
response to heat
ribosomal large subunit biogenesis
transcription by RNA polymerase III
conjugation
chromosome segregation
RNA catabolic process
mitochondrion organization
organelle fission
peroxisome organization
regulation of translation
Gene ontology by process
C
1-Low-pass
2-Linear
3-Band-pass
4-Saturation
5-Band-stop
Fold change ($\log_{10}$)
VP16 only
ies4p
rxt2p
3-Band-pass
ecm5p
gcn5p
n=4
n=4
n=4
n=3
1.6
1.2
0.8
0.4
0
1
0
-1
1.5
1
0.5
0
2
1
0
6x109
1.2x1010
6x1010
6x10^9
1.2x10^10
6x10^10
Amplitude (au)
D
Fold change ($\log_{10}$)
VP16 only
taf9p
ecm5p
3-Band-pass
gcn5p
hda3p
n=4
n=4
n=3
n=4
1.6
0.8
0
2
1
0
-1
3
2
1
0
5
120
600
PW (sec)
E
Frequency (sec-1)
6.7e-4
3.3e-2
1.0e-1
$\log_{10}$(CV)
1.2
0.8
0.4
0
-0.4
-2
0
2
4
$\log_{10}$ (fold change)

Figure S7. Related to Figure 7. Characterization of filtering behavior. (A) CV for the example CRs shown in Figure 7. (B) Fold enrichment for gene ontology based on complex (top, from Keung et al., 2014), function (middle), or process (bottom). *p<0.05, two-sided Fisher exact test with Bonferroni-Holm correction. (C) Examples of CRs that fit into the filtering clusters for AM (Panel C) and PWM (Panel D). Note that there were no CRs that fit into the band-pass cluster for AM and PWM. n=4 biological replicates. List of clusters for all CRs is given in Table S3.