Causes and Consequences of Gene Expression Noise

Thesis by

Chiraj K. Dalal

In Partial Fulfillment of the Requirements for the

degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2010

(Defended July 21, 2009)

ACKNOWLEDGEMENTS

One of the things I have never understood about life as a graduate student was the graduation itself. You're supposed to stand at a podium and present your results, defending that the work you have done is worthy of a degree. The part I don't understand is the practice of a single person presenting their own results. I haven't been doing lab work very long, but as far as I can tell, it is very rare for a single person to generate their own results. Scientific ideas are the result of interaction, discussion and inspiration, usually divided up into several small conversations with a wide variety of people including oneself. Experiments are usually done after someone has shown you how to do bench work, how to use equipment, and how to analyze data. The process is not easy and it takes a lot of time, but more importantly, it takes a lot of help. The most important ingredient then, of course, is to work in an environment which is conducive to discussion and education.

I was incredibly lucky to work in a lab where such an environment was fostered and I think all the credit for that must go to my advisor, Michael Elowitz, and my labmates throughout the years for their innumerable implicit contributions to this work. I want to start by thanking what I refer to as the 1st generation labowitzers, Jonathan Young, Sidney Cox and Gurol Suel for helping to set the tone and culture of the laboratory. Ensuing members of the lab, David Sprinzak, Avigdor Eldar, Joe Levine, Shaunak Sen, Randall Wald, Michelle Fontes & Georg Seelig, added diversity to the lab, slowly making it more multilingual and multicellular. As the lab grew, so did our competence in several biological systems, and having members such as Fred Tan, Lauren LeBon, John Yong, Leah Santat, Julia Tischler, Long Cai, James Locke, Rajan Kulkarni, Murat Acar and Amit Lakhanpal to bounce ideas off of was incredibly useful. In addition to the educational and open environment, the lab was also a hub of fun and time-wasting activities. I think a special note should be made to Joe Levine and Jon Young for propagating such activities. In a similar vein, I think it's only fair to thank our neighboring labs, especially the Pierce lab for their patience and tolerance.

Additionally, I would also like to specifically thank members of other labs and facilities at Caltech for all their help in helping me get certain projects started. Most notably, I would like to thank Rochelle Diamond of the flow cytometry facility for teaching me flow cytometry, Robert Oania of the Deshaies lab for helping getting started with yeast genetics, Allison Ross for making sure all my paperwork was up to date and Ernesto Mercado, for nourishment. Of course, I would also like to thank my thesis committee, Professors Scott Fraser, Paul Sternberg and Barbara Wold, for generously spending time with me, and giving me constructive criticism about both science and graduate school itself.

Of course, most of the work described was part of a collaboration. Dr. David Sprinzak helped quite a bit with the analytic model described in chapter 1. Dr. Long Cai was my partner in crime for the works described in Chapters 3 and 4. Kasra Rahbar, a SURF student, also helped with the work described in Chapter 4 and Appendix C. It is absolutely critical for me to thank all of them for their hard work that contributed towards the described work.

Yet, despite having a strong support system in lab, on campus and amongst collaborators, graduate school is still hard because no matter how many experiments you do, it is virtually guaranteed that 73% of them will utterly fail. In this case, it is nice to have a support system outside of the lab which enables sanity. For that, I have to thanks my family, especially my parents and sister Radha, all of whom are local, as well as various friends, including Ming Gu, Steven Kuntz, Michael Krout, Ryan Zeidan, and Habib Ahmad, for their helps in keeping me engaged in other disciplines, especially in sports, where I was a member of two IM sports teams on campus, Playmakers, in football, and How we Roll, in basketball. Both activities were absolutely vital in getting me out of the lab, a critical ingredient to getting me to this stage.

Finally, I'd like to thank Pa Ra Ma, a nickname for someone who is incredibly dear and important to me, for whom I have no words to describe, to whom I owe everything.

ABSTRACT

Genetically identical cells harvested in the same environment exhibit heterogeneity in gene expression. This phenomenon, termed gene expression noise, has been measured in several model organisms under various conditions. However, we still do not have a clear understanding of (1) the factors responsible for generating gene expression noise, or (2) the potential consequences noise can have on cellular processes. In an attempt to investigate these issues, we have determined the effects of 1) directional selection, 2) promoter mutation and 3) fluctuations in transcription factor localization on gene expression noise. First, we have used analytic and computational modeling of the effects of directional selection on gene expression noise to discover that, assuming expression can be described by up to two independent parameters, μ, mean, and σ, noise, strong directional selection yields an increase in noise. Next, we generated mutant promoter libraries and measured gene expression to determine the effects of *cis*-regulatory mutations on gene expression noise. Here we found that the expression noise can indeed be modulated by mutation independent of mean expression levels, lending credence to the previously mentioned analytical result. Based on this result, that mutations can harness noise, we wanted to determine whether the binding and unbinding of transcription factors to promoter regions also contributed to gene expression noise. To do so, we analyzed the localization dynamics of a transcription factor Crz1. We determined that Crz1 translocates to the nucleus in coherent bursts of localization in response to calcium. The frequency, but not the duration, of these bursts increases with the concentration of extracellular calcium. This frequency modulation propagates downstream of Crz1, enabling proportional regulation of target genes. Intrigued by this result, we characterized different types of localization dynamics used by the yeast proteome. We have found several classes of localization behavior, including proteins that burst on several timescales, exhibit static heterogeneity, and show amplitude modulation. Strikingly, several of these dynamic localization systems must co-exist in the same cell under the same conditions. Amongst the proteins that burst on a fast timescale like Crz1, Msn2 and Mig1 are transcription factors that both burst when deprived of glucose. Furthermore, both regulate a common set of target genes. Interestingly, when

imaged together, the proteins exhibit correlations on two timescales, a positive correlation that typically lasts for an hour and an anti-correlation that lasts a few minutes. We are continuing to investigate the potential regulatory impact of these correlations by measuring the expression of their combinatorial target genes.

TABLE OF CONTENTS

NOMENCLATURE

Amplitude Modulation (AM): A technique used in electronic communication to transmit information through a radio carrier wave. Here the strength of the transmitted signal varies in relation to the information being sent. We have adapted this term to nuclear localization to describe a phenomenon in which the nuclear intensity of a protein varies in response to a biological signal.

Directional Selection: Selective pressure favoring the increasing (or decreasing) values of a quantitative trait- causing allele frequency to continue to shift in one direction.

Frequency Modulation (FM): A technique used in electronic communication in which information is conveyed over a carrier wave by varying its frequency. Note that this in contrast to amplitude modulation in which the frequency is constant. We have adapted this term to nuclear localization to describe a phenomenon in which the frequency of nuclear localization bursts varies in response to a biological signal.

Gene Expression Noise: Copy number variability in mRNA or proteins across a clonal population of cells.

Green Fluorescent Protein (GFP): A fluorescent marker frequently used in molecular biology. Spectral variants in several different colors also exist.

Flow Cytometry: A technique to count and characterize (measure gene expression of) microscopic particles (including yeast cells) suspended in liquid

Promoter: A region of DNA that activates transcription of a particular gene, its typically located upstream of the gene.

Pulse-Width Modulation (PWM): A widely employed technique in electronics in which an analog signal is controlled digitally. Here the duty cycle, or width, of a square wave (an all-or none phenomenon) is modulated to encode a specific analog level. This phenomenon has been co-opted to localization dynamics. In this case, the all or none phenomenon is a burst and its duty cycle, or width, is the duration of the burst.

Static Heterogeneity: When cells burst, they can do so independently of other cells around them. As a result, only a small percentage of cells burst at any given time; this dynamic phenomenon creates creating population heterogeneity. However, it is also possible to create such heterogeneity without any dynamics. In this case, cells respond with nuclear localization and that localization remains constant over time, but only a fraction of cells respond as several cells remain cytoplasmic.

I N T R O D U C T I O N

One of the most striking aspects of biology is in the diversity of cellular life. Although much of this variability has been attributed to genetic and environmental factors, recent studies have shown4 that genetically identical organisms in the same environment exhibit heterogeneity in gene expression. This phenomenon, gene expression noise, has been observed and measured across species as divergent as prokaryotes and mammalian cells.[1-4]

We have been keenly interested in understanding more about the origin of this heterogeneity as well as wondering what potential functional consequences it may have. We have begun addressing these questions with a suite of experiments described in the following chapters.

In Chapter 1, we examine the effects selection may have on gene expression noise *in silico*. We find that under fairly general conditions, directional selection for the value of a quantitative phenotypic trait can yield an increase in the noise in that trait.

In Chapter 2, we begin to look experimentally at the effects *cis*-regulatory mutations have on gene expression noise. Through the creation of promoter mutant libraries, we look for mutations that change noise; furthermore, by also looking at how these same mutations affect mean levels of gene expression noise, we test the independence of these two parameters. By measuring mean and noise levels of gene expression in three mutant yeast promoter libraries, we see that the two parameters are independent in some cases. More importantly, we see a wide variety of mean and noise values, suggesting that *cis*-regulatory mutations can control gene expression noise independently of mean.

Upon finding out that specific promoter mutations can impact gene expression noise, we were interested in finding out if the binding and unbinding of transcription factors to promoter regions had a similar impact. In Chapter 3, we observe the localization dynamics of a calcium-responsive yeast transcription factor, Crz1. We find that it localizes to the nucleus in short coherent bursts in response to calcium. The frequency, but not duration, of Crz1 localization bursts increases with extracellular calcium. This frequency-modulated nature of the bursts enables proportional control of target genes. Interestingly, these localization bursts also lead to transcriptional bursts of target genes, leading to noisy

gene expression. This is especially intriguing because it suggests a trade-off of sorts. The bursts lead to noisy gene expression but they also lead to proportional regulation. Hence the localization dynamics of a yeast transcription factor provide a mechanism to increase gene expression noise and provide proportional control of a regulon of target genes.

Since the strategy of looking at protein localization dynamics had yielded such interesting results, we decided to take a more comprehensive look at protein dynamics in yeast. In Chapter 4, we describe a screen in which we take movies of all available tagged proteins in yeast. We find that there are a few other fast bursting proteins, like Crz1. However, several other types of localization behavior exist, including slow bursts, amplitude modulation, and static heterogeneity. Amongst the other fast bursting proteins, we have found two transcription factors, Msn2 and Mig1, which burst in response to glucose deprivation. When observing both proteins in the same cell, it is clear that their bursts are not completely independent, as seen in the cross-correlation function. We hope to use chemical perturbations to track down this lack of independence and measure expression levels of combinatorial target genes, hoping it will help uncover mechanistic insights into the observed fast bursts and their downstream consequences.

To summarize, we have looked at different potential mechanisms for generating gene expression noise, namely selection, *cis*-regulatory mutations, and the localization of *trans*-acting factors. All appear as if they could contribute to gene expression noise; most notably, the bursting of transcription factors in and out of the nucleus, in addition to leading to increased amounts of gene expression noise, also yields proportional control of a suite of genes. These results also provide a mechanism for proportional control, a phenomenon which may be important in several contexts and may even provide a rationale for some of the gene expression noise that we observe. We anticipate that further investigation of the functional consequences of noise will continue to provide insights into cellular behavior.

Chapter 1

EFFECTS OF EVOLUTIONARY SELECTION ON NOISE

Motivation

Many biological traits are quantitative: levels of gene expression, sizes of appendages, and abundances of cellular components can vary over a wide range. Mean values of such phenotypes are generally genetically encoded; therefore, they are subject to the forces of selection. Recently, it has become clear that such phenotypes are fundamentally noisy: that is, genes specify a distribution of possible values for the trait, rather than a precise value. More significantly, the variance of this distribution, like the mean, is under genetic control,[5] and the variance, at least in stress response genes, may be genetically independent of mean.[1,3] This data suggests that both the mean and variance of a quantitative phenotype are influenced independently by selective pressures that act on the phenotypes expressed in individual cells. However, an individual in a population is subject to selection only on its particular phenotypic value, not on the mean and variance that specify its phenotypic distribution. Here we ask how positive directional selection affects variance in a simple quantitative trait.

A notable feature of this question is that there are only three possible outcomes. The variance can increase, decrease, or stay the same. Conceptually, one might expect that as you continue to select for a specific phenotype, you select against other phenotypes, decreasing the variance in the population[6] (Fig.1A). Conversely, one can imagine a scenario in which selection on the tail end of a distribution can yield an extremely heterogeneous population. Here the initial population is very unlikely to pass the selection threshold unless there is a large variance (Fig. 1B). Finally, it is possible that as the selection increases the mean, the phenotypic variance remains the same (Fig. 1C). To distinguish between possibilities, we conducted an *in silico* experiment.

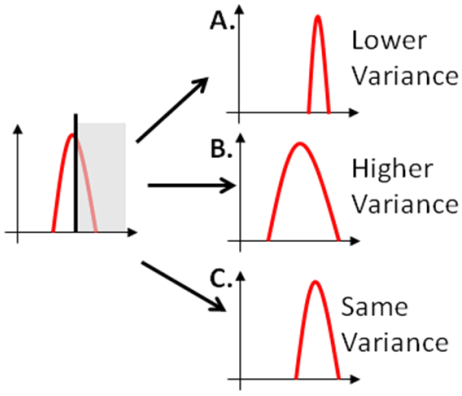


Fig. 1. **How does Selection Affect Variance?** There are three possibilities: Variance can decrease (A), increase (B), or stay the same (C), in response to directional selection.

***In Silico* Result**

We considered a phenotype *x*, whose distribution is a Gaussian, parameterized by two genetically determined independent values (μ, σ) (other distributions, including log-normal and gamma, work as well):

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Here, μ and σ represent the mean phenotypic level, and variance, respectively. Either can vary under mutation. We asked how the population means of these parameters, <μ> and <σ>, change under rounds of mutation, selection, and growth (Fig. 2B). The number of cells, or potentially organisms, was kept large to minimize any effects of genetic drift. The initial population was left homogeneous, i.e. both genetic traits were identical in all cells, for simplicity. During mutation, the σ and μ for a given cell are assigned new values from a distribution centered on their previous values. During selection, we assume that only a certain percentage of cells with the highest phenotypes can survive. For example, under tight selection, the threshold may be chosen to allow only the highest phenotypes, e.g., the top 5%, to survive, while weak selection permits the top 55% to survive. After selection, surviving cells replicate to restore the original population size, completing one cycle of directional selection (Fig. 2B).

The result, shown in Fig. 2C, indicates that tight selection favors increasing variance while weak selection favors decreased variance. This result can be understood qualitatively by considering two individuals with different values of μ and σ, as shown in

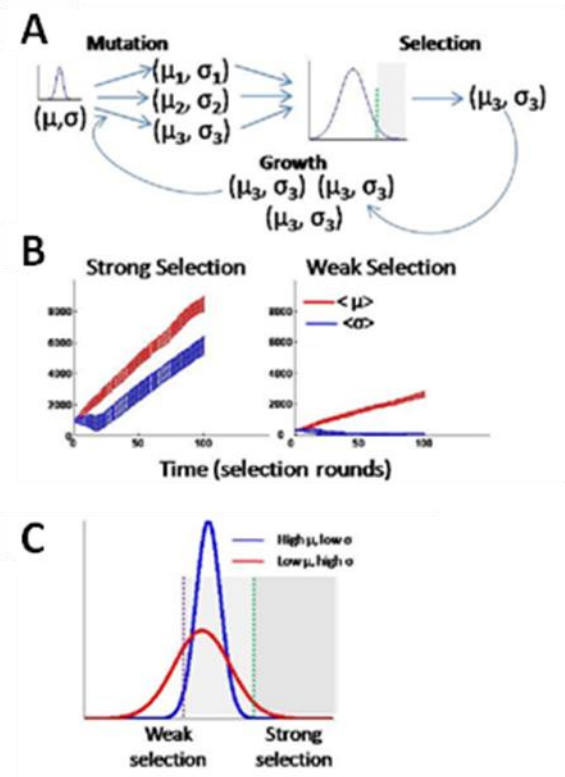


Fig. 2. Phenotypic Selection
A) A flow-chart of the *in silico* experiment. An isogenic population (μ,σ) is mutated so that now each cell has a different genotype (μ_n,σ_n), giving a broader distribution of expression levels than the original population. A threshold selection is imposed (green dashed line). Cells are grown to the original population size. The selected cells are then mutated again and the cycle continues. B) In a simulation, 20,000 isogenic cells with a mean of 1000 and a variance of 300 have been mutated and selected such that the top 5% (left) or 55% (right) of phenotypes are allowed to survive for 100 selection rounds. Strong selection favors increasing $<\sigma>$ (red lines) while weak selection favors decreasing $<\sigma>$ (blue lines). $<\mu>$ increases in both scenarios. Error bars indicate one standard deviation over the population. C) Two individual distributions, one with a high mean and low variance (blue) and one with a lower mean and higher variance (red) are shown. If strong selection, (cells with the top 5% of all phenotypes are selected, (green dashed line)) is imposed on both distributions, the red distribution will be favored. However, if weak selection (the top 55% of all cells will be selected, shown with the magenta dashed line) is imposed, the red distribution will be favored.

Fig. 2C. The tighter selection (dotted green line) favors the longer "tailed," or high variance individual, despite its lower mean level. Meanwhile, the weaker selection (dotted purple line) favors the individual with the higher mean and lower variance.

This finding has several insights. As expected, the mean phenotype, $<\mu>$, always increases under directional selection, irrespective of the strength of selection, and the rate of increase is proportional to the mutability, the amount that the mean can increase in each round of mutation. More interestingly, $<\sigma>$ behaves differently depending on the strength

of selection. When more than half the cells survive selection, variance increases, but when less than half survive, phenotypic variance decreases to a basal level.

In summary, assuming that a phenotypic distribution is the product of two independent genes, computational analysis predicts that directional selection in which less than half of all cells survive, yields an increase in phenotypic variance.

Analytic Model

Although the intuitive diagram shown in Fig 2c explains the aforementioned *in silico* result, it does not explain the precise effect of selection strength. Why is it that selection pressures in which <50% of cells yield an increase in variance while those in which >50% of cells yield a decrease in variance? What is so special about 50%? We constructed an analytic model hoping to resolve this question.

We assumed the quantitative phenotype (x) controlled by two independent genetic traits, has a phenotypic response is given by (μ,σ). Hence, for a given value of genetic traits the phenotype distribution is given by:

$$(1)\quad P(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The genotype distribution function $\rho_t(\mu,\sigma)$ is defined as the probability to find the genotype (μ,σ) within the population at generation *t*. (In more formal terms, the probability is given by $\rho_t(\mu,\sigma)d\mu d\sigma$).

The goal here is to obtain an expression for $\rho_t(\mu,\sigma)$ after many generations of mutation, selection, and growth. More specifically, we are interested in what happens to $\langle\sigma\rangle$ and $\langle\mu\rangle$, defined as the mean and variance of the phenotype, respectively, after many generations under the influence of different types of selection.

In order to do that, we need to write the dependence of $\rho_{t+1}(\mu,\sigma)$ on $\rho_t(\mu,\sigma)$. The process which occurs in each generation is as follows:

(2) $\rho_t(\mu,\sigma) \to \rho_t^m(\mu,\sigma) \to \rho_{t+1}(\mu,\sigma)$,

where $\rho_t^m(\mu,\sigma)$ is the distribution after the mutation phase. The first arrow in (2) corresponds to the mutation phase and the second arrow corresponds to the selection phase.

Mutation phase

In the mutation phase, we allowed the values of μ and σ of each genotype to change a little. We can define a mutation function:

(3) $M(\mu,\sigma;\Delta\mu,\Delta\sigma) = \frac{1}{\sqrt{2\pi}\Delta\mu} e^{-\frac{\mu^2}{\Delta\mu^2}} \frac{1}{\sqrt{2\pi}\Delta\sigma} e^{-\frac{\sigma^2}{\Delta\sigma^2}}$,

where $\Delta\mu$ and $\Delta\sigma$ are fixed parameters that correspond to mutation ranges in μ and σ in each generation.

To obtain $\rho_t^m(\mu,\sigma)$, one needs to convolve $\rho_t(\mu,\sigma)$ with the mutation function:

(4) $\rho_t^m(\mu,\sigma) = \int \rho_t(\mu-\mu',\sigma-\sigma') M(\mu',\sigma',\Delta\mu,\Delta\sigma) d\mu' d\sigma'$.

(Please note that one can also write the continuous version of that with a free diffusion equation.) The new genotype distribution is broadened by the convolution with mutation.

Selection Phase

In the selection stage, we essentially look at the total phenotypic distribution and apply a phenotypic selection function S(x) to select for the surviving population. So:

(5) $\rho_{t+1}(\mu,\sigma) = \rho_t^m(\mu,\sigma)\int_x P(x;\mu,\sigma)S(x;\mu,\sigma)dx$,

Let's assume, for example, that we select by introducing a threshold T which is defined such that only a fixed fraction of the population ε_T remains after the selection. Equation (5) then becomes

(6) $\rho_{t+1}(\mu,\sigma) = \rho_t^m(\mu,\sigma)\frac{1}{\varepsilon_T}\int_{x>T} P(x;\mu,\sigma)dx = \rho_t^m(\mu,\sigma)\frac{1}{2\varepsilon_T}erfc\left(\frac{T-\mu}{\sqrt{2}\sigma}\right)$.

Note that the term on the right hand side was normalized by ε_T. This corresponds to letting the population grow back to its original size after selection. The threshold T is defined by requirement that $\rho_{t+1}(\mu,\sigma)$ should be normalized, i.e,

(7) $\int \rho_t^m(\mu,\sigma)\frac{1}{2\varepsilon_T}erfc\left(\frac{T-\mu}{\sqrt{2}\sigma}\right)d\mu d\sigma = 1$.

The relation between selection rounds

Substituting Eq. (4) into Eq. (6) provides the required relation between two consecutive generations:

(8) $\rho_{t+1}(\mu,\sigma) = \frac{1}{2\varepsilon_T}erfc\left(\frac{T-\mu}{\sqrt{2}\sigma}\right)\left[\int \rho_t(\mu-\mu',\sigma-\sigma')M(\mu',\sigma',\Delta\mu,\Delta\sigma)d\mu' d\sigma'\right]$.

The right hand side is just a mutated distribution (in square brackets) multiplied by a term we'll describe as the selection function. The function erfc(z) is the complementary error function:

(9) $erfc(z) = \frac{2}{\sqrt{\pi}}\int_z^\infty e^{-t^2}dt$.

The selection function becomes

(10) $erfc\left(\frac{T-\mu}{\sqrt{2}\sigma}\right)$.

From the analysis of this expression, we have learned several important lessons. First, if the variance, or σ, is small and the mean phenotype, μ, is less than the selection threshold, T, there is no chance that individual will be selected. However, if the mean phenotype stays the same and the variance increases, that individual has some chance of being selected. The 50% selection threshold, at least in a Gaussian distribution is the point where the mean and the threshold are equal. Here the selection expression automatically becomes *erfc(0)*= 1. There is no longer any dependence on the variance in the expression.Next, in the cases when selection is stringent (ε_T <<1) and only a few cells pass selection, it is typical for most of the mean phenotypes to be less than the threshold value. Here it appears that the optimal strategy to pass selection is to have a large variance. In fact, the high variance might even act as an insurance policy, enabling cells to withstand mutations that lower their means and still pass selection. Finally, when selection is lenient, (ε_T >0.5) individuals with higher means and lower variances outcompete those with higher variances. High variance populations also have some very low phenotypic values far below the threshold, making them less fit than their low variance counterparts.

Discussion

This result is general; it applies to any quantitative phenotype under selection. For many selection experiments, the effects of variance generally have not been measured. However, for those experiments, in which distributions are plotted, it seems as if the variance does increase. For example, in 1957 Clayton and Robertson selected for increased and decreased abdominal bristle number in *Drosophila melanogaster*.[7] The resulting phenotypic distributions from their selection appear to be broader than their initial population (Figure 3).

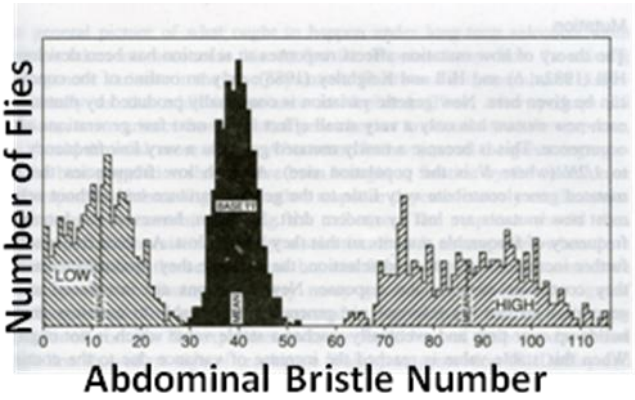


Fig. 3. **A Selection Example.** Clayton and Robertson selected for high and low abdominal bristle number in flies. They showed that flies under directional selection, exhibited higher variance than the base population, consistent with the proposed computational model.

Additionally, Peter and Rosemary Grant have been measuring several quantitative phenotypes in Darwin's finches on the Galapagos Islands for almost 40 years. They have found that after one drought, during a single monsoon season, the beak depth of the finches increases. After just one generation, they found that the distribution of beak depths broadened.[8]

In a long-term evolution experiment with *E. coli*, Richard Lenski and colleagues found that while only a few mutations evolved at first, 6 of 12 replicate cultures independently mutated a DNA replicase creating hypermutable strains.[9] Due to selection pressures, this hypermutable mutation was only allowed to survive because it had the same fitness, or mean phenotype, as non-hypermutable strains. However, in half of the cell lines evolution selected for this increased ability to mutate, or higher variance in fitness.

Although these long-term evolution experiments are incredibly complicated because of the variety of selection pressures, the fact that we still see variance increasing in response to directional selection reiterates the potentially generality of our result. Furthermore, it shows that in the examination of how selective pressures affect phenotypes, it is important to consider how both the mean and variance of a given phenotype are affected.

Future Directions

As discussed in the introduction, our interest lies in gene expression noise and its potential causes and consequences. Using gene expression as a quantitative phenotype, we can apply our computational and analytic results towards noise. Specifically, we predict directional selection may provide a mechanism that can explain high levels of biological fluctuation. This seems even more plausible in lieu of recent computational[10] and experimental work.[11] Most notably, Kaneko and colleagues used stringent selection (top 0.1%) of a specific GFP to increase expression noise.[12]

Furthermore, we have designed a forward experiment to verify the strong evolutionary prediction: strong and weak directional selection for high expression level will select for high or low noise, respectively. One can take two clonal populations of cells with overlapping gene expression distributions and artificially impose different types of selection using flow cytometry. After several rounds of selection and re-growth, the library should be enriched for the clone with low or high noise in the weak or strong threshold cases, respectively (Fig 4). The only potential drawback of such an experiment is the inability to account for other types of selection that could potentially interfere with the resulting data. For example, before and after every flow cytometry-imposed selection, there would be ample time for the populations to grow. During this interim period, when both clonal populations of cells would have to be cultured together, any small growth differences between the populations could have a tremendous impact on the experiment. To account for such errors, one could conduct a similar experiment without the flow cytometry-based selection. Ideally, in this experiment, both clonal populations would be tagged with a specific fluorophore, one with a CFP and one with a cherry. This way, one could use a plate reader or flow cytometer to measure the fraction of cells that are labeled with each color and determine the impact, if any, growth had on the ratios of the two populations. If there was an impact, its measured value could be used to deconvolve its effects from the flow cytometry imposed selection. Combined, both experiments could be used to measure the effect of directional selection on phenotypic distributions and, specifically, on phenotypic variance. We anticipate that such selection experiments can

yield insight into the role different selective pressures play in causing phenotypic variance in gene expression and other quantitative phenotypes.

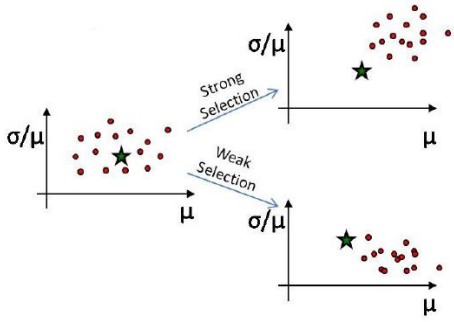


Fig. 4. How selection affects phenotypes.
We propose to impose artificial selection on gene expression via flow cytometry on pooled promoter mutant libraries. If we impose strong selection, we expect both the mean and cv of the pooled population to increase. If we impose weak selection, we expect the mean to increase, but the cv to decrease.

Acknowledgements

The analytic model presented in this Chapter was done with amazing assistance from Dr. David Sprinzak, a postdoctoral fellow in the Elowitz lab.

Chapter 2

EFFECTS OF CIS-REGULATORY MUTATIONS ON NOISE

Background

It is well established that levels of gene expression are quantitative and noisy.[13] Although *cis*-regulatory regions of DNA, called promoters, are known to control mean levels of expression, it is unclear whether these same regions affect gene expression noise. Here we randomly mutate promoter regions of a few genes to determine the impact, if any, they have on gene expression noise.

Gene expression noise has been operationally defined as the standard deviation divided by the mean of expression levels in a population of genetically identical cells grown in the same environment. Furthermore, noise can be subdivided into two components, extrinsic and intrinsic noise. The best way to understand these terms is to think of the case where a single promoter is controlling the expression of two genes, each of which expresses a spectrally distinct color. If a factor were to affect the noise expression of both colors in the same way, due to fluctuations in upstream signaling or cell cycle times, it is extrinsic. However, if the factor were intrinsic, it would only affect the expression of one of the colors. Such factors include promoter kinetics and mRNA turnover.[2,14]

Previous work analyzing gene expression levels in yeast showed that (1) variance is under genetic control[2,5] and (2) that there appears to be a scaling relationship between mean and variance across most yeast proteins.[1,3] More specifically, mean expression levels are approximately inversely proportional to expression variance levels across all genes in yeast. Furthermore, the proportionality constant appears to be the same irrespective of gene or condition in which the gene was measured. Given that the noise in several genes is affected similarly, this suggests that the predominant source of noise is extrinsic and that a common extrinsic factor is responsible for the observed scaling behavior. Interestingly, there are

many proteins, including those involved in stress response, metabolism, and/or chromatin remodeling that do not exhibit this scaling.[1,15]

Motivation

This scaling phenomenon suggests that the abundance of a protein is sufficient information to determine its noise; in other words, both μ and σ are not necessary to determine the distribution of expression of a given gene. However, there are certain subsets of genes that do not scale. Do the promoters of these genes have special properties that enable scaling-independent expression? Is it because mutations in their promoters affect both μ and σ in an uncorrelated fashion? To address this question, we have characterized μ and σ in random promoter mutants of three genes who did not observe the scaling effect.

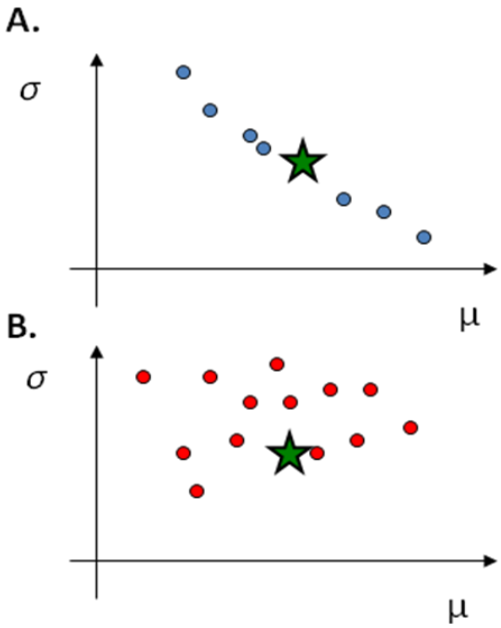


Fig. 1. **Potential Mutagenesis Results.** A) If all mutants in the promoter mutant library exhibit a scaling effect between μ and σ, that means mutations affect both parameters in a correlated fashion. However, if the mutants are uncorrelated as depicted in B) then the parameters are uncorrelated.

From random promoter mutagenesis, we expect two limiting results; in one case all mutant promoters exhibit the scaling effect (Fig 1a) and lie on the same line. This is interesting in that it means promoters are restricted to a single line on the μ-σ axes. This would mean that all mutations affect both parameters in the same way. Moreover, it would imply that gene expression in yeast is determined by one parameter irrespective of whether the gene exhibits the scaling affect or not. Perhaps in the latter case, there are other extrinsic factors affecting gene expression noise, thus, changing the scaling affect, but maintaining the single parameter-dependence on expression.

In the other case, the mutant promoters would be uncorrelated, forming a cloud of points on the same graph (Fig. 1b). This result would suggest that both parameters can be controlled independently, at least in certain mutations in certain promoters. If this were true of promoters that did observe the scaling effect, it would imply that evolution is selecting for promoter architectures in which most noise sources are suppressed. As a result, wild-type promoters would be controlled by a single extrinsic factor, yielding the scaling effect.

Of course, it is also possible to see a mixture of the two limiting cases, in which some mutations appear correlated in both μ and σ while others do not.

Methods/Results

We created 3 promoter-YFP reporter genes in yeast, starting with the relatively noisy genes Tim17, Pir1 and Gal1. These promoter fusions were cloned into low-copy plasmids (YCp) and the entire plasmid was mutagenized by transformation into an *E. coli* mutator strain (XL1-Red, Invitrogen).[16] Mutagenesis was tuned by varying the number of cell cycles the plasmids spent replicating in this cell line. After plasmid purification, the entire plasmid library was transformed into yeast. Single yeast colonies were inoculated and mean and variance of YFP expression was assayed in a flow cytometer. Each dot in Figure 2 represents a clonal population of at least 10,000 cells with the depicted μ and σ. The statistics were calculated as depicted in Appendix B Figure 13. As shown in Fig. 2, the 3 genes show diverse behaviors after 24 hours in the mutator cell line. For Tim17, μ and σ are generally uncorrelated. On the other hand, Pir1 mutants show an anti-correlation. Finally, Gal1 shows two sub-populations with different μ-σ relationships. These results are a combination of the two limiting cases described in Fig. 1. It is noteworthy though, that the fact that some mutations have independent effects on μ and σ provides a proof-of principle for the future experiments proposed in Chapter 1.

A critical part of the model described in chapter 1 is the distribution of mutational effects on mean expression level, μ and noise, σ. Although distributions of mutational effects have been measured systematically for protein coding sequences[17] a corresponding analysis of regulatory mutations has not been reported. This information is necessary both in the context of the evolutionary model and more generally as the basis for understanding how quantitative levels of gene expression can evolve. Furthermore, the data collected here

will allow us to answer which sequence elements contribute to noise and expression level in these promoters?

Additionally, these data show that mutations generally tend to decrease mean expression levels; however, when mutagenesis is tuned to lower levels, it is possible to increase mean expression levels.

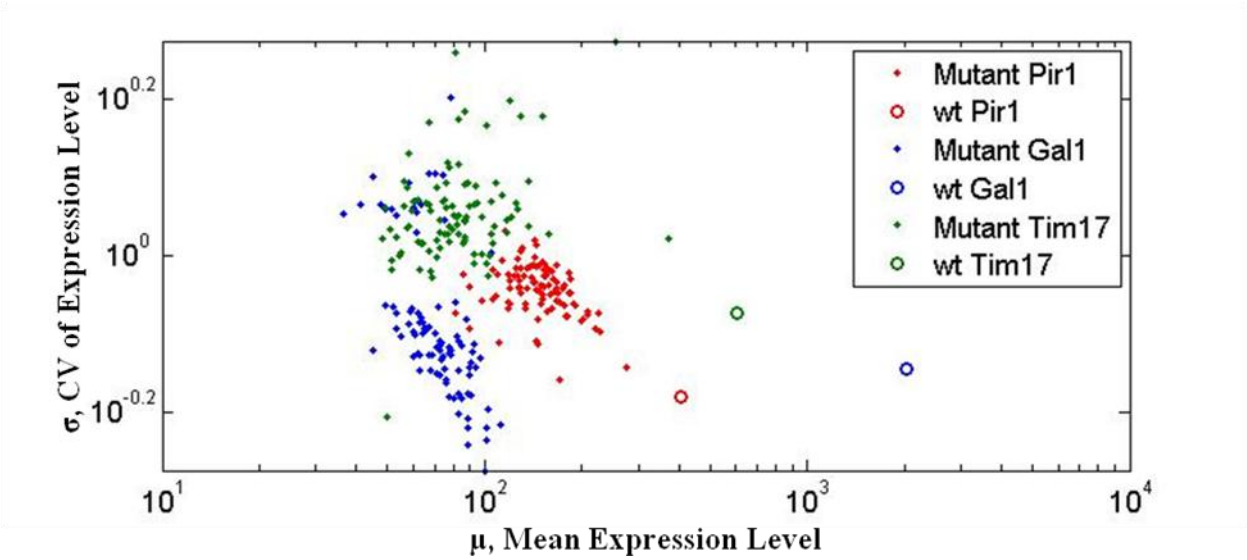


Fig. 2. **Promoter Mutagenesis Results.** After 24 hours of growth in a mutagenesis cell line, three promoter mutant libraries show some intriguing results. Mutants of Pir1 seem to scale while mutants of Tim17 are fairly uncorrelated. Meanwhile, Gal1 mutants fall into 2 categories, those that are uncorrelated and noisy and those that scale and are less noisy.

The mutational load depicted in Figure 2 was chosen for the large change in expression levels and noise. This was not the case at lower levels of mutagenesis (Figure 3) where most mutant promoters, grown in the mutator strain for 12 hours, were very similar to the wild-type promoter. Generation of similar μ-σ datasets at varying levels of mutagenesis for a larger collection of reporter strains as well as large-scale sequence analysis of interesting mutant promoters can yield useful insight into yeast promoter architecture. Here the hope would be to understand the phenotypic effects single mutations have on both μ and σ.

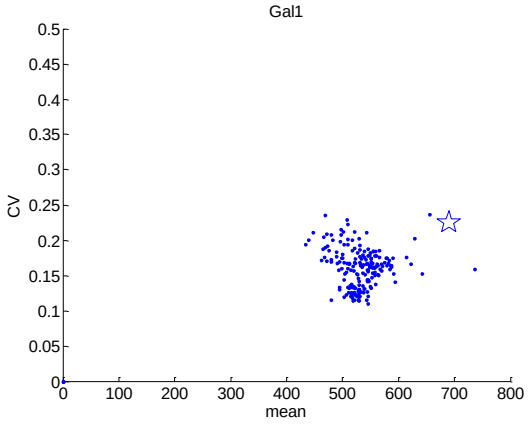


Fig. 3. **Mild Promoter Mutagenesis Results.** After 12 hours of growth in a mutagenesis cell line, the Gal1 mutant promoters (blue dots) are remarkably similar to that of the wild-type promoter (blue star). In fact, the figure is displayed on a linear plot to accentuate the minor differences.

Discussion and Future Experiments

The results described above reveal the effects of mutation on promoters in the absence of selection. Although the most exciting result for us is the potential independence of μ and σ in determining gene expression, there are several fundamental gene regulatory questions that can be addressed with this experimental system.

First, a fundamental aspect of mutations is their interactions. Depending on the mechanism by which regulatory mutations influence expression and noise, they could interact in a neutral (additive), aggravating, or alleviating fashion. As a simple example, consider a mutation that increases expression level and one that reduces it. If they are combined, what will the new expression level be?

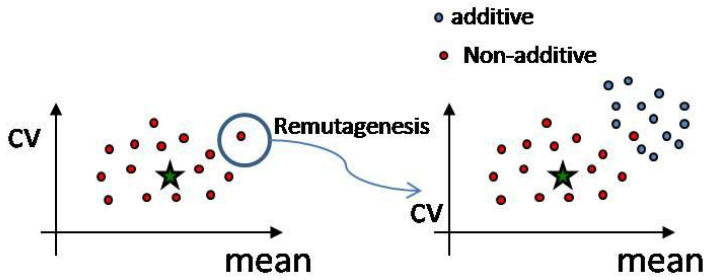


Fig. 4. **Epistatic Affect of Mutations.** From our initial dataset we can pick a mutant clone with high mean and/or high noise and subject it to a second round of mutangenesis. If the mutations are additive, the mutant clone will continue increasing in mean and/ or noise (blue). If they are not additive, they will cluster around the wild-type promoter (green star) similarly to the mutants after the first round of mutagenesis (red).

To answer this fundamental question, we can select individual mutants with high expression level or high noise from the mutant libraries. We will subject these individual clones to additional rounds of random mutagenesis (Fig. 4). Alternatively, we will use site-directed mutagenesis to combine specific mutations with known effects (as determined

from the dataset in Fig. 2). The resulting clones will be analyzed by flow cytometry so that epistasis in both mean expression level and noise phenotypes can be determined

An integral part of the aforementioned experiments is to sequence the mutant promoters in an effort to determine the molecular determinants of gene expression noise and changing expression levels. Using well-annotated promoters, such as the ones described in Fig. 2, we will be able to use the location of these mutations (e.g. in known binding sites) to generate hypotheses for the mechanism by which they operate. These hypotheses will be tested by site-directed mutagenesis and analysis of epistatic interactions with other sequenced mutations. With these experiments we hope to determine the molecular origins of the individual mutations and epistatic interactions identified above. But even before we begin to do that, we can already make predictions from the datasets we already have.

Although our model of gene expression implies a normal distribution with two parameters, other models have predicted that the distribution is better fit by the gamma distribution.[17] With this model, the two parameters controlling expression are transcriptional bursts size (*a*) and transcriptional bursts frequency (*b*). Plotting mutant promoters on the *a/b* axis (Fig. 5) shows that the Pir1 mutants all seem to have the same smaller bursts size while many of the Gal1 mutants have the same amount of less frequent transcriptional bursts. In these cases, there might be specific dominant mutations that cause the changes in burst size and/or frequency; presumably these same mutations also have a dominant impact on gene expression mean and noise. Meanwhile, the Tim17 mutants seem to vary in both burst size and frequency.

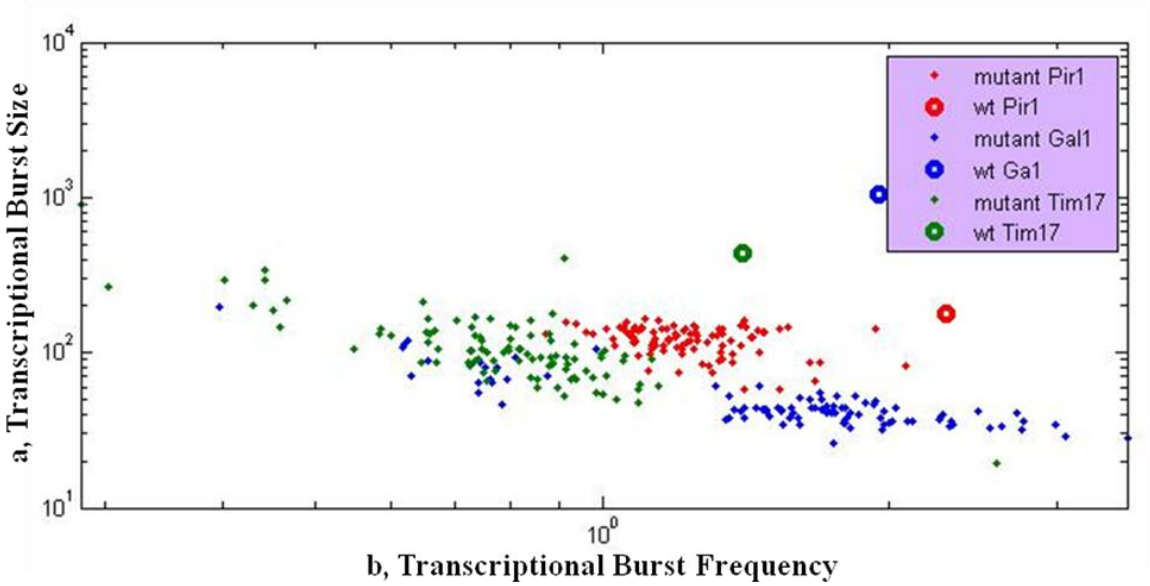


Fig. 5. **Transformed Promoter Mutagenesis Results.** After 24 hours of growth in a mutagenesis cell line, the promoter mutant library data shows interesting mechanistic results. Mutants of Pir1 have a lower burst frequency than wild-type promoters which most Gal1 mutants have a burst size smaller than their wild-type counterpart. Meanwhile, Tim17 mutants vary in both parameters.

This result, showing the potential dominance of certain mutations, suggests that specific mutations can restrict promoter landscapes. For example, by measuring expression levels across varying mutational loads, one can get a sense of the dynamic range of a promoter. However, upon a specific mutation, does the promoter retain that dynamic range or is it limited? Does this depend on the mutation? How so? With the current set of mutants and assay conditions, such experiments can begin to tell us about promoter architecture, its evolution and how it translates into gene expression noise.

However, it is worth nothing that there are some problems with the proposed experimental setup, namely that the mutagenesis is not restricted to the promoter and that the copy number variation from the plasmid can also affect our measurements. To correct for these flaws, we envision adding a strong constitutive RFP to the plasmid which can be used to normalize for plasmid copy number. Furthermore, after each round of mutagenesis, the promoter will be amplified and ligated into the wild-type plasmid, ensuring that the mutations are relegated to the promoter region. With these minor

adjustments, the sets of experiments described and proposed in this chapter can help disentangle how promoter mutations affect gene expression noise.

Acknowledgements

The work described in this chapter benefitted from several discussions with various members of the Elowitz lab.

Chapter 3

FREQUENCY-MODULATED NUCLEAR LOCALIZATION BURSTS PROPORTIONALLY REGULATE GENE EXPRESSION

Chapter Overview

In yeast, the transcription factor, Crz1, is dephosphorylated and translocates into the nucleus in response to extracellular calcium. Using time-lapse microscopy, we found that Crz1 exhibited short bursts of nuclear localization (~2 minutes) that occurred stochastically in individual cells and propagated to the expression of downstream genes. Strikingly, calcium concentration controlled the frequency, but not duration, of localization bursts. Using an analytic model, we found that this frequency modulation (FM) of bursts ensures proportional expression of multiple target genes across a wide dynamic range of expression levels, independent of promoter characteristics. We experimentally confirmed this theory with natural and synthetic Crz1 target promoters. Another stress response transcription factor, Msn2, exhibits similar, but largely uncorrelated, localization bursts under calcium stress. These results suggest that FM regulation of localization bursts may be a general control strategy utilized by the cell to coordinate multi-gene responses to external signals.

Introduction

Cells sense extracellular signals and respond by regulating the expression of target genes.[18,19,20] This process requires two stages of information processing: First, cells encode extracellular signals internally, in the states and localization of transcription factors. Second, transcription factors activate the expression of downstream genes that will implement cellular responses.[18,19,20] Although many signal transduction systems have been studied extensively, it often remains unclear how signals are encoded dynamically in transcription factor activities at the single-cell level. In addition, cellular responses often involve many proteins acting in concert, rather than individually. But it is not known in general how the expression levels of target genes are coordinated, allowing them to be regulated together, despite diverse promoter architectures.[3] Here we investigate how signal encoding and protein coordination are achieved in individual cells.

Results

We examined the calcium stress response pathway in *Saccharomyces cerevisiae,* or budding yeast. Cellular response to extracellular calcium is mediated by Crz1, the calcineurin-responsive zinc finger transcription factor.[21] The activity of Crz1 is modulated by phosphorylation and dephosphorylation, resulting in changes in Crz1 nuclear localization[21] (Fig. 1a), rather than changes in its protein abundance (Appendix B, Fig. 1). To understand how Crz1 phosphorylation dynamics respond to calcium and regulate the more than 100 different targets necessary for calcium adaptation,[22] we acquired time-lapse movies of Crz1 localization dynamics, using a strain in which the Crz1 protein was tagged with GFP.[23] In each movie, we tracked the response of Crz1 localization in individual cells to step changes in extracellular calcium concentration. We find that Crz1 dynamics connect the encoding of signals and the coordination of target gene expression.

In the absence of calcium, Crz1 was cytoplasmic in all cells. Upon the addition of calcium, individual cells exhibited a rapid, synchronized burst of Crz1 nuclear localization, similar to behavior observed by the yeast osmosensor Hog1.[24,25] However, unlike Hog1, this initial burst was followed by sporadic unsynchronized localization bursts, typically lasting about 2 minutes (Fig. 1 b-d) and persisting throughout the course of the movie (up to 10 hours). Moreover, these single-cell Crz1 dynamics are consistent with micro-array studies performed on cell populations:[22] After a step change in calcium, an initial overshoot in mRNA levels of Crz1 target genes results from the initial synchronous burst of Crz1, while the subsequent elevated average expression levels are due to sustained unsynchronized bursts in individual cells (Fig. 1c-e).

We next addressed how the amount of calcium affects the dynamics of nuclear localization. We observed that the fraction of cells with nuclear-localized Crz1 increased with calcium concentration. Because Crz1 localizes in bursts, this calcium-dependence could in principle result from increases in burst frequency or duration.

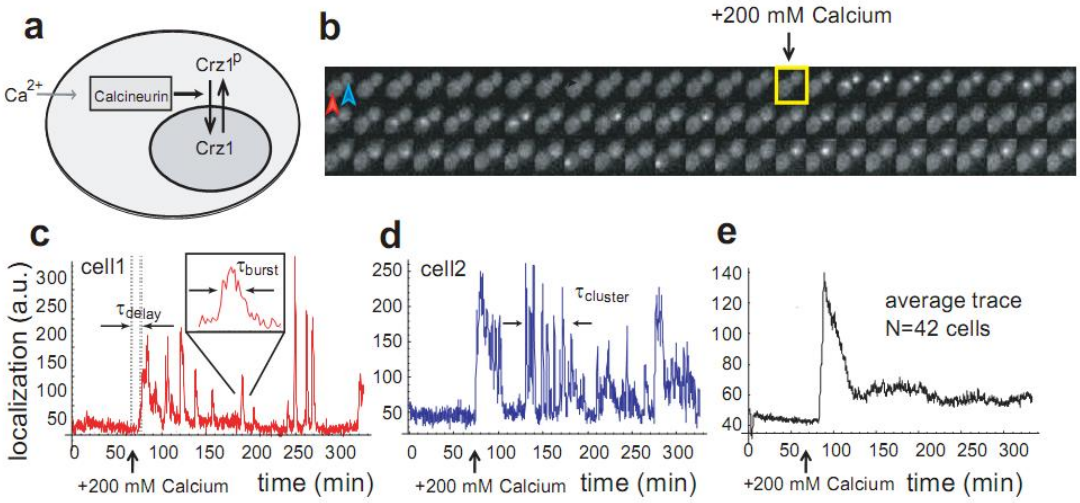


Fig. 1. **Crz1 undergoes bursts of nuclear localization in response to calcium**. **a.** In the presence of extracellular Ca^{2+}, Crz1 is dephosphorylated and translocates into the nucleus. **b.** Filmstrip showing yeast cells with Crz1-GFP before and after addition of 200mM extracellular calcium (yellow square). Frames displayed here are separated by 4.5 minutes, but actual time resolution is higher. **c,d.** Two single cell time traces showing Crz1 localization behavior of the two cells in **b**. Note that there is a synchronized initial burst of nuclear localization followed by subsequent unsynchronized isolated and cluster bursts of localization. Individual burst duration, τ_{burst}, and cluster duration, $\tau_{cluster}$, as well as the delay between calcium addition and the initial response, τ_{delay}, are defined on the traces. Averaged localization trace shows how single-cell burst dynamics yield partial adaptation across a population of cells.

Strikingly, analysis of the movies revealed that only the burst frequency increased (Fig. 2a), while the distribution of burst durations remained constant at all calcium concentrations (Fig. 2b). This distribution was consistent with two rate-limiting stochastic steps, each with a timescale of ~70 seconds (Fig. 2b). Thus, cells use stereotyped Crz1 localization bursts in a frequency-modulated fashion to encode and respond to extracellular calcium. This contrasts with amplitude-modulation control, in which the fraction of Crz1 molecules found in the nucleus would change with calcium, but remain constant over time.

The movies also revealed two modes of nuclear localization bursts: isolated individual bursts and clusters of bursts, analogous to spike trains in neurons (Fig. 1c,d). At calcium concentrations less than 100 mM, only isolated bursts were observed and the averaged autocorrelation function of the localization trajectories from individual cells was well-fit by a single exponential. However, as clustered bursts emerged at calcium concentrations greater than 100mM (Fig. 1c,d), the averaged autocorrelation function was

better fit by a sum of two exponentials, whose timescales matched the typical durations of isolated bursts and burst clusters, respectively (Fig. 2d). Higher levels of calcium led to an increasing proportion of bursts occurring in clusters (Fig. 2d inset). Eventually, at the highest calcium levels, Crz1 nuclear localization trajectories appeared more similar to sustained oscillations[26] than to the isolated stochastic bursts seen at lower calcium concentrations.

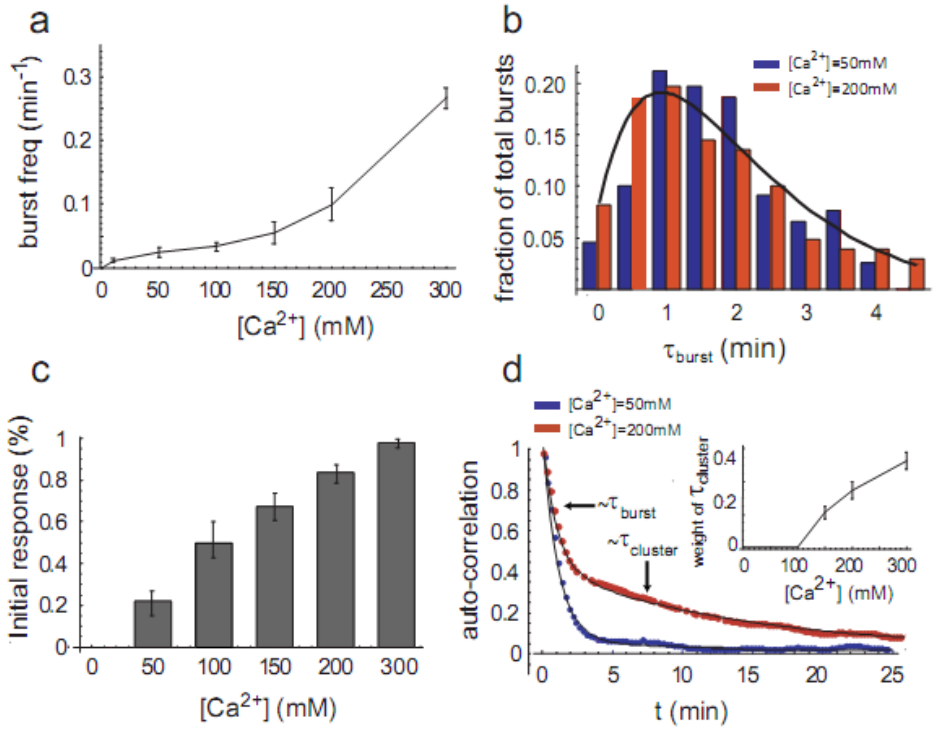


Fig. 2. **Calcium modulates the frequency, but not the duration, of Crz1 nuclear localization bursts. a**. Frequency of bursts increases with calcium concentration (error bars calculated by using different thresholds for burst determination). **b.** Burst duration is independent of calcium concentration. Normalized histograms, $h(t)$, of total burst duration at two calcium concentrations are both well-fit by $h(t) = t\, e^{-t/\tau}$, with $\tau = 70$ seconds (black line). **c.** The proportion of cells that respond rapidly to extracellular calcium increases with the calcium concentration. Bars represent the fraction of cells with nuclear-localized Crz1 within 15 minutes of addition of calcium. **d.** Average auto-correlation functions of localization trajectories (N=58 and 85 cells respectively) from a population of cells at two calcium concentrations. At low Ca^{2+} concentrations (blue, 50mM), the auto-correlation can be well-fit by a single exponential with timescale $\tau_{burst} \approx 60$ seconds, whereas at high Ca^{2+} concentrations (red, 200mM), two time scales of fluctuations emerge, $\tau_{burst} \approx$ 60 seconds and $\tau_{cluster} \approx$ 720 seconds, corresponding to isolated and clustered bursts, respectively. Inset shows the relative weight of the clustered bursts, which appear at Ca^{2+} concentrations greater than 100mM and increase in frequency as calcium increases. Error bars are estimated from bootstrap.

An additional level of quantization emerged in the initial response to a change in calcium, which exhibited an all-or-none response at the single-cell level. At calcium concentrations ≥ 100 mM, most cells displayed a synchronous initial burst of Crz1 nuclear localization. This resulted in a sharp histogram of values of τ_{delay}, defined as the time interval between calcium addition and the first burst of Crz1 nuclear localization (Fig. 1c). In contrast, at lower calcium levels of about 10mM, no synchronous initial response occurs, as reflected in the much broader distribution and larger mean of τ_{delay}. Intermediate calcium concentrations produced a mixture of the two discrete behaviors (Appendix B, Fig. 2). Thus, rather than controlling the amplitude of the initial response, calcium modulates the proportion of fast-responding cells (Fig. 2c). Critically, in both the initial response and the subsequent bursts, we never observed persistent intermediate levels of localization. Taken together, these results reveal that Crz1 burst activity is 'quantized' at multiple levels in cells and that only the frequencies of burst events are modulated by extracellular calcium.

Since nuclear localization bursts occurred stochastically in single cells and were unsynchronized within a population on the same slide, they could not be driven by fluctuations in external conditions. These bursts also cannot be explained simply by independent fluctuations in the phosphorylation or localization state of individual Crz1 molecules because each burst involves coherent translocation of a large fraction of the ~1000 copies of Crz1 present per cell.[3] In order to gain insight into these dynamics, we next asked whether Crz1 bursts were related to, or driven by, other dynamic cellular phenomena.

First, we acquired movies of cells expressing both the G1 cell cycle phase marker Whi5-GFP[27] and Crz1-mCherry. We found no evidence for cell cycle regulation of Crz1 localization bursts, nor were bursts in daughter cells correlated with those in corresponding mother cells (Appendix B, Fig. 5).

Second, we examined the role of intracellular calcium, which in other cell types has been shown to exhibit spike-like dynamics.[28] We acquired movies of yeast cells expressing both a FRET-based calcium sensor[29] and Crz1-mCherry. We observed sporadic transient spikes in intracellular calcium lasting about 38 seconds (Appendix B, Fig. 7). These spikes

coincided with some Crz1 localization bursts (Fig. 3a and Appendix B, Fig. 6). However, Crz1 localization bursts also occurred without corresponding calcium spikes, indicating that localization bursts may be stimulated by calcium spikes, but are not exclusively determined by them.

Third, we asked whether Crz1 bursts were driven by fluctuations in calcineurin, the upstream phosphatase that initiates Crz1 nuclear localization.[30] We analyzed a Crz1 mutant[31]with higher affinity to calcineurin, GFP-Crz1^{high}, together with Crz1-mCherry simultaneously in the same cell. If Crz1 bursts were just passive readouts of spikes in calcineurin activity, one would expect both types of Crz1 to burst simultaneously, with amplitudes related to their relative calcineurin affinities. Instead, we observed that Crz1^{high} exhibited an increased burst frequency. Wild-type Crz1 bursts were a subset of these Crz1^{high} bursts (Fig. 3b). Similarly, sub-saturating concentrations of the calcineurin inhibitor, FK506,[32] reduced the frequency of bursts, but did not affect their amplitude (Appendix B, Fig. 8). As the calcineurin-Crz1 interaction controls only the frequency of burst initiation, Crz1 localization dynamics do not simply follow upstream fluctuations in calcineurin activity.

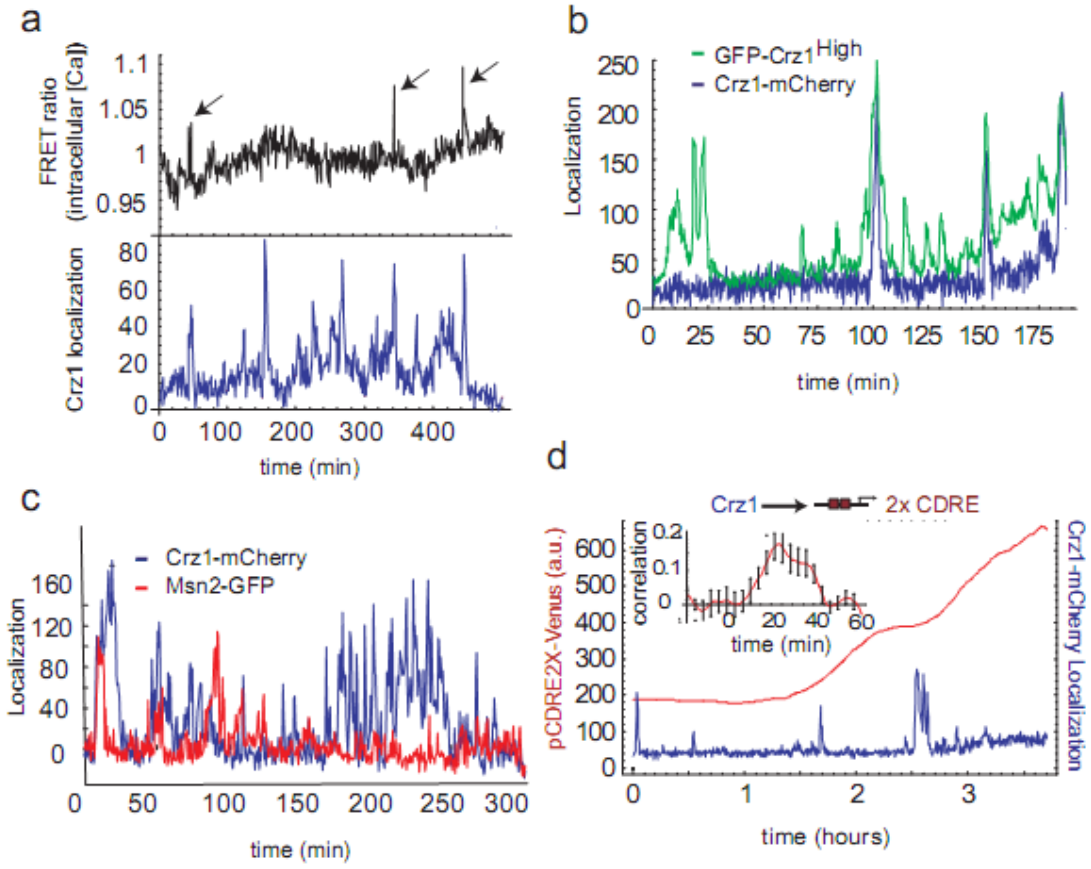


Fig. 3. **Crz1 localization bursts are partially independent of other cellular processes and affect downstream gene expression**. **a.** Time traces of Crz1-mCherry localization (blue) in arbitrary units and FRET ratio changes (black), indicating intracellular calcium levels. Arrows indicate spontaneous calcium spikes coincident with Crz1 localization bursts. **b.** Single-cell time traces of Crz1-mCherry (blue) and the Crz1^{high}-GFP mutant (green) with increased affinity to calcineurin. Both Crz1 proteins are expressed and measured simultaneously in the same cell. **c.** Single-cell traces of Crz1-mCherry and Msn2-GFP in the same cell. Note that the two proteins exhibit statistically similar burst-like behavior, but only weak correlation. **d.** Crz1-mCherry localization (blue) increases expression of the Crz1 target synthetic promoter (2X CDRE) (red). Transcriptional bursts in p2XCDRE-venus are preceded by corresponding Crz1 localization bursts. Inset shows positive cross-correlation (n=9 cells) between the promoter activity and Crz1 localization with a delay corresponding to target protein maturation. Error bars are estimated from bootstrap.

Finally, to investigate the generality of localization bursts, we examined Msn2, a general stress response transcription factor that was previously reported to exhibit nuclear localization oscillations.[33,34,35] We found that Msn2-GFP localization is induced by calcium stress. Like Crz1, it localized in short bursts on the timescale of 1.5-2 minutes, and also exhibited clustered bursts (Fig. 3c and Appendix B, Fig 9). However, despite these statistical similarities, bursts of the two proteins were largely uncorrelated when observed simultaneously in the same cells under calcium stress (Fig. 3c and Appendix B, Fig. 10). 18.4±2.0% of Msn2-GFP bursts coincided with Crz1-mCherry bursts, a fraction only slightly higher than the 14.0±1.2% of overlapping events expected by chance if the two proteins burst independently (Appendix B). Furthermore, bursts of Msn2 nuclear localization can occur in the absence of calcium, while Crz1 bursts cannot. Similarly, we observed that the glucose-responsive repressor Mig1 exhibited bursts of nuclear localization at low glucose concentrations. These results suggest that cells operate multiple transcription factor localization burst systems in a largely independent manner.

What effect do FM-regulated Crz1 localization bursts have on downstream genes? We analyzed the transcriptional activity of a synthetic Crz1-dependent promoter containing calcineurin-dependent response elements (2X CDRE),[36] driving expression of the fluorescent protein Venus. We also monitored Crz1-mCherry localization simultaneously in the same cell (Fig. 3d). We found that the promoter was expressed in transcriptional bursts that followed Crz1 localization bursts (Fig. 3d). Interestingly, not all Crz1 localization bursts resulted in observable transcriptional bursts, suggesting that transcription initiation is probabilistic. Nevertheless, the rate of Venus production (time derivative of fluorescence) was correlated with Crz1-mCherry bursts with a time-delay comparable to the maturation time of the Venus fluorophore, as expected (Fig. 3d inset and Appendix B, Fig. 12). We observed similar results with a natural Crz1 target gene, Cmk2 (Appendix B, Fig. 11).[22] Thus, transcription factor localization bursts propagate to downstream targets, and represent a general mechanism for generating "transcriptional bursting" [37,38,39] in downstream gene expression.[2 1 13 17 40 41 42 15 43 4 44]

Standard models of gene regulation involve amplitude modulation, or analogue, changes in transcription factor concentration in response to external signals. What

biological functions could the frequency modulation form of regulation observed here provide for the cell? Crz1 regulates more than 100 different target genes,[22] including Ca^{2+} pumps and structural proteins necessary for calcium adaptation. The target promoters of these genes may differ in their input functions, defined as the dependence of transcription rate on the concentration of transcription factor in the nucleus. Input functions vary widely in their minimal and maximal levels of expression, the concentration of transcription factor at which they reach half-maximal activity, and in the sharpness, or cooperativity, of their response.[45] We used an analytic model to show that frequency-modulation regulation of nuclear localization bursts could allow transcription factors to modulate the expression of multiple target genes in concert, keeping their relative abundances fixed over a wide dynamic range, regardless of the shapes of their input functions.

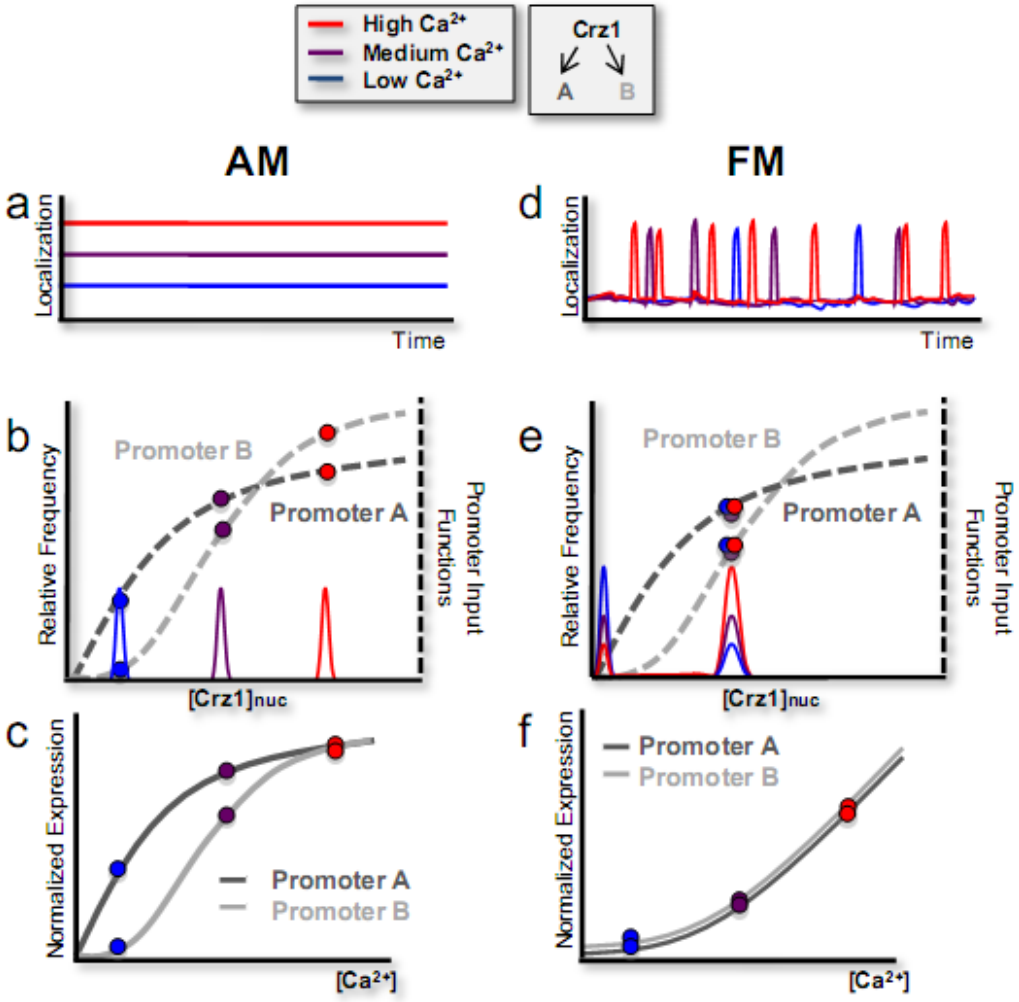


Fig. 4. **Frequency-versus amplitude–modulation regulation of two hypothetical target genes, labeled A and B (schematic).** **a.** In the amplitude-modulation regulation system, the fraction of nuclear Crz1 ($Crz1_{nuc}$) changes with calcium, but remains constant over time. **b.** As such, the histogram of $Crz1_{nuc}$ yields single peaks at calcium-dependent positions. Target gene expression level is proportional to the input functions at these peak positions. **c.** Because their input functions differ, the normalized rates of A and B expression vary differently with nuclear Crz1, and hence with calcium, yielding different (uncoordinated) expression profiles as a function of calcium. **d.** In frequency modulation, where Crz1 molecules collectively move between nuclear and cytoplasmic compartments, $Crz1_{nuc}$ is either high or low, during or between bursts, respectively. This graph depicts the limiting case of rapid and complete transitions between two states, but results do not depend on this assumption (see Supplementary Information). **e.** This yields a bimodal histogram in which the height, but not the position, of the peak is calcium-dependent. **f.** Consequently, the expression levels of A and B are each proportional to burst frequency, and hence to each other, yielding coordinated expression, as shown.

This hypothesis can be understood by comparing the effects that amplitude- and frequency-modulation regulation systems have on two hypothetical target promoters, labeled A and B, with different input functions (Fig. 4b, e, gray dashed curves). In the AM regulation system, the fraction of Crz1 molecules in the nucleus, $Crz1_{nuc}$, would increase with calcium, but remain constant over time (Fig. 4a, b). Accordingly, histograms of $Crz1_{nuc}$ would exhibit a single peak whose position is calcium-dependent (Fig. 4b). The expression level of each gene would be proportional to its input function evaluated at this peak position. Because it is expected that these input functions would, in general, differ in shape, the normalized rates of A and B expression would vary differently with $Crz1_{nuc}$, and hence their ratio would vary with calcium (Fig. 4c). Thus, in the amplitude modulation model, A and B would be expressed in an 'uncoordinated' fashion.

In contrast, frequency-modulation regulation (Fig. 4d) would control the fraction of time that Crz1 is nuclear, rather than the concentration of nuclear Crz1. In the limiting case of fast switching between two localization states, this would cause $Crz1_{nuc}$ to be either high, during a burst, or very low, between bursts, but rarely in between, resulting in a bimodal $Crz1_{nuc}$ histogram (Fig. 4e). The expression level of each target promoter would be determined mainly by the value of its input function near the location of the higher histogram peak, and by the fraction of time Crz1 spends in the localized state, which determines the height of the peak. In frequency modulation regulation, higher levels of calcium would increase the relative height of this peak by increasing the frequency of bursts, but would not change its position (Fig. 4e). The expression levels of the two genes would therefore be individually proportional to the burst frequency, and consequently remain proportional to each other, as calcium is varied. Thus, A and B would be coordinated, expressed at a constant ratio at all calcium levels, regardless of the shapes of their input functions (Fig. 4f).

This regulatory strategy is general: It does not require bursts to saturate target promoter input functions, nor does it require that the $Crz1_{nuc}$ histogram be bimodal (see Appendix B for a more detailed treatment). Thus, it functions with bursts, such as those observed here, that span a range of localization amplitudes, giving rise to a non-bimodal histogram. Finally, this strategy is more general than FM regulation; it can also work by

varying the duration, rather than the frequency, of localization bursts. We refer to this principle as FM-coordination.

To test the FM-coordination hypothesis experimentally, we first analyzed three synthetic promoters containing one, two, and four CDREs. These promoters were previously shown to exhibit Crz1-dependent expression in response to calcium.[36] We used flow cytometry to analyze strains containing each of these three promoters fused to yellow fluorescent protein (YFP). The promoters varied in their expression strength over two orders of magnitude, in the ratio of 1:15:450, respectively (Fig. 5a inset). Nevertheless, they exhibited identical dependence on calcium, collapsing onto a single curve when normalized, as predicted by FM-coordination (Fig. 5a).

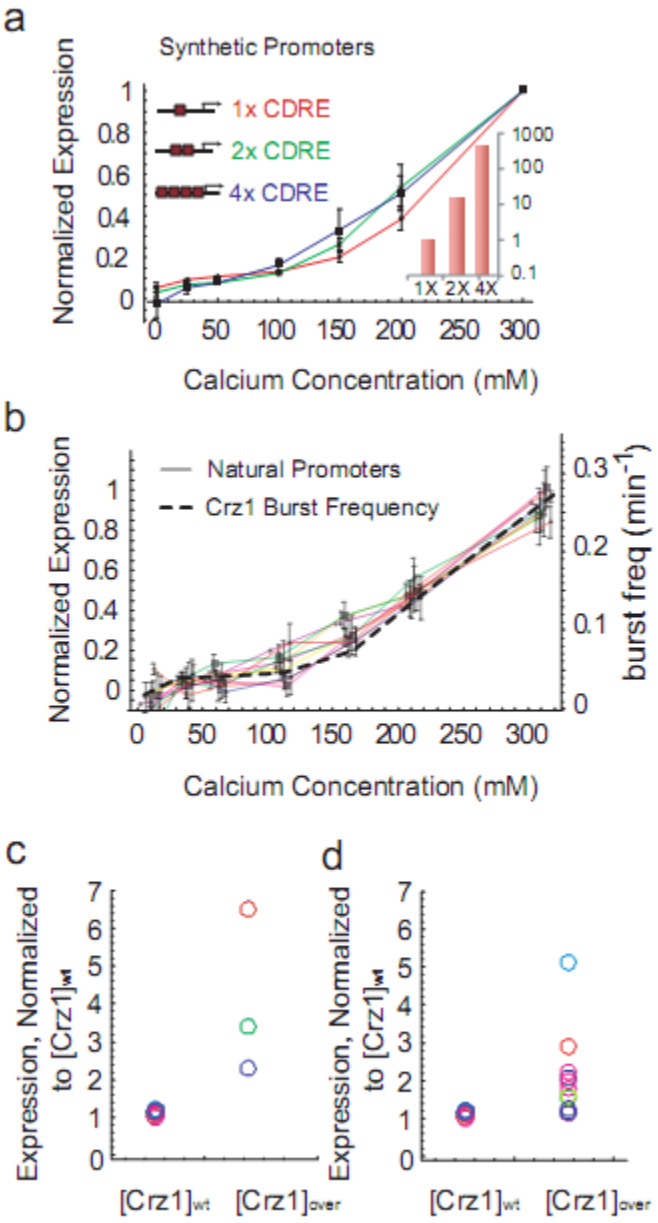


Fig. 5. **Frequency-modulated bursts coordinate gene expression.** **a.** Measured expression levels of synthetic Crz1 target promoters containing one, two and four CDREs. In each case, expression levels are normalized by their maximum. Note similarity of curves to each other and to burst frequency. Bottom inset shows expression at 200mM calcium normalized by expression of the 1x CDRE. **b**, Expression profiles of natural target promoters (solid lines) exhibit similar inductions as the Crz1 burst frequency (dashed line). Only ten are plotted here for clarity. Error bars indicate standard error from repeated experiments. **c,** Expression level of synthetic target genes at wild-type versus overexpressed levels of Crz1. Each promoter is normalized by the maximum expression in wild-type Crz1 levels. Thus the normalized expression levels of one, two and four CDREs at $[Crz1]_{wt}$ each equal 1. If all promoter input functions were identical, then their normalized expression should increase by the same factor when Crz1 is overexpressed. However, the data show a range of fold changes in expression levels, excluding identical input functions as an explanation for proportional regulation. **d**. Same as **c**. for the natural promoters in **b.**

Natural eukaryotic promoters exhibit more diverse architectures than these pure synthetic promoters. Therefore, we also investigated whether FM-coordination could explain the behavior of natural Crz1 target genes. A previous microarray study identified 163 potential Crz1 target genes,[22] 83 of which were available as fluorescent protein fusions.[23] Of these, 40 showed a measurable and predominantly Calcineurin-dependent response to calcium under our media conditions (Appendix B, Fig. 15). These natural promoters, like their synthetic counterparts, exhibited a broad distribution of expression levels (Appendix B, Figs. 14 and 17). Notably, a large subset of these target genes (34 out of 40) exhibited normalized calcium response curves identical to each other and to those of the synthetic promoters. They were thus regulated in proportion to burst frequency, in accordance with FM-coordination (Fig. 5b and Appendix B, Fig. 14a).

In principle, target promoters could respond proportionally because they share identical input functions. If so, increasing total Crz1 concentration would increase the expression of all promoters by the same factor (Fig. 5c, d). Conversely, if the expression levels of the promoters change by different factors when Crz1 concentration is increased, then their input functions must differ in shape. This is what we observed when we compared the expression of both the synthetic and natural promoters in wild-type strains to their expression in a Crz1-overexpressing strain.[46] The increase in expression varied between one-fold and seven-fold over this set of promoters (Fig. 5c, d), confirming that synthetic and natural promoters are proportionally regulated by the FM-coordination mechanism, despite non-identical input functions.

Discussion

Diverse perturbations including calcium, pharmacological inhibitors, and genetic mutations vary mean Crz1 activity across a broad dynamic range exclusively by affecting the frequency of localization bursts, rather than their duration or amplitude. Thus, Crz1 activity is quantized both in its initial response to a step change in calcium, which is all-or-none at the single-cell level, and also in its sustained response, which is composed of stereotyped bursts of localization. These data suggest that eukaryotic cells can encode information about the extracellular environment in the frequency of stochastic intracellular

events, rather than in the concentrations of molecular species. This contrasts with canonical signal transduction pathways, in which concentrations of activated proteins are used to convey information to the nucleus. Using localization as a single-cell reporter for post-translational modifications, we observed qualitatively similar, but largely uncorrelated, behavior in the Msn2 stress response transcription factor and the Mig1 glucose repressor, suggesting that localization bursts are broadly employed in the cell.

In engineering, frequency modulation appears in diverse signal processing and control applications such as broadcasting[47] and rocket thruster