UNIVERSITY OF CALIFORNIA
SANTA CRUZ

SPECIFICITY IN TRANSCRIPTIONAL REGULATION

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOMOLECULAR ENGINEERING & BIOINFORMATICS

by

Robert I. Shelansky

September 2020

The Dissertation of Robert Shelansky is
Approved:

_____
Professor Todd Lowe, Chair

_____
Professor Ed Green

_____
Professor Mark Akeson

_____
Professor Hinrich Boeger

_____
Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

SPECIFICITY IN TRANSCRIPTIONAL REGULATION

Robert Shelansky

Abstract

Gene-specific regulation of transcription is achieved through the binding of transcription factors to DNA sequences. Many Eukaryotic transcription factors maintain affinity differences between target and non-target sequences that appear too small to explain the specificity observed for the genes they regulate. How is specificity achieved in Eukaryotic gene expression? In eukaryotes, DNA is spooled around histone protein octamers to form nucleosomes. The nucleosome represses transcription by acting as a barrier to the binding of transcription factors. Thus, gene activation requires the recruitment of ATP-dependent chromatin remodelers which remove nucleosomes covering important regulatory sequences. However, promoter nucleosome structure is heterogeneous even under activating conditions. Why does the cell expend energy to maintain heterogenous promoter chromatin in the promoters of actively transcribing genes?

In Chapter 1, I present a model of gene transcription which represents a unified solution to these questions, among others. I show that activator mediated ATP dependent stochastic removal and reformation of nucleosomes on promoter DNA may be used for the kinetic proofreading of activator-DNA interactions. The specificity enhancement due to kinetic proofreading is an archetype that, in part, can be used to explain the observed specificity in Eukaryotic gene expression. I show that contrary to expectation, heterogeneity in promoter chromatin structure reduces

the variation observed in gene expression. Additionally, I provide insight into the necessity of transcriptional bursting for regulated, highly expressed genes.

In Chapter 2, I present a number of experimental tests of the proofreading model. We observe transcriptional bursting, chromatin remodeling and activator binding at a classic model gene, *PHO5,* in *Saccharomyces cerevisiae*. I show that transcriptional bursting of *PHO5* occurs in at least two distinct timescales, an expectation of the proofreading model. In addition, I show that mutation of a single chromatin remodeler, Isw2, is sufficient to disrupt correlation at the longer timescale. I present a model of kinetic proofreading of activator specificity by Isw2 and test conjectures such a model purports.

In chapter 3, I present a technique for studying eukaryotic gene expression by generating and testing the expression of >400,000 permuted synthetic cassettes generated from 26 genes from *Saccharomyces cerevisiae*.

This thesis is dedicated to the pursuit of objective truth,

and the accompanying loss of certainty.

Acknowledgments

I want to take this opportunity to thank the battalions of mentors, colleagues, mentees, family, friends and those who are yet anonymous to me. You all have made this experience possible and successful. First, I want to thank my defense committee for taking the time to critically review this document. It seems extensive, though knowing the extent of work dedicated over the last 6 years to the projects described, in part, in this thesis, projects described in other manuscripts or projects lost to time, it could have been worse. Thank you, Mark Akeson, Ed Green, Hinrich Boeger and Todd Lowe. Mark, you were the first to call me Dr. Shelansky, though almost a decade too early.

To my army of undergraduate researchers I have trained and seen grow, many of whom have become graduate students at universities across the world, I say good luck! Thank you: Saketh Akula, Kevin Chen, Fernanda Gonzales, Kurt Henderson, Shayal Narayan, Nikhila Kadiyala, Heta Patel, Lillian Wang and Znala Williams. Each of you brought a unique and bright atmosphere to our lab. I have so many memories of each of you, some good and some great. I do not know if I have ever had better experiences forcing biology students to learn to program. Thank you to the “little guys”, two high school students brave enough to analyze single molecule chromatin structure. Scott and William, wherever you are, I wish well.

To all of my colleagues and friends who also made this life choice (or not), to those who have moved on and those who struggle on, each of you is an inspiration. Brian Lin, you may have beat me out but I will cherish our time together, even if it

was spent arguing over the most minute semantics. Michael Doody, I could not have asked for a better step-colleague. My work was made better because of your diligence and critical thinking. I promise you will likely never have to do another gene ring prep... except for that Isw2 mutant. Edwardo Hirata and Gustav Peterson without your expertise and diligence the MFM could not have been built, it was a pleasure working with you and a joy to become friends, I wish you both the best of luck. Brandon Saint-John, I regret not getting to include our work together in this thesis. Thank you for working on the add-seq project with me, it is such exciting work and I wish you the best of luck. Thank you Jay, Lenaig, Ole and Verena, quarantine would be a lot more lonely alone.

There have been so many mentors that have had a huge impact on my experience at UCSC. Chris Brown and Stacy Harvey, each of you dedicated real time to train a computational scientist in wet lab techniques, thank you; it must have been challenging. I hope you remember our time together as fondly as I do. I would like to share a few anecdotes from various Professors that I took to heart (which are taken entirely out of context). “Why invest time studying anything that will not win you the nobel prize.” -Ed Green (I hear this quote is borrowed?); “If you do not have a reagent, find it.” -Grant Hartzog; “All swans are white.” -Hinrich Boeger; Grant, when I was an undergraduate, I was an MCD student. I applied for your honors viral genomics course for freshmen. I pestered you weekly about it, until one day you informed me that I was not selected. I switched majors to BME that quarter; but somehow, for better or worse, I made my way back to an MCD lab. Thank you for all

of the time you invested in me, I greatly appreciate it, you impacted my outlook on experimental biology greatly. Rohinton Kamakaka, thank you for giving me the opportunity to work on the permutational genes project. It was a much needed escape from the work I was drowning in and it was a privilege to learn from you. Sara Abrahamson, you came into my scientific life at a low point, bringing technical knowledge, intellectual insight and a much needed breath of fresh air. This project would have been impossible without you and I am exceedingly grateful for your attention and time. Tineke Lenstra and Dan Larson, It has been a long time since I ventured to Bethesda to learn alongside you. However, your investment into us and our project was essential to the advances in understanding we have now achieved. You may not remember but all hopes of my project were invested in my ability to capture transcriptional dynamics of *PHO5* during a two-week stay in the Larson lab at the NIH. The second day I came down with the most severe flu of my life. I was bed ridden at an air BnB for days. I remember wandering around Bethesda completely delirious trying to find accommodations for food and medicine. I was so embarrassed; though I imagine you did not want me anywhere near the lab. Thank you for allowing me to have an extended stay and thank you for your support and time. Angela Brooks, thank you for allowing me to work alongside you, Brandon and Eva. It has been an honour and I am looking forward to seeing how that project turns out. It has been a fantastic experience to work alongside such great scientists and people.

Hinrich, I know you know this story but it is worth retelling. 1.5 years into my PhD at UCSC, after the rejection of multiple fellowship applications the funding for my project was rescinded. I was scared and uncertain about my future. Students in this situation are often lost in the cracks of the institution. These students usually leave of their own free will.

Hinrich, you invited me in, enticing me with an interesting technical problem and the promise of support. I want to thank you for this and for all you have done in mentoring me. I want to thank you for the countless hours you dedicated to teaching me biology but also later the countless hours we spent in your office arguing about everything. There were days, I swear, I would pop into your office at 9 am to ask about some random yeast strain and we would look out the window and it would be dark outside and we would be discussing the search for objective truth and how knowledge is gained. I know this must be an exaggeration because I rarely skip lunch. Thank you for your time and dedication. Thank you for your scientific thoughtfulness, your willingness to make bold conjectures and then your willingness to reject them. I still remember what you said to me that first day. I am sorry I did not accomplish more. But I am truly grateful for the opportunity you gave me and the support you have shown me.

Thank you Martha Zuniga, Mrs Pam Swift and Robert A. Ludwig for the Robert A. Ludwig Memorial Graduate Fellowship. This award came at a economically troubling time in my tenure as a student and afforded me the

opportunity to speak publicly about my work at BPS 2020 in San Diego. The purpose of this award was inspirational to me and receiving it an honour. Thank you for the opportunity.

Nedda Faye Saremi, you must be so tired of hearing me talk about nucleosomes. I know you are tired of explaining puma and bear population genetics to me. Thank you for listening and thank you for explaining even when I make you angry. I hope every experience we share together can be at least as good as the PhD.

Hi Mother, Father and Brother, Shelansky. There is more to thank you all for then could fit in this thesis. I do not think that you all expected me to be an academic, considering my experiences in early education. I must say Andrew, I imagine all those nights you spent lecturing me about math, science and computation before bed probably helped. Thank you all for your love, support and so much more.

I know I am missing many, many people who deserve to be thanked by name here. Thank you all. Also, thank you to whoever reads this document. I hope you learn something. I know I did writing it.

The text of this thesis includes reprint and adaptation of the following previously published material:

1) Shelansky R, Boeger H. Nucleosomal proofreading of activator-promoter interactions. Proc Natl Acad Sci U S A. 2020;117: 2456–2461.
2) Dhillon N, Shelansky R, Townshend B, Jain M, Boeger H, Endy D, et al. Permutational analysis of Saccharomyces cerevisiae regulatory elements. Synth Biol. 2020 [cited 13 Jul 2020]. doi:10.1093/synbio/ysaa007

The co-author listed in these publications directed and supervised the research which forms the basis for the thesis.

Chapter 1: Nucleosomal proofreading of activator–promoter interactions

Robert Shelansky, Hinrich Boeger

Specificity in transcriptional regulation is achieved by the binding of gene-specific transcriptional activators to specific DNA sequences, “enhancers”. Mutation of a single activator binding site may abolish regulation [1]. The activator, once bound, stimulates transcription indirectly by recruitment of other, gene agnostic factors such as chromatin remodelers, histone modifying enzymes, Mediator and SAGA, eventually leading to the formation of the pre-initiation complex and finally polymerase release [2–5].

Activators find their DNA binding sites by trial and error [6]. To quickly find their target sequence activators maintain high on-rates. Thus, the affinity for nonspecific DNA is high, equilibrium dissociation constants fall into the micromolar range [7]. Different enhancers and core promoters may be mixed and matched [8]. Suggesting that activators may stimulate transcription from both target enhancers, which bear specific binding sites for the activator, and non-target enhancers, which do not. Thus, the question arises of how binding at target and not target enhancers triggers transcription?

Activator specificity is achieved through differences in affinity between correct and incorrect sequence binding. High regulatory specificities may be obtained either

by activator on-rates close to zero, or large affinity differences between correct and incorrect sequence binding. Higher affinities generally imply extended DNA dwell times; the observed rate constants of the on-reaction are closely similar for different DNA sequences of the same activator [7,9,10]. Notably, measured affinity differences between correct and incorrect sequences are small, -3 kcal/mol, about three times the average kinetic energy of a molecule at 25°C or less [9]. The cell may tune activator concentration to decrease on-rate but small activator on-rates result in long search times for target sequences. Gene regulation requires both high specificity *and* openness to change, "controllability". The specificity problem is thus introduced, how is activation both specific and controllable given the apparent affinities of activators for their target and non-target sequences. The specificity problem is further exacerbated by the fact that non-target enhancers greatly outnumber target enhancers.

Molecular biological enzymes (DNA and RNA polymerases, aminoacyl-tRNA synthetases, the spliceosome and ribosome) face a similar specificity problem, as they must discriminate between correct and incorrect substrates on the basis of small differences in binding energy, ΔΔG°. Remarkably, molecular biological enzymes exhibit error frequencies well below the lower limit imposed by the energetics of substrate-enzyme binding of $\varepsilon_0 = e^{\frac{\Delta\Delta G^\circ}{RT}}$, which is attainable only in the asymptotic limit of infinitely slow catalysis [11]. Specificity enhancement is achieved by the insertion of an additional energy consuming delay step, the 'proofreading reaction', into the Michaelis Menten pathway. Differences in sojourn

time of the activator at the catalytic center are exploited twice, before and after the proofreading reaction, reducing the minimal error frequency from ε_0 to ε_0^2 [12]. Kinetic proofreading is a biochemical mechanism of “double-checking” substrate recognition.

The dynamics of promoter chromatin structure, inferred from analysis of single gene molecules [13–15], suggests a comparable solution to the activator specificity problem. The nuclear DNA of eukaryotic cells is spooled onto octamers of histone proteins [16,17]. These spools, or 'nucleosomes', the basic structural unit of chromatin, impede access to DNA and thus are universal repressors of transcription [18].

Promoter nucleosomes have been viewed as an impediment to transcription that is overcome once, during the transition from transcriptionally repressed to active chromatin [19]. Analysis of the *PHO5* gene of yeast led to a different conclusion: promoter nucleosomes are continually removed and reformed as the promoter stochastically transitions between alternative nucleosome configurations, including the fully nucleosomal and nucleosome-free promoter [13,20]. Activator binding increases the transition probability from configurations with more to those with fewer nucleosomes, increasing the structural heterogeneity of promoter chromatin [13].

Initially, this probabilistic theory of promoter chromatin dynamics was conceived to reconcile apparently contradictory experimental findings that suggested both loss and presence of nucleosomes at transcriptionally active promoter

sequences [21,22]. The theory was subsequently employed to explain the statistical distribution of promoter nucleosome configurations observed by electron microscopy [13]. However, the biological question remained of why cells allow for, or perhaps prefer, structurally heterogeneous over homogeneous promoter chromatin. Heterogeneity is caused by the random transitioning between nucleosome configurations, some conducive to transcription and others not. This random transitioning should increase noise and thus reduce the signal-to-noise ratio in gene expression [23].

Here we show that, to the contrary, stochastic structural dynamics in promoter chromatin, when maintained at a distance from equilibrium, may both increase specificity and attenuate transactional noise. We propose the kinetic proofreading of activator-DNA interactions by nucleosomes as a partial solution to the specitity problem introduced above.

Results

A framework for characterizing specificity in transcription

To model activator specificity, we consider two genes that are identical, except that one copy bears the binding site for a specific activator (figure 1**A**) whereas the other does not. We define 'regulatory specificity' or 'activator fidelity', f, as a measure of the activator's ability to distinguish between target and non-target promoters, where

$$f = \frac{v_t}{v_n} \qquad (1)$$

where v_t is the average steady-state rate of transcription for target promoter binding and v_n for non-target promoter binding. The fidelity is defined by the ratio of transcriptional output from target and non-target promoters. Thus, when the activator promotes transcription indiscriminately, or transcription is activator-independent, $f = 1$.

Free-energy differences for activator-DNA binding reactions, ΔΔG°, are generally determined by differences in the activator's DNA-residence time [7,9,10]. Thus, $\Delta\Delta G^\circ = -RTln(\frac{k_n}{k_t})$, where k_t is the off-rate of the activator at target sequences and k_n is the off-rate at non-target sequences, and $k_t < k_n$.

For all following calculations, we set $k_t = 1$, by normalizing all rate constants by k_t. In addition, for ease of argument we set the ratio $\frac{k_n}{k_t} = 100$. $\frac{k_n}{k_t}$ reflects the effective energy difference between target and non-target promoter binding. A value of 100 corresponds to an estimate of the energy difference, $\Delta\Delta G^\circ$ = -2.7 kcal/mol, between target and non-target sequences for a typical activator of yeast, e.g. Pho4 the activator of Pho5 [9]. For stochastic simulations and noise calculations we assumed an average expression level of 50 transcripts per cell, corresponding to a strongly transcribed gene of yeast, *e.g.* the fully induced *PHO5* gene [1]. While the numerical results of our calculations depend on the choice of specific parameter values, our principal conclusions do not (see Appendix).

In the simplest case, the standard model of transcriptional regulation (Model 1; figure 1**A**), the promoter transitions between two states: activator-bound and

unbound, where only the activator-bound state is transcriptionally active [24]. We assume that v_t and v_n linearly depend on the steady-state probability of finding the promoter in its transcriptionally active state, the activator bound state. Activator fidelity for Model 1, a two-state promoter model, thus, is given by

$$f_1 = \frac{k_n + \kappa}{k_1 + \kappa} \tag{2}$$

(see Appendix 3). The maximal fidelity, f_0 , is attained as the activator on-rate, κ , which linearly depends on the concentration of the activator, tends to zero (figure 1**B**). We will refer to the upper limit to regulatory specificity imposed by the energetics of activator-DNA binding as the 'Hopfield barrier' to activator fidelity [25].

As k tends to zero, the activator's search time for its binding sequence tends to infinity. Thus, increasingly long pauses of inactivity in the activator unbound state are interrupted by periods of transcriptional activity with short pauses between initiation events in the activator-bound state. As periods of inactivity grow, to maintain a constant average rate of transcription, v_t, the rate of transcription in the activator-bound state, μ, must increase. Increasing μ further exacerbates the discrepancy between short pauses, with average length $\frac{1}{\mu}$, and long pauses, with average length $\frac{1}{k}$. As long pauses grow longer and short pauses shorter, transcription in stochastic bursts becomes increasingly manifest (figure 1**C**). The strength of bursting or magnitude of 'transcription noise' may be expressed in terms of a population statistic, the Fano factor, the variance of mRNA abundance normalized by its mean (see Appendix 3). As activator fidelity approaches the

Hopfield barrier, the noise of transcription tends to infinity (figure 1**D**). Approaching this limit, of course, is unrealistic: The activator concentration, cannot and does not come arbitrarily close to zero (figure 1**B**), nor may transcript initiation-rate tend to infinity. This simple model of specificity illustrates the important trade-offs in balancing specificity, expression and noise. The biological problem to be solved is how to reconcile the requirements for finite activator on-rates and finite off-rates with the need for high regulatory specificity.

The kinetic proofreading of activator promoter interactions

We now discuss a model, model 2 (the proofreading model) which is a potential solution to this biological problem. For simplicity, we consider a promoter (enhancer plus core promoter) with a single nucleosome position and single activator binding site (model 2, figure 2**A**). The nucleosome is subject to stochastic removal and reformation [13]. Removal may occur by nucleosomes sliding away from promoter sequences [26] or disassembly [27]. Our promoter model encompasses four states: with nucleosome but without activator (state 1), with nucleosome and activator (state 2), with activator but without nucleosome (state 3), and without both (state 4).

We assume: that transcription requires both activator binding and nucleosome removal, transcription occurs in state 3 alone; the activator recruits chromatin remodeling activities to the promoter that catalyze removal of the nucleosome, nucleosome removal in the presence of the activator occurs at a faster

rate than its absence, $\alpha > \lambda$ in figure 2**A**. In addition, for simplicity, we assume that the kinetics of activator binding are not altered by the nucleosome, the energetics of activator binding are the same for all models discussed here. However, similar conclusions may be drawn from more complex models where activator binding dynamics are altered by the nucleosome.

Our model, thus, entails two types of promoter states, activator-bound and unbound, with different nucleosome removal kinetics. This kinetic asymmetry implies the system is not in equilibrium; the system maintains a closed loop of reactions where transition cycles in one direction (clockwise in figure 2**A**) are more probable than transition cycles in the reverse direction (see Appendix 3).

For every clockwise cycle of promoter state transitions, transcription requires the bound activator twice: for removal of the nucleosome *and* initiation of transcription. This sequential twofold requirement allows for kinetic discrimination between target and non-target promoter binding twice, transitions $2 \rightarrow 1$ and $3 \rightarrow 4$ in figure 2**A**. As a consequence, without changing the energetics of activator-DNA binding, the upper limit of activator fidelity increases from f_0 to f_0^2 (see Appendix 3). Thus, activator fidelity of this model may significantly surpass the Hopfield barrier; and it always exceeds the fidelity afforded by Model 1 (figure 2**B**). An analogous result was obtained by Hopfield for enzyme kinetics [12]. Following Hopfield, we call the mechanism which affords this increase in activator fidelity 'kinetic proofreading' [25].

With kinetic proofreading fidelities close to the Hopfield barrier may be attained with faster activator on-rates, k, than without. This is because the nucleosome 'filters out' many incorrect activator binding events. Fast on-rates for the activator markedly dampen temporal fluctuations in transcript number (compare blue trace and grey trace in figure 2**C**). Thus, at equal fidelity, kinetic proofreading affords lower transcription noise (compare blue and stippled curve in figure 2**D**).

Activator-dependent nucleosome removal, $\alpha > \lambda$, entails the preference of clockwise cycles over counterclockwise cycles and is essential for nucleosome-mediated kinetic proofreading. The preference for cycles in one direction over another requires energy expenditure and deviation of the steady state from equilibrium. If, $\alpha = \lambda$, the system steady state is also thermodynamic equilibrium or 'detailed balance': forward and reverse transition of all reactions are equally probable (see Appendix 3). Clockwise and anticlockwise cycles are equally probable and kinetic proofreading is lost: $f_1 = f_2$ (figure 2**B**). Activator fidelity, again, is limited from above by the Hopfield barrier, (see Appendix 3).

Therefore, in equilibrium the transcript number wildly fluctuates for fidelities close to f_0 (figure 2**C**) since fidelities close to f_0 require activator on-rates, k , close to zero, reference equation for f_1. Random transitioning between transcriptionally conducive and inconducive states, exacerbates the noise in transcription, as expected (figure 2**D**), the slower the nucleosome dynamics the higher the noise.

Maintenance of nucleosome dynamics away from equilibrium requires entropy production, free energy dissipation. The greater the distance from equilibrium, the more energy must be dissipated per unit time (figure 2**E**). Although kinetic proofreading in Model 2 requires non-equilibrium promoter dynamics, $\alpha = \lambda$, increasing free energy expenditure to increase the rate of nucleosome removal does not monotonically improve activator fidelity (figure 2**B**). The reason is simple: kinetic proofreading requires nucleosome removal kinetics that are slow relative to the dissociation kinetics of activators that bind the promoter nonspecifically. As nucleosome removal speeds up past this point, specificity approaches, f_0, while the energy required to remove nucleosomes faster ever increases.

Effective kinetic proofreading requires that initiation of transcription is tied to activator binding. If transcription persists after dissociation of the activator, model 3 (figure 3**A**), kinetic discrimination between correct and incorrect activators occurs only once: before removal of the nucleosome, transition $2 \rightarrow 1$. The double checking mechanism is lost. Since transcription is partially uncoupled from activator binding, activator fidelity remains well below the fidelity afforded by Model 1, despite nucleosome dynamics away from equilibrium (figure 3**B**; see Appendix 3). In addition, the noise in transcription rises faster in Model 3 than Model 1 as fidelity increases with decreasing activator on-rate (figure 3**C**). Energy expenditure alone is insufficient to yield kinetic proofreading, coupling of the presence of the activator in multiple steps is also a requirement.

Additional steps of kinetic proofreading may further increase activator fidelity. The central component for core promoter recognition of all genes, the TATA box binding protein, TBP, is thought to be recruited to promoters by activators as a subunit of either the SAGA or TFIID complex [28]. The enzyme Mot1 couples ATP hydrolysis to the removal of TBP from DNA [29,30]. Thus, Mot1 must drive the TBP-DNA binding reaction away from equilibrium. This may afford a second step of activator proofreading. If transcriptional initiation requires continued activator binding for steps downstream of TBP binding, Model 4 (figure 3**D**), the system may discriminate between target and non-target activator binding thrice, before nucleosome removal, transition $2 \rightarrow 1$, after nucleosome removal, transition $3 \rightarrow 4$, and after TBP binding, transition $5 \rightarrow 6$. The second proofreading step increases the upper limit of fidelity to f_0^3 (figure 3**E**) and, as may be expected, affords further noise suppression (figure 4).

Discussion

In equilibrium, there is a fundamental limit, the 'Hopfield barrier', to how well any information processing task can be undertaken, *e.g.* transcription of specific genes in response to an environmental signal [25]. However, the Hopfield barrier may be surpassed at the expense of free energy to maintain the system away from equilibrium. Outside of equilibrium the system dynamics are irreversible, for some sequence of events, forward and reverse direction are statistically distinguishable (see Appendix 3); in the proofreading model, Model 2 (figure 2**A**), clockwise cycles

are more probable than counterclockwise cycles, indicating irreversibility and maintenance away from equilibrium.

Non-equilibrium dynamics, irreversibility, corresponds to entropy production (figure 2**E**). Therefore, nucleosomal proofreading of activator-DNA binding calls for enzymes that couple the catalysis of nucleosome dynamics to exergonic, entropy producing, reactions. That such enzymes indeed exist, ATP-dependent chromatin remodelers [31], fulfills a critical demand of our theory.

Irreversibility is a necessary but by no means sufficient condition for kinetic proofreading. For instance, in the Model 2 if nucleosome removal was much faster than the off-rate of the incorrect activator, $k_n << \alpha$, the promoter dynamics would still be irreversible. Yet, the expenditure of energy would afford no increase in activator fidelity. The ability to distinguish between target and non-target promoter binding in state 2 no longer exists. This explains the monotonic decrease in activator fidelity beyond an optimal rate for activator-controlled nucleosome removal (figure 2**B**). Activators must promote nucleosome removal to increase activator fidelity, however, if nucleosome removal is too effective the fidelity gain is lost. Therefore, transcriptionally active promoters must not be nucleosome-free, on average, in good agreement with experimental observation [13,32].

Our theory implies that ATP hydrolysis by one or more chromatin remodelers recruited to the promoter is not used to speed up the approach to equilibrium but to maintain nucleosome dynamics away from equilibrium. This demand may be difficult

to satisfy if removal of nucleosomes occurred by sliding alone. For the same remodeler may use ATP hydrolysis both to slide nucleosomes away from the promoter and back. Thus, removal of nucleosomes by sliding is an inefficient use of ATP hydrolysis to drive nucleosome dynamics away from equilibrium. Although, if activators propagate uni-directional sliding this inefficiency would be assuaged. In contrast, nucleosome removal by ATP-dependent nucleosome disassembly couples ATP hydrolysis only to removal. The reverse reaction, nucleosome reassembly, entails synthesis of ATP from ADP and phosphate, which under physiological conditions, renders the reverse reaction highly improbable. This may explain why nucleosomes are removed from transcriptionally active promoters and not simply slid away [27].

Are the dynamics of activated promoter nucleosomes non-equilibrium dynamics? In our theory, irreversibility is engendered by activator-stimulated nucleosome removal. Consistently, Pho4, the transcriptional activator of the *PHO5* gene, promotes loss of *PHO5* promoter nucleosomes [19,22], not by occluding nucleosomes but by recruitment of ATP-dependent chromatin remodelers [13,33,34]. Therefore, It is likely no accident that assumptions of irreversibility, the unidirectional sliding and ordered removal of nucleosomes [13,20], helped to explain the observed statistical frequencies of *PHO5* promoter nucleosome configurations. Accordingly, nucleosome occupancy at many promoter sequences in yeast is not explained by the thermodynamics of nucleosome formation [35].

Kinetic proofreading involves continual reactivation of transcription: to test for the continued presence of the (correct) activator, the promoter stochastically returns to nucleosome configurations that suppress transcription, despite environmental conditions that induce gene activity. Thus, the variation in promoter chromatin structure must be intrinsic, independent of the environment. This prediction of the theory has been tested [23]. If the nucleosomal variation was imposed by environmental variation, the nucleosome configuration of one promoter copy would be stochastically dependent on the configuration of another copy within the same cell. Contrary to this expectation, electron microscopic analyses of *PHO5* promoter pairs in single cells showed that both copies were stochastically independent [23]. The heterogeneity of promoter chromatin cannot be reduced to environmental variation; it must arise 'intrinsically'. Thus, promoter nucleosome dynamics fulfill another critical demand for kinetic proofreading.

A theory of kinetic proofreading by nucleosomes suggests that activators promote multiple steps toward transcription. Notably, eukaryotic activators do not bear specific activities to stimulate transcription, but promiscuously recruit other factors instead [2,3]. Promiscuous recruitment easily affords the same activator the ability to promote multiple, biochemically distinct, steps toward transcription. In addition, recruitment delays the activator's effect on transcription, a critical requirement for effective kinetic proofreading (figure 2**B**). Thus, we predict activation by promiscuous recruitment is a feature of eukaryotic activators which enhances specific activation, contrary to intuitive expectation.

It is possible that transcription may not strictly be limited to activator-bound promoter states because past activator binding events are remembered in form of other factors, *e.g*. TBP, that are recruited by the activator but may remain at the promoter after dissociation of the activator. This uncoupling of transcription from activator binding diminishes the effectiveness of kinetic proofreading, and in fact reduces specificity in general (figure 3**B**). We predict that the cell makes an effort to ensure that transcription only occurs when the activator is present. Thus, our theory requires the existence of an enzyme that removes TBP from promoter DNA: Mot1. TBP removal by Mot1 erases the memory of past activator binding events. Erasure maintains a close relationship between transcription and activator binding; as a consequence, maintenance of transcription requires continual reactivation.

Mot1 may contribute to activator fidelity in yet another way. Energy expenditure by Mot1 affords a second kinetic proofreading step, Model 4 (figure 3**D**). Mot1 must drive the TBP-DNA binding reaction away from equilibrium, which may be used by the cell to significantly increase activator fidelity (figure 3**E**). In this context it is of interest that both Mediator and TBP are recruited to promoters in association with repressing factors [36–38], which may provide additional proofreading steps. However, whether relief of this repression requires free-energy expenditure is not known. Multiple kinetic proofreading steps are likely required to solve the activator specificity problem in eukaryotes.

We note, the free energy for regulatory specificity may be provided by activator-DNA binding alone. However, higher binding affinities entail longer dwell

times of the activator on its target sequence. Thus, activated genes are not easily turned off again. Activator fidelity may be increased by decreasing activator concentration. In this case, the price to be paid is long activator search times. Long search times imply erratic promoter activity, blurring the correspondence between regulatory signal and transcriptional response (figure 2**C**). Kinetic proofreading of activator-promoter interactions may be used to resolve this dilemma. The free energy required to increase specificity and dampen transcription noise is provided by ATP hydrolysis and not by increasing the binding energy for activator-promoter recognition, by increasing the average lifetime of the activator on the DNA. Thus, kinetic proofreading reconciles high specificity with fast promoter state kinetics, controllability.

Archaea possesses a precursor of the nucleosome [39], but eubacteria lack nucleosome-like structures entirely. How do eubacteria solve the problem of regulatory specificity? Surprisingly, many regulators of transcription in eubacteria, *e.g*. the TetR and lac repressors, recognize their target sequences with greater specificity than most eukaryotic activators: the equilibrium dissociation constants often fall into the picomolar range and energy differences between target and non-target promoter binding are two or three times larger compared to most eukaryotic activators [7,9,40,41]. Eubacteria solve the specificity/controllability problem in a different way. The entailed problem of long DNA dwell times is solved by controlling the activity of transcriptional regulators *via* allosteric effectors, *e.g*. tetracycline and allolactose, that upon binding induce large changes in the affinity of

the transcription factor for its target sequence [42]. Thus, the free energy required to reconcile controllability with specificity is spent creating or localizing allosteric effector molecules. The evolution of transcriptional activators in the presence of a ubiquitous repressor of transcription, the nucleosome [43], may have proffered kinetic proofreading as a solution to the specificity problem instead.

Irreversibility is a probabilistic and not deterministic phenomenon [44]. Transcription and its regulation may be fully understood only on the basis of probabilistic theories, whether kinetic proofreading is employed to: increase activator fidelity, as suggested here, sharpen the gene regulatory function [25], increase the fidelity of substrate recognition by RNA polymerase [45] or promoter recognition by general transcription factors [46]. The randomness of molecular interaction is built into biological function. Rigorous testing of theories that assert random molecular behavior requires methods for the analysis of gene expression at the level of single gene molecules [47].

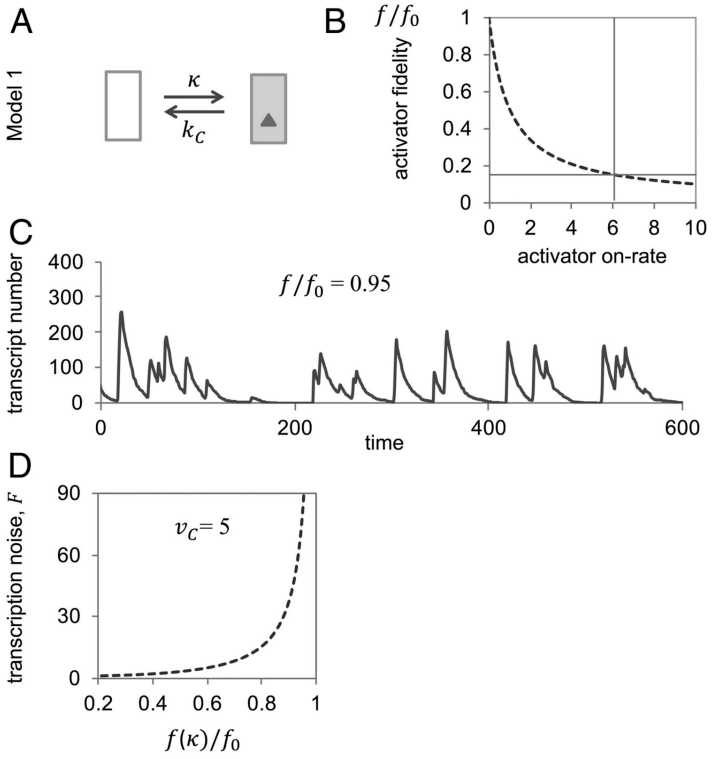

Figure 1**. Specificity and noise in the two-state promoter model:** Activator fidelity is bounded by the Hopfield barrier. **(A)** Transition graph of Model 1. **(B)** Activator fidelity approaches its upper limit or Hopfield barrier, $f_0 = e^{\frac{-\Delta\Delta G^\circ}{RT}}$, as the activator on rate, k, tends to zero. To calculate the graph, we assumed $k_t = 1$, $k_n = 100$. Actual fidelities must be markedly lower than f_0: For instance, measured off-rates for Pho4 of yeast (the activator of PHO5) for target and non-target sequences are ~0.01 s^{-1} and 1 s^{-1}, respectively [7,9]. From Pho4's equilibrium dissociation constant for correct binding of K_d=11 nM [9], and nuclear concentration of ~60 nM [48] (assuming a nuclear volume of 4 femtoliters), both the on-rate, k =0.06 s^{-1} (indicated by a vertical line) and relative fidelity (indicated by horizontal line) may be calculated; the unit on the abscissa, then, is 0.01 s^{-1}. **(C)** Representative 'sample path' (single cell trajectory of mRNA abundance) at relative activator fidelity of 0.95; the sample path was obtained with Gillespie's stochastic simulation algorithm [49] with $k = 0.05$, $k_t = 1$, $\delta = 0.1$ (rate constant for mRNA degradation) and average rate of transcription, $v_t = 5$. **(D)** The Fano factor tends to infinity as activator fidelity, f_1 , approaches the Hopfield barrier, $\frac{f_1}{f_0}$. Calculations were based on rate constants from C. Both Fano factor and fidelity were calculated as functions of the activator on-rate, k (see Appendix 3).

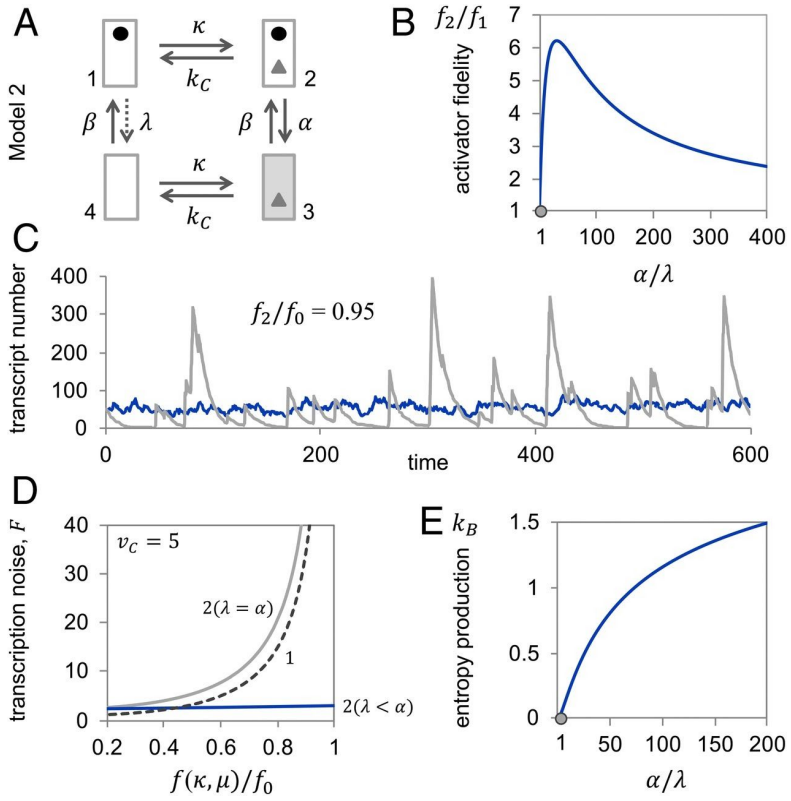

Figure 2**. Nucleosome dynamics away from, but not in, equilibrium allow for increased activator fidelity and attenuation of transcription noise: (A)** Transition graph of Model 2. **(B)** Activator fidelity of Model 2, f_2, normalized by the fidelity of Model 1, f_1, as a function of the rate of nucleosome removal in the activator-bound state, α, normalized by the rate of removal in the unbound state, λ. For calculations we assumed $k = 1$, $k_c = 1$, $k_n = 100$, $\beta = 2$, and $\lambda = 0.1$. The gray dot indicates the equilibrium state. **(C)** Representative sample paths at relative activator fidelity $\frac{f_2}{f_0} = 0.95$ and $v_t = 1$ for non-equilibrium nucleosome dynamics (dark gray; $\alpha = 2,\ \lambda = 0$), which required $k = 4.26$ and $\mu = 13.26$; and equilibrium dynamics (light gray; $\alpha, \lambda = 2$), which required $k = 0.053$ and $\mu = 198.68$. For both simulations we assumed $\delta = 0.1$, $v_t = 5$. **(D)** Transcription noise as a function of relative activator fidelity, $\frac{f(k,\mu)}{f_0}$, for Model 2 in equilibrium (light gray, $\alpha, \lambda = 2$), away from equilibrium (blue, $\alpha = 2,\ \lambda = 0.1$), and Model 1 (stippled line, 1; same as in Fig. 1D). For all calculations we assumed, as above, figure 1. Fano factor and activator fidelity were calculated as functions of the activator on-rate, k, and the rate of transcription in the active state, μ (see Appendix 3). **(E)** Entropy production (in units of k_b, the Boltzmann constant) as a function of nucleosome removal rate in the activator-bound state, α, relative to the rate in the unbound state, $\lambda = 2$, for $k = 1$. The gray dot indicates the equilibrium state.

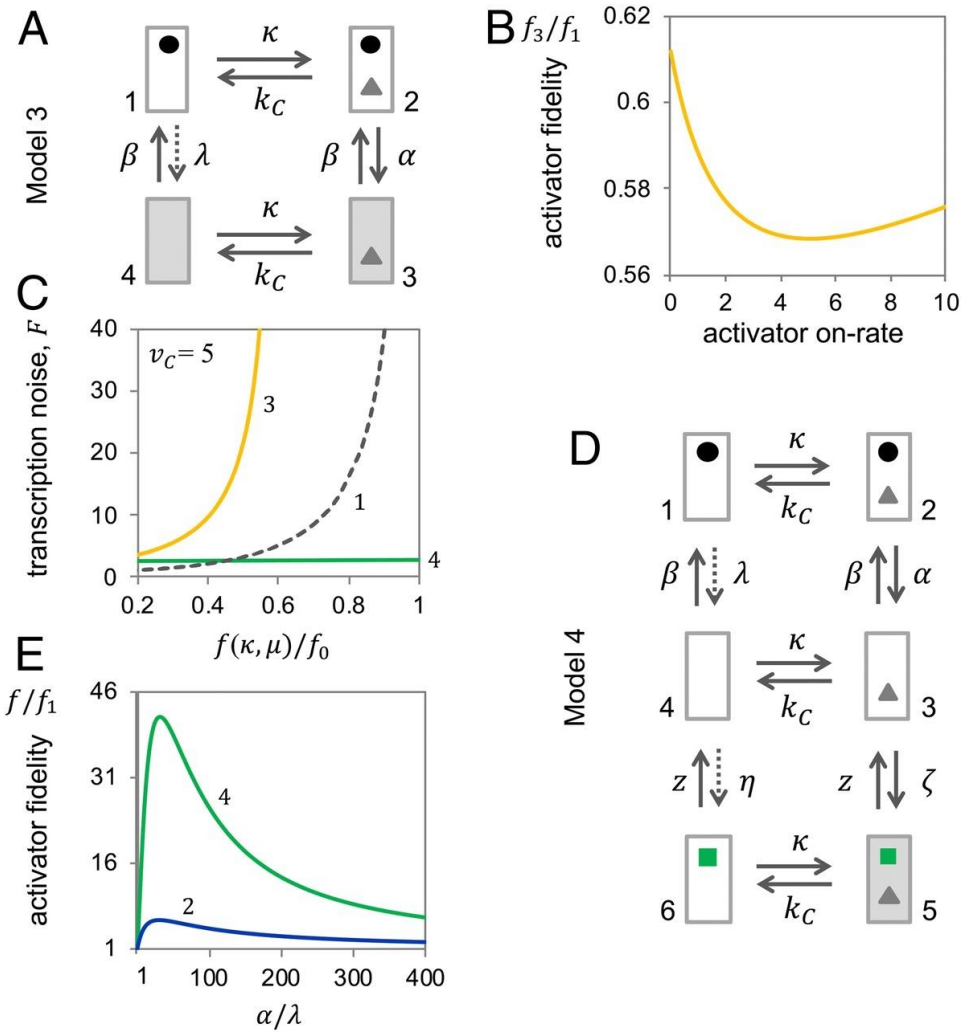

Figure 3. **Kinetic proofreading requires coupling of transcript initiation to activator binding; multiple proofreading steps improve fidelity: (A)** Transition graph of Model 3. **(B)** Activator fidelity of Model 3, f_3, relative to the fidelity of Model 1, f_1, as a function of activator on-rate, k; with $k_c = 1$, $k_n = 100$, $\alpha, \beta = 2$, and $\lambda = 0$ (for $\lambda > 0$, fidelities further decrease). **(C)** Transcription noise as a function of relative activator fidelity, $\frac{f(k,\mu)}{f_0}$, for Model 3 (yellow, 3), Model 4 (green, 4), and Model 1 (gray stippled, 1). As in calculations for figure 2, we assumed $\delta = 0.1$ and $v_t = 5$; for Model 4 alone: $\lambda = 0.1$ and $z, \zeta = 10$ to reflect both active removal (by Mot1) and high concentration of TBP; all other parameters were as indicated above. Fano factor and fidelity were calculated as functions of the activator on-rate, k, and the rate of transcription in the active state, μ (see Appendix 3). **(D)** Transition graph of Model 4. **(E)** Activator fidelities of Model 4 (green, 4) and Model 2 (blue, 2) relative to fidelity for Model 1 as a function of $\frac{\alpha}{\lambda}$. For Model 4, we assumed $\alpha = \zeta$ and $\lambda, \eta = 0.1$ and $z = 10$.

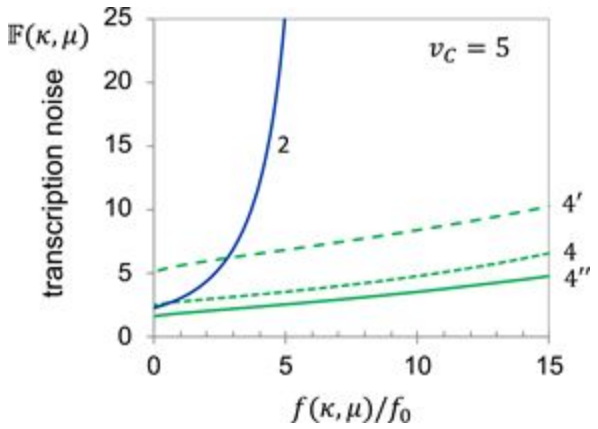

Figure 4**. Additional proofreading steps further attenuate noise:** Transcription noise as a function of relative activator fidelity, $\frac{f(k,\mu)}{f_0}$, for Model 4 (green, 4,4',4'') and Model 2 (blue, 2). All parameter values were as indicated in the legends to figure 2 and figure 3, except 4 ($\zeta = 10,\ z = 10$) and 4' ($\zeta = 2,\ z = 2$) and 4' ($\zeta = 2,\ z = 10$).

Chapter 2: Loss of isw2 abolishes temporal correlation between bursts generated by nucleosome dynamics

Robert Shelansky, Chris Brown, Michael Doody, Kevin Chen, Stacy Harvey, Eduardo Hirata, Heta Patel, Gustav Peterson, Sara Abrahamson, Dan Larson, Tineke Lenstra, and Hinrich Boeger

Within each Eukaryotic cell, DNA is spooled around histone protein octamers to form nucleosomes. The nucleosome acts as a general repressor of gene expression. Repression is established by the occlusion of DNA from transcription factors [50]. Occlusion of transcription factors is maintained passively, by masking or destabilizing interactions with DNA [51,52], or actively, by eviction via catalysis of nucleosome sliding by ATP-dependent chromatin remodelers [53]. To relieve repression nucleosomes must be removed. Gene activation engenders a quantitative change in promoter chromatin structure; a transition from a nucleosome-rich structure in repressing conditions to a nucleosome-depleted, but not nucleosome free, structure in activating conditions [54,55].

When promoter chromatin structure is interrogated through methods which rely on statistical averaging, structural variation between molecules becomes obfuscated [56]. For example, transcriptionally active genes across eukaryotes maintain nucleosome depleted regions (NDRs) [57–59] flanked by a pair of well positioned nucleosomes, the '-1' and '+1'. Surprisingly, when viewed at single

molecule precision in activating conditions, promoter chromatin varies drastically from gene molecule to gene molecule [55]. This may manifest by a complete collapse of the NDR or a more subtle shift in the position of a nucleosome over important regulatory sequences. Thus, assuming ergodicity, individual gene molecules either randomly return to or remain in a repressed state under activating conditions. While the purpose of nucleosome removal is evident, the persistence and/or reformation of promoter nucleosomes under activating conditions is not. Why does the cell expend so much energy to maintain heterogenous promoter chromatin? Which regulatory structures are “active” structures? The specific biochemical steps in NDR formation, activator binding, and transcription remains elusive [60,61].

Regulation of transcription including the removal of nucleosomes is achieved through specific activation by transcriptional activators or repressors which bind to specific DNA sequences in the promoter region of their target genes. Specificity in gene selection therefore requires that transcription factors maintain a higher affinity for their target sequences over other non-target sequences. Differences in transcription factor affinity are primarily determined by differences in the off-rate constant. However, A stably bound activator is not easily removed. The cell must make a compromise between specificity and controllability (see Chapter 1).

Sparse activator binding is a requirement for high specificity. However, as activator search-times tend to zero, the search time for specific binding sequences tends to infinity. It becomes increasingly difficult to quickly activate target-genes and controllability is lost (see Chapter 1). In addition, highly expressed genes face yet

another challenge. To compensate for long time intervals between binding events, the rate of transcript initiation while the activator is bound on the DNA must be high. Long periods of inactivity are interrupted by shorter periods of high activity. Transcription occurs in random bursts. The correspondence between gene activity and physiological state becomes blurred. Thus, increasing specificity and reducing gene expression noise are competing requirements (see Chapter 1).

Specificity and fast regulatory kinetics may be reconciled through the insertion of an energy-dissipating delay step into the activation process which kinetically distinguishes between target and non-target activator binding twice (see Chapter 1). The activator's presence is “double checked”: once prior to the removal of a transcriptional repressor, the nucleosome, and once again subsequently [62]. As activators bind and unbind, so do promoter nucleosomes. This dynamic process which occurs on single gene molecules produces structural heterogeneity of individual gene molecules. Thus, the structural variation observed across a population of gene molecules is pivotal to regulation, rather than a hindrance.

The increase in activator fidelity is attained without increasing the affinity between activator and target promoter. This allows for the retention of both fast activator on- and off-rates. Thus, fast regulatory kinetics are achievable while maintaining high activator fidelities close to or above the limit imposed by the “hopfield barrier”, the maximal fidelity dictated by the energetics of activator-DNA binding. Furthermore, fast regulatory kinetics reduce gene expression noise.

Enhanced fidelity and fast regulatory kinetics come at a cost, continual free-energy consumption (entropy production).

Activators are thought to recruit ATP-dependent chromatin remodelers to promoters, coupling nucleosome transitions to both ATP hydrolysis and activator binding. An irreversible cyclical process ensues (figure 5, the proofreading model), same model as Chapter1, figure 2**A**. This cyclical process maintains a sequence of molecular transitions with cycles in one direction of rotation which are statistically preferred over cycles in the reverse direction: activator binding → nucleosome removal → activator loss → nucleosome reformation, is more likely than, nucleosome loss → activator binding → nucleosome reformation → activator loss. Free energy is spent to create this statistical preference in one direction over the other, affording kinetic proofreading. The free energy for irreversibility is provided by ATP-dependent chromatin remodeling. Additionally, kinetic proofreading of activator-promoter interactions requires that transcription is linked to the presence of the activator on the DNA (in figure 5, only state 3 is transcriptionally active). Activators must facilitate additional steps toward transcription following nucleosome removal.

As discussed above, linking transcription to the transient DNA-binding of activators engenders transcriptional bursting, periods of correlated initiation events. Similarly nucleosome removal and reformation dynamics should introduce additional temporal correlations, possibly at longer timescales. Slow proofreading kinetics are more effective in increasing the specificity of activator-promoter interactions than fast

kinetics. On the other hand, proofreading kinetics must not be too slow, or else the kinetics of regulation, again, become slow and the noise of expression high. The phenomena of transcriptional bursting may be viewed as a byproduct of the biochemical processes required to maintain specificity in activation for highly expressed genes.

Whether chromatin remodelers effect transcriptional bursting and if so at which timescale is unknown (see Appendix 1). To address this question, we analyzed the transcriptional activity, *in vivo*, of single gene molecules of a classic model of eukaryotic promoter chromatin structure, the yeast *PHO5* locus (see Appendix 2). Using multifocus fluorescence microscopy, we found that transcription signals are correlated at two distinct timescales, consistent with transcriptional bursting. Mutations in the activation domain of Pho4, the activator of *PHO5*, affected both burst frequency and size. This is consistent with the multiple roles the activator plays in transcription. Surprisingly, loss of a single ATP-dependent chromatin remodeler, Isw2, abolished long-timescale correlations but not transcription, suggesting that the absence of Isw2 locks the promoter in a state conducive to transcription. We hypothesize that Isw2 is a critical component for the kinetic proofreading activator-DNA interactions.

In addition, we wished to investigate the relationships between activator binding, and chromatin remodeling at the *PHO5* locus. Using Chromatin Endogenous Cleavage Cross Linking Kinetic (ChEC-CLK) analysis, we found that activator binding kinetics correspond to “fast” timescale bursting and that the activation

domain of the activator Pho4 modulates its own binding. We then test a model by which Pho4's regulation of nucleosome dynamics at the PHO5 locus modulates activator binding in the promoter using electron microscopy of single gene molecules.

Results

Transcriptional bursting of *PHO5* occurs at distinct time scales

To attain high regulatory specificity, promoters must randomly transition between transcriptionally conducive ('ON') and inconducive (OFF') states. The transitioning between ON and OFF states entails the grouping of transcription events into bursts and, potentially, groups of bursts. Thus, transcription events become temporally correlated, possibly at different timescales.

To observe temporal correlations between transcription events, we analyzed nascent transcript number dynamics for individual *PHO5* gene molecules by performing live-cell RNA visualization using fluorescence MultiFucus Microscopy (MFM). Using MFM we captured a 7 frame z-stack spanning the volume of the yeast nuclei as quickly as 5 frames per second [63]. Faster acquisition is possible by MFM; however, acquisition speed is limited by our use of a whole cell RNA labeling technique. We necessarily analyze nascent RNA, and so must "blur" cytoplasmic transcripts using a long, >150ms, exposure. Without "blurring" it becomes difficult to differentiate cytoplasmic and nascent transcripts.

Labeling of *PHO5* RNA was achieved with the PP7 labeling technique [64]. The PP7 technique relies on the insertion of sequential repeats into a gene of

interest (figure 6). These repeats encode RNA hairpin motifs, PP7, which are recognized and bound specifically by a RNA binding protein, PP7 coat protein (PCP), fused to a fluorescent protein, GFPenvy [65]. We inserted 14 PP7 repeats, each containing two binding sites for PCP, into the 5' end of the *PHO5* locus and expressed PCP endogenously under control of the yeast *RPS2* promoter.

We tracked fluorescent puncta associated with the site of transcription (TS) from individual cells (figure 7**A,** blue trace) and partitioned these individual time series into periods of transcriptional activity (≥1 nascent transcripts were detected) and inactivity (transcript absence) by change point detection (CPD) (figure 7, grey shaded areas). On average active periods were short, 124s ±13s. These short active periods were separated by longer periods of inactivity that lasted 449 ± 66s, on average. Thus, the probability of transcriptional activity was 0.19. Active periods appeared to correspond to bursts of transcription, *i.e*. multiple initiation events in short sequence, rather than single initiation events. Several observations support this conclusion. i) The average lifetime of the active periods exceeds the expected lifetime of a nascent transcript by a factor 2. ii) The statistical distribution for the length of active periods was better described by an exponential distribution, as expected for the sojourn in an ON state, and not the distribution for the lifetime of the nascent transcript. iii) The intensity profiles of many active periods appeared to increase and decrease in a stepwise fashion, suggesting the release of multiple transcripts in short sequence (figure 19**B**). iv) If active periods corresponded to single initiation events, the cell would synthesize only one transcript every 7.5 minutes on average. This low initiation-rate would make it difficult to account for the large

number of cytoplasmic transcripts observed by single molecule fluorescent in situ hybridization (smFISH), ~40 on average (figure 11**A,** blue histogram). v) Active periods with an average of 1 nascent transcript contradicts TS intensity distributions determined by smFISH (figure 19**A**). Thus, we infer that periods of activity and inactivity correspond to bursts of transcription (see appendix 1).

Ignoring signal fluctuations during active and inactive periods, fluorescence-time series may be replaced by step functions that alternate between two values only: 1 (active) and 0 (inactive) (figure 7**A,** black line). This approximation captured temporal correlations in our data well, the autocorrelation function (*acf*) calculated from step functions closely resembled the *acf* obtained from fluorescence-time series. This implies that much of the correlation observed in our nascent transcript fluctuations corresponds to the correlation introduced by active/inactive transitioning alone and not by variation in transcript number. This finding significantly simplified both the mathematics of the autocorrelation function (see appendix 3) and interpretation of the data, the number of model parameters was reduced. Using this approximation we demonstrated the logical consistency of results obtained by autocorrelation and change point analysis.

A simple bursting process (a two-state promoter model) produces temporal correlations determined by the length of both active and inactive periods. The active period lifetimes are set by the rate of transition from active to inactive, κ_A. While the inactive period lifetimes are set by the rate of transition between inactive to active, κ_I. The theoretical *acf* of a two-state promoter model is that of a bernoulli process,

represented by a single exponential function $e^{-\lambda_1 t}$, where $\lambda_1 = \kappa_A + \kappa_I$ (see appendix 3). However, we found that the observed *acf* did not fit a single exponential but the sum of two weighted exponentials. The two exponential distributions had markedly different kinetic parameters, fast (1.6 X 10^{-2} s^{-1}) and slow (4.4 X 10^{-3} s^{-1}). Thus, we conclude that the standard 2-state model is insufficient to account for our observations. The data suggest at least two routes into the active state indicated by the two exponential fit.

In contrast, the proofreading model, which encompasses more than one inactive state and thus allows for both bursts and the groups of bursts, was consistent with our observations; a close fit between experimental and theoretical *acf* could be attained that also recapitulates the results from CPD, and smFISH and activator binding dynamics.

Each additional promoter state adds an additional weighted exponential to the *acf* (see appendix 3). For the proofreading model, the parameters of the exponentials are polynomial functions of all rate constants with the exception of the first exponential parameter $\lambda_1 = \kappa_{ON} + \kappa_{OFF}$, the sum of the activator's on- and off-rate. However, if the dynamics of nucleosome removal, β, and reformation, α, are significantly slower than the dynamics of activator-promoter binding, the additional parameters simplify to $\lambda_1 \approx \lambda_2$ and $\lambda_3 \approx \frac{1}{2}\alpha + \beta$ (see appendix 3); thus, while one exponential parameter reflects the kinetics of activator binding and unbinding, λ_1, the other, λ_3, corresponds to the kinetics of ATP-dependent nucleosome removal and reformation. Slow proofreading kinetics may engender long-timescale

correlation, and thus explain the appearance of a second exponential with a slower kinetic parameter (figure 7**C**, blue curve). We therefore tentatively conclude that bursting occurs in at least two distinct timescales, a “slow” timescale generated by slow proofreading dynamics, and a second “fast” timescale generated by the dynamics of the activator. We note here that many more sophisticated multi-state models with a wide variety of topologies may explain the dynamics observed at PHO5.

It must be kept in mind that the classifying of transitions between model states as distinct physical transitions is at best an approximation; for practical and epistemic reasons models must be less detailed than reality. As long as we conflate observed active periods with the sojourn in the unobserved ON state, the rate constants that determine the kinetics of transitioning from OFF to ON (figure 5), the on-rate of the activator, k_{ON}, and the rate constant for nucleosome removal, α, must encompass additional steps toward transcription, such as the binding of general transcription factors, open complex formation, transcript initiation, and promoter clearance. For example, the assumption of a single activator on-rate for both the nucleosomal and nucleosome free promoter (figure 5) imply that kinetics of additional steps toward transcription that only occur after nucleosome removal must be subsumed into the rate constants available in the the proofreading model; it is likely not the case that binding of the activator is the only step required in the absence of the nucleosome for the gene to initiate transcription. This assumption is expedient rather than realistic. The specific rates inferred are thus likely false. However, the proofreading model, surprisingly, explains our observations while predicting a

problem to be solved by its implementation: activator-promoter specificity. We use this simplified model as a tool to better understand regulation, specificity, and the emergence of transcriptional bursting.

Nucleosome dynamics generate correlation at a "slow" timescale In PHO5 transcription

To test the notion that ATP-dependent chromatin remodeling engenders long-timescale correlations, we recorded fluorescence-time series of *PHO5* in various chromatin remodeler mutants (figure 17**B-C**, figure 18**B-C**). We expected loss of specific chromatin remodelers to impact the slow exponential of the acf. Surprisingly, we found that loss of the ATP-dependent chromatin remodeler Isw2, virtually abolished long-timescale correlations (figure 9**C**, gold curve), but not transcription. Although expression from the *PHO5* locus in the isw2 delete was reduced to 72% of wild type (WT).

Evidently, loss of Isw2 abolished nucleosomal repression under activating conditions, the acf of the proofreading model for WT closely fit the experimental acf of the mutant simply by setting $\beta = 0$. However, the Isw2 phenotype was more complex. Loss of nucleosome repression should increase burst frequency. However, burst frequency seemingly decreased, the inactive periods significantly increased in length to 612s ±13s, while burst duration decreased, the active period times were 111s ±13s (figure 9). In the context of the proofreading model and in the absence of nucleosomal repression, these changes must be attributed to a decrease in the rate

of activator binding, κ_{ON}, and an increase in the rate of activator dissociation, κ_{OFF}. The opposing effect on both rate constants explained why their sum, inferred from the *acf*, remained virtually unaltered. Additionally, the deviation in activator binding kinetics observed between Isw2 delete and WT may be attributed to the simplification of the model by setting β = 0; the model becomes a two-state model. Rate constants that had previously been encapsulated within the nucleosome dynamics in a simpler model now impacted inferred activator binding kinetics. Thus, loss of Isw2 traps the promoter into a subset of states, in the proofreading model states without the nucleosome, while at the same time inhibiting activator binding and leading to increased activator removal.

Loss of other chromatin remodelers, however, did not have the same impact. For instance, loss of Chd1 had little effect on the *acf* but did lead to reduced *PHO5* expression similarly to that of Isw2 (figure 10). In contrast to Isw2, the *acf* of the Chd1 deletion was fit well by two exponentials with parameters closely similar to WT, fast (1.6 X 10^{-2} s^{-1}) and slow (4.1 X 10^{-3} s^{-1}). There was only a mild effect on the weights of the exponential terms, causing a slight shift of the *acf* toward shorter correlation times (figure 10**C,** green curve). CPD analysis indicated both a shortening of active periods, 107s ±13s, and extension of inactive periods to 538 ± 50s. CPD was in good agreement with the observed decrease in *PHO5* expression to about 78% of WT. Loss of Chd1 (like Isw2) reduced both burst frequency and duration. However, unlike loss of Isw2, Chd1 had no discernible effect on the kinetics of the proofreading step, long-timescale signal correlations were virtually unchanged. Loss

of function of other remodelers had either no effect on *PHO5* expression (Isw1) or abolished *PHO5* expression entirely (Swi/Snf).

We have made two deletions in different chromatin remodelers, Isw2 and Chd1, which impact PHO5 promoter chromatin structure and likely play different roles in regulation of transcription and chromatin architecture. These impacts on chromatin structure propagate to changes in transcriptional dynamics, changes in burst frequency and duration. We conclude Isw2 is required for generation of groups of bursts in Pho5 transcription. Loss of Isw2 abolished nucleosomal repression under activating conditions. We infer Isw2 establishes a repressed state under activating conditions.

Pho4 regulates burst frequency and burst duration

To test the proofreading model further various mutations were introduced into the *Pho4* activation domain. These sequentially larger deletions of the activation domain result in a continuous loss of expression (figure 11**A-E**) [66]. Since activators are thought to accelerate multiple steps toward transcription, *e.g*. nucleosome removal and recruitment of GTFs to the promoter etc., we expected that activator mutations slow down promoter transitions and thus decrease one or more exponential parameters in the *acf*. We were surprised, therefore, to find the opposite. The *acf* obtained from mutant cells that bore a deletion mutation in the activation domain of Pho4, Pho4Δ75-90, showed 85% reduced expression but the *acf* was shifted toward shorter correlation times with exponential parameters that were larger than for WT (figure 12**C**, red curve). This observation implied that some rate constants increased.

In support of this notion, the proofreading model for WT could be refit to the experimental *acf* of the activator mutant by increasing the rate of nucleosome removal, β alone. However, this contradicts observations by CPD which indicated that some rate constants must have decreased; CPD suggested a decrease in burst frequency indicated by a marked increase in the lifetime of inactive periods, 1261 $\pm$ 110s (figure 12**A**, red vs. blue bar plot). Although, CPD does recapitulate the observation of an increased rate parameter. The mutant maintained shortened active periods, 87 $\pm$15s, indicative of faster transitions from active to inactive (figure 12**B**, red vs. blue bar plot). This apparent contradiction is solved with the hypothesis that the activator regulates rates of both nucleosome removal and reformation. An equally good fit could be obtained by both decreasing the rate of nucleosome removal, α, and increasing the rate of nucleosome reformation, β. Our findings suggest that activators both increase the rate of nucleosome removal, as suggested earlier, and inhibit nucleosome reformation, possibly by recruitment of GTFs. Additionally, decreased active time and increased inactive time may be explained on the basis of changing activator binding kinetics. It is feasible, increased nucleosome occupancy over the promoter due to mutation in the activation domain, increased activator off-rate and a decreased activator on-rate (see below).

As suggested previously, the active time encompanses the sojourn time of all of the biochemical steps which coincide with the lifetime of the nascent transcript on the DNA [67]. In the regime of a two state model, as burst size becomes small, gene expression appears more and more Poissonian [68]. The active time becomes more

and more dominated by the dynamics of transcript polymerization, as observed at the Gal3 locus under UAS mutation [67]. Under a weakly expressing *PHO4* mutation it becomes exceedingly rare to observe more than one nascent transcript at *PHO5*, the observed variance in transcript intensity decreased drastically. It is therefore likely that the active time in the activator mutant is more dominated by the average transcript polymerization time for a *PHO5* transcript. The full length *PHO5* construct is 2.3kb implying a transcription-rate of ~26 bp/s. However, it is unclear when the transcript becomes visible; it is unknown how many PP7 stem loops must be transcribed and bound by PCP-GFP to detect the signal. Thus, the effective visible size of a nascent transcript varies from 1.4 - 2.3 kbs implying a transcription-rate 16 - 26 bp/s, on the low end of observations for transcription-rate for RNA polymerase II. However, even in the activator mutant the distribution of active times remains well fit by an exponential decay function. It seems likely that transcriptional kinetics must play a significant role in the sojourn of the active state in the activator mutant. As a consequence, it is likely our estimate of model rate parameters maintain higher error in the activator mutant. Additionally, because bursts of transcription were rare in the activator mutant we have fewer samples of active and inactive period lengths.

From live-cell microscopy we conclude that the Pho4 activation domain modulates both burst size and burst frequency. Mechanistically this can be understood on the basis that the activator mutant is less able to recruit chromatin remodelers and GTFs. The loss of recruitment of chromatin remodelers explains reduced burst frequency, it is harder to remove nucleosomes and therefore get into an active state; while loss of recruitment of remodelers and GTFs explains the

decreased burst duration, nucleosome reformation is inhibited, possibly, by the presence of GTFs and certain chromatin remodelers. We conclude that activators antagonistically regulate the forward and reverse reaction of the proofreading step. The activator regulating multiple distinct steps toward initiation is a key prediction of high specificity and kinetic proofreading. Notably, the dynamics of *PHO5* expression were similarly affected upon deletion of Chd1, suggesting that the activator recruits Chd1 to the promoter. Loss of Chd1 decreases the average number of nucleosomes over PHO5.

Pho4 activator modulates its own binding dynamics

Specificity requires that active periods, bursts, are nested within the promoter's sojourn in the activator-bound state; and because proofreading, on average, delays the start of transcriptional activity, active periods are expected to be significantly shorter than the lifetime of the activator-bound state (figure 5). Live cell fluorescent MFM data of nascent *PHO5* RNA suggest that activator binding dynamics occur at the timescale of “fast” decay in the *acf*. In addition, the reduced expression and changes in promoter dynamics upon mutation of the activation domain and chromatin remodelers, Chd1 and Isw2, are explained by a hypothesis of altered activator binding kinetics, decreased on-rate and increased off-rate. To test the relationship between activator binding dynamics and transcriptional dynamics, we measured Pho4 binding dynamics at UAS1, the weak Pho4 binding site of *PHO5*, by employing Chromatin Endogenous Cleavage Cross-Linking Kinetic (ChEC-CLK) analysis. ChEC-CLK, like CLK measures DNA binding dynamics of activators by taking advantage of the time dependence of formaldehyde cross linking of DNA

bound proteins (figure 13**B**) [69]. ChEC-CLK is a modified CLK where CheC is employed in place of Chromatin Immunoprecipitation (ChIP). ChEC employs the tethering of an activatable nuclease to a DNA binding protein [70]. After fixation by formaldehyde crosslinking the nuclease is activated and only DNA crosslinked by the fusion protein is cleaved. In this way, bound DNA molecules may be accounted for by northern analysis. Thus, activator occupancies are measured at specific genomic locations, en-vivo, at varying formaldehyde crosslinking times by CheC. Then binding dynamics are inferred by explicitly modeling both DNA binding and formaldehyde crosslinking (figure 13**A**). From ChEC-CLK activator on-rate, absolute initial occupancy and absolute maximum occupancy may be inferred.

Inference of both the maximal and initial activator occupancy depends on the number of total available binding sites. Nucleosomes occlude DNA binding proteins; therefore, we predict that for most activators the binding site availability is significantly less than 100% (figure 13**B**, rightmost column). Thus, the estimated relative maximum occupancy measured by CLK should be strictly less than the absolute activator occupancy; while the relative initial occupancy should be strictly greater than the absolute initial occupancy. We note here that relative occupancy measurements are sufficient for correct inference of on and off-rates [69]. However, to measure occupancy directly and to observe how changing nucleosome occupancy impacts binding a method like ChEC that accounts for total bound molecules is required.

To perform ChEC-CLK at the UAS1 binding site of the *PHO5* Promoter, we inserted a HA tagged micrococcal nuclease (MNase) into the 3' end of *PHO4* to create a Pho4-MNase fusion. The Pho4-MNase fusion activates *PHO5* expression equally well as WT [51]. Formaldehyde crosslinking does not significantly deplete the total pool of unbound Pho4, an important assumption of CLK. However, crosslinking slowly abrogates the ability of a subpopulation of Pho4-MNase to cleave DNA. Loss of activity is made evident at long crosslinking times, >20 minutes. Thus, we conclude occupancy measurements to be lower bounds.

Cells containing Pho4-MNase fusion were crosslinked for various time lengths and quenched with glycine to inhibit further crosslinking. To activate the nuclease cell extracts were incubated with Ca^{2+} ions. DNA was isolated and digested with restriction enzymes, Apa1 and Pst1, to create a full length ubound *PHO5* fragment, 2800 bps. Then, northern analysis was performed to determine cleavage frequencies. As seen previously, activated *PHO5* promoter is cleaved at two sites corresponding to the two Pho4 binding sites, UAS1 and UAS2. The probe used binds upstream of UAS1 which allows for specific quantification of UAS1 occupancy but not UAS2 occupancy; cleavage at UAS1 cleavage "erases" cleavage at UAS2.

From CLK we infer the average time between Pho4 binding events at UAS1 is ~30s and once bound the average lifetime of the bound state is ~200s (figure 13**C**, blue curve). The lifetime of the activator bound state is comparable with the lifetime of the active gene, 124s, as observed by live cell microscopy. In addition, the active fraction, 0.19, as measured by live-cell microscopy, is smaller than the

occupancy of Pho4 at UAS1, 0.24, suggesting that a bound UAS1 is required for entrance into the actively transcribing state; a requirement for activator specificity. This suggests activator dynamics play an important role in regulating at the timescale of the "fast" decay component observed in the *acf*. The Pho4 on-rate was fast enough to be near the edge of detection for CLK using our experimental setup, in this regime parameter inference is highly error prone. However, using inferred kinetics and an estimate for the nuclear volume for haploid yeast with the equilibrium dissociation constant we back calculated the expected concentration of Pho4 in the nucleus which closely matched published Pho4 molecule counts. In addition, the observed kinetics for Pho4 binding are comparable with what has been observed for other yeast specific activators measured by alternate techniques [9,67]. Notably, the activator on-rate is faster than we expect it to be when inferred from live cell microscopy of nascent transcripts. We must conclude that activator binding is not rate limiting for entrance into the actively transcribing state. However, it seems likely that the presence of the activator is a requirement for maintenance of the active state.

Under activating conditions most binding sites remain unavailable for Pho4 binding. The occupancy of Pho4 at *PHO5* UAS1 under activating conditions is 0.24 (figure 13**C,** blue curve meets the y-axis) while the maximal occupancy is 0.29 (figure 13**C,** blue dashed line). Since unbound Pho4 was not depleted by crosslinking this suggests that 70% of *PHO5* UAS1 DNA is not available for binding, even under activating conditions. In addition, of the available UAS1 binding sites, most, 81%, are occupied by Pho4. The small maximum occupancy of Pho4 at UAS1 is

unexpected given the observations of nucleosome occupancy at that location (see below). It is likely that not all crosslinking events lead to DNA cleavage and occupancy measurements are at best lower bounds. Notably, nucleosome loss is not observed in the time course. The maximal occupancy stabilizes well below 100% and remains constant for >15 minutes, suggesting that either nucleosome removal dynamics are too slow to be observed by CLK or crosslinking prevents nucleosome dynamics entirely as depicted in figure 13**B** (middle row).

Upon mutation of the activation domain of Pho4, nucleosome occupancy over important regulatory sequences increases (see below). We predict that the activator modulates its own activator binding dynamics mediated by recruitment of chromatin remodelers. Decreased on-rate and increased off-rate of the activator are predicted from mutations to the activation domain of Pho4 on the basis of the proofreading model fits to live-cell microscopy data under various mutations. We completed CheC-CLK of the *PHO5* UAS1 in a Pho4Δ85-99 context, to investigate the role of chromatin structure on maximal activator occupancy and activator dynamics (figure 13**C,** red curve). A reduction in the fraction of total available UAS1 binding sites, maximal occupancy, to 0.25 was observed (figure 13**C,** blue to red dashed line). From this we infer a decrease in the fraction of available UAS1 sites by 0.16. In addition, upon mutation, the average Pho4 occupancy at UAS1 is reduced from 0.24 to 0.16 (figure 13**C,** red curve meets the y-axis). Thus, both the average number of available Pho4 binding sites and the average number of bound Pho4 decreases. The average lifetime of the activator bound UAS1 was shorter than WT, 93s, and the lifetime of the unbound UAS1 was longer than WT, 55s. We conclude the Pho4

activation domain impacts Pho4 binding dynamics. Changing activator binding kinetics supports the reduction in active time and the increase in inactive time observed with live cell microscopy for comparable mutations in the activation domain of Pho4. Effects on the binding dynamics of the activator are observed by ChEC-CLK for activator mutations that severely reduce expression of *PHO5*. Low expression made analysis of these same mutations impossible by live-cell microscopy.

Promoter chromatin structure is dynamic, and these intrinsic dynamics generate discrete promoter nucleosome structures [55,71]. Each of these structures is likely differentially capable of binding activators; obvious in the case of the fully nucleosomal and nucleosome free promoter. Activators recruit chromatin remodeling activities by virtue of their activation domain. Thus, activation domain mutations inhibit chromatin remodeling and push the promoter toward structures that are more nucleosomal. These structures in turn hinder activator binding, by either obstruction or destabilization. It has been shown for *PHO5* that Pho4 activator strength correlates with Pho4 occupancy and is anticorrelated with nucleosome occupancy and dynamics [55,66]. Changing nucleosome occupancy leads to a change in absolute activator occupancy and potentially a change in activator binding kinetics. It remains unclear what features in promoter chromatin structure are important for determining this relationship. To test models by which *PHO5* promoter nucleosome structure and dynamics alter Pho4 binding dynamics we observed the nucleosome architectures of single *PHO5* gene molecules.

Shortened NDR correlates with decreased activator on-rate and increased activator off-rate

To explain the changes in activator binding kinetics at UAS1 due to mutation in the activator binding domain of Pho4, we hypothesize that the Pho4 activation domain prepares its own binding sites with enhanced binding kinetics by recruiting chromatin remodelers which leads to NDR formation, altering both the rate of binding and the stability of the Pho4-DNA complex. To explore the role of promoter chromatin structure and dynamics on activator kinetics we mapped nucleosomes across chromatin rings containing the *PHO5* locus isolated from living cells [55]. Briefly, site specific recombination sequences and an affinity tag are inserted flanking the PHO5 locus allowing for the formation, isolation and purification of chromatin rings formed in vivo (figure 14**A**) [54]. Nucleosome positions are "etched" onto the DNA by psoralen crosslinking and visualized by EM as depicted in figure 14**B** [72]; psoralen is a DNA intercalator that when exposed to UV irradiation preferentially cross links between nucleosomes. This technique allows for an estimate of both the absolute nucleosome occupancy along *PHO5* as well as the correlation in occupancy between multiple positions within individual *PHO5* gene molecules.

From single molecule data we estimated the nucleosome occupancy over the entire PHO5 locus (figure 14**C**) (see methods). To look at the correspondence between activator occupancy and nucleosome occupancy, we interrogated nucleosome occupancy over UAS1. Under activating conditions, the average nucleosome occupancy over UAS1 was 0.34; UASp1 was considered occupied if

mapped within a nucleosome-sized single-stranded DNA bubble (>90 bps) (figure 14**C,** red curve in between the dashed vertical lines). While EM analysis indicated that 66% of all UAS1 sequences were nucleosome-free, ChEC-CLK analysis suggested that only 30% were accessible to Pho4 binding. Our analysis only included nucleosomes which completely covered UAS1. It is likely that proximal nucleosomes additionally inhibit Pho4 binding in multiple ways. However, as discussed above, it is also likely that we do not observe every Pho4 binding event by ChEC-CLK. It is likely that this discrepancy occurs due to measurement error from CheC-CLK.

To identify which factors in promoter chromatin structure impact activator binding kinetics, we analyzed NDR formation on individual gene molecules. We characterized the NDR from each gene molecule by determining the length of available DNA between UAS1 and its most proximal nucleosomes upstream and downstream (see methods). As was observed previously, we found large variation in NDR size, ~10 - 800 bps, including molecules that have the same size NDR as repressed *PHO5 [55]*. The median length of the linker spanning UAS1 under activating conditions is ~350 bps, a 4-fold increase from repressing conditions (figure 15**A,B**). In line with this observation, the closest nucleosome to UAS1 on average, is ~100 bps away, a 5-fold increase from repressing conditions. Notably, nucleosome proximity is unequal on either side of UAS1. The 5' edge of the downstream nucleosome, under activating conditions, is generally shifted toward the TS (figure 15**B**, cartoon). This is likely due to cooperativity in nucleosome removal between sequence regulatory elements; *PHO5* possesses multiple regulatory sequences

downstream of UAS1 that each may bind transcription factors which recruit chromatin remodeling activities. Directionality in transcription is instilled at *PHO5* by the TATA box; loss of the TATA box abrogates transcription. However, loss of the TATA box does not lead to loss of NDR extension toward the TS. The presence of a second PHO4 binding site, UAS2, is likely the source of stretching of the NDR toward the TS. The location of the 5' edge of the downstream nucleosome seems continuous across the promoter, suggesting that this nucleosome is being slid away from UAS1 (figure 15**D**). However, when the average occupancy of identified downstream NDR nucleosomes was analyzed, evidence for distinct positioning emerges as suggested in [55]. Under activating conditions two distinct positions for downstream nucleosomes occur ~200 bp apart from one another (figure 15**E,** blue area). The positions are distinct from the single position found in repressing conditions. This suggests that while being slid the downstream nucleosome is "trapped" in new positions and/or that an entire nucleosome is occasionally removed. The 5' edge of the downstream nucleosome in activating conditions often coincides with the nucleosome position traditionally called "+1", defined as the first nucleosome whose dyad axis occurs after the TS. Notably, under repressing conditions, the nucleosome in the "+1" position occupies the TATA box (figure 16**A**, grey curve), however, when activated the "+1" position shifts sufficiently (figure 16**A**, blue curve) to uncover this important regulatory sequence, suggesting a potential mechanism for Isw2-mediated repression of transcription under activating conditions. Notably, strong positioning away from the TATA box appears to be lost in TATA delete cells

(figure 16**B**, pink curve). Suggesting that recruitment of TBP to the promoter is required for repositioning of the nucleosome occupying the "+1" position.

The contradiction between occupancy data, the observation of distinct nucleosome positions, and 5' edge data, perceived continuous positioning, may arise due to variance in identifying the 5' edge of each nucleosome which is less apparent in occupancy data. Error in edge detection is introduced by "leaking" of crosslinks into the edges of a nucleosome; we employed multiple long rounds of crosslinking to ensure complete crosslinking. While, the position of nucleosomes estimated by overall occupancy was better preserved. In addition, single stranded DNA has variable length. Thus, Inference of *PHO5* coordinates based on single stranded DNA must maintain this variance.

Upon mutation of *PHO4* with Δ85-99, Pho4 occupancy at UAS1 is reduced. From ChEC-CLK we determined the reduction in occupancy is the result of a change in both stability and binding rate of Pho4. This recapitulates a prediction generated by fits of the proofreading model to the *acf* from live-cell microscopy data of various mutants that kinetics are both sped up and slowed down; we observe a decrease in on-rate and an increase in off-rate as nucleosome occupancy increases. The change in activator off-rate may be explained by a change in proximity of adjacent nucleosomes to activating sequences. We therefore predict a change in nucleosome proximity to UAS1. The closest nucleosome to UASp1, upstream or downstream, was ~50 bps away on average, a 2-fold reduction from WT (figure 15**C,** cartoon). To explain the change in binding rate we analyzed overall NDR size. Under the

hypothesis that activators first bind DNA weakly then traverse in one dimension along the DNA to find their binding site [73], we expect that the more available DNA for binding the larger the observed activator binding rate. Thus, the NDR length and activator on-rate should increase together. This theory explains, in part, why nucleosomes must be removed from eukaryotic promoters. Long NDRs act as a scaffold to recruit DNA binding proteins. As NDR size increases so does available DNA to bind the activator, specifically or non-specifically. The average NDR size decreases ~1.5 fold. As described above, most of this change is due to repositioning of the downstream nucleosome (figure 15**E,** cartoon red line).

Importantly, large amounts of variation was observed for NDRs across conditions measured (figure 15**A-C**); small and large NDRs were observed in both activated and repressed conditions. Activator on and off-rates should vary greatly from cell to cell and over time; we expect that activator binding in a context of a large NDR to be significantly faster than the average. Measurements of activator binding that average across a population, like ChECLK, average across all of these chromatin contexts.

Discussion

Using live cell fluorescence microscopy of nascent PHO5 transcripts, we show that correlation in *PHO5* transcription under activating conditions is generated on at least two distinct timescales, one at ~100s and one at ~3 minutes. Correlation at the “fast” timescale is caused by brief periods of activity separated in time by longer periods of inactivity. Active periods are not fully defined by the elongation-rate of individual

transcripts, rather windows of activity encompass multiple transcription events, suggesting bursts of transcript initiation. This is made most evident upon Pho4 mutation, when the average number of transcripts in the active window is reduced to ~1 and the average active time is reduced to the elongation time of ~1 *PHO5* transcript; a similar phenomenon has been observed and verified by elongation-rate measurements for a GAL regulated gene [74]. The exponential decay component observed at the “slow” timescale is generated by correlation introduced between bursts of transcription.

Deletion of an individual chromatin remodeler Isw2 is sufficient to completely abolish “slow” timescale correlation. This suggests that chromatin dynamics over the *PHO5* promoter are responsible for generating correlation between bursts of transcription. We hypothesize that correlation is introduced through dynamic repression by Isw2. However, we must also conclude that Isw2 does more to regulate transcription. We observed a decrease in active times and an increase in inactive times of the TS by live-cell microscopy suggesting altered activator binding kinetics. This may be explained by the hypothesis that alterations of chromatin structure and dynamics imparted by an *isw2* deletion impact activator binding dynamics directly. Loss of other remodelers either changed expression but had little effect on long-timescale correlation (Chd1), had no effect on *PHO5* expression (Isw1) or abolished *PHO5* expression entirely (Swi/Snf).

Electron microscopic analyses of *PHO5* molecules supported the notion that the activator on-rate decreases and off-rate increases with increasing promoter nucleosome density (see below). This corroborates the notion that chromatin

dynamics are a likely source of correlation in *PHO5* transcription. To explain live-cell microscopy data we hypothesize that dynamic repression by Isw2 is responsible for generating correlation between bursts of transcription. However, the nature of the biochemical relationship between Isw2 and transcriptional activation remains elusive. *ISW2* is implicated as an antagonist to RISC and SWI-SNF, shrinking the NDR and shifting the +1 nucleosome toward the promoter [75]. The +1 nucleosome often occupies the TATA box, obstructing the transcriptional machinery. We observe a repositioning, away from the promoter, of the +1 nucleosomes when *PHO5* is activated. It is reasonable to expect that dynamic repositioning in the +1 nucleosome leads to stochastic availability of crucial DNA binding elements. Loss of the antagonist Isw2 could lead to trapping of the system in the active state, abolishing the transition between nucleosome configurations which are conducive for transcription to those that are inconducive to transcription. Thus, correlation in transcription is lost. This mechanistic understanding of Isw*2* assumes the trapping of promoter chromatin in an active state. Naively, we would predict increased expression. However, *PHO5* expression in the Isw2 delete decreases 2-fold. We must therefore resort to the auxiliary hypothesis that Isw2 plays multiple roles in regulating PHO5 expression; a likely hypothesis given our observation that alterations to NDR size has a significant impact on both activator stability and binding rate. This model implies that nucleosome heterogeneity on promoter elements produces heterogeneity in gene expression. This work suggests that Isw2 plays an essential role in proofreading activator binding by acting as a repressive force to the assembly of the pre-initiation complex. Thus, the activator is constantly required to

re-remodel chromatin, expending energy to continually form promoter structures conducive to expression. In potentially a serendipitous coincidence, energy expenditure is used in the erasure of promoter memory [76].

It is also plausible that the rate limiting step which causes long time scale correlation is actually acetylation and/or methylation of nucleosomes required for the appropriate remodeling by ISW2 [77]. In this case the energy for kinetic proofreading of activator/DNA recognition would come from the chromatin modifiers and the remodelers together. We take the opportunity to note here that energy expenditure of ISWI family chromatin remodelers has previously been implicated in kinetic proofreading [78–82]; in this case the energy is used toward increasing the specificity of recognition of Isw's target substrate, modified nucleosomes. It will be interesting to see if/how the energy of ATP hydrolysis is shared between enhancing ISW/nucleosome recognition and activator/DNA recognition. It may be a general rule for ATP hydrolyzing chromatin remodelers that energy expenditure is used toward achieving multiple goals.

We find that mutations to the activation domain of Pho4, the specific activator of PHO5, modulated both burst frequency and burst duration. However, these mutations appear to have minimal impact on transcriptional correlation in general. These two observations combined imply that two separate rate constants must be altered, one which is increased and one which is decreased. Thus, the changes in inactive and active time may be explained by the hypothesis that the activator recruits chromatin remodelers and the GTFs to the promoter. By increasing the rate of removal of nucleosomes the rate of entrance into the active state is increased.

Similarly, the presence of GTFs and remodelers may stabilize the active state by preventing nucleosome assembly. Alternatively, changes in rate constants may be due to changes in the activators ability to bind key UAS elements. We observe changing binding dynamics of the activator upon mutation. It is likely a combination of these two regulatory forces that lead to the transcriptional dynamics we observe at PHO5.

The proofreading model requires close coupling between activator binding and transcription, burst duration must not exceed the lifetime of the activator. Otherwise, activator specificity is lost. Using ChEC-CLK we show that Pho4's binding dynamics correspond to the timescale faster than our "fast" correlations, <30s. However, the on fraction of the activator recapitulates the on fraction of the active state. We found that the average lifetime of the Pho4 activator at UASp1 was ~200s, ~75s longer than the average active period inferred from microscopy. Thus, as predicted the active time is shorter than the dwell time of the activator.

The discrepancy between burst duration and sojourn of the activator on the promoter may be explained on the assumption that one or more additional slow steps are required after activator binding but before initiation of the first transcript. Thus, these additional steps delay a burst of transcription by ~75s. We speculate that, in our proofreading model, the transition from an activator-less state to an activator-bound state must encompass the sojourn of these additional steps. Our ChEC-CLK analysis indicated a faster activator on-rate than predicted by fitting the proofreading model to microscopy data. The difference between the inference of time

between activator binding events from microscopy and the observed time between activator binding events determined by CheC-CLK agreed nearly perfectly with the observed difference between active time and activator sojourn time, ~75s. Thus, we conclude that the inferred transition kinetics between the activator unbound promoter and the activator bound but actively transcribing promoter must include additional steps with sojourn ~75s. We suspect that this quirk in the model fit is due to the requirement in the proofreading model of state 3 to be an actively transcribing state; the model assumes, likely incorrectly, that transition from an activator-less promoter to an activator bound promoter, in the absence of a nucleosome, is sufficient to be actively transcribing.

Some of the time between activator binding and the initiation of transcription is likely consumed by assembly of the G-lobe of the preinitiation complex, composed of promoter DNA with general transcription factors TBP, TFIIA, TFIIE and TFIIH. However, once the G-lobe is formed, activator-mediated recruitment of TFIIB may lead to rapid RNA polymerase binding and transcription; the persistence of the G-lobe following promoter clearance by RNA polymerase allows for rapid reinitiation of transcription. The G-lobe complex is sedimented as a single entity in a glycerol gradient, attesting to its stability. While the dissociation of TFIIB upon promoter clearance accounts for the persistent need of the activator for transcript initiation. Thus, subsequent to G-lobe formation, activator binding events lead to recruitment of TFIIB and bursts of transcript initiation. Otherwise, once the G-lobe was assembled activator presence would become redundant and specificity would be lost.

Similarly, after dissociation of the activator, the complex of G-lobe proteins with promoter DNA, must disintegrate quickly, or else fast rebinding of the activator would extend the burst. Burst duration would exceed the sojourn of the activator on its specific binding sequence and specificity would be lost; we however found the opposite. Therefore, it seems likely that the majority of activator binding events lead to the reformation of the G-lobe complex. It is also feasible that the delay between binding and formation of the G-lobe complex may increase regulatory specificity by filtering out short activator binding events that do not successfully lead to formation of the G-lobe. In addition, active disassembly of the G-lobe complex by Mot1, which uses the free energy of ATP hydrolysis to dissociate TBP from DNA, may explain short burst duration. Alternatively, the activities of chromatin remodelers like the RSC complex, as suggested for Ace1 at the *CUP1* locus [83], or ISW2 as we have suggested may be responsible for disassembly of the G-lobe. Notably, the proposed mechanism of RSC toggling of the -1 nucleosome in both directions to rip activators from DNA would inhibit activator specificity [83].

From both CheCCLK and live cell microscopy it seems clear that activator dynamics play a significant role in generating correlation in transcription at the fast timescale. This supports previous work suggesting Gal4 binding dynamics generate GAL transcriptional dynamics [74]. However, we do not observe that the dwell time of the activator coincides perfectly with the sojourn of the active state. Rather, we propose assembly of the G-lobe to be rate limiting. Notably, “slow” time scale

correlations are also observed for GAL and other genes [67,84]. Our work suggests that these correlations are likely produced by slower promoter chromatin dynamics.

We observe by ChEC-CLK that mutations in the activation domain of Pho4 impact both on-rate and off-rate of the activator. We show that these differences correspond to structural changes in promoter chromatin structure by analyzing single gene molecules by psoralen crosslinking and EM. We predict that nucleosome proximity to UAS binding sites decreases activator stability. Upon mutation of the activator, we observe the proximity of the most adjacent nucleosomes to UAS1 shrinks two-fold, from ~100 bps to ~50 bps on average. It is unclear if this decrease in stability is due to the presence of a single nucleosome or interactions between adjacent nucleosomes. We hypothesize that disruption in Pho4 stability is likely achieved by interactions between nucleosomes both upstream and downstream from the UAS1.

The on-rate of Pho4 at UAS1 is seemingly also impacted by NDR length. It has been conjectured that activators find their binding sites through facilitated diffusion [6,10]. Thus, Pho4 likely transitions between 3-dimensional searching and 1-dimensional searching along linker DNA between nucleosomes. We predict nucleosomes inhibit 1-dimensional searching [73]; long stretches of naked DNA are rare on nucleosomal DNA. Upon activation the stretch of naked DNA containing target sequences increases, for the *PHO5* NDR this is roughly ~5 fold. Thus, the size of the available DNA to initiate 1 dimensional searching increases upon activation. This increased size enhances the apparent target site on-rate. A similar change in activator on-rate was observed for Ace1 at the *CUP1* locus in response to deletion of

the RSC complex [83]. However, it appears that disruption of NDR positioning over the UAS is only observed under inactivating conditions. We predict that the average length of the *CUP1* NDR shrinks and nucleosome proximity to the *CUP1* UAS increases even under activating conditions.

Creating a large “landing pad” over the promoter increases the on-rate of all DNA binding transcription factors; however, it seems not for all activators. The on-rate of the activator Rap1, which is classified as a “pioneering” transcription factor for its ability to bind DNA within a nucleosome, is insensitive to removal of a nucleosome within its binding site [85]. We speculate that pioneering factors maintain an alternative search strategy which allows them to efficiently find and bind their target sequences in the presence of nucleosomes.

It has been suggested recently that promoter sequences generally act as “antennas” which actively recruit transcription factors [86]. General recruitment enhances factor specificity for target sequences inside promoters. It makes functional sense that “antenna” sequences coincide with NDR formation. Promoter DNA that is made available to enhance factor on-rates also contains sequences which promote factor recruitment. This model also unifies the observation of general cooperativity between activators and transcription factors, without the need for them to interact directly. A similar model has been suggested for equilibrium binding of many transcription factors internal to a single nucleosome which bind and cooperatively destabilize the nucleosome [87]. Our model differs from this type of cooperativity, slightly, in that it relies on the system being away from equilibrium; the recruitment of ATP dependent chromatin remodelers and the removal of

nucleosomes by an individual activator can increase the on-rates and reduce the off-rates of all activators at all binding sites within the NDR including itself, without the need for direct interaction. Importantly, it allows one binding site to maintain 'cooperativity' with itself, kinetic proofreading. Notably, we observe that the NDR lengthens toward the TSS, uncovering key regulatory sequences. Such a mechanism taken together with the proofreading model suggests the major hurdle the cell overcomes to specifically regulate gene activation.

These observations highlight a major challenge, and potentially suggests some solutions, faced in fabricating transcriptional regulatory networks; in-vitro in equilibrium, it is challenging to generate simple regulatory networks with many specific regulators without inhibitively high levels of crosstalk between regulators and their targets [88]. Ofcourse, one solution is increasing the specificity of the regulator, however, this solution does not scale. As discussed in chapter 1, as specificity is increased controllability is lost. Thus, only in setting up a system away from equilibrium, be it with kinetic proofreading or allosteric effector molecules or some other non-equilibrium mechanism, may high specificity and controllability be realized.

It is clear from single molecule analysis of PHO5 gene rings that there is extensive cell to cell variation in promoter chromatin architecture. Under activating conditions we see NDR lengths ranging from 10 - 800 bps. We have also shown a correspondence between average NDR size and activator on-rate, and nucleosome proximity and off-rate. Thus, we predict that each cell experiences distinct activator binding kinetics due to its distinct promoter chromatin architecture. With application of an ergodic hypothesis of chromatin remodeling [56], we expect that individual cells

as they remodel chromatin, transition stochastically between different activator kinetics in time. Therefore, the activator binding we observe by ChEC-CLK is representative of the average Pho4 binding kinetics at UAS1. We predict varied binding dynamics for Pho4 at UAS1. The on-rate should very from almost no binding when a nucleosome occludes activating sequences to on-rates faster than we observed by CheC-CLK, in cells with a full sized NDR. We predict, ~34% of cells should have almost no binding under activating conditions; cells with UAS1 occupied by a nucleosome. While, the off-rates should vary from short sojurns when UAS sites are adjacent or within a nucleosome, observed in-vitro for Gal4 binding [74], to off-rate comparable to in-vitro expectations on naked DNA [7]. These observations, in part, may explain the unexpectedly fast activator dynamics observed by single molecule tracking in vivo [74]. Thus, reconciling the need for specificity and the observation of short sojourn times of specific activators. Similarly, heterogeneity in activator kinetics caused by chromatin dynamics may serve as the source of the "slow" decay component in the *acf* observed by live cell microscopy and may be yet another mechanism by which kinetic proofreading of activator binding is achieved.

By gene ring analysis we observed heterogeneity in nucleosome position over other important regulatory elements; In activating conditions a nucleosome is sometimes observed occupying the TATA box and in other cells it is observed to be adjacent to the TATA box (figure 16**A**, blue curve). Notably, positioning is lost upon mutation of the TATA box (figure 16**B**, pink curve). This heterogeneity is an expectation of the proofreading model by Isw2 and is suggestive that indeed the

theory that positioning of the “+1” nucleosome by ISW2 generates long timescale correlations in PHO5 transcription.

Recent studies rely on the two state model to infer bursting dynamics genome wide [89]. Following the two state model, It is common to define burst frequency as the rate at which active periods occur, burst duration as the average length of active periods, as we have done here. However, If a two state model is applied when modeling transcription of a system with more than two states, observed change of bursting parameters does not imply, generally, modulation of the same biochemical steps. For example, in our own proofreading model, modulation of either activator on-rate or nucleosome off-rate impacts burst frequency. In general, it can be shown that modification of any rate upstream of the rate limiting step of burst formation may appear as burst frequency modulation (data not shown). In practice, the biochemical topologies regulating transcription are much larger than our simple models. Therefore, it is often unknown what set of chemistries are encapsulated within burst frequency and burst duration when modeling with a two state promoter. Here we show experimentally that a priori assumption of the the two state model when inferring bursting parameters from population data, e.g. smFISH, may yield “incorrect” bursting parameters and lead to drawing false conclusions. The two state model of transcription fits our *PHO5* smFISH perfectly. However, when live-cell microscopy is completed, it becomes clear that bursting occurs at multiple timescales and that the rate constants inferred by smFISh cannot explain. This contradiction between population data and single molecule data highlights the necessity for single molecule experiments when interrogating mechanisms. More theoretical work is

needed to determine how mechanistic topologies map to bursting dynamics and, maybe, more importantly, why the cell selects specific topologies over others to regulate different loci.

To regulate gene expression the cell controls a number of specific parameters. Average expression, noise, gene regulatory function, inducibility and controllability have each been shown to be dictated by the biochemical network regulating expression [62,90–92]. The topologies of the biochemical networks regulating gene expression have evolved to meet these requirements. Articulating the specifications the cell required for proper regulation is pivotal in understanding the complexity of regulation; form follows function. Of course, biological networks evolve under specific constraints and in certain contexts. However, the structure of a biochemical network limits its function. By identifying problems the cell faces we may better understand the biochemistries which solve those problems. In this work we show that a requirement for high specificity is a driving force in producing heterogeneity in Pho5 expression. Nucleosome dynamics held away from equilibrium may be used to proofread binding of the specific activator Pho4. These dynamics, because of the constraints of specificity, produce bursts of Pho5 transcription at the minute timescales. In addition, we identify a reason eukaryotes might adapt toward using nucleosome occlusion of transcription factors from DNA to regulate gene expression. Not only to repress transcription generally, but to enable specific activation.

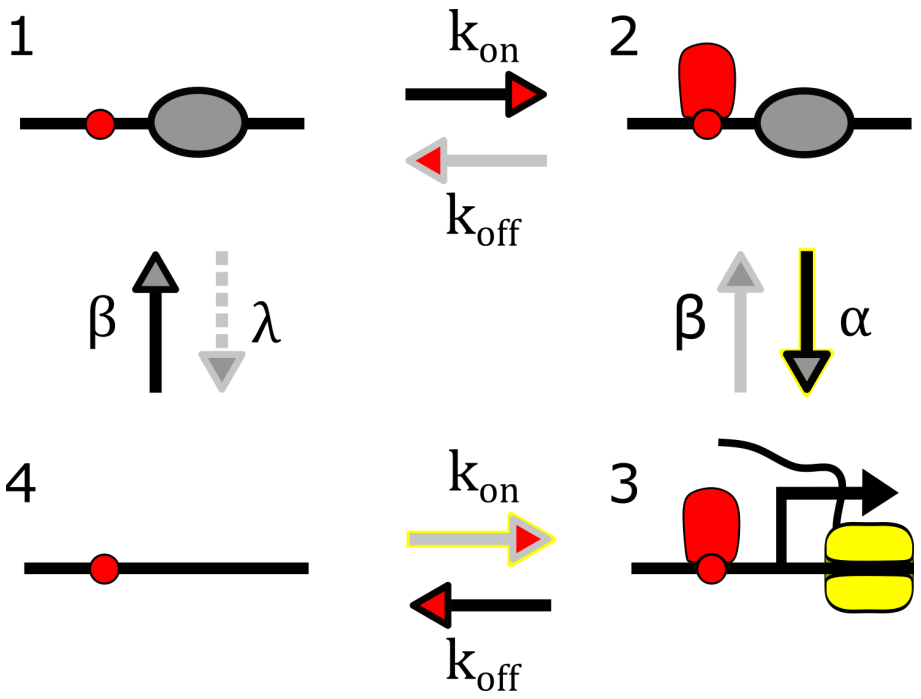

Figure 5**. The proofreading model:** The 4-State model of kinetic proofreading of activator binding mediated by activator coupled nucleosome removal. An activator (red) binds to a target sequence (red circle) to remove a nucleosome (grey oval). Present are four types of transitions: nucleosome removal/assembly transitions (grey arrowhead) as well as activator binding/unbinding transitions (red arrowhead). The dashed arrow highlights nucleosome removal in the absence of the activator which exemplifies the loss of nucleosome disassembly activities recruited by the activator, in the extreme λ = 0. Black arrows (clockwise) indicate transitions which are statistically preferred due to kinetic proofreading while grey arrow lines (counter clockwise) indicate the statistically less likely cycle. The only transcriptionally active state is the state containing an activator but no nucleosome (State 3), RNA polymerase II is indicated in yellow. Yellow highlighting marks transitions (arrows) which lead into the active state.

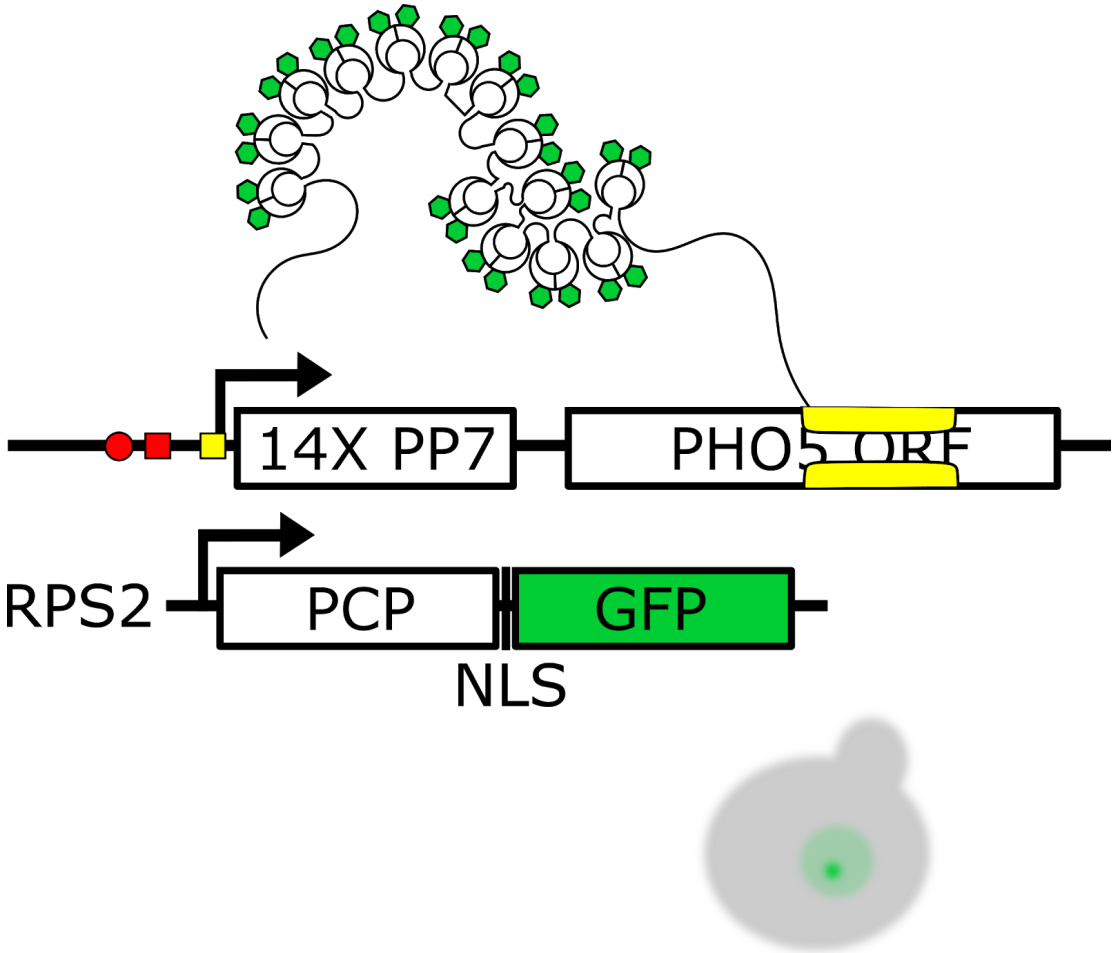

Figure 6**. PP7 labeling technique: (top)** To visualize individual nascent RNA 14X PP7 binding motifs were inserted in the 5' UTR of PHO5. Shown are key regulatory sequences: UAS1 (red circle), UAS2 (red square), TATA box (yellow square) and the transcription start site (bent arrow). **(bottom)** The PCP-GFP fusion expression was driven by the RPS2 promoter and targeted to the nucleus by a nuclear localization signal (NLS). A cartoon of a yeast cell with a nascent transcript tagged with localized GFP is depicted within the nucleus haze, caused by localization of PCP-GFP to the nucleous.

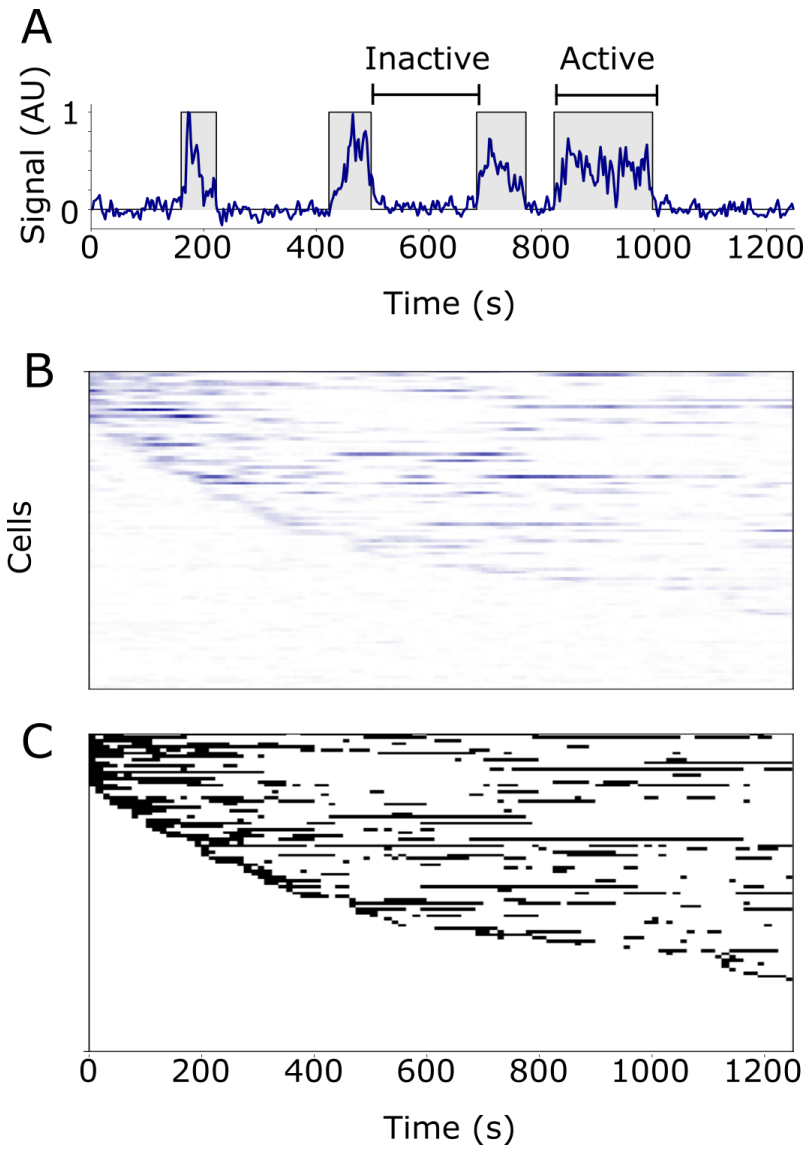

Figure 7**. Live-cell microscopy of nascent *PHO5* RNA:** (**A**) Example time series of transcription at the WT *PHO5* locus (blue line) under activating conditions. Fluorescent intensity is normalized to the maximum intensity observed in this series. Grey boxes indicate predicted active periods using change point detection, while inactive times are represented as gaps between grey boxes. (**B**) Collection of 100 randomly selected time series of WT *PHO5* transcription. Each row represents one cell in time. Rows are sorted by occurrence of the first transcript detected in each cell. TS intensity is normalized by maximum intensity measured across all cells presented. (**C**) Same cells and cell order as described in **B**, however, intensity values are replaced with active periods (black) and inactive periods (white) inferred from change point analysis.

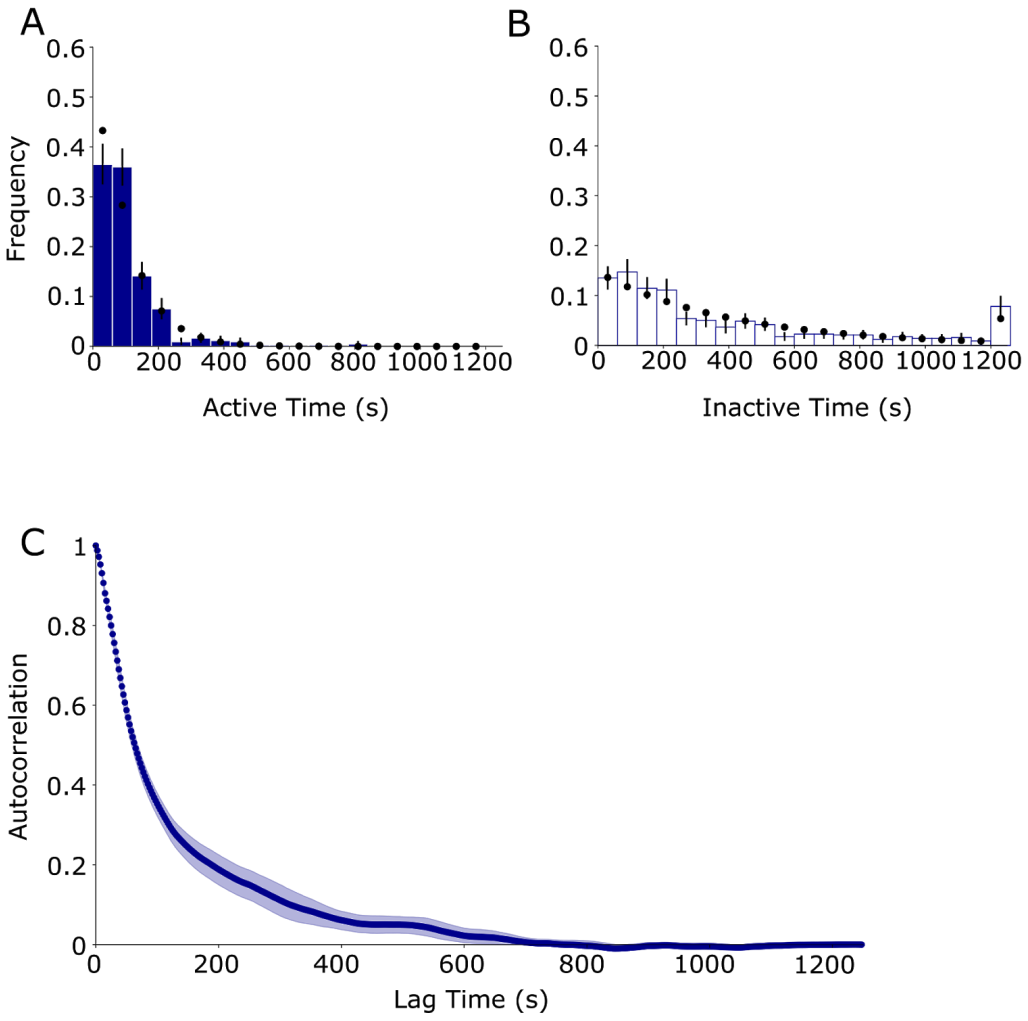

Figure 8**. WT active cells exhibit short active times but maintain a long lag in the autocorrelation function:** (**A**) Histogram of active times determined by change point analysis normalized such that the sum of all bins is 1. Error bars (black) represent the 90% confidence interval of 10000 bootstrapped samples from the data. The expected value for each bin (black dots) was calculated from a fit of all active times across all cells to a single exponential decay. (**B**) Same as **A**, except histogram is of inactive times. (**C**) The autocorrelation function of transcriptional start site intensity computed at each time point (blue dots). The shaded area represents the 90% confidence interval of 10000 bootstrapped calculations of the autocorrelation function.

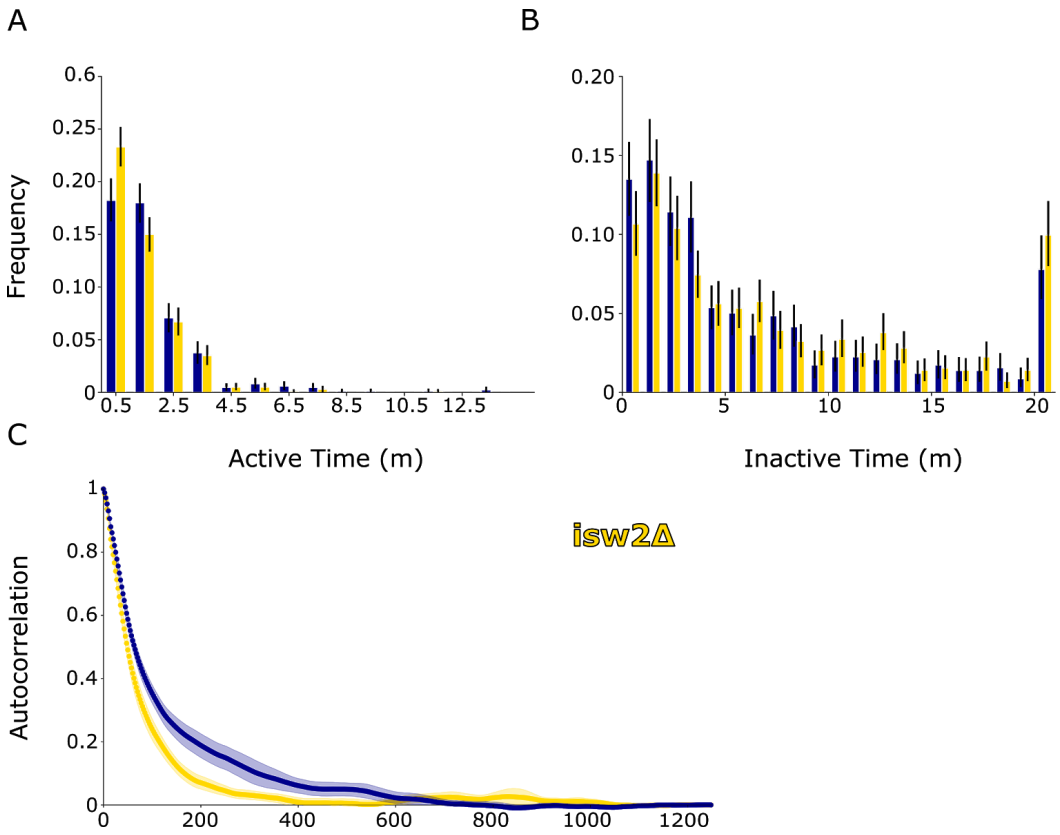

Figure 9 **Deletion of Isw2Δ is sufficient to abrogate long timescale correlation in *PHO5* transcription:** (**A**) Histogram of active times from WT (blue) and isw2Δ (gold) cells, as described in figure 8. (**B**) Histogram of inactive times from WT (blue) and isw2Δ (gold) cells as described in figure 8**E**. (**C**) The autocorrelation functions as described in figure 8**C**, but of WTs (blue) and isw2Δ (gold) cells.

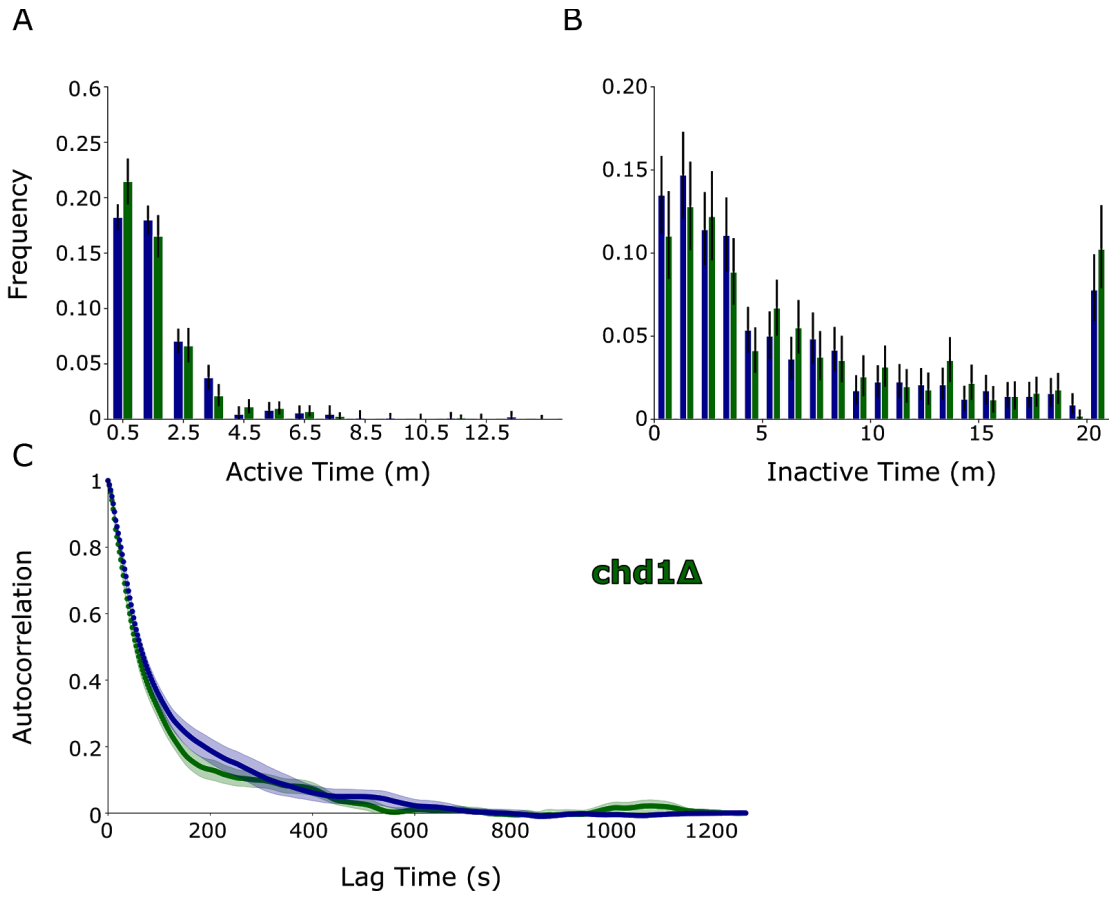

Figure 10**. Deletion of Chd1Δ alters both burst duration and frequency of *PHO5* transcription:** (**A**) Histogram of active times from WT (blue) and chd1Δ (green) cells, as described in figure 8**A** (**B**) Histogram of inactive times as described in figure 8B. (**C**) The autocorrelation functions as described in figure 8**C**.

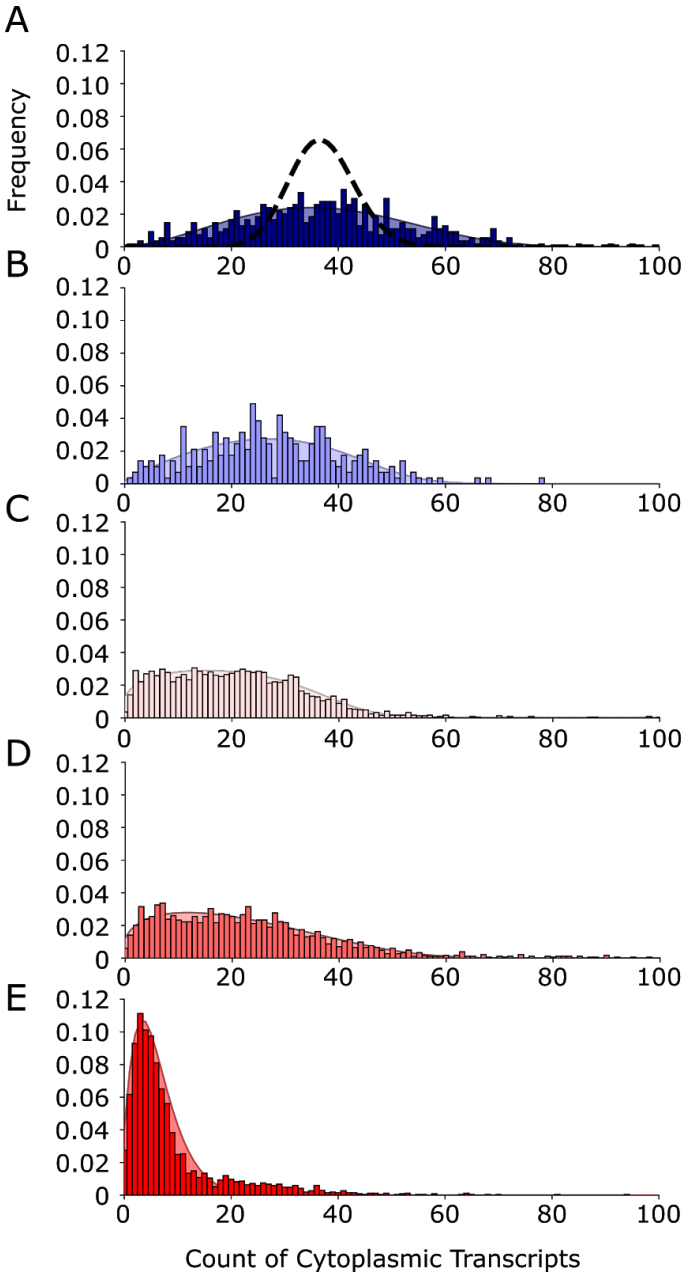

Figure 11**. smFISH of PHO5:** (**A**) Histogram of cytoplasmic RNA counts per cell determined by smFISH of PHO4 WT cells (blue bars). Also plotted are the best fit to a poisson process (dashed line) with mean of 37 transcripts and the fit to the random telegraph model (blue filled area). (**B**) Same as **A**, but for Pho4[Δ75-78] cells (light blue). (**C**) Same as **A**, but for Pho4[Δ79-90] cells (pink). (**D**) Same as **A**, but for Pho4[Δ91-99] cells (light red). (**E**) Same as **A**, but for Pho4[Δ79-92] cells (red).

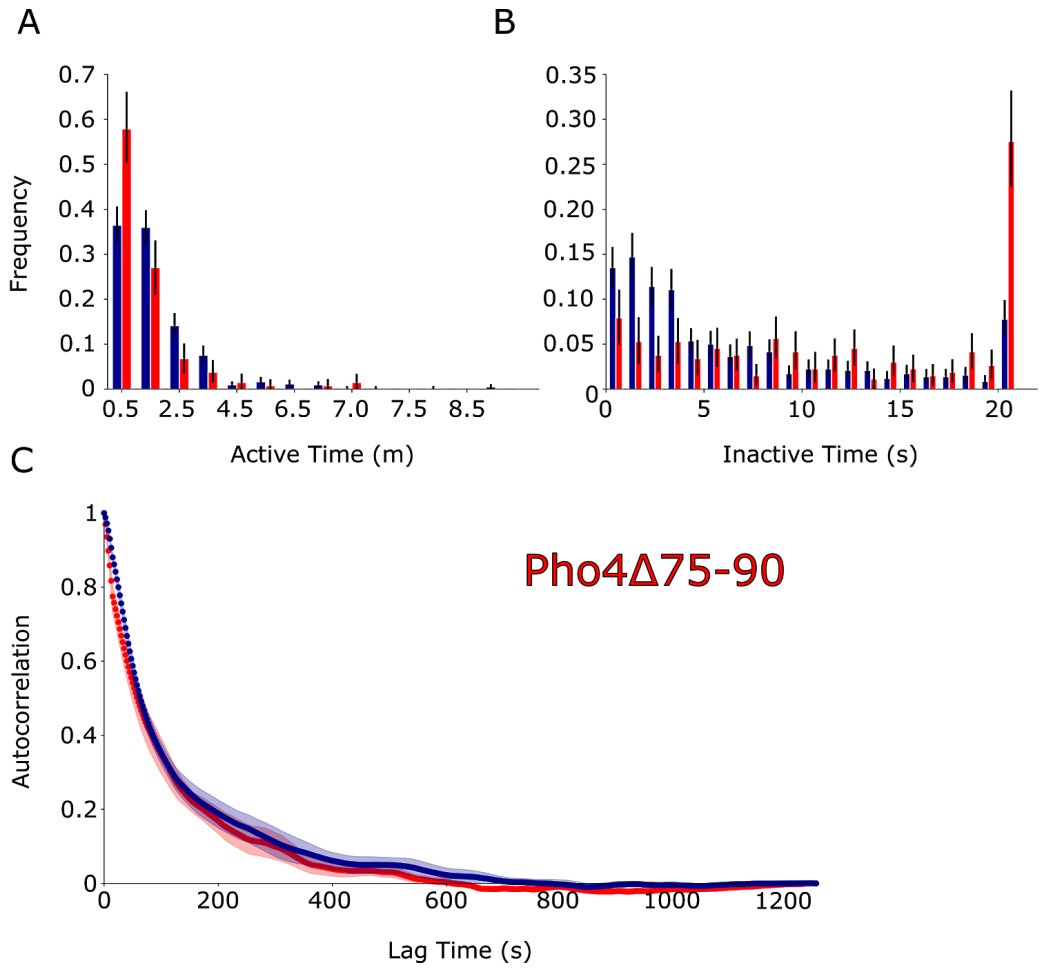

Figure 12**. Pho4 regulates both burst duration and frequency of *PHO5* transcription:** (**A**) Histogram of *PHO5* active times as described in figure 8**A**, from PHO4 WT (blue) and Δ75-90 (red) cells (**B**) Histogram of inactive times as described in figure 8**B**. (**C**) The autocorrelation functions for *PHO5* transcription as described in figure 8**C**.

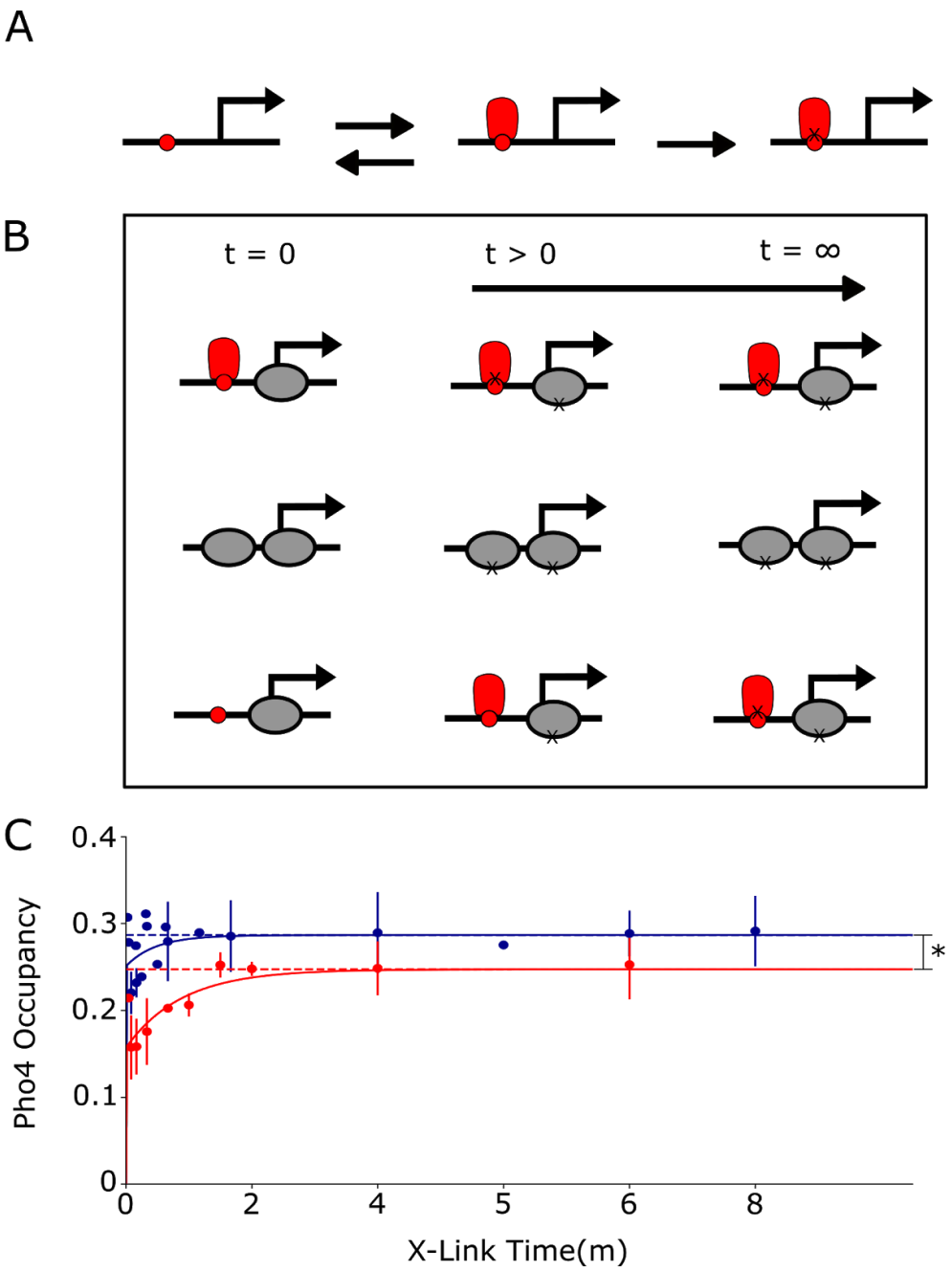
A
B
t = 0
t > 0
t = ∞
C
Pho4 Occupancy
0.4
0.3
0.2
0.1
0
0
2
4
5
6
8
X-Link Time(m)
*

Figure 13**. ChEC-CLK of Pho4 binding dynamics at UAS1 of *PHO5*:** (**A**) To compute activator dynamics a transition in steady state upon addition of formaldehyde is modeled by the CLK Model [69]. An activator (red) stochastically binds to a target sequence (red circle). The formaldehyde crosslinking (marked by an X) of the activator to DNA is irreversible indicated by an unidirectional transition. (**B**) Diagram depicting theoretical samples collected at various time points by ChEC. Prior to addition of formaldehyde (t = 0), activator and nucleosome binding are in steady state; some average number of target sequences are bound by the activator. Upon addition of formaldehyde (t > 0), the activator bound fraction is crosslinked. Nucleosomes additionally become locked in place by crosslinking. As crosslinking time continues more and more activators bind the available target sequence (target sites without nucleosomes) and become crosslinked. Thus, as crosslinking time approaches infinity (t = ∞), every available target site becomes saturated with crosslinked activators. (**C**) ChEC data of Pho4-MNase binding at UAS1 of the *PHO5* promoter for various formaldehyde crosslinking times (points) with standard deviations (error bars) for both Pho4-MNase (blue) and Pho4[Δ85-99]-MNase (red) cells. The inferred maximum activator occupancy (dashed line) as well as best fit by the CLK model (solid curve) is also indicated for each condition. The difference in maximal activator occupancy (*) between WT and mutant cells indicates the difference in the number of total available UAS1 binding sites between conditions.

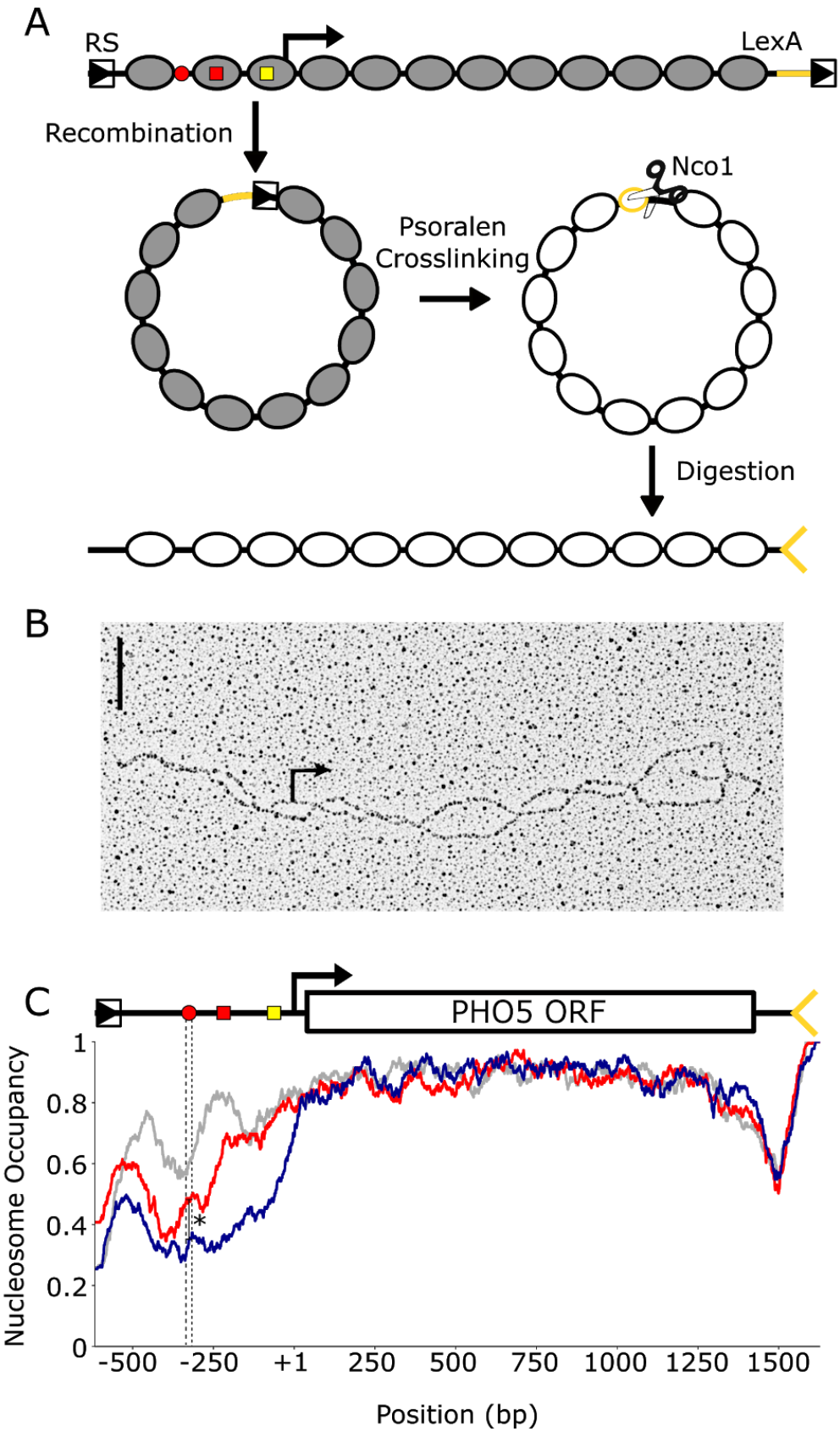
A
RS
LexA
Recombination
Nco1
Psoralen
Crosslinking
Digestion
B
C
PHO5 ORF
Nucleosome Occupancy
1
0.8
0.6
0.4
0.2
0
-500
-250
+1
250
500
750
1000
1250
1500
Position (bp)

Figure 14**. Gene ring analysis and average nucleosome occupancy across the PHO5 locus:** (**A**) Diagram depicting psoralen crosslinking and gene ring purification for EM analysis of individual *PHO5* gene molecules: i) nucleosomal *PHO5* prior to recombination (top), ii) nucleosomal gene ring post recombination (middle left), iii) psoralen cross linked and purified gene ring (middle right), psoralen preferentially crosslinks linker DNA such that nucleosome positions become “etched” onto the DNA as single stranded bubbles when the nucleosomes are removed and the DNA is denatured, iv) Nco1 digested psoralen cross linked gene molecule (bottom), LexA binding motifs used in the purification process generates a single stranded “fork” at the 3’ end of the *PHO5* locus. Using this distinguishing feature, individual molecules may be oriented to one another. (**B**) EM image of chromatinized gene ring from *PHO* active WT cells from [55]. The position of the TS is indicated. In addition, the fork is visible, left. Scale bar indicates 100 nm. (**C**) Plot of average nucleosome occupancy per base inferred from r-value measurements from EM images of psoralen cross linked PHO5 gene molecules in a repressive (grey), active (blue line), Pho4[Δ85-99] mutant (red line) context. A diagram of the *PHO5* locus (top) is depicted to scale. Indicated is the position of UAS1 (dashed lines) highlighting the difference in average nucleosome occupancy between PHO4 WT and PHO4[Δ85-99] cells at UAS1 (*).

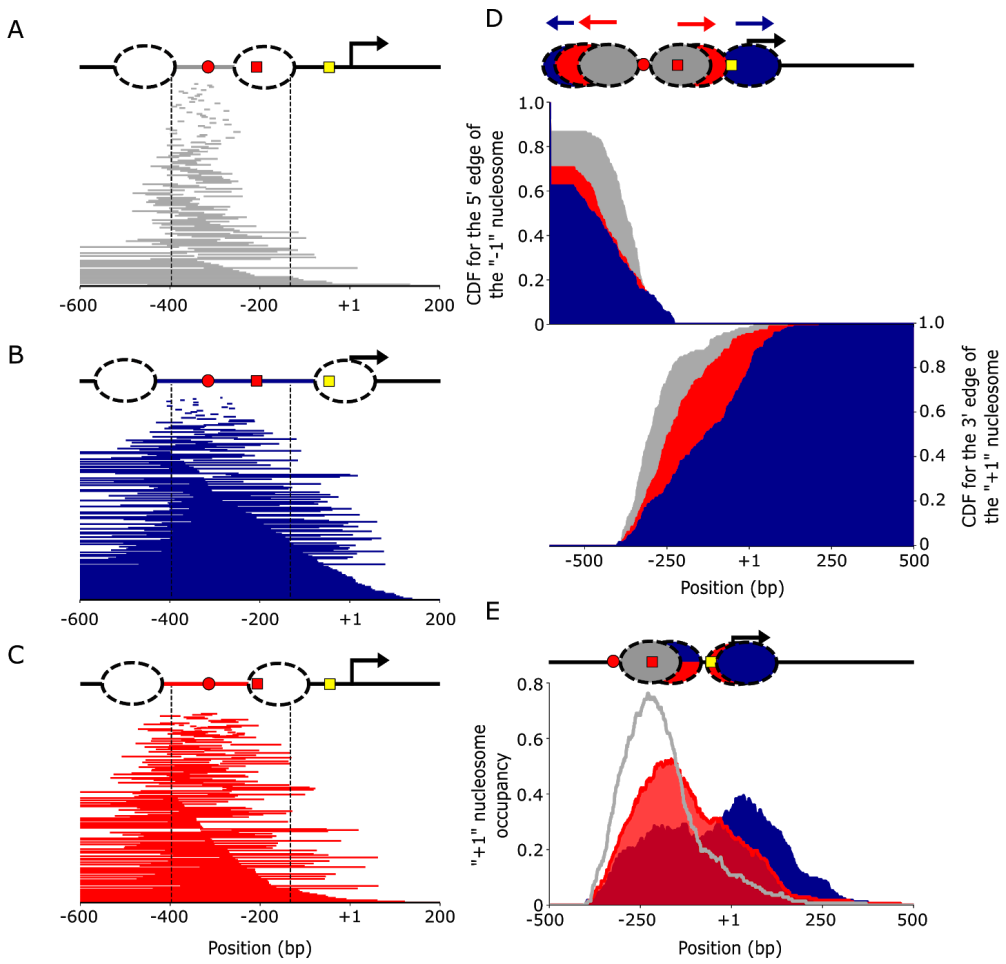

Figure 15**. Single molecule analysis of individual NDRs in the PHO5 promoter:** (**A**) The "NDR" within each *PHO5* gene molecule, under repressing conditions (grey), is identified and plotted. Individual cells (rows) are sorted by the length of the "NDR". The "NDR" is identified as the longest contiguous linker region which overlaps activating sequences (vertical dashed lines). A diagram of the PHO5 locus is depicted to scale (in the x dimension), (top) showing the average "NDR" length (grey line) and the average position of the most proximal nucleosomes (dashed oval), the "-1" and "+1" nucleosomes, on either side of the "NDR". (**B**) Same as for **A**, but under activating conditions (blue). (**C**) Same as for **B**, but in the Pho4[Δ85-99] cells (red). (**D**) Plotted is the inverse cumulative density function of the end position of the "-1" nucleosome (top) and the cumulative density function of the start position of the "+1" nucleosome (bottom) for the three conditions (top). A diagram of the *PHO5* locus with average "-1" and "+1" nucleosome positions is depicted (top). (**E**) Plotted is the average occupancy of identified "+1" nucleosomes; the "+1" nucleosome is identified in the same way as **A**. A diagram of the *PHO5* locus depicting two subpopulations for positions of the "+1" nucleosome (top).

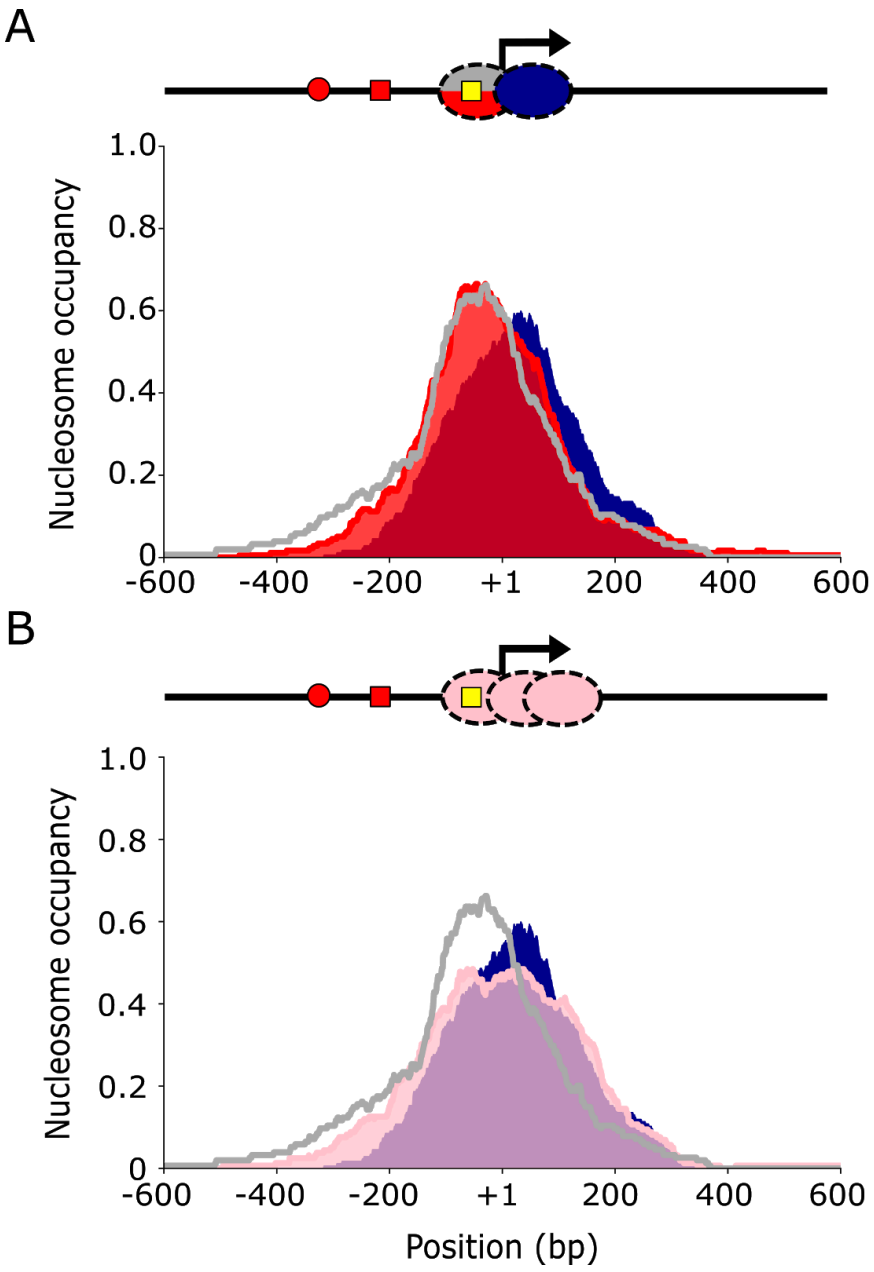

Figure 16. **Nucleosomes shift downstream from the TATA box upon activation.** (**A**) Nucleosome occupancy inferred after selecting only the first nucleosome which overlaps the TS of each molecule; isolating nucleosomes which occupy the 3' position described in figure 15**E.** A diagram of the *PHO5* locus (top) depicting the shift in average nucleosome position, which uncovers the TATA box. (**B**) Same as **A**, however, in cells with a TATAΔ delete (pink). Positioning of the nucleosome is disrupted. Each peak is depicted as a representative cartoon nucleosome (top).

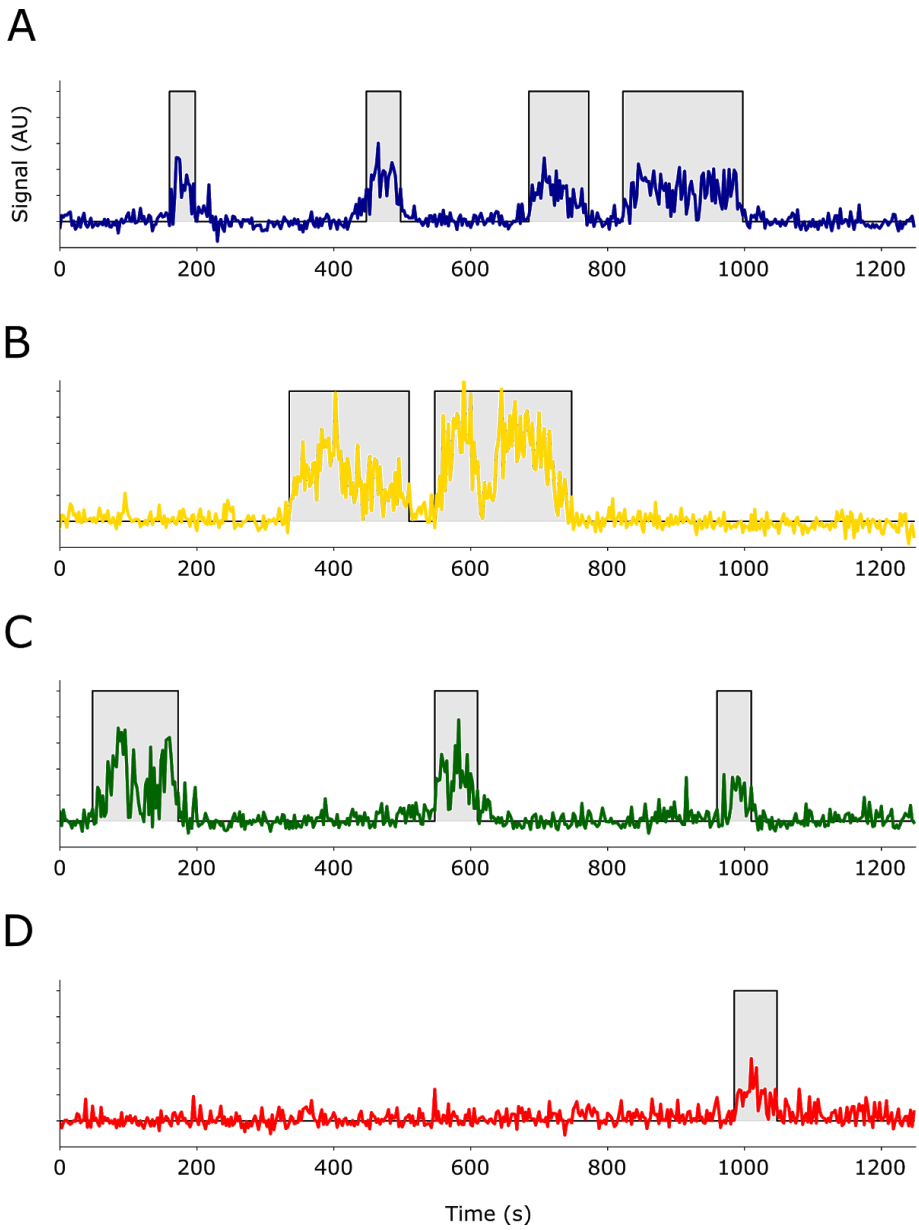

Figure 17**. Single sample paths of the PHO5 TS under varying mutations:** (**A-D**) *PHO5* transcription site intensities for a single cell, same as figure 7**A**. Except with WT (blue), isw2Δ (gold), chd1Δ (green) and Δ75-90 cells (red).

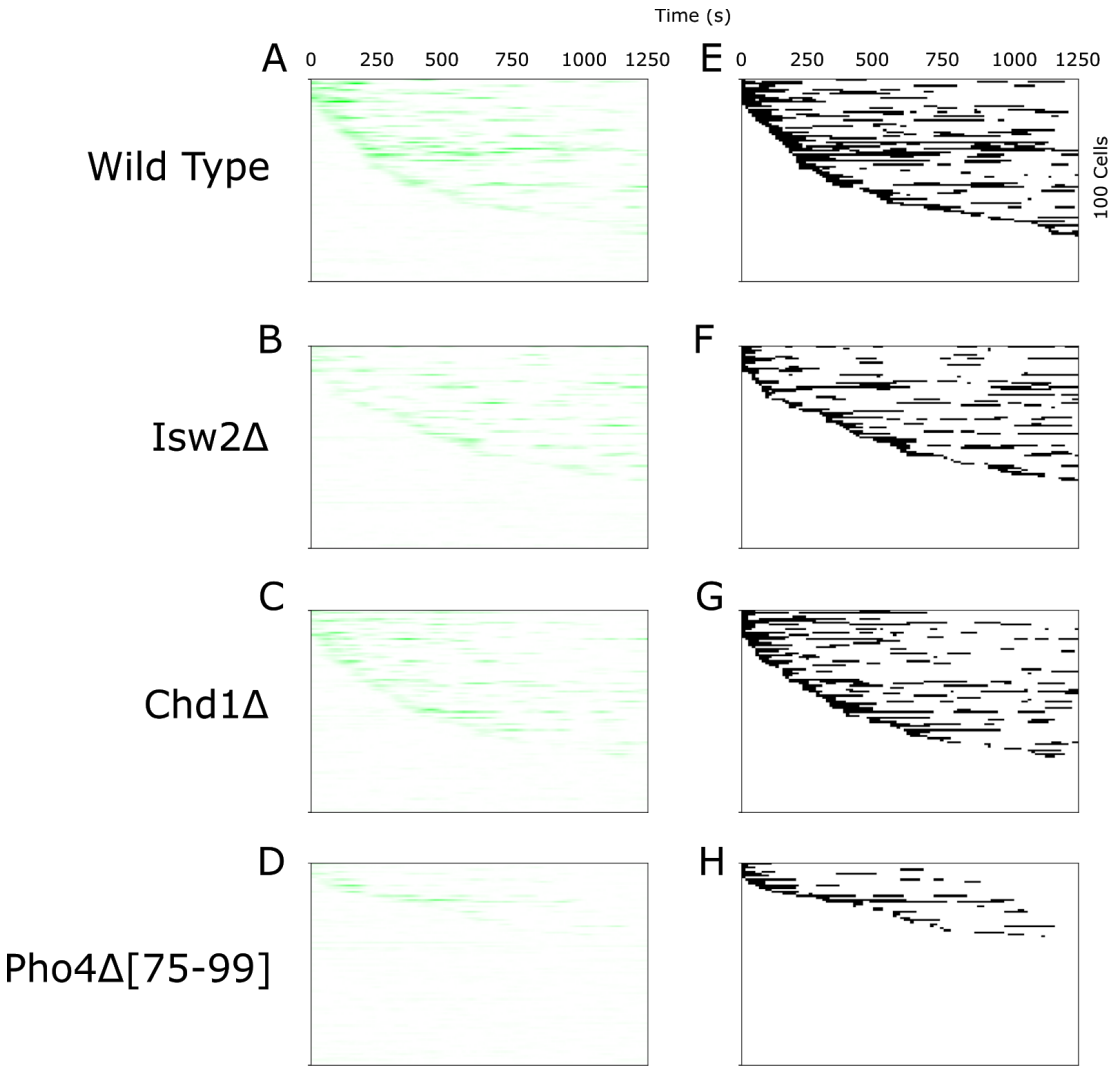

Figure 18**. 100 randomly sampled cell trajectory of PHO5 transcription under varying mutations:** (**A-D**) *PHO5* transcription start site intensities for 100 randomly selected cells, same as figure 7**C**, however, under varying conditions. (**E-H**) Inferred active and inactive periods same as figure 7**D**. Cells are the same as **A-B**.

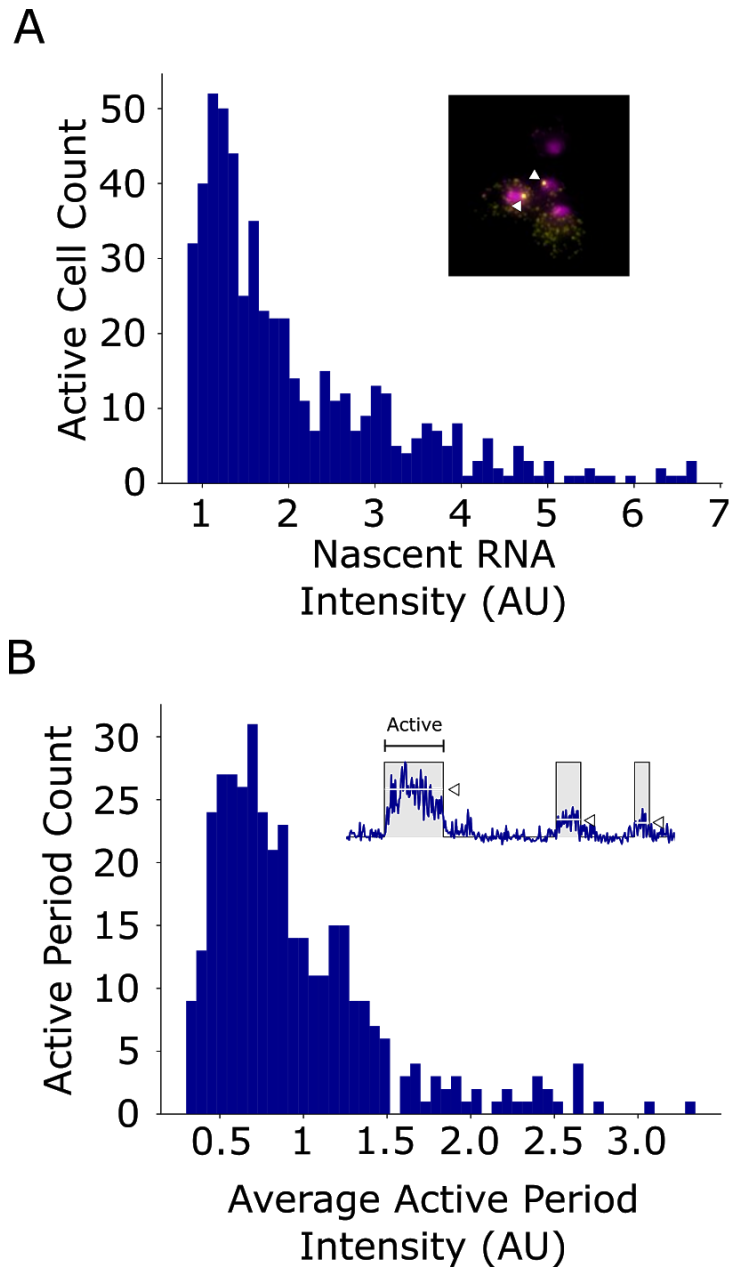

Figure 19**. Distribution of transcript number in active periods:** (**A**) Histogram of TS intensity from active WT cells determined by smFISH normalized by the mode of the distribution of cytoplasmic transcripts. An example smFISH image (inset) is also depicted with a DAPI stained nucleus (magenta) and *PHO5* transcripts tagged with CY3 (yellow). The nascent transcript at the TS is marked by a white triangle. (**B**) Histogram of average TS intensity from individual active periods determined by live-cell microscopy using CPD. An example trace (inset) is depicted with active periods indicated (grey boxes). The average of each active period in the example trace (white bars) are indicated with a white triangle. Variance in transcript intensity indicates the variable number of transcripts present at the TS in time.

Chapter 3: Permutational analysis of *Saccharomyces cerevisiae* regulatory elements

Namrita Dhillon, Robert Shelansky, Brent Townshend, Miten Jain, Hinrich Boeger, Drew Endy, Rohinton Kamakaka

Gene expression requires the integration of many signals at specific regulatory elements at a diverse set of locations at different times. As discussed above, regulation is often mediated by sequence specific regulators that recognize their cognate binding sites in both DNA. However, regulation is also achieved through the recognition of regulatory elements in RNA. Mutational analysis and genomic/epigenomic mapping of proteins to specific genomic locations has led to the classification and delineation of seemingly modular sequence elements: enhancers, promoters, 5' untranslated regions (5' UTRs), open reading frames (ORFs), 3' untranslated regions (3' UTRs), as well as transcriptional terminators (TTs). In yeast, the promoter and enhancers are often conflated together; in this document, we will use the term promoter to refer to the DNA elements that are bound by the general transcription factors and the polymerase and the term enhancer to refer to sequences containing UAS elements which are recognized and bound by specific activators [93]. The enhancer, made up of UAS elements, directs the initiation of transcription in response to signals while the core promoter, made of the TATA box and initiator elements, functions as a scaffold for the binding for GTFs and RNA polymerase II. The 5' UTR is involved in the association of the mRNA with the

ribosome. While the 3' UTR is involved in mRNA stability and turnover. The TT signals to initiate poly-adenylation of nascent transcripts which leads to transcript release from the DNA.

Together these elements, enhancer, promoter, 5' UTR, 3'UTR and TTS, form a regulatory cassette responsible for the regulation of expression. While these protein bound elements are modular and interchangeable, they do not function in isolation. Thus, expression ultimately requires effective integration of all of the signals via functional communication between different regulatory elements leading to a defined output. Early studies on gene regulation investigated regulatory elements of single genes via directed mutagenesis [94–96]. These analyses gave way to saturation mutational studies of a single well-defined element. For example, mutating one UAS element in an enhancer [97]. These approaches studied the function of these elements in isolation but did not systematically study the ability of an element to functionally communicate with other elements in a regulatory cassette. To study functional interactions between elements in prokaryotes an alternative approach was developed which relied on the construction of large permutational libraries of promoters and 5'UTRs [98–100]. This method used a fluorescent reporter introduced downstream of each cassette in the permuted library to sort cells, on the basis of signal intensity, into bins of various expression levels. Then the sorted subpopulations of cells were sequenced to identify which sequence elements and collections of sequence elements correspond to a given level of expression.

We describe a simple and rapid approach to build a combinatorial library of regulatory elements and have used this library to study regulation of 26 different

genes. We develop a similar approach as described above to characterize regulatory elements in the eukaryote, Saccharomyces cerevisiae. To this end, we annotated 26 different yeast genes with four different classes of regulatory elements: enhancers, promoters, 5' UTRs and 3' UTR/TT based on published data [101–107]. Then, we constructed a permuted library with a maximal number of 456976, complete synthetic genes that drove the expression of a fluorescent reporter, using a modular cloning system [108]. Cells containing the synthetic genes were then sorted by fluorescence intensity and sequenced by Nanopore sequencing. Thus, the collective contribution of each sequence element to overall expression of each cassette was inferred.

Methods

Annotation of regulatory elements

Approximately, 60 different yeast genes of various expression levels in glucose were initially screened [102,103,109]. From this set we selected 26 genes for experimental analysis based on the availability of data mapping various epigenetic marks and transcription factors. This final gene set endogenously maintains various levels of expression in glucose. Notably, a quarter of the genes are perceived as inactive in glucose. The regulatory elements of these 26 genes were then annotated using various databases [101–107]. We identified the 3' UTR and TT sequences [101]. The TS was identified from RNA-seq data generated from cells grown in glucose [101–103]. Using TS information and the location of the most proximal start codon we demarcated the 5'UTRs. The core promoter was defined by the binding of TATA

binding protein (TBP) [103–107]; we selected genes classified as both TATA-less and TATA containing genes [103]. The enhancer was defined by the negative space between the most proximal upstream gene and the core promoter. The chromatin architecture of the upstream regulatory regions of these 26 different genes was mapped using ATAC-seq data [107]. These accessible sites have previously been shown to occur at protein binding sites in chromatin (figure 29). The permuted library was constructed from these individual annotated elements, using the Golden Gate ligation protocol to combine them in a directed but random manner [108].

Quantification of expression

Cells containing the library were grown to log phase in glucose containing media and sorted using a fluorescence assisted cell sorter (FACS) based on expression of both mRuby2 and EBFP2-2. We sorted the library into 4 expression fractions (no expression, low, medium and high expression) (figure 21). The gates for fluorescent cell sorting were based on various control strains. 63% of the sorted cells were in the no expression sorted fraction, 26% were in the low expression fraction, 8% were in the medium expressing fraction and 3% were in the high expressing fraction.

DNA was isolated from the four sorted pools and the entire regulatory cassette, including mRuby2, was PCR amplified. Barcodes were ligated to the amplified cassettes to distinguish the four sorted pools and the PCR products were subsequently sequenced using an Oxford Nanopore MinION sequencer. Long read sequencing is a requirement for this method in order to identify every element across each gene cassette, >500 bps. Otherwise, it would be impossible to identify the interaction between individual regulatory elements. Of the total mapped reads, 61%

were from the no mRuby2 expression fraction, 19% were from the low expression fraction, 5% were from the medium mRuby2 expression fraction and 14% were from the high mRuby2 expressing fraction. To ensure a sequencing depth across cell fractions comparable to the number of cells sorted into each fraction, samples were pooled with concentrations which corresponded to their ratio in the cell fraction. We then estimate the mean expression of a specific element or a collection of elements from the elements read distributions. For ease of communication we define a gene fragment as a contiguous piece of DNA which contains one or more elements. Thus, a fragment containing all 4 elements, enhancer, promoter, 5' UTR and 3' UTR/TT, represents one regulatory cassette; while a fragment with a specific enhancer and promoter represent the average expression from 676, 26^2, regulatory cassettes.

We approximate the fraction of cells from a particular FACS fraction for a fragment using the number of reads observed for that fragment in each FACS fraction. We determined the ratio of a specific regulatory fragment in each sorted cell fraction by estimating the number of cells observed for each regulatory fragment in each cell fraction. The estimate, $x_{i,b}$, of the number of cells containing fragment, i, sorted into each fraction, b, was determined by normalizing the number of reads, $r_{i,b}$, by multiplication with,

$$x_{i,b} = \frac{r_{i,b} C_b}{R_b \sum C_b} \tag{3}$$

where C_b and R_b are the total number of cells sorted and reads mapped from bin b, respectively; i.e. we calculated the fractional representation of fragment i, in bin b and subsequently scaled that fraction by the fraction of cells observed in bin b by FACS.

We then inferred the mean expression exhibited by each regulatory element by fitting an estimate of cell counts for each fragment, described above, across each sorted fraction to a log normal distribution [110]. Thus, the minimum expression that could be achieved by a fragment would occur if that fragment was solely present in the no-expression FACS pool. Similarly, maximum expression by a fragment would be achieved if that fragment was solely present in the highly-expressing FACS pool. We assume that $x_{i,b}$ are random variables sampled from binned log-normal distributions where the bins are determined by the FACS fraction boundaries,

$$x_i = \begin{bmatrix} \int_{log(A_3)}^{+\infty} N(x, log(\mu_i), \sigma_i)dx \\ \int_{log(A_2)}^{log(A_3)} N(x, log(\mu_i), \sigma_i)dx \\ \int_{log(A_1)}^{log(A_2)} N(x, log(\mu_i), \sigma_i)dx \\ \int_{-\infty}^{log(A_1)} N(x, log(\mu_i), \sigma_i)dx \end{bmatrix}. \quad (4)$$

Where, x_i, is the vector of ratios for all bins described above, μ_i, is the mean expression, σ_i, is the standard deviation of expression and A_i, is the expression value for the upper boundary of bin b determined by FACS.

Results

Contribution of individual elements to expression

We compared the distribution of enhancers in the four sorted fractions (figure 22). We expect individual elements to maintain similar expression as the endogenous expression for the gene they came from. For example, enhancers from genes known to be active in cells grown in glucose rich media, like *TDH3* and *RPL28* that were enriched in the high and medium expressing fractions [102,103,109]. Similarly, among the genes included in our combinatorial library were a set of inducible genes, *GAL1*, *ADH2*, *CUP1* and *ICL1*, that are inactive in glucose rich media. The enhancers from these inducible genes are present almost exclusively in the non-expressing fraction; enhancers that were repressive in glucose conditions remained repressed regardless of other regulatory elements present. However, promoter elements show a much different result. Promoters for *PGI1* and *CDC19* are both highly expressed in glucose containing media [102,103,109], but their promoter fragments are present to a greater extent in the non-expressing fractions. While, promoters of genes not active in glucose containing media such as *HXT2*, *PHO5* and *ADH2* are enriched in the highly expressing fractions. The UTR fragments have similar distributions across the four sorted fractions, with a few exceptions, suggesting that, in general, they play a lesser role in regulating gene expression.

We then analyzed the average inferred expression of each element. Enhancers affect expression over a large range of expression values (figure 23). The inducible enhancers, like GAL1, were inactive in glucose rich media. While the

glucose induced and housekeeping enhancers, TDH3 and RPL28, were active to varying extents under these growth conditions (figure 23). The promoter fragments also modulate expression level, but the variation in expression is less than that observed with the enhancers. A similar profile was seen with the 3' UTRs. Most of the 5' UTRs cluster together indicating that the different 5' UTRs function more or less equivalently. We note here that we are measuring protein levels and the fold difference between the highly expressed genes and the inactive genes is not as great as the fold difference reported for these genes via measurements of mRNA. This could be due to protein homeostasis dampening the expression levels of mRuby2. In addition, Inferred expression levels are averaged across every other element in the permuted collection. Thus, the average inferred expression is expected to be significantly different from the expression of each endogenous locus. Though, as suggested above we expect the expression rank to be conserved.

Pairwise interactions between regulatory elements

To identify interactions between pairs of regulatory elements we analyzed the expression of all possible pairwise interactions of two elements within our data set. For example, the mean expression of each enhancer element when paired on the same fragment with one of the 25 different promoter elements was determined against a background of all 25 5' UTRs and all 25 3' UTRs/TTs. We sorted enhancers based on mean expression and identified strong, intermediate and inactive enhancers in glucose containing media (figure 24**A**). This rank distribution is consistent with published measurements of mRNA levels from these genes

[102,103,109] but there are some differences in the ranking presumably because the other regulatory elements alter enhancer mediated expression levels.

The *TDH3* gene is one of the most highly expressed genes in yeast. When we analyzed the *TDH3* enhancer across promoters we found several promoters (*HXT2*, *PDC1*, *ADH1* etc.) that were able to increase gene expression over the native *TDH3* enhancer/promoter pair. Similar increases in expression were observed for the other active enhancers as well. This change in expression was not only in one direction. Several promoters dampened expression from even the strongest enhancers. For instance, expression from the *TDH3* enhancer was significantly reduced by the *LEU9*, *CUP1* and *ICL1* promoters. This suggests that promoters functionally communicate with enhancers to modulate enhancer mediated transcription. This also suggests a method of communication that is multiplicative and not additive [91]. An additive interaction cannot yield repression. Enhancers and promoters regulate distinct steps toward gene expression.

Clustering was used to classify and subcategorize different regulatory elements and pairs of regulatory elements. We identify three main clusters of enhancers, a high and a low expressing cluster and a cluster where the enhancer is not active (figure 24**B**). The difference in inferred expression between enhancer clusters is made by increasing the level of expression of all promoters. However, certain enhancers, *TDH3* and *CDC19*, appear capable of activating promoters no other enhancer can activate even though they are not the highest expressing enhancers. Clustering of promoters, though not robust, suggests that “TATA-less” promoters are weak promoters while the TATA containing core promoters cluster to

some extent as strong promoters which is consistent with studies of core promoters in yeast and human cells [111].

To test the hypothesis that expression of the endogenous locus corresponds to strength of individual regulatory elements that make up that locus we compared the rank correlation between enhancer expression and promoter expression. There was no observed correlation between the rank order of enhancer activity and the rank order of promoter activity. This is likely due to the inclusion of inducible genes into our library which are repressed in glucose media but highly expressed when activated. When their native promoters are separated from their cognate enhancers and paired with other enhancers, these promoters' innate ability to foster high expression manifests itself. It is for this reason that *HXT2* and *ADH2* enhancers are inactive but their core promoters are among the strongest. This again highlights the distinct roles enhancers and promoters play in regulation of expression.

Analysis of expression for the regulatory elements indicated that the principal driver of variation in expression was generated by enhancers and promoters. Individual strong/weak promoters and enhancers dictate overall expression of a particular gene fragment. We sought to identify groups of enhancers and promoters which communicated with each other. A PCA using enhancers as samples and promoters as features, effectively distributed enhancers in promoter space (figure 25**A**). This analysis found that one predominant axis of variation across the 26 different promoters explained ~90% of the total variance (figure 25**A** Inset graph). Plotting the 26 enhancers across the first two principal components shows no distinct clusters, rather a gradient emerges. This distribution mirrors the rank order of these

enhancers based on expression. We therefore infer that PC1 reflects the ability of an enhancer to amplify expression, we call this enhancer strength. However, we observed some outliers to this general trend. The RPL28 enhancer occupies a distinct position on PC2, while *TDH3* is an outlier along PC1 demonstrating its ability to mediate very high levels of expression. The same analysis with promoters as samples and enhancers as features also showed that one principal component explains ~90% of the total variance (figure 25**B** inset graph). Thus, we conclude PC1 likely reflects promoter strength, with "TATA-less" core promoters at one end and strong TATA containing promoters at the other. It was surprising to us that the presence of the TATA box simply dictated expression level and did not impart more sophisticated structural communication between enhancers and promoters.

Interestingly, PC1 for 5' UTRs and 3' UTR's (data not shown) explains only ~60% of the total variance suggesting a more complex regulatory relationship between the UTRs though the exact nature of this variation remains to be determined.

Validation of expression of gene fragments

To verify the results obtained by FACS and Nanopore sequencing of various enhancer/promoter pairs, we selected 9 genes, of various strengths, from the full set of 26 to interrogate by an alternate technique. We built a smaller 81 cassette set with only permuted enhancers and promoters combined with the promoters' cognate 5'UTR controlling the fluorescent reporter Venus. The *PGK1* 3' UTR was selected as a control for these constructs. We measured the expression of the fluorescent Venus reporter, directly, using a fluorescent plate reader. The fluorescence intensity was

normalized to the total intensity observed across the sum of all 81 constructs (figure 26). A better solution would be to use an internal standard, i.e. control cells driving a known expression level of Venus, to normalize expression, as was done by FACs. In the absence of such a standard, normalizing by the sum was the best way to determine relative expression values for each gene fragment.

In glucose containing media, the *PGK1* enhancer with its cognate promoter generates 2.23% of the total fluorescence. This value almost doubles when the PGK1 enhancer is combined with either the *TDH3* or the *PDC1* core promoters (figure 26). Similarly, the *TDH3* enhancer with its cognate promoter generates approximately 5.95% of the total fluorescence. This value increases to 7.85% when the *TDH3* enhancer is combined with the *PDC1* promoter. Analysis of moderately strong enhancers shows vast increases in expression when paired with strong promoters from other active genes. For example, the *ACO1*, *RPL28* and *TPI1* enhancers/promoter combinations generate high levels of protein but the levels can be increased significantly by swapping their native promoters with the strong promoters.

Using these data, we show both enhancer and promoter elements positively and negatively influence expression. For example, the native *TDH3* and *PDC1* cassettes are ranked 1st and 2nd in overall expression (figure 24**A**). However, when the enhancer and promoter are separated, the *PDC1* promoter increases expression from the *TDH3* enhancer while the *TDH3* promoter dampens expression from the *PDC1* enhancer (figure 26) and the highest expressing cassette is the *TDH3* enhancer combined with the *PDC1* promoter. Similarly, the native *ICL1* cassette is

inactive in glucose rich media. However, analyzing its promoter separated from its native enhancer, we find that the *ICL1* promoter is a moderately strong promoter.

Total noise scales with mean expression

FACS sorting and Nanopore sequencing allow large numbers, >400000, of permuted fragments to be analyzed simultaneously, however, with reduced resolution in evaluation of expression mean and variance. When maintaining a small number of bins it is difficult to infer the amount of variation in expression present in individual fragments. To study expression noise, we therefore investigated the smaller 81-cassette yeast cell library using cytometry (figure 27). Using cytometry we calculated the mean fluorescence and the Fano factor values for each of the 81 cassettes. There was an increase in Fano factor with a corresponding increase in the mean and the relationship appears to track a universal curve. The universal relationship could be caused by a number of different noise sources, both biological and technical but the data are consistent to what has been observed before [112–116]. Since we tested TATA-containing and TATA-less promoters, our data indicate that the relationship is not simply dependent on the presence or absence of a TATA box. Notably, we observe a significant outlier to this general relationship between mean and noise, the fragment containing a *TDH3* enhancer and the *RPL28* promoter. We hypothesize that this relationship between total noise and mean expression is generated by translation. Due to the expectation that differential regulation generates different levels of intrinsic RNA noise it is unlikely that this trend is dominated by intrinsic RNA noise. However, noise generated by translational

bursting should scale with the expression of a gene. Alternatively, this relationship could be generated by extrinsic factors to expression (see appendix 1).

Enhancers and promoters act independently to respond to environmental stimuli

We used the combinatorial library to study gene activation and repression under varying growth conditions (figure 28**A-C**). To test whether repression and activation are achieved by enhancers and promoters independently, we chose to study expression of our 81-construct library in media containing glucose, galactose (a fermentable sugar), glycerol (a non-fermentable carbon source), as well as media lacking adenine.

When comparing changes in gene expression in glucose compared to galactose containing media, we find that the *ICL1* enhancer becomes derepressed in galactose. Notably, the expression of the *ICL1* promoter does increase as well. Similarly, in glycerol containing media both the *ICL1* and *ACO1* enhancers become active while genes involved in fermentation show reduced activity. Media containing or lacking adenine shows a similar effect for the *ADE2* enhancer. In contrast to *ICL1*, the *ACO1* promoter shows decreased expression. This is likely due to the general repression of most enhancers in the library. In addition, because we normalize to the sum of total intensity our estimate of expression is relative to the expression of the whole set. However, even without a normalizing control we observe repression and derepression. Thus, we conclude that in some instances both the promoter/5'UTR and enhancer play a cooperative role in regulation. This is unexpected because

activation is typically explained through the binding of specific activators to UAS. It is possible that uncharacterized UAS elements within the promoter generate this regulation.

Chromatin structure alterations generate promoter specific regulation

We investigated the specific change in regulation for enhancers and promoters in response to changes in chromatin structure by mutation of Rsc2 and Isw2. In a Rsc2 mutant the fold expression of most genes is reduced (figure 28**D-E**). Though, the opposite effect is observed at the *LEU9* promoter suggesting that the repressive effects of the *LEU9* promoter may be Rsc2 dependent. The same change in expression is observed in an Isw2 mutant. Risc and Isw2, as discussed above maintain a seemingly antagonistic behavior over the position of the "+1" nucleosome. A model of activation and repression of transcription that relies on the dynamic movement of the "+1" nucleosome would explain these results. These data suggest that this remodeling activity is essential for repression at some promoters but not others. In addition, we hypothesize that at most promoters, chromatin remodelers are required for proper gene activation.

Discussion

We observe that the *RPL28* and the *CDC19* enhancers have distinct expression patterns when paired with the different core promoters compared to the patterns observed with other glucose induced enhancers (such as *TDH3*, *PGK1*, *PDC1*). The *RPL28* and *CDC19* genes are regulated by the transcription activators Rap1p and

Abf1p. Rap1p binds 300-400 bp upstream from the transcription start site and has the ability to evict nucleosomes ~400bp from its binding site [117–119]. Genes required for growth in glucose containing media (such as *TDH3*, *PGK1*, *PDC1*) are regulated in part by the transcription activators Reb1p and Gcr1p, which bind near the "-1" nucleosome and promote *RSC* mediated nucleosome mobility immediately downstream of their binding sites [117–120]. It is possible that the ability of Rap1p to mobilize nucleosomes over a greater distance translates into its ability to activate genes from a more diverse set of promoters. Further mutagenic and molecular analysis of these synthetic constructs is needed to test the mechanisms underlying these interactions.

We observe that the promoters affect expression independent of specific regulators bound to the enhancers. This suggests that the promoter acts to integrate signals emanating from the enhancer and to modulate overall expression. In yeast, there are two promoter architectures: TATA containing promoters and TATA-less promoters [96,97,111,121–123]. However, our data suggest that the presence of a TATA box is likely to increase the levels of expression emparted by weak enhancers. The underlying molecular mechanism is most likely modulation of TBP binding. The presence of a TATA box at a core promoter likely increases the probability of the formation of a functional pre-initiation complex at the promoter since TATA boxes are high affinity binding sites for TBP/TFIID [111]. Thus, weak activators stimulate transcription via a molecular mechanism that benefits from enhanced affinity of TFIID binding to the promoter while strong activators can mediate high-levels of transcription even in the presence of a sub-optimal core promoter. This begs the

question, why do certain genes maintain weak enhancers but strong promoters. The answer likely lies in the requirements for gene activation, specificity, gene regulatory function, etc. described in Chapters 1 and 2. From PCA, we hypothesize that, in general, the enhancers impact on a promoter, and vice versa, may be described by that element's “strength”. Thus, the expected expression of a fragment is mostly imparted by the combination of enhancer and promoter “strength”. Notably, this relationship is seemingly multiplicative. Additive relationships do not explain how an element may repress other elements, which we observe for both promoters and enhancers. This type of mechanistic relationship may only be explained with cooperativity by activation of distinct reaction steps [91]. Thus, it is likely that enhancers and promoters play distinct biochemical roles in regulation.

The standardization of regulatory elements and their characterization under varying growth conditions is necessary in order for regulatory elements to be routinely mixed and matched for use in synthetic circuits [124]. We have generated a large, >4000000 fragment library of synthetic regulatory elements that exhibited varying activity levels similar to approaches previously used to explore enhancer-promoter combinations in prokaryotes [93,125,126]. Using this library, we have identified combinations of regulatory elements that generate a large spectrum of activity which is not present in each element's endogenous context. We also show that these cassettes respond to external stimuli. Additionally, we have created a catalog of elements that is a valuable resource for the design of synthetic regulatory circuits in yeast.

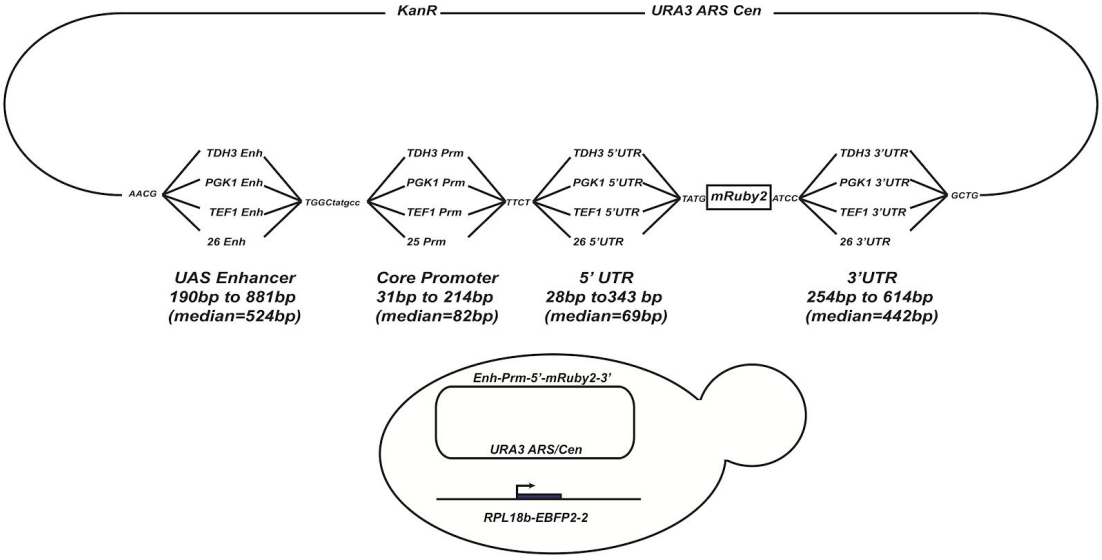

Figure 20**. Construction of the permuted library:** Each of the delineated regulatory elements (enhancer, promoter, 5'UTR and 3' UTR) from the 26 genes were PCR amplified using specific primers and cloned into the recently described parts vector pYTK001 [108]. Using the Golden Gate ligation protocol, the parts plasmids were then used to create a permutational library such that the elements would combine in a directed but random manner [108]. This resulted in ~400,000 recombinant plasmids containing different permutations of the four regulatory elements, UAS enhancer, core promoter, 5'UTR and 3' UTR, controlling the expression of a fluorescent reporter, mRuby2 (Figure 2A). The purified library was transformed into W-303 yeast cells, ROY5634. This strain also contained a fluorescent protein mTagEBFP2-2 under the control of the RPL18b promoter.

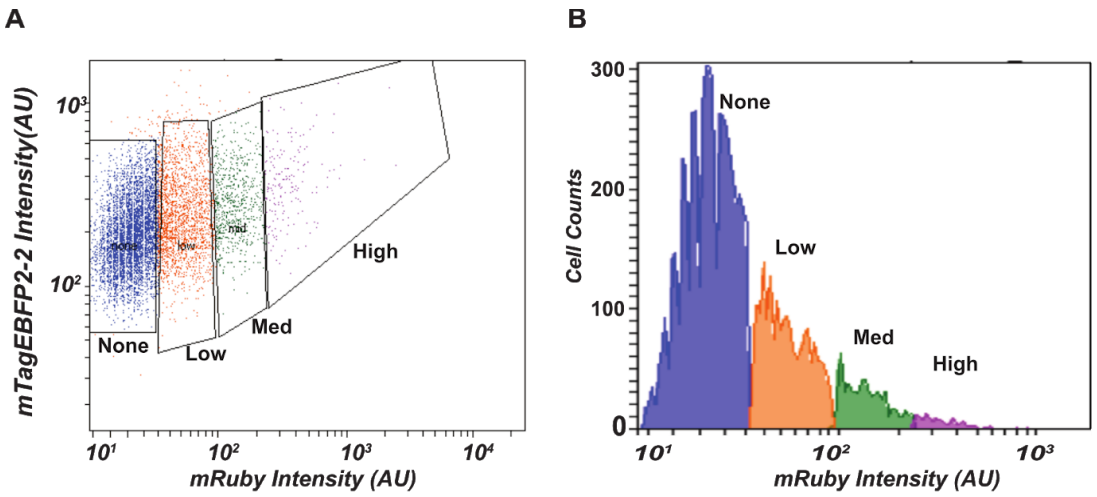

Figure 21**. Cytometry traces and sorted bins of yeast cells transformed with the permutation library:** Cells with no, low, medium and high mRuby2 expression were collected in four fractions. Blue: no mRuby2 expression, Orange: low mRuby2 expression, Green: medium mRuby2 expression and Purple: high mRuby2 expression. (**A**) mRuby vs mTagEBFP2-2 fluorescence, of individual cells (dots) by FACS. Bin cutoffs are drawn indicated (quadrilaterals). (**B**) mRuby2 fluorescence vs. cell count.

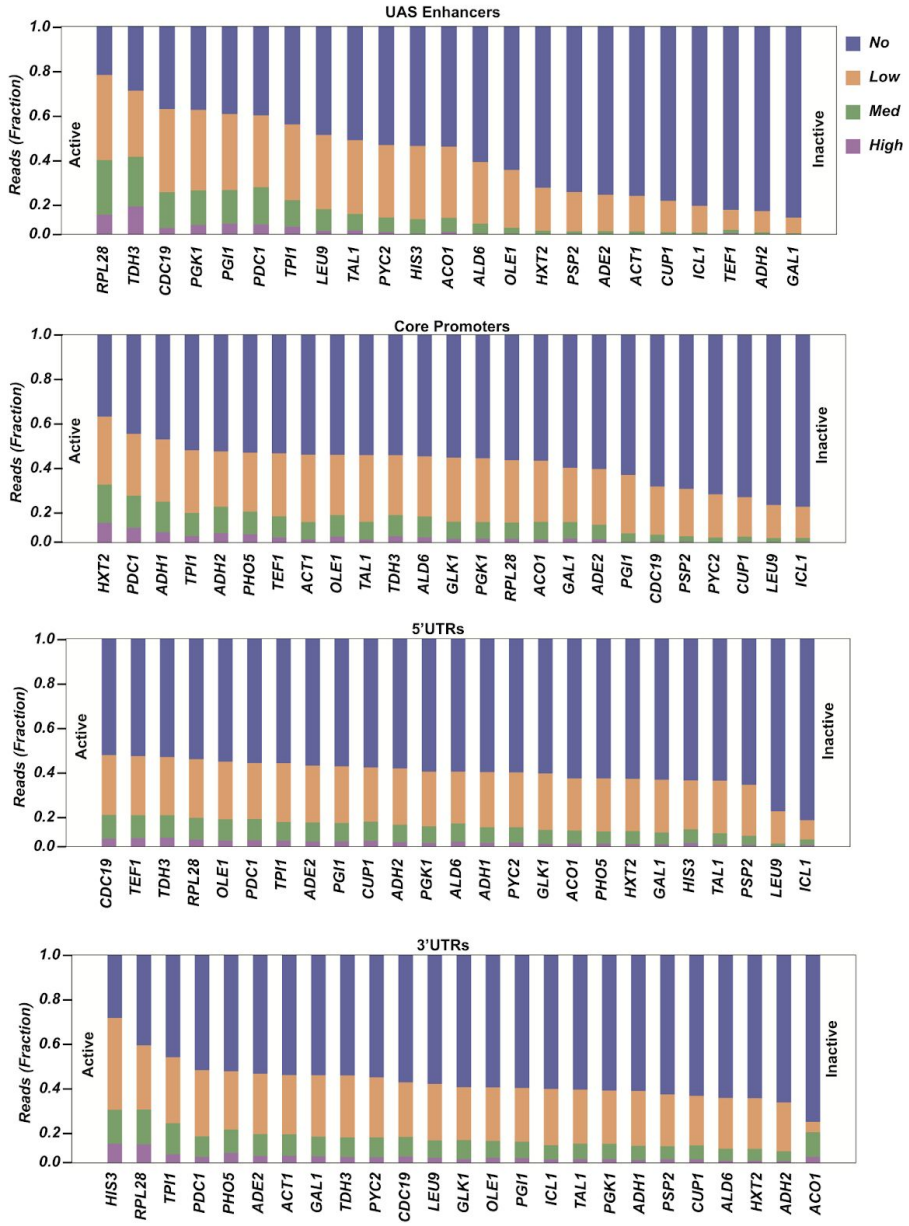

Figure 22. **Read histograms of regulatory elements:** Stacked histograms of the fraction of each regulatory element (enhancer, promoter, 5' UTR and 3' UTR/TT) present in each of the four sorted fractions. The elements are rank ordered based on read fraction in the no expression bin. Color code is the same as figure 21.

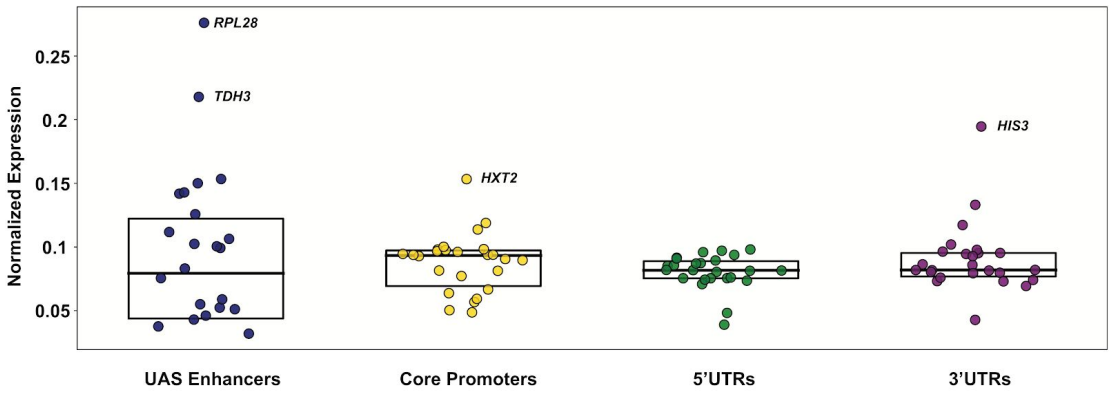

Figure 23**. Expression analysis of each regulatory element:** The mean expression of a regulatory fragment was calculated as discussed in the methods. Box plots depicting the normalized expression of all 26 regulatory elements are shown. The box plot edges represent the 25th to 75th percentiles and the line across the box represents the median. Note, this is the average expression for a particular element across every other possible combination of elements.

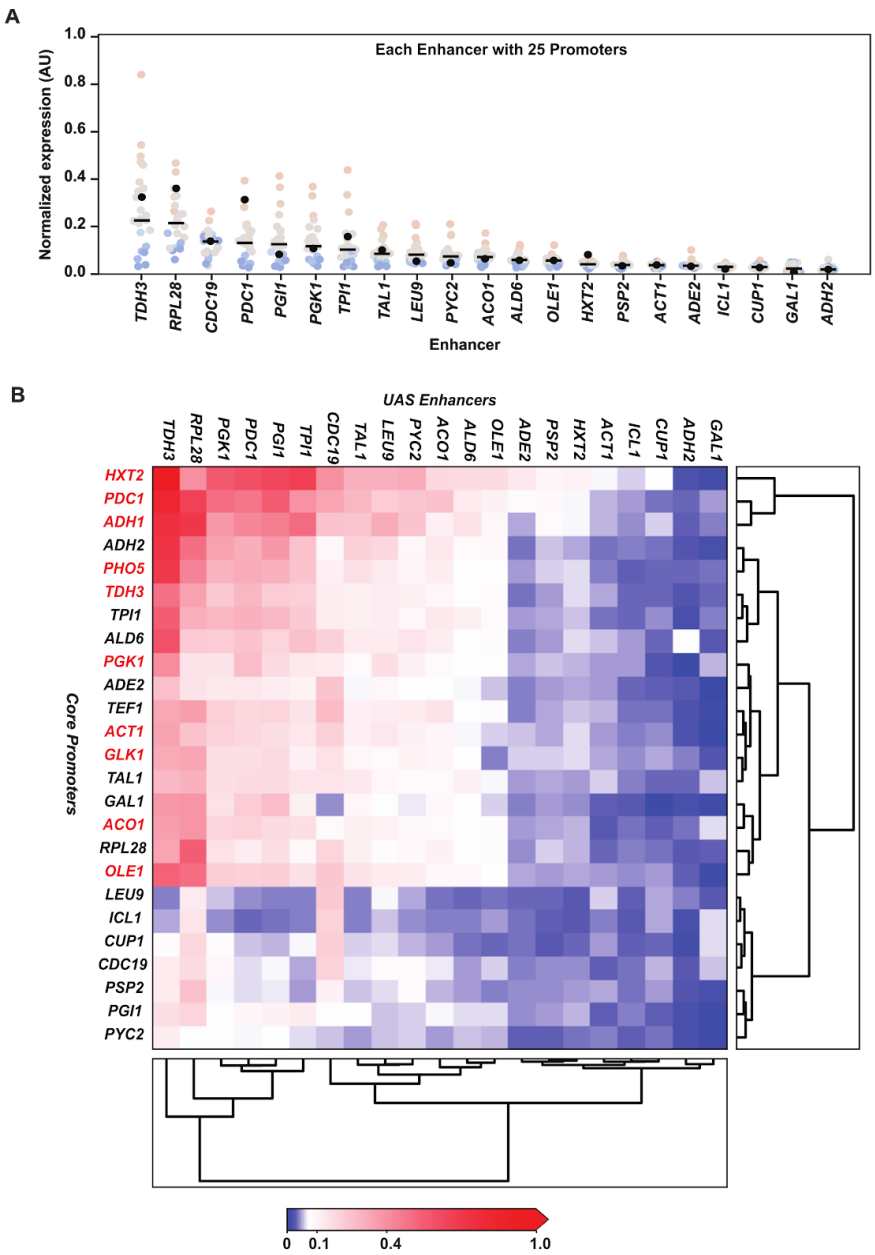

Figure 24. **Expression analysis of pairwise combinations of regulatory elements:** (**A**) The mean expression of a regulatory fragment was calculated as described in methods. Pairwise comparison of Enhancers (x-axis) with Promoters is shown. Each dot in the box plot represents one specific core promoter element. The color (red-blue) of the dot indicates the average expression of the corresponding promoter. Notably, the rank order of the colored dots remains constant across enhancers. The black dot represents the expression level mediated by the enhancer in combination with its native promoter. The black bar is the average enhancer expression. (**B**). Heat map and clustering of enhancers and promoters expression values. TATA containing promoters are labeled in red.

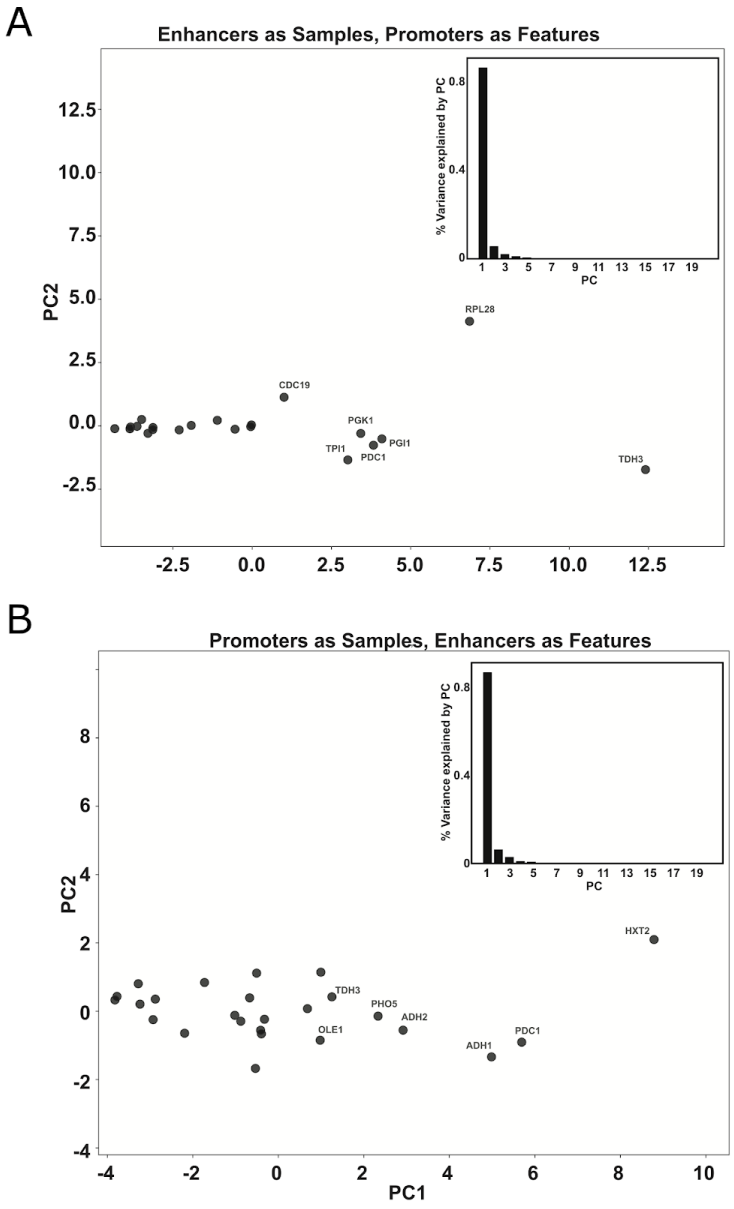

Figure 25. **Principal component analysis of pairwise combinations of regulatory elements:** (**A**) The plot displays each sample with respect to the first two principal components with UAS enhancers as samples and core promoters as features and was used to determine the relationship among the samples. The percent variance present within each principal component is plotted in the inset. (**B**) The score plot displays each sample with respect to the first two principal components with promoters as samples and enhancers as features and was used to determine the relationship among the samples. The percent variance present within each principal component is plotted in the inset.

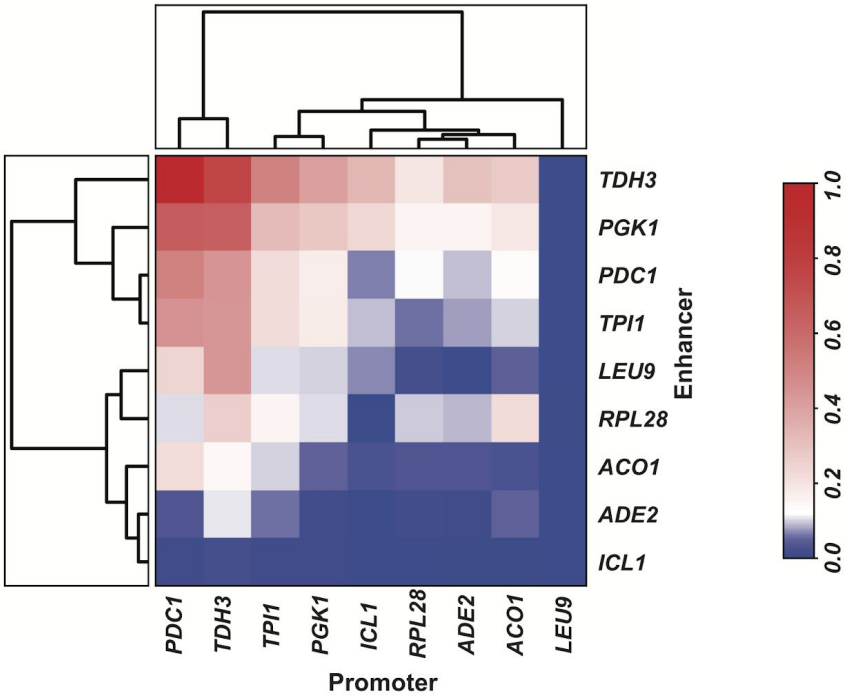

Figure 26**. Expression of enhancers with promoters/5'UTRs in glucose containing medium by direct observation of fluorescence:** Depicted is a 9x9 heatmap of different combinations of enhancers, promoters/5'UTRs driving expression of Venus with a PGK1 3' UTR/TT. Cell color is determined by Venus fluorescent intensity using a flourometer. The expression of each individual pairwise combination was listed as a percentage of the sum of the expression values of all 81 constructs. Clustering of elements by Ward's method is also depicted.

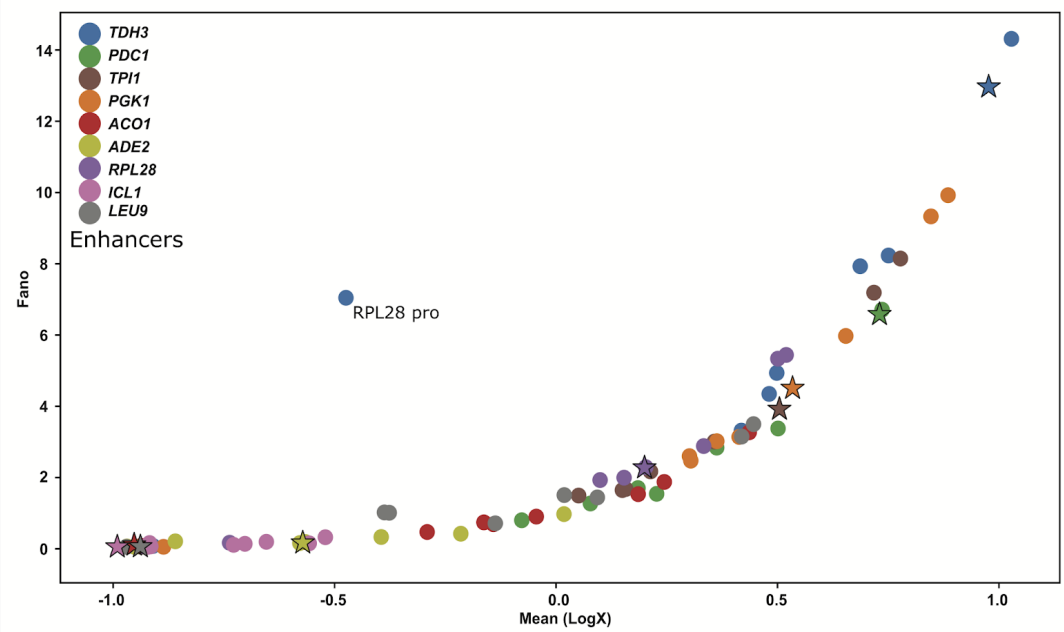

Figure 27**. Mean vs. noise determined by cytometry:** Cells containing the 81 constructs were grown to log phase and analyzed in a flow cytometer to measure expression levels in individual cells. The mean and Fano factor, a metric which describes noise, were calculated for the population of cells for each construct. Experiment was performed in triplicate.

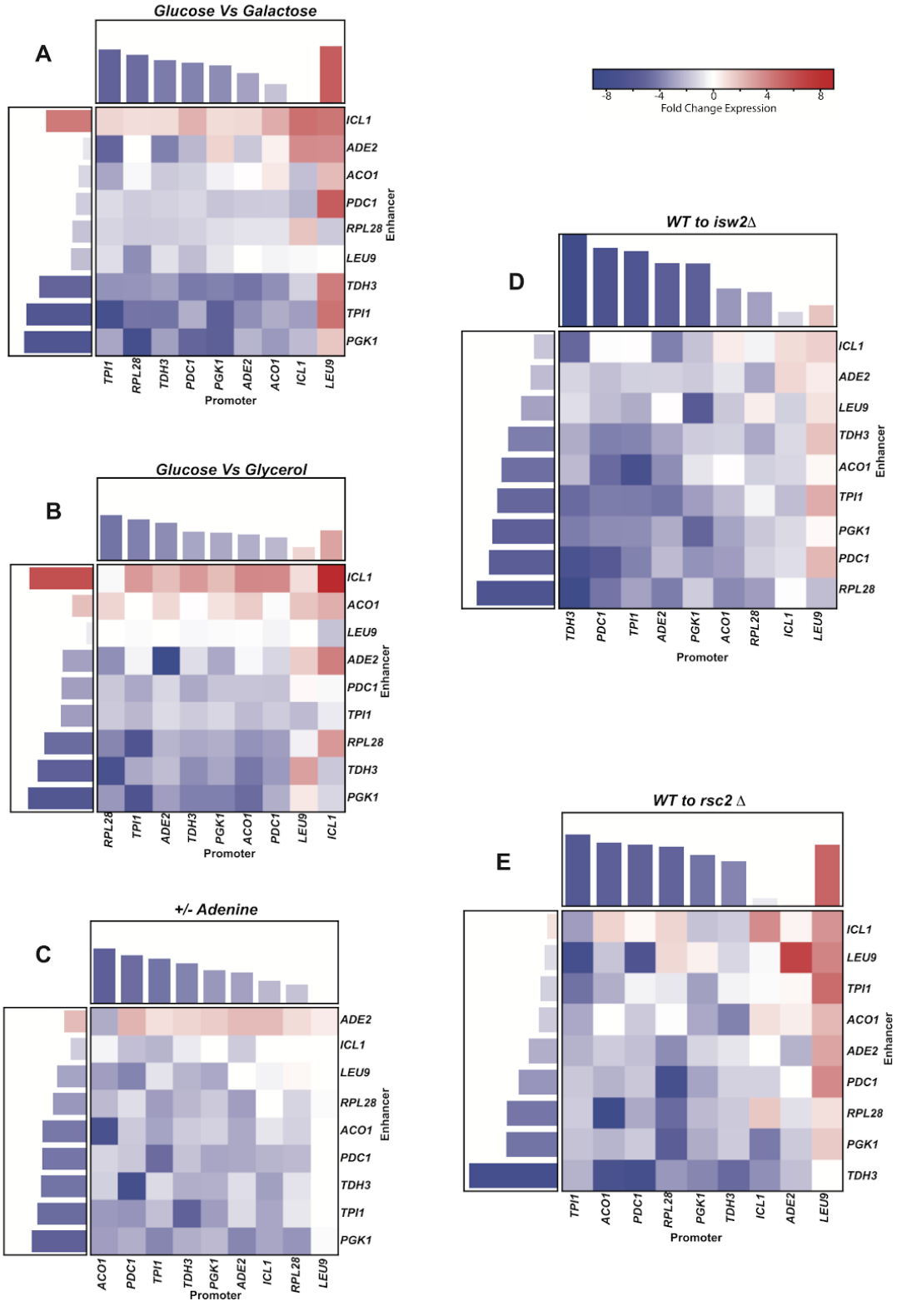
A
Glucose Vs Galactose
ICL1
ADE2
ACO1
PDC1
RPL28
LEU9
TDH3
TPI1
PGK1
Enhancer
TPI1 RPL28 TDH3 PDC1 PGK1 ADE2 ACO1 ICL1 LEU9
Promoter
-8 -4 0 4 8
Fold Change Expression
B
Glucose Vs Glycerol
ICL1
ACO1
LEU9
ADE2
PDC1
TPI1
RPL28
TDH3
PGK1
Enhancer
RPL28 TPI1 ADE2 TDH3 PGK1 ACO1 PDC1 LEU9 ICL1
Promoter
C
+/- Adenine
ADE2
ICL1
LEU9
RPL28
ACO1
PDC1
TDH3
TPI1
PGK1
Enhancer
ACO1 PDC1 TPI1 TDH3 PGK1 ADE2 ICL1 RPL28 LEU9
Promoter
D
WT to isw2Δ
ICL1
ADE2
LEU9
TDH3
ACO1
TPI1
PGK1
PDC1
RPL28
Enhancer
TDH3 PDC1 TPI1 ADE2 PGK1 ACO1 RPL28 ICL1 LEU9
Promoter
E
WT to rsc2 Δ
ICL1
LEU9
TPI1
ACO1
ADE2
PDC1
RPL28
PGK1
TDH3
Enhancer
TPI1 ACO1 PDC1 RPL28 PGK1 TDH3 ICL1 ADE2 LEU9
Promoter

Figure 28**. Expression analysis of the 81 constructs under varying growth conditions.** Cells containing the 81 combinations of 9 enhancers with 9 promoters/5' UTR were grown in different growth conditions and expression of the Venus cassette was measured in a fluorometer. (**A-E**) Each heatmap depicts log-fold change in expression that was calculated as from each condition to WT cells grown in glucose containing media. Three biological replicates were measured for each construct. Bar graphs above and on the left of the heat map are the summation of the nine individual values in the rows (enhancers) or columns (promoters). (**A**) Cells grown in medium containing galactose. (**B**) Cells grown in medium containing glycerol. (**C**) Cells grown in medium lacking adenine. (**D**) Cells containing an isw2 deletion grown in glucose. (**E**) Cells containing an rsc2 deletion grown in glucose.

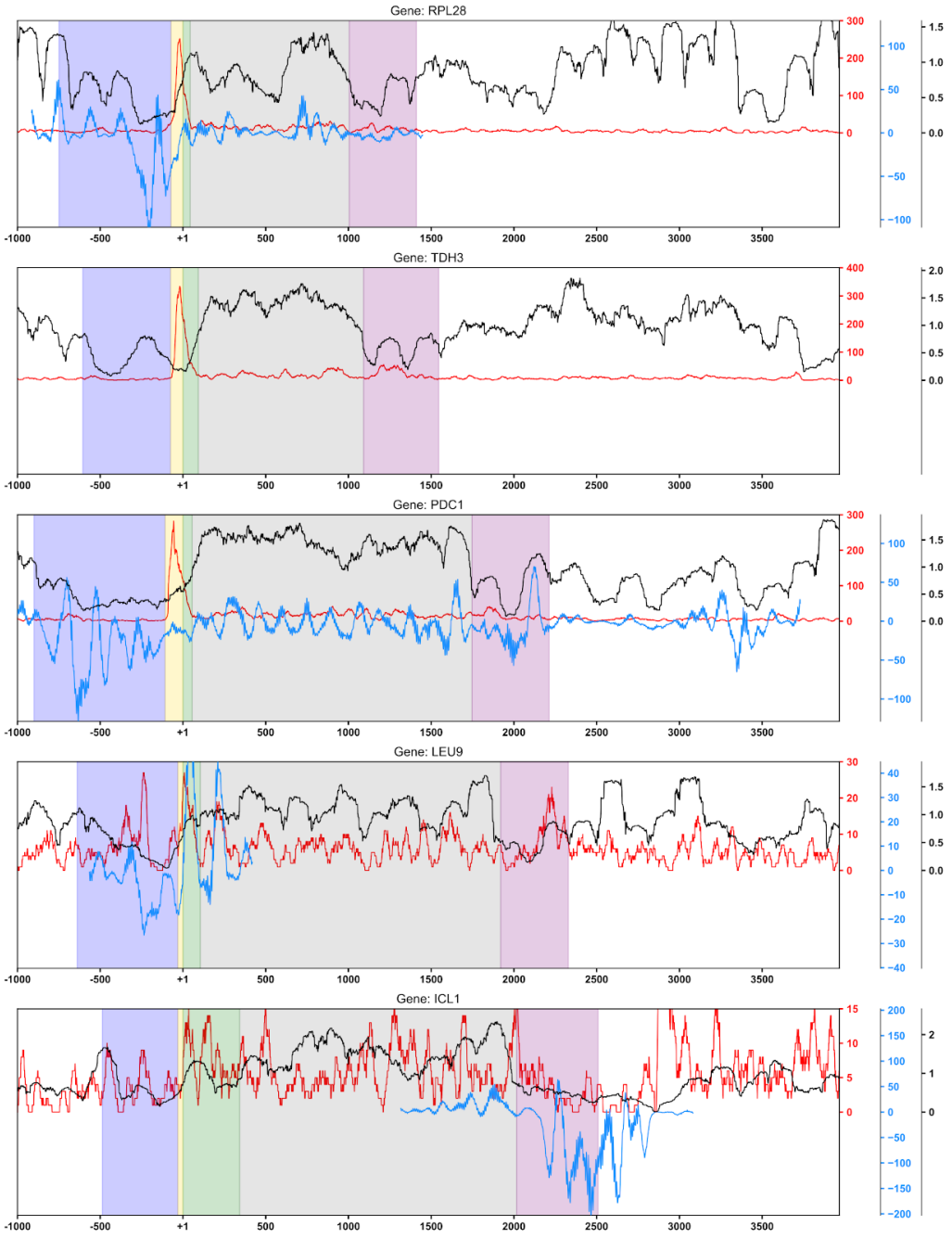

Figure 29**. Plots of nucleosomes and transcription factor distribution at a subset of genes analyzed:** The x-axis indicates the coordinate in the genome relative to the TS. The elements of each gene are indicated by shaded areas, enhancer (blue), promoter (yellow), 5' UTR(green), endogenous ORF (grey) and 3' UTR (purple). Nucleosome occupancy as determined by mnase seq (black line), ATACseq (blue line). TBP occupancy determined by ChIP (red line).

Concluding Remarks

In this thesis I presented a framework for describing the specificity of transcriptional regulation using continuous time markov processes, allowing for the description of the specificity of regulatory systems away from equilibrium. In doing so, I showed that stochastic removal and reformation of nucleosomes on promoter DNA may be used for the kinetic proofreading of protein-DNA interactions. This proposed proofreading model explains a number of observed biological phenomena, at least in part: i) why promoter nucleosome structure is heterogeneous even under activating conditions, ii) why transcription occurs in stochastic bursts, iii) how eukaryotic activators maintain high specificities while also maintaining high theoretical on-rates, among others.

In addition, I presented multiple experimental and computational techniques to observe transcriptional bursting, chromatin remodeling and activator binding at a classic model gene, *PHO5,* in *Saccharomyces cerevisiae*. Using these techniques I showed that transcriptional bursting of *PHO5* occurs in at least two distinct timescales. I presented a model of kinetic proofreading of activator binding by the chromatin remodeler Isw2 and showed that correlation between bursts of transcription is lost upon deletion of Isw2. I also tested a model by which activators find their target sequences by 1-dimensional searching across promoter DNA between nucleosomes. I showed that the distance between nucleosomes in the *PHO5* promoter correlates with activator on and off rates at a specific binding site

within the *PHO5* promoter. This hypothesis potentially explains indirect cooperativity observed between activators. In addition, this model suggests why nucleosome removal is required for gene activation but not transcript elongation, to enhance activator on-rates.

I also presented a novel technique for studying eukaryotic gene expression by generating and testing the expression of >400,000 permuted synthetic constructs generated from 26 genes from *Saccharomyces cerevisiae*.

Appendix A: Background

Pho5 and the *PHO* Pathway

The yeast PHO5 Gene is a classic model of eukaryotic promoter chromatin structure and gene expression noise [127]. Pho5 is an acid phosphatase that is secreted from the cell into the periplasm [128]. PHO5 expression is tightly regulated by activation of the PHO pathway, which responds to depletion of intracellular phosphate concentrations [122]. Upon phosphate starvation the PHO pathway becomes active, deploying a myriad of mechanisms to boost and maintain internal phosphate stores. This includes the expression Pho5, which scavenges phosphate from the extracellular environment [129].

The PHO5 gene responds to two upstream activating sequences (UAS) which are recognized and bound by the primary specific activator of the PHO pathway, Pho4. Pho4 binds cooperatively with the activator Pho2 [130]. Pho4 is constitutively expressed, however, its presence in the nucleus is regulated through a phosphorylated state controlled by the cyclin, Pho80, and cyclin dependent kinase, Pho85 [131]. Active Pho85 leads to a phosphorylated Pho4 which is shuttled out of the nucleus repressing PHO regulated genes. We use a deletion in pho80Δ to inactivate Pho85 and force an active PHO5 gene. This has the experimental benefit of allowing PHO5 to achieve a steady state expression independent of phosphate concentration; as opposed to an induction of the PHO pathway by phosphate starvation where expression varies in time. Within this thesis discussion of activating

conditions usually refers to a pho80Δ as repression conditions refers to WT cells grown in phosphate containing media

Conflating active and inactive periods with ‘On’ and ‘Off’ transitioning of a two state model

The correspondence between the observable, activity and inactivity, and the theoretical, ON and OFF transitioning, can be tenuous. An active period must encompass both the lifetime of individual transcripts on the DNA as well as the lifetime of the ON state. The lifetime of an active period cannot be shorter than the lifetime of a single transcript on the DNA. Thus, the lifetime of the transcript on the DNA sets the lower limit for the lifetime of the active period. While, the edges of the inactive periods are demarcated by the lifetime of the OFF state but may be obscured by the lifetime of the transcript on DNA. In addition, Short lifetimes in ON states may pass without activity. Liftimes in OFF states that are shorter than the lifetimes of the nascent transcript are undetectable. A window of activity must always start after the beginning of an ON state. While an ON state may end before or after the end of a window of activity. Beyond this, the determination of active and inactive periods is not free of arbitrary decisions (e.g. thresholding, CPD). However, CPD, high frequency MFM, and the use of a strongly transcribed gene with short ON sojourn and short transcript length, PHO5, may together provide a good estimate of the lifetimes for ON and OFF states through the observation of active and inactive periods. However, in general conflating periods of activity with the On/Off transition assumed from a two state promoter model yields false results.

Noise In Gene Expression

Gene expression for a particular gene varies from cell to cell within a population and even between gene homologues within individual diploid cells [132]. Intercellular and intracellular variability in gene expression, called transcriptional noise or simply noise, is caused by two types of variation: “extrinsic” and “intrinsic” [133]. Extrinsic noise is generated by variation in the cellular environment where transcription occurs and intrinsic noise is generated within the biochemical process of transcription [134]. Hence, extrinsic noise is the measured variation in expression between identical genes within two distinct cellular environments; intrinsic noise is the measured variation in expression between identical genes within the same cellular environment. Eukaryotic genes exhibit vastly different levels of intrinsic noise [135,136]. Divergent intrinsic noise levels for a pair of genes, determined by context specific biochemistry, are explained by alternative modes of regulation.

Noise may be quantified by the Fano factor -- a statistic that measures the ratio of transcript variability to transcript mean, $\frac{\sigma^2}{\mu}$. We may call the relationship between the Fano factor and mean transcript copy number the noise profile of a gene. [122]. Divergent noise profiles are explained by alternative modes of transcription: “Poissonian” and “bursty” [137]. Poissonian transcription is observed when transcript initiation events are uncorrelated to one another; bursty expression is exemplified by discrete pulses, or bursts, of correlated transcription initiation events followed by periods of inactivity [138].

A linear noise profile at unity is well explained by a stochastic birth-death process [138]. The birth-death model is a process in which transcripts are made at rate, (l), and degraded at rate, (d); this implies regulation only occurs at the level of transcript initiation and degradation. Noise profiles that are not linear or leave unity are often modeled by a random-telegraph process [135,136,139]. This model expands on the birth-death process by formalizing the promoter state: a promoter may either be “active” or “inactive” [133,138,140,141]. Transcripts are degraded at rate, (d), but can only be produced at rate, (l), if and only if the promoter is active. This model introduces two new regulatory parameters: (a), the rate at which the promoter becomes active and (b), the rate at which the promoter becomes inactive again; the parameter (a) describes the burst frequency, how often the promoter turns “on”, and the parameter (b) describes the burst size, how long a promoter remains “on” once activated. Note, noise produced within a random-telegraph process may reduce to that of the birth-death process [68,122].

Appendix B: Detailed Methods

Microscopy

Live-cell imaging

Yeast Cells were grown in liquid culture to mid log-phase (3 - 5 X10^7 cells/ml). Cells were then concentrated and spotted onto an agar patch and #1.5 coverslip as described in [142] using the mold described in [143]. Cells were then incubated at 30°C for 30 minutes. A stage top incubator was used to maintain constant temperature of 30°C during imaging. MFM Images were taken in 2.5 second

increments over 20 minutes with an exposure time of 250ms and a laser power of 30% (get exact power measurement).

smFISH

smFISH was accomplished as previously described [144]. Yeast cells were grown in liquid culture to mid log-phase. Cells were then cross-linked with formaldehyde(product), lysed with liticase(product) and adhered to poly-l-lysine coated coverslips (product). Coverslips were then hybridized with 2.5nM probe for 5 hrs at 37℃. Probes were targeted to the PP7 hairpin repeats and labeled with Quasar570 and Quasar670 (biosearch technologies). Coverslips were then mounted onto glass slides with mounting media containing DAPI(ProLong Gold, Life Technologies).

Image Analysis

Live-cell imaging

TS fluctuation analysis is achieved in four steps: i) Identification of candidate RNA, ii) assignment of candidate RNA to nuclei, iii) quantification of the TS intensity, and iv) tracking each TS in time. Candidate RNA are identified by finding local maxima in the band-pass filtered maximum projection of the z-stack at each time point; the maximum projection minimizes fluctuations in puncta intensity due to RNA movement in the z direction during imaging and the filter reduces the false positive rate due to cytoplasmic RNA and unbound coat protein. Once identified, a candidate RNA is then assigned to a nucleus by colocalization with an estimate of the nuclear

boundary. The brightest nuclear RNA is assumed to be the TS; A 'trace' is formed by tracking an individual TS in time. If no puncta is assigned at a given time point the TS position is interpolated. TS intensity quantification is achieved using an established gaussian mask algorithm [145]. Photobleach correction is achieved by trace detrending as described in [146]. Background normalization is achieved by subtracting the mode of the kernel density estimate for each trace; this assumes the most likely number of transcripts on the TS is zero. In autocorrelation analysis, median smoothing with a window size of 6, 15s (shorter than the lifetime of a single transcript), was used to reduce noise due to transcript movement in the z-dimension. As a transcript leaves focus between planes its intensity is reduced. Segmentation of traces into 'active' and 'inactive' periods was accomplished by CPD using a windowed approach with a normal cost function detailed in [147]. CPD has benefits over traditional thresholding methods due to its ability to better approximate boundary frames while reducing error in transcript detection. Transcript edges often have reduced intensity. Thresholding techniques are tuned to miss classify these frames in order to reduce background noise generally. CPD combines multiple frames across a segment to infer transcript presence of a single frame. Thus, noise in transcript detection is reduced while maintaining the ability to detect transcript edges. Autocorrelation functions were computed with the global mean and variance as suggested in [148]. Autocorrelation functions were fit to the 1-step promoter, 2-step promoter and a 4-state promoter.

smFISH

Cytoplasmic and nascent RNA were Identified as described above; however, cytoplasmic transcripts were required to lie within an estimate of the cytoplasmic boundary. The number of nascent RNA was calculated by normalizing the TS intensity with the average intensity of cytoplasmic transcripts. All analysis was accomplished using custom built scripts and previously established code [148]. RNA count distributions were fit by a Poisson process and the random telegraph process [149,150].

ChEC-CLK

Clk and Chec were adapted from [151], and [152]. Cultures were grown to mid log phase in YPD. Cultures were concentrated and aliquots of ~3 X10^9 cells were made. Cell aliquots were crosslinked in 1% formaldehyde for varying time lengths and quenched in 2.27M (final) Glycine. Fixed cells were washed in 125mM glycine, pelleted, frozen in liquid nitrogen and stored at -80°C for subsequent cleavage.

Thawed cells were washed by bufferA_d (15mM Tris-HCl ph 7.5, 80mM KCl, 2mM EDTA, 0.2mM spermine, 0.5mM spermidine, 5mM β-Mercaptoethanol) prepared fresh and supplemented with 1X PIs and 1X PMSF. Cells were washed a second time with bufferA_g (15mM Tris-HCl ph 7.5, 80mM KCl, 0.1mM EGTA, 0.2mM spermine, 0.5mM spermidine, 5mM β-Mercaptoethanol) supplemented with 1X PIs and 1XPMSF, also prepared fresh. Cell aliquots were split in half for the no cleavage control. Cleavage reactions were carried out by addition of 2.2mM (final) $CaCl_2$ and incubated at 30°C for 30 minutes. Reactions were quenched in stop buffer (1% SDS, 200mM NaCl, 10mM EDTA, 2mM EGTA). RNA and protein were then degraded with

RNAse A and Proteinase K treatment and DNA was subsequently purified by phenol/chloroform extraction and ethanol precipitation. Southern analysis was performed with a DNA probe upstream of PHO5's UAS1. Pho4 occupancy was determined computationally by quantifying relative band intensities for bound and unbound PHO5 gene molecules.

Gene Ring Analysis

Yeast culture, purification, psoralen crosslinking and electron microscopic imaging of gene rings was accomplished as described in [153]. R-value analysis of each molecule was determined by hand tracing electron micrographs and computationally assigning base coordinate and occupancy status using custom built python code. In brief: i) Orientation of a gene molecules were determined by presence of a downstream "fork" [55]. ii) Each DNA strand was identified and aligned to its complement by closest distance. iii) A coordinate was said to be nucleosome free if the distance between its position on the two DNA strands exceeded a threshold distance, determined empirically.

Strain	Genotype
YS18	MATalpha; his3-11; his3-15; leu2-3; leu2-112; can1-100; ura3D5
yM8.14	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; PHO5[GC]
yM19.2	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[GC, tata]
yM2.1	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; PHO5[GC]
yM7.8	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3
yR22	MATa; ade2-1; ura3-1; trp1-1; leu2-3,112; his3-11; can1-100; PHO4-MNase-3xHA-KANMX6
yC76.1	MATa; ade2-1; ura3-1; trp1-1; leu2-3,112; his3-11; can1-100; PHO4-MNase-3xHA-KANMX6; Pho4:D[85-99]
yM249.1	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[14xPP7-loxPKanMXloxP, insterted before start codon]
yM255.3	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[GC, 14xPP7, inserted at start codon]
yRS102	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[GC, 14xPP7, inserted at start codon]; pRPS2_PCP_GFPEnvy::pSIVURA
yRS97.1	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[GC, 14xPP7, inserted at start codon]; isw2d:KANMX
yRS118	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[GC, 14xPP7, inserted at start codon]; isw2d:KANMX; pRPS2_PCP_GFPEnvy::pSIVURA
yM265.1	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[GC, 14xPP7, inserted at start codon]; chd1:KanMX
yRS99.1	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[GC, 14xPP7, inserted at start codon]; chd1:KanMX; pRPS2_PCP_GFPEnvy::pSIVURA
yRS103	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[GC, 14xPP7, inserted at start codon]; pho4:URA3
yRS113	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[GC, 14xPP7, inserted at start codon]; Pho4 D75-90
yRS120	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[GC, 14xPP7, inserted at start codon]; Pho4 D75-90; pRPS2_PCP_GFPEnvy::pSIVURA
yRS19.7	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[14xPP7, inserted at start codon]
yTL362D2	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[14xPP7, inserted at start codon]; (PP7-PS-NLS-2x::URA)
yRS23.3	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[14xPP7, inserted at start codon]; pho4D:URA
yRS25.4	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[14xPP7, inserted at start codon]; Pho4D79-92
yRS38.1	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[14xPP7, inserted at start codon]; Pho4D79-92; (PP7-PS-NLS-2x::URA)
yRS30.5	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[14xPP7, inserted at start codon]; Pho4D91-99
yRS41.1	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[14xPP7, inserted at start codon]; Pho4D91-99; (PP7-PS-NLS-2x::URA)
yRS26.3	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[14xPP7, inserted at start codon]; Pho4D79-90
yRS39.1	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[14xPP7, inserted at start codon]; Pho4D79-90; (PP7-PS-NLS-2x::URA)
yRS27.1	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[14xPP7, inserted at start codon]; Pho4D75-78
yRS40.1	MATalpha; his3-11; his3-15; leu2-3; leu2-112; canR; ura3D5; pho80D:HIS3; pho5[14xPP7, inserted at start codon]; Pho4D75-78; (PP7-PS-NLS-2x::URA)
yRS64.1	MATa; ade2-1; ura3-1; trp1-1; leu2-3,112; his3-11; can1-100; PHO4-MNase-3xHA-KANMX6; Pho4:D[85-99]; pho80D:HIS3
yRS60.3	MATa; ade2-1; ura3-1; trp1-1; leu2-3,112; his3-11; can1-100; PHO4-MNase-3xHA-KANMX6; pho80D:HIS3

Table 1. Yeast strain construction for Chapter 2: Genotype changes are indicated in bold.

Strain	Parent	Transforming DNA
yM249.1	yM8.14	PCR(pTL31, P446/P447)
yM255.3	yM249.1	pSH47
yRS102	yM255	pTL205:Pac1
yRS97.1	yM255	pCM122:Xba/Kpn1
yRS118	yRS97.1	pTL205:Pac1
yM265.1	yM255	pCM123.1: Xba1/Kpn1
yRS99.1	yM265.1	pTL205:Pac1
yRS103	yM255	pCM4.5:Bamh1/Sal1
yRS113	yRS103	pCM64.2:Bamh1/Sal1
yRS120	yRS113	pTL205:Pac1
yRS19.7	yM7.8	PCR(pTL31, P446/P447) + pSH47
yTL362D23	yRS19.7	(pTL92)
yRS23.3	yRS19.7	pCM4.5:Bamh1/Sal1
yRS25.4	yRS23.3	pCM71.1:Bamh1/Sal1
yRS38.1	yRS25.4	(pTL92)
yRS30.5	yRS23.3	pCM63.2:Bamh1/Sal1
yRS41.1	yRS30.5	(pTL92)
yRS26.3	yRS23.3	pCM6.3:Bamh1/Sal1
yRS39.1	yRS26.3	(pTL92)
yRS27.1	yRS23.3	pCM67.1:Bamh1/Sal1
yRS40.1	yRS27.1	(pTL92)
yRS64.1	yC76.1	pCM67.6:EcoR1
yRS60.3	yR22.1	pCM67:EcoR1

Table 2. Strain Construction Strategy.

Primert	Sequence
p446	CTTCATCTCTCATGAGAATAAGAAACAACAAACAAATAGAGCAAGCAAATTCGAGATTACCAcaaagtgggagcgaggagatcc
p447	AATGGTACCTGCATTGGCCAAAGAAGCGGCTAAAATTGAATAAACAACAGATTTAAACATgcataggccactagtggatctg

Table 3. Primer List.

Appendix C: Continuous Time Markov Processes

A Graphical Representation of Biochemical Processes

We model stochastic molecular dynamics of *PHO5* as a continuous time homogeneous markov process, which may be represented by a 'directed' graph [20,21,154]. Let $Y = \{Y_t : \Omega \to S | t \in \mathbb{R}\}$ be a continuous time homogeneous Markov process with countable state space, $S = \{1, ..., n\}$. The sample space Ω corresponds to the set of all *PHO5* genes, while the state space, S, is the set of specific *PHO5* promoter configurations. The transition graph consists of nodes, which correspond to elements of S, and directed edges, which indicate allowed transitions between states. Thus, Individual biochemical states are represented as nodes and transitions between biochemical states are represented as directed edges. Y_t represents the state of the promoter at a specific point in time t. Thus, we can describe the probability of a promoter being in a specific state, i, at time, t, $P(Y_t = i)$. The relative concentrations of individual biochemical states are represented by a column vector of probabilities (which sum to 1) of the system being in each node at time t, $\mathbf{p}(t)$, while the rate of probability flow between nodes is described by the derivative of vector of probabilities, $\frac{d\mathbf{p}(t)}{dt}$.

By applying a Markov assumption, that the process is time invariant, it can be shown,

$$\frac{d\mathbf{p}(t)}{dt} = W\mathbf{p}(t) \tag{1}$$

Where W is a matrix of transition rate constants, called the 'generator', which defines the process. Equation 1 is often referred to as 'the master equation' of the process. The master equation is a series of differential equations relating the concentrations of individual biochemical states, probabilities of nodes, to the rates of chemical reactions, probability flow between nodes.

Definition of the transition function

The stochastic process may be defined by a matrix of transition functions with individual elements, $p_{ji}(t, t+h)$, for all pairs of nodes, (j, i), in the graph including self pairs. Where, $p_{ji}(t, t+h)$, represents the probability of being in state j after time t given the process started in state i at time t+h. The Markov assumption requires that the rate constants of a chemical reaction at any given time t are the same at any other time t+h. Applying a time homogeneous Markov assumption we get,

$$p_{ji}(t+h, t) = p_{ji}(h, 0) \equiv p_{ji}(h) \qquad (2)$$

The transition function, $p_{ji}(h)$, is time invariant; the probability of transitioning from i to j only depends on being in state i at time t and not the process' history prior to t. With respect to our process Y,

$$p_{ji}(t) = P(Y_t = j | Y_0 = i).$$

We then may define $P_t = \left(p_{ji}(t)\right)$, the square matrix of transition functions.

$$P_t = \begin{bmatrix} p_{11}(t) & p_{12}(t) & \cdots & p_{1n}(t) \\ p_{21}(t) & p_{22}(t) & \cdots & p_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1}(t) & p_{n2}(t) & \cdots & p_{nn}(t) \end{bmatrix}.$$

From equation 1. We see that,

$$\mathbf{p}(t) = P_t \mathbf{p}(0). \tag{3}$$

Where, $\mathrm{p}(0)$ are the initial conditions of the process. The probability of being in any state at time t depends only on the transition matrix and the initial conditions. The probability of being in state j at time t, $p_j(t)$, is equal to the sum of probability flow into and out of state j in time t, $\sum_i p_{ji}(t) p_i(0)$. The derivative of equation 3 yields the rate of probability flow between nodes,

$$\frac{d\mathbf{p}(t)}{dt} = \frac{d}{dt} P_t \mathbf{p}(0). \tag{4}$$

Equation 4 is a set of linear differential equations which describe the process. To solve this equation we must solve $\frac{d}{dt} P_t$.

Derivation of the Master equation

We derive the master equation from $p_{ji}(t+h)$ using a trick where we insert an additional state, k, in between the transition from $i \to j$, $i \to k \to j$, such that the transition from $i \to k$ occurs at time t, $p_{jki}(t+h, 0)$; where k can be i or j or any other state. Applying the Markov assumption, we can seperate $p_{jki}(t+h)$ into $p_{jk}(h)\, p_{ki}(t)$

because $p_{jki}(t+h,0) = p_{jk}(t+h,t)\,p_{ki}(t,0) \equiv p_{jk}(h)\,p_{ki}(t)$. The probability of transitioning from state i to state j in time t+h, $p_{ji}(t+h)$ is equal to the sum of the probabilities of transitioning from i to k at time t and then from k to j at time t+h over all k. Thus, we derive the Chapman-Kolmogorov equation:

$$p_{ji}(t+h) = \sum p_{jk}(h)p_{ki}(t)$$

The Chapman-Kolmogorov equation may be written in matrix form:

$$P_{t+h} = P_h P_t \tag{5}$$

We then take the derivative of P_t. By definition of the derivative,

$$\frac{dP_t}{dt} = \lim_{h\to 0} \frac{P_{t+h} - P_t}{h}.$$

From equation 5, we get,

$$\frac{dP_t}{dt} = \lim_{h\to 0} \frac{P_h P_t - P_t}{h}.$$

Then we may pull out P_t,

$$\frac{dP_t}{dt} = \lim_{h\to 0} \frac{(P_h - I)P_t}{h};$$

where I is the identity matrix, whose diagonal elements are all 1. We can show $P_0 = I$. Intuitively, the probability of transitioning from any state, i, to any other state, j, where $i \neq j$ is 0, $p_{ji}(0) = 0;\ i \neq j$. Additionally, if $i = j$, the probability must be 1; no other transitions were possible. Thus, $P_0 = I$, a matrix whose diagonal entities,

the $i = j$ entries, are all 1's and all other elements, $i \neq j$, are 0's. From this follows with substitution,

$$\frac{dP_t}{dt} = \lim_{h \to 0} \frac{P_{0+h} - P_0}{h} P_t,$$

because P_t is independent of h. From the definition of the derivative again, we get:

$$\frac{dP_t}{dt} = \frac{dP_0}{dt} P_t.$$

We then define W, where $W = \frac{dP_0}{dt}$,

$$\frac{dP_t}{dt} = W P_t. \tag{6}$$

The generator W is a constant matrix, whose entries are the transition 'rate constants' of the process. The diagonal elements are equal to the inverse of the sum of their columns, a byproduct of the conservation of probability mass; The rate out of a node is equal to the sum of the rates into other nodes from that node. These rate constants dictate the sojourn time before each transition. For the solution of P_t, see below.

Substituting in Equation 4 to Equation 6 we get,

$$\frac{d\mathbf{p}(t)}{dt} = W P_t \mathbf{p}(0).$$

Which substituting from Equation 2 simplifies into Equation 1,

$$\frac{d\mathbf{p}(t)}{dt} = W \mathbf{p}(t).$$

■

Steady state, equilibrium and detailed balance

In steady state, when the probabilities of each state stop changing in time, $\frac{d\mathbf{p}(t)}{dt} = 0$,

the master equation becomes,

$$0 = W\boldsymbol{\pi}; \tag{7}$$

Where, **π**, is the column vector of steady state (stationary) probabilities. The 0 here is the column vector of 0s. The generator W is singular, therefore, a non-trivial solution to Equation 7, **π ≠ 0** , always exists. If the graph is strongly connected, every node is reachable by every other node by some path, **π** is uniquely defined [155]. Therefore, **p**(t)→**π** for t → ∞ [156]. Equation 7 shows that **π** is the basis vector, whose sum is one, of the kernel of W. Alternatively, **π** may be calculated using the Matrix-Tree theorem [154,155].

A process is in 'detailed balance' if and only if, $w_{ji}p_i(t) = w_{ij}p_j(t)$, for all i,j, and t; the flux of probability mass between each node is the same. Detailed balance implies steady state, $\frac{d\mathbf{p}(t)}{dt} = 0$. This is intuitively understood: if the rate of probability flow between all nodes is equally balanced, detailed balance, then there can be no change in probability in any node in time, steady state. The reverse is not true. Detailed balance is analogous to the concept of equilibrium in biochemistry.

In detailed balance, there is no way to determine the direction of time, for there is no net flux of probability mass between any two nodes. Both forward and backward direction of any sequence of events are stochastically indistinguishable. A process in detailed balance is therefore called 'reversible'; a process that violates detailed balance is called 'irreversible' [156].

Whether a process in steady state is also in detailed balance may be inferred from the generator of the process [156]; A stationary process on a directed graph is in detailed balance, if and only if, for any closed loop of transitions, multiplication of the rate constants (generator elements, edge labels) going around the loop yields the same product regardless of direction (clockwise and counterclockwise). This is called Kolmogorov's criterion or ‘cycle condition’ for detailed balance [154,156]. We note that stationary processes on graphs without closed loops are necessarily in detailed balance; for the absence of loops trivially ensures that Kolmogorov's criterion is fulfilled.

Transcriptional Specificity

To model activator specificity, we consider two genes that are identical, except that one copy bears the binding site for a specific activator whereas the other does not. We define 'regulatory specificity' or 'activator fidelity', f, as a measure of the activator's ability to distinguish between target and non-target promoters, where

$$f = \frac{v_t}{v_n}$$

where v_t is the average steady-state rate of transcription for target promoter binding and v_n for non-target promoter binding. The fidelity is defined by the ratio of transcriptional output from target and non-target promoters. Thus, when the activator promotes transcription indiscriminately, or transcription is activator-independent, $f = 1$.

We infer v_t and v_n from promoters modeled as continuous time homogeneous Markov processes. First we define which promoter states are transcriptionally active, ON states. We assume the rate of transcript initiation is linearly dependent on the probability of being in an ON state, $\boldsymbol{\pi}^{ON}$. In steady state, $\boldsymbol{\pi}^{ON}$, is simply the sum of every element which corresponds to an on state, $\sum^{ON} \pi_i$. Thus, the specificity, f, is simply the ratio of $\boldsymbol{\pi}_t^{ON}$ for the target sequence over the non-target sequence, $\boldsymbol{\pi}_n^{ON}$.

$$f = \frac{\boldsymbol{\pi}_t^{ON}}{\boldsymbol{\pi}_n^{ON}}$$

If activator fidelity was being calculated we assumed that the only difference in rate between target and non-target binding was the off-rate, k_{off}. Thus, to calculate activator fidelity with changing activator off-rate, k_{off} was substituted for k_t and k_n to calculate $\boldsymbol{\pi}_t^{ON}$ and $\boldsymbol{\pi}_n^{ON}$.

Autocorrelation of Active Periods

Let $\{Y_t : \Omega \to S | t \in \mathbb{R}\}$ be a continuous time homogeneous Markov process with countable state space, $S = \{1, ..., n\}$, transition graph, G, and generator, W. Let G be strongly connected. We define a function, $f : Y_t \to X_t$, where.

$$X_t = \begin{cases} 1 : Y_t = k \\ 0 : Y_t \neq k \end{cases}$$

Similar to the definition of the on state for specificity we define a set of nodes where the promoter is transcriptionally active, present the value 1. While the set of inactive states present the value 0. We assume only one state $k \in S$ as the active state and all other states are inactive. A similar result can be determined for multiple states, k, and even for a process which models transcription explicitly.

We will now show that the *acf* of X_t is:

$$acf(t) = \sum_{i=2}^{n} \frac{A_i(k,k)}{1 - A_1(k,k)} e^{\lambda_i t}$$

Where λ_i are the eigenvalues of W and A_i are n X n matrices that are obtained from the eigenvectors of W, described below.

Because G is strongly connected, a uniquely determined stationary distribution **π** exists, we will label the vector **π** elements as p_i. For a stationary process the expectation of X_t:

$$\mathbb{E}(X_t) = \sum_{i \in S} f(i) p_i = p_k$$

The autocovariance is:

$$\begin{aligned} acv(t) &= \mathbb{E}(X_t X_0) - \mathbb{E}(X_t)\mathbb{E}(X_0) \\ &= \mathbb{E}(X_t X_0) - p_k^2 \end{aligned}$$

Given the Markov assumption and the definition of the process:

$$p_{ji}(t) = P(Y_t = j | Y_0 = i)$$

Thus by definition of the joint probability distribution and expected value,

$$\begin{aligned} \mathbb{E}(X_t X_0) &= \sum_{j \in S} \sum_{i \in S} f(j) f(i) P(Y_t = j | Y_0 = i) P(Y_0 = i) \\ &= \sum_{j \in S} \sum_{i \in S} f(j) f(i) p_{ji}(t) p_i \\ &= p_{kk}(t) p_k \end{aligned}$$

Thus follows,

$$acv(t) = p_{kk}(t) p_k - p_k^2 = p_k (p_{kk}(t) - p_k) .$$

From the *acv* we compute the variance of X_t, is $acv(0) = p_k(1 - p_k)$.

Thus, the *acf* defined as the $\frac{acv(t)}{acv(0)}$ is,

$$acf(t) = \frac{p_{kk}(t) - p_k}{1 - p_k} . \tag{8}$$

The transition function, P_t, including $p_{kk}(t)$, and stationary probability, p_t, may be obtained as follows. If W is diagonalizable (*i.e.*, if W has n linearly independent eigenvectors corresponding to n eigenvalues), P_t may be written as,

$$P_t = \sum_{i \in S} e^{\lambda_i t} A_i \quad .$$

Where, λ_i are the eigenvalues of W and A_i are n X n matrices that are obtained from the eigenvectors of W as follows, for proof see [157] appendix.

Let B be the matrix whose i[th] column vector, b_i is the eigenvector with corresponding eigenvalue of λ_i. Let B^{-1} be the inverse of B. The inverse always exists because $B's$ column vectors (and row vectors) are linearly independent. Matrix A_i is obtained by multiplication of the i[th] column vector of B with the i'th row vector of B^{-1}:

$$A_i = \begin{pmatrix} b_{1i} \\ \vdots \\ b_{ni} \end{pmatrix} \begin{pmatrix} b_{i,1}^{-1} & \dots & b_{i,n}^{-1} \end{pmatrix} \quad .$$

The transition function, $p_{kk}(t)$, is the k[th] diagonal element of P_t. Thus, with $A_i(k,k)$ the k[th] diagonal element of A_i, we obtain,

$$p_{kk}(t) = \sum_{i \in S} e^{\lambda_i t} A_i(k,k) \quad .$$

The generator always has the eigenvalue 0. This is a consequence of the conservation of probability mass. Let this eigenvalue be $\lambda_1 = 0$. Thus,

$$p_{kk}(t) = A_1(k,k) + \sum_{i=2}^{n} A_i(k,k)e^{\lambda_i t}$$

.

For all $i \in S$, $p_{ki}(t) \to p_k$ for $t \to \infty$. As time goes to infinity, the likelihood of ending up in state k is simply p_k. It follows that $p_k = A_1(k,k)$. Because all exponential terms go to 0 as $t \to \infty$. With substitution into Equation 8,

$$acf(t) = \sum_{i=2}^{n} \frac{A_i(k,k)}{1 - A_1(k,k)} e^{\lambda_i t}$$

.

■

Two-state promoter model

The generator for the Two-state promoter model was:

$$W = \begin{bmatrix} -k_{on} & k_{off} \\ k_{on} & -k_{off} \end{bmatrix}$$

The probability of the transcriptionally active state was:

$$\boldsymbol{\pi}^{ON} = \frac{k_{on}}{k_{on} + k_{off}}$$

The activator fidelity was:

$$f = \frac{k_{on} + k_t}{k_{on} + k_n}$$

The *acf* was $e^{-(a+b)t}$.

Proofreading model

The generator for the Proofreading promoter model was:

$$W = \begin{bmatrix} -k_{on} - \lambda & k_{off} & 0 & \beta \\ k_{on} & -k_{off} - \alpha & \beta & 0 \\ 0 & \alpha & -k_{off} - \beta & k_{on} \\ \lambda & 0 & k_{off} & -k_{on} - \beta \end{bmatrix}$$

The specificity was determined for the proofreading model by assuming that the activators presence and the removal of the nucleosome was required for transcription. Thus only state 3 is transcriptionally active. If activation does not require the activator but instead only requires the removal of the nucleosome, like in model 3 in chapter 1, then both state 2 and state 3 are active states.

References

1. Mao C, Brown CR, Falkovskaia E, Dong S, Hrabeta-Robinson E, Wenger L, et al. Quantitative analysis of the transcription control mechanism. Mol Syst Biol. 2010;6: 431. doi:10.1038/msb.2010.83

2. Ptashne M, Gann A. Transcriptional activation by recruitment. Nature. 1997;386: 569–577. doi:10.1038/386569a0

3. Bryant GO, Ptashne M. Independent recruitment in vivo by Gal4 of two complexes required for transcription. Mol Cell. 2003;11: 1301–1309. doi:10.1016/s1097-2765(03)00144-8

4. Cosma MP, Tanaka T, Nasmyth K. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. Cell. 1999;97: 299–311. doi:10.1016/s0092-8674(00)80740-0

5. Kuras L, Borggrefe T, Kornberg RD. Association of the Mediator complex with enhancers of active genes. Proc Natl Acad Sci U S A. 2003;100: 13887–13891. doi:10.1073/pnas.2036346100

6. Halford SE, Marko JF. How do site-specific DNA-binding proteins find their targets? Nucleic Acids Res. 2004;32: 3040–3052. doi:10.1093/nar/gkh624

7. Geertz M, Shore D, Maerkl SJ. Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. Proc Natl Acad Sci U S A. 2012;109: 16540–16545. doi:10.1073/pnas.1206011109

8. Shen WC, Green MR. Yeast TAF(II)145 functions as a core promoter selectivity factor, not a general coactivator. Cell. 1997;90: 615–624. doi:10.1016/s0092-8674(00)80523-1

9. Maerkl SJ, Quake SR. A systems approach to measuring the binding energy landscapes of transcription factors. Science. 2007;315: 233–237. doi:10.1126/science.1131007

10. Suter DM. Transcription Factors and DNA Play Hide and Seek. Trends Cell Biol. 2020;30: 491–500. doi:10.1016/j.tcb.2020.03.003

11. Pauling L. The probability of errors in the process of synthesis of protein molecules. Birkhauser; 1957. pp. 597–602.

12. Hopfield JJ. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. Proc Natl Acad Sci U S A. 1974;71: 4135–4139. doi:10.1073/pnas.71.10.4135

13. Brown CR, Mao C, Falkovskaia E, Jurica MS, Boeger H. Linking stochastic fluctuations in chromatin structure and gene expression. PLoS Biol. 2013;11: e1001621. doi:10.1371/journal.pbio.1001621

14. Jessen WJ, Hoose SA, Kilgore JA, Kladde MP. Active PHO5 chromatin encompasses variable numbers of nucleosomes at individual promoters. Nat Struct Mol Biol. 2006;13: 256–263. doi:10.1038/nsmb1062

15. Small EC, Xi L, Wang JP, Widom J, Licht JD. Single-cell nucleosome mapping reveals the molecular basis of gene expression heterogeneity. Proc Natl Acad Sci U S A; 2014.

16. Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. Science. 1974;184: 868–871. doi:10.1126/science.184.4139.868

17. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 1997;389: 251–260. doi:10.1038/38444

18. Kornberg RD, Lorch Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell. 1999;98: 285–294. doi:10.1016/s0092-8674(00)81958-3

19. Almer A, Rudolph H, Hinnen A, Hörz W. Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. EMBO J. 1986;5: 2689–2696. Available: https://www.ncbi.nlm.nih.gov/pubmed/3536481

20. Boeger H. Nucleosomes, transcription, and probability. Mol Biol Cell. 2014;25: 3451–3455. doi:10.1091/mbc.E14-02-0753

21. Boeger H, Griesenbeck J, Kornberg RD. Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription. Cell. 2008;133: 716–726. doi:10.1016/j.cell.2008.02.051

22. Boeger H, Griesenbeck J, Strattan JS, Kornberg RD. Nucleosomes unfold completely at a transcriptionally active promoter. Mol Cell. 2003;11: 1587–1598. doi:10.1016/s1097-2765(03)00231-4

23. Brown CR, Boeger H. Nucleosomal promoter variation generates gene expression noise. Proc Natl Acad Sci U S A. 2014;111: 17893–17898. doi:10.1073/pnas.1417527111

24. Kepler TB, Elston TC. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophys J. 2001;81: 3116–3136. doi:10.1016/S0006-3495(01)75949-8

25. Estrada J, Wong F, DePace A, Gunawardena J. Information Integration and Energy Expenditure in Gene Regulation. Cell. 2016;166: 234–244. doi:10.1016/j.cell.2016.06.012

26. Hamiche A, Sandaltzopoulos R, Gdula DA, Wu C. ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF. Cell. 1999;97: 833–842. doi:10.1016/s0092-8674(00)80796-5

27. Boeger H, Griesenbeck J, Strattan JS, Kornberg RD. Removal of promoter nucleosomes by disassembly rather than sliding in vivo. Mol Cell. 2004;14: 667–673. doi:10.1016/j.molcel.2004.05.013

28. Taatjes DJ. The Continuing SAGA of TFIID and RNA Polymerase II Transcription. Mol Cell. 2017;68: 1–2. doi:10.1016/j.molcel.2017.09.028

29. Adamkewicz JI, Hansen KE, Prud'homme WA, Davis JL, Thorner J. High affinity interaction of yeast transcriptional regulator, Mot1, with TATA box-binding protein (TBP). J Biol Chem. 2001;276: 11883–11894. doi:10.1074/jbc.M010665200

30. Auble DT, Hansen KE, Mueller CG, Lane WS, Thorner J, Hahn S. Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism. Genes Dev. 1994;8: 1920–1934. doi:10.1101/gad.8.16.1920

31. Zhou CY, Johnson SL, Gamarra NI, Narlikar GJ. Mechanisms of ATP-Dependent Chromatin Remodeling Motors. Annu Rev Biophys. 2016;45: 153–181. doi:10.1146/annurev-biophys-051013-022819

32. Griesenbeck J, Boeger H, Strattan JS, Kornberg RD. Affinity purification of specific chromatin segments from chromosomal loci in yeast. Mol Cell Biol. 2003;23: 9275–9282. doi:10.1128/mcb.23.24.9275-9282.2003

33. Mao C, Brown CR, Griesenbeck J, Boeger H. Occlusion of regulatory sequences by promoter nucleosomes in vivo. PLoS One. 2011;6: e17521. doi:10.1371/journal.pone.0017521

34. Brown CR, Mao C, Falkovskaia E, Law JK, Boeger H. In vivo role for the chromatin-remodeling enzyme SWI/SNF in the removal of promoter nucleosomes by disassembly rather than sliding. J Biol Chem. 2011;286: 40556–40565. doi:10.1074/jbc.M111.289918

35. Lorch Y, Maier-Davis B, Kornberg RD. Role of DNA sequence in chromatin remodeling and the formation of nucleosome-free regions. Genes Dev. 2014;28: 2492–2497. doi:10.1101/gad.250704.114

36. Sermwittayawong D, Tan S. SAGA binds TBP via its Spt8 subunit in competition with DNA: implications for TBP recruitment. EMBO J. 2006;25: 3791–3800. doi:10.1038/sj.emboj.7601265

37. Jeronimo C, Others. (2016) Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. Mol Cell. 64: 455–466.

38. Anandapadamanaban M, Others. (2013) High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. Nat Struct Mol Biol. 20: 1008–1014.

39. Malik HS, Henikoff S. Phylogenomics of the nucleosome. Nat Struct Biol. 2003;10: 882–891. doi:10.1038/nsb996

40. Hillen W, Gatz C, Altschmied L, Schollmeier K, Meier I. Control of expression of the Tn10-encoded tetracycline resistance genes. Equilibrium and kinetic investigation of the regulatory reactions. 1983;169: 707–721.

41. Forde GM, Others. (2006) LacO-LacI interaction in affinity adsorption of plasmid DNA. Biotechnol Bioeng. 95: 67–75.

42. Schleif RF. Modulation of DNA binding by gene-specific transcription factors. Biochemistry. 2013;52: 6755–6765. doi:10.1021/bi400968e

43. Lorch Y, LaPointe JW, Kornberg RD. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell. 1987;49: 203–210. doi:10.1016/0092-8674(87)90561-7

44. Boltzmann L. Ueber die Beziehung zwischen dem zweiten Hauptsatz der mechanischen Waermetheorie und der Wahrscheinlichkeitsrechnung, respective den Saetzen ueber das Waermegleichgewicht. Wiener Berichte. 1877;75: 373–435.

45. Mellenius H, Ehrenberg M. Transcriptional accuracy modeling suggests two-step proofreading by RNA polymerase. Nucleic Acids Res. 2017;45: 11582–11593. doi:10.1093/nar/gkx849

46. Liu X, Bushnell DA, Silva D-A, Huang X, Kornberg RD. Initiation complex structure and promoter proofreading. Science. 2011;333: 633–637. doi:10.1126/science.1206629

47. Chen H, Larson DR. What have single-molecule studies taught us about gene expression? Genes Dev. 2016;30: 1796–1810. doi:10.1101/gad.281725.116

48. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nat Methods. 2014;11: 319–324. doi:10.1038/nmeth.2834

49. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. 1976.

50. Kornberg RD. The molecular basis of eukaryotic transcription. Proc Natl Acad Sci U S A. 2007;104: 12955–12961. doi:10.1073/pnas.0704138104

51. Mao C, Brown CR, Griesenbeck J, Boeger H. Occlusion of regulatory sequences by promoter nucleosomes in vivo. PLoS One. 2011;6: e17521. doi:10.1371/journal.pone.0017521

52. Luo Y, North JA, Rose SD, Poirier MG. Nucleosomes accelerate transcription factor dissociation. Nucleic Acids Res. 2014;42: 3017–3027. doi:10.1093/nar/gkt1319

53. Li M, Hada A, Sen P, Olufemi L, Hall MA, Smith BY, et al. Dynamic regulation of transcription factors by nucleosome remodeling. Elife. 2015;4. doi:10.7554/eLife.06249

54. Boeger H, Griesenbeck J, Strattan JS, Kornberg RD. Nucleosomes unfold completely at a transcriptionally active promoter. Mol Cell. 2003;11: 1587–1598. doi:10.1016/s1097-2765(03)00231-4

55. Brown CR, Mao C, Falkovskaia E, Jurica MS, Boeger H. Linking stochastic fluctuations in chromatin structure and gene expression. PLoS Biol. 2013;11: e1001621. doi:10.1371/journal.pbio.1001621

56. Boeger H, Shelansky R, Patel H, Brown CR. From Structural Variation of Gene Molecules to Chromatin Dynamics and Transcriptional Bursting. Genes . 2015;6: 469–483. doi:10.3390/genes6030469

57. Lee C-K, Shibata Y, Rao B, Strahl BD, Lieb JD. Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet. 2004;36: 900–905. doi:10.1038/ng1400

58. Yuan G-C, Liu Y-J, Dion MF, Slack MD, Wu LF, Altschuler SJ, et al. Genome-scale identification of nucleosome positions in S. cerevisiae. Science. 2005;309: 626–630. doi:10.1126/science.1112178

59. Kubik S, Bruzzone MJ, Jacquet P, Falcone J-L, Rougemont J, Shore D. Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast. Mol Cell. 2015;60: 422–434. doi:10.1016/j.molcel.2015.10.002

60. Rodriguez J, Larson DR. Transcription in Living Cells: Molecular Mechanisms of Bursting. Annu Rev Biochem. 2020. doi:10.1146/annurev-biochem-011520-105250

61. Brahma S, Henikoff S. Epigenome Regulation by Dynamic Nucleosome Unwrapping. Trends Biochem Sci. 2020;45: 13–26. doi:10.1016/j.tibs.2019.09.003

62. Shelansky R, Boeger H. Nucleosomal proofreading of activator-promoter interactions. Proc Natl Acad Sci U S A. 2020;117: 2456–2461. doi:10.1073/pnas.1911188117

63. Abrahamsson S, Chen J, Hajj B, Stallinga S, Katsov AY, Wisniewski J, et al. Fast multicolor 3D imaging using aberration-corrected multifocus microscopy. Nat Methods. 2013;10: 60–63. doi:10.1038/nmeth.2277

64. Larson DR, Zenklusen D, Wu B, Chao JA, Singer RH. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. Science. 2011;332: 475–478. doi:10.1126/science.1202142

65. Slubowski CJ, Funk AD, Roesner JM, Paulissen SM, Huang LS. Plasmids for C-terminal tagging in Saccharomyces cerevisiae that contain improved GFP proteins, Envy and Ivy. Yeast. 2015;32: 379–387. doi:10.1002/yea.3065

66. Mao C, Brown CR, Falkovskaia E, Dong S, Hrabeta-Robinson E, Wenger L, et al. Quantitative analysis of the transcription control mechanism. Mol Syst Biol. 2010;6: 431. doi:10.1038/msb.2010.83

67. Donovan BT, Huynh A, Ball DA, Poirier MG, Lenstra TL. Single-molecule imaging reveals the interplay between transcription factors, nucleosomes, and transcriptional bursting. 2018 [cited 24 Feb 2020]. doi:10.1101/404681

68. Munsky B, Neuert G, van Oudenaarden A. Using Gene Expression Noise to Understand Gene Regulation. Science. 2012;336: 183–187. doi:10.1126/science.1216379

69. Poorey K, Viswanathan R, Carver MN, Karpova TS, Cirimotich SM, McNally JG, et al. Measuring chromatin interaction dynamics on the second time scale at single-copy genes. Science. 2013;342: 369–372. doi:10.1126/science.1242369

70. Schmid M, Durussel T, Laemmli UK. ChIC and ChEC; genomic mapping of chromatin proteins. Mol Cell. 2004;16: 147–157. doi:10.1016/j.molcel.2004.09.007

71. Brown CR, Boeger H. Nucleosomal promoter variation generates gene expression noise. Proc Natl Acad Sci U S A. 2014;111: 17893–17898. doi:10.1073/pnas.1417527111

72. Cech T, Pardue ML. Cross-linking of DNA with trimethylpsoralen is a probe for chromatin structure. Cell. 1977;11: 631–640. doi:10.1016/0092-8674(77)90080-0

73. Mirny L, Slutsky M, Wunderlich Z, Tafvizi A, Leith J, Kosmrlj A. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. J Phys A: Math Theor. 2009;42: 434013. doi:10.1088/1751-8113/42/43/434013

74. Donovan BT, Huynh A, Ball DA, Patel HP, Poirier MG, Larson DR, et al. Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting. EMBO J. 2019;38. doi:10.15252/embj.2018100809

75. Kubik S, Bruzzone MJ, Challal D, Dreos R, Mattarocci S, Bucher P, et al. Opposing chromatin remodelers control transcription initiation frequency and start site selection. Nat Struct Mol Biol. 2019;26: 744–754. doi:10.1038/s41594-019-0273-3

76. Bennett CH. Demons, Engines and the Second Law. Scientific American. 1987. pp. 108–116. doi:10.1038/scientificamerican1187-108

77. Rodriguez J, Ren G, Day CR, Zhao K, Chow CC, Larson DR. Intrinsic Dynamics of a Human Gene Reveal the Basis of Expression Heterogeneity. Cell. 2019;176: 213–226.e18. doi:10.1016/j.cell.2018.11.026

78. Blossey R, Schiessel H. Kinetic proofreading of gene activation by chromatin remodeling. HFSP J. 2008;2: 167–170. doi:10.2976/1.2909080

79. Narlikar GJ. A proposal for kinetic proof reading by ISWI family chromatin remodeling motors. Curr Opin Chem Biol. 2010;14: 660–665. doi:10.1016/j.cbpa.2010.08.001

80. Blossey R, Schiessel H. Kinetic proofreading in chromatin remodeling: the case of ISWI/ACF. Biophys J. 2011;101: L30–2. doi:10.1016/j.bpj.2011.07.001

81. Florescu A-M, Schiessel H, Blossey R. Kinetic control of nucleosome displacement by ISWI/ACF chromatin remodelers. Phys Rev Lett. 2012;109: 118103. doi:10.1103/PhysRevLett.109.118103

82. Blossey R, Schiessel H. Histone mark recognition controls nucleosome translocation via a kinetic proofreading mechanism: Confronting theory and high-throughput experiments. Phys Rev E. 2019;99: 060401. doi:10.1103/PhysRevE.99.060401

83. Mehta GD, Ball DA, Eriksson PR, Chereji RV, Clark DJ, McNally JG, et al. Single-Molecule Analysis Reveals Linked Cycles of RSC Chromatin Remodeling and Ace1p Transcription Factor Binding in Yeast. Mol Cell. 2018;72: 875–887.e9. doi:10.1016/j.molcel.2018.09.009

84. Corrigan AM, Tunnacliffe E, Cannon D, Chubb JR. A continuum model of transcriptional bursting. Elife. 2016;5. doi:10.7554/eLife.13051

85. Mivelaz M, Cao A-M, Kubik S, Zencir S, Hovius R, Boichenko I, et al. Chromatin Fiber Invasion and Nucleosome Displacement by the Rap1 Transcription Factor. Mol Cell. 2020;77: 488–500.e9. doi:10.1016/j.molcel.2019.10.025

86. Castellanos M, Mothi N, Muñoz V. Eukaryotic transcription factors can track and control their target genes using DNA antennas. Nat Commun. 2020;11: 540. doi:10.1038/s41467-019-14217-8

87. Mirny LA. Nucleosome-mediated cooperativity between transcription factors. Proc Natl Acad Sci U S A. 2010;107: 22534–22539. doi:10.1073/pnas.0913805107

88. Schaffter SW, Schulman R. Building in vitro transcriptional regulatory networks by successively integrating multiple functional circuit modules. Nat Chem. 2019;11: 829–838. doi:10.1038/s41557-019-0292-z

89. Larsson AJM, Johnsson P, Hagemann-Jensen M, Hartmanis L, Faridani OR, Reinius B, et al. Genomic encoding of transcriptional burst kinetics. Nature. 2019;565: 251–254. doi:10.1038/s41586-018-0836-1

90. Estrada J, Wong F, DePace A, Gunawardena J. Information Integration and Energy Expenditure in Gene Regulation. Cell. 2016;166: 234–244. doi:10.1016/j.cell.2016.06.012

91. Scholes C, DePace AH, Sánchez Á. Combinatorial Gene Regulation through Kinetic Control of the Transcription Cycle. Cell Syst. 2017;4: 97–108.e9. doi:10.1016/j.cels.2016.11.012

92. Li C, Cesbron F, Oehler M, Brunner M, Höfer T. Frequency Modulation of Transcriptional Bursting Enables Sensitive and Rapid Gene Regulation. Cell Syst. 2018;6: 409–423.e11. doi:10.1016/j.cels.2018.01.012

93. Blazeck J, Alper HS. Promoter engineering: recent advances incontrolling transcription at the most fundamental level. Biotechnol J. 2013;8: 46–58.

94. Maniatis T, Goodbourn S, Fischer JA. Regulation of inducible and tissue-specific gene expression. Science. 1987;236: 1237–1245. doi:10.1126/science.3296191

95. Smale ST, Kadonaga JT. The RNA polymerase II core promoter. Annu Rev Biochem. 2003;72: 449–479. doi:10.1146/annurev.biochem.72.121801.161520

96. Hahn S, Young ET. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics. 2011;189: 705–736. doi:10.1534/genetics.111.127019

97. Lubliner S, Regev I, Lotan-Pompan M, Edelheit S, Weinberger A, Segal E. Core promoter sequence in yeast is a major determinant of expression level. Genome Res. 2015;25: 1008–1017. doi:10.1101/gr.188193.114

98. Kosuri S, Goodman DB, Cambray G, Mutalik VK, Gao Y, Arkin AP, et al. Composability of regulatory sequences controlling transcription and translation in Escherichia coli. Proc Natl Acad Sci U S A. 2013;110: 14024–14029. doi:10.1073/pnas.1301301110

99. Mutalik VK, Guimaraes JC, Cambray G, Lam C, Christoffersen MJ, Mai Q-A, et al. Precise and reliable gene expression via standard transcription and translation initiation elements. Nat Methods. 2013;10: 354–360. doi:10.1038/nmeth.2404

100. Mutalik VK, Guimaraes JC, Cambray G, Mai Q-A, Christoffersen MJ, Martin L, et al. Quantitative estimation of activity and quality for collections of functional genetic elements. Nat Methods. 2013;10: 347–353. doi:10.1038/nmeth.2403

101. Yamanishi M, Ito Y, Kintaka R, Imamura C, Katahira S, Ikeuchi A, et al. A genome-wide activity assessment of terminator regions in Saccharomyces cerevisiae provides a terminatome''toolbox. ACS Synth Biol. 2013;2: 337–347.

102. Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, et al. Bidirectional promoters generate pervasive transcription in yeast. Nature. 2009;457: 1033–1037. doi:10.1038/nature07728

103. Rhee HS, Pugh BF. Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature. 2012;483: 295–301. doi:10.1038/nature10799

104. Dion MF, Kaplan T, Kim M, Buratowski S, Friedman N, Rando OJ. Dynamics of replication-independent histone turnover in budding yeast. Science. 2007;315: 1405–1408. doi:10.1126/science.1134053

105. Brogaard K, Xi L, Wang J-P, Widom J. A map of nucleosome positions in yeast at base-pair resolution. Nature. 2012;486: 496–501. doi:10.1038/nature11142

106. Rhee HS, Bataille AR, Zhang L, Pugh BF. Subnucleosomal structures and nucleosome asymmetry across a genome. Cell. 2014;159: 1377–1388. doi:10.1016/j.cell.2014.10.054

107. Hamdani O, Dhillon N, Hsieh T-HS, Fujita T, Ocampo J, Kirkland JG, et al. tRNA Genes Affect Chromosome Structure and Function via Local Effects. Mol Cell Biol. 2019;39: 8. doi:10.1128/MCB.00432-18

108. Lee ME, DeLoache WC, Cervantes B, Dueber JE. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. ACS Synth Biol. 2015;4: 975–986. doi:10.1021/sb500366v

109. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. 2008;320: 1344–1349. doi:10.1126/science.1158441

110. Townshend B, Kennedy AB, Xiang JS, Smolke CD. High-throughput cellular RNA device engineering. Nat Methods. 2015;12: 989–994. doi:10.1038/nmeth.3486

111. Lubliner S, Keren L, Segal E. Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. Nucleic Acids Res. 2013;41: 5569–5581. doi:10.1093/nar/gkt256

112. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, et al. Noise in protein expression scales with natural protein abundance. Nat Genet. 2006;38: 636–643. doi:10.1038/ng1807

113. Hornung G, Bar-Ziv R, Rosin D, Tokuriki N, Tawfik DS, Oren M, et al. Noise-mean relationship in mutated promoters. Genome Res. 2012;22: 2409–2417. doi:10.1101/gr.139378.112

114. Jones DL, Brewster RC, Phillips R. Promoter architecture dictates cell-to-cell variability in gene expression. Science. 2014;346: 1533–1536. doi:10.1126/science.1255301

115. Sanchez A, Golding I. Genetic determinants and cellular constraints in noisy gene expression. Science. 2013;342: 1188–1193. doi:10.1126/science.1242975

116. Sanchez A, Choubey S, Kondev J. Regulation of noise in gene expression. Annu Rev Biophys. 2013;42: 469–491. doi:10.1146/annurev-biophys-083012-130401

117. Fermi B, Bosio MC, Dieci G. Promoter architecture and transcriptional regulation of Abf1-dependent ribosomal protein genes in Saccharomyces cerevisiae. Nucleic Acids Res. 2016;44: 6113–6126. doi:10.1093/nar/gkw194

118. Reja R, Vinayachandran V, Ghosh S, Pugh BF. Molecular mechanisms of ribosomal protein gene coregulation. Genes Dev. 2015;29: 1942–1954. doi:10.1101/gad.268896.115

119. Knight B, Kubik S, Ghosh B, Bruzzone MJ, Geertz M, Martin V, et al. Two distinct promoter architectures centered on dynamic nucleosomes control ribosomal protein gene transcription. Genes Dev. 2014;28: 1695–1709. doi:10.1101/gad.244434.114

120. Rossi MJ, Lai WKM, Pugh BF. Genome-wide determinants of sequence-specific DNA binding of general regulatory factors. Genome Res. 2018;28: 497–508. doi:10.1101/gr.229518.117

121. Vo Ngoc L, Wang Y-L, Kassavetis GA, Kadonaga JT. The punctilious RNA polymerase II core promoter. Genes Dev. 2017;31: 1289–1301. doi:10.1101/gad.303149.117

122. Raser JM, O'Shea EK. Control of Stochasticity in Eukaryotic Gene Expression. Science. 2004;304: 1811–1814. doi:10.1126/science.1098641

123. Zhang Z, Dietrich FS. Mapping of transcription start sites in Saccharomyces cerevisiae using 5' SAGE. Nucleic Acids Res. 2005;33: 2838–2851. doi:10.1093/nar/gki583

124. Endy D. Foundations for engineering biology. Nature. 2005;438: 449–453. doi:10.1038/nature04342

125. Blazeck J, Garg R, Reed B, Alper HS. Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. Biotechnol Bioeng. 2012;109: 2884–2895. doi:10.1002/bit.24552

126. Rajkumar AS, Liu G, Bergenholm D, Arsovska D, Kristensen M, Nielsen J, et al. Engineering of synthetic, stressresponsive yeast promoters. Nucleic Acids Res. 2016;44: 17.

127. Korber P, Barbaric S. The yeast PHO5 promoter: from single locus to systems biology of a paradigm for gene regulation through chromatin. Nucleic Acids Res. 2014;42: 10888–10902. doi:10.1093/nar/gku784

128. To-E A, Ueda Y, Kakimoto SI, Oshima Y. Isolation and characterization of acid phosphatase mutants in Saccharomyces cerevisiae. J Bacteriol. 1973;113: 727–738. Available: http://www.ncbi.nlm.nih.gov/pubmed/4570606

129. Oshima Y. Regulatory Circuits for Gene Expression: The Metabolism of Galactose and Phosphate. Cold Spring Harbor Monograph Archive. 1982;11B: 159–180. doi:10.1101/087969180.11B.159

130. Barbaric S, Münsterkötter M, Goding C, Hörz W. Cooperative Pho2-Pho4 interactions at the PHO5 promoter are critical for binding of Pho4 to UASp1 and for efficient transactivation by Pho4 at UASp2. Mol Cell Biol. 1998;18: 2629–2639. Available: http://www.ncbi.nlm.nih.gov/pubmed/9566882

131. Lenburg ME, O'Shea EK. Signaling phosphate starvation. Trends Biochem Sci. 1996;21: 383–387. doi:10.1016/S0968-0004(96)10048-7

132. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;297: 1183–1186. doi:10.1126/science.1070919

133. Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci U S A. 2002;99: 12795–12800. doi:10.1073/pnas.162041399

134. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet. 2005;6: 451–464. doi:10.1038/nrg1615

135. Zenklusen D, Larson DR, Singer RH. Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol. 2008;15: 1263–1271. doi:10.1038/nsmb.1514

136. Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. Science. 2011;332: 472–474. doi:10.1126/science.1198817

137. Lionnet T, Singer RH. Transcription goes digital. EMBO Rep. 2012;13: 313–321. doi:10.1038/embor.2012.31

138. Larson DR, Singer RH, Zenklusen D. A single molecule view of gene expression. Trends Cell Biol. 2009;19: 630–637. doi:10.1016/j.tcb.2009.08.008

139. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell. 2005;123: 1025–1036. doi:10.1016/j.cell.2005.09.031

140. Peccoud J, Ycart B. Markovian Modeling of Gene-Product Synthesis. Theor Popul Biol. 1995;48: 222–234. doi:10.1006/tpbi.1995.1027

141. Kepler TB, Elston TC. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophys J. 2001;81: 3116–3136. doi:10.1016/S0006-3495(01)75949-8

142. Zawadzki K, Broach J. A rapid technique for the visualization of live immobilized yeast cells. J Vis Exp. 2006; 84. doi:10.3791/84

143. Skinner SO, Sepúlveda LA, Xu H, Golding I. Measuring mRNA copy number in individual Escherichia coli cells using single-molecule fluorescent in situ hybridization. Nat Protoc. 2013;8: 1100–1113. doi:10.1038/nprot.2013.066

144. Trcek T, Chao JA, Larson DR, Park HY, Zenklusen D, Shenoy SM, et al. Single-mRNA counting using fluorescent in situ hybridization in budding yeast. Nat Protoc. 2012;7: 408–419. doi:10.1038/nprot.2011.451

145. Thompson RE, Larson DR, Webb WW. Precise nanometer localization analysis for individual fluorescent probes. Biophys J. 2002;82: 2775–2783. doi:10.1016/S0006-3495(02)75618-X

146. Wu Z, Huang NE, Long SR, Peng C-K. On the trend, detrending, and variability of nonlinear and nonstationary time series. Proc Natl Acad Sci U S A. 2007;104: 14889–14894. doi:10.1073/pnas.0701020104

147. Truong C, Oudre L, Vayatis N. Selective review of offline change point detection methods. Signal Processing. 2020;167: 107299. doi:10.1016/j.sigpro.2019.107299

148. Coulon A, Larson DR. Fluctuation Analysis: Dissecting Transcriptional Kinetics with Signal Theory. Methods Enzymol. 2016;572: 159–191. doi:10.1016/bs.mie.2016.03.017

149. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 2006;4: e309. doi:10.1371/journal.pbio.0040309

150. Shahrezaei V, Swain PS. Analytical distributions for stochastic gene expression. Proc Natl Acad Sci U S A. 2008;105: 17256–17261. doi:10.1073/pnas.0803850105

151. Hoffman EA, Zaidi H, Shetty SJ, Bekiranov S, Auble DT. An Improved Method for Measuring Chromatin-binding Dynamics Using Time-dependent Formaldehyde Crosslinking. Bio Protoc. 2018;8. doi:10.21769/bioprotoc.2905

152. Babl V, Stöckl U, Tschochner H, Milkereit P, Griesenbeck J. Chromatin Endogenous Cleavage (ChEC) as a Method to Quantify Protein Interaction with Genomic DNA in Saccharomyces cerevisiae. Methods Mol Biol. 2015;1334: 219–232. doi:10.1007/978-1-4939-2877-4_14

153. Brown CR, Eskin JA, Hamperl S, Griesenbeck J, Jurica MS, Boeger H. Chromatin structure analysis of single gene molecules by psoralen cross-linking and electron microscopy. Methods Mol Biol. 2015;1228: 93–121. doi:10.1007/978-1-4939-1680-1_9

154. Gunawardena J. A linear framework for time-scale separation in nonlinear biochemical systems. PLoS One. 2012;7: e36321. doi:10.1371/journal.pone.0036321

155. Mirzaev I, Gunawardena J. Laplacian dynamics on general graphs. Bull Math Biol. 2013;75: 2118–2149. doi:10.1007/s11538-013-9884-8

156. Kelly FP. Reversibility and stochastic networks. Cambridge University Press; 2011. Available: https://www.cambridge.org/gb/academic/subjects/statistics-probability/applied-probability-and-stochastic-networks/reversibility-and-stochastic-networks

157. Cinlar E. Introduction to Stochastic Processes. Courier Corporation; 2013. Available: https://play.google.com/store/books/details?id=uq8AAQAAQBAJ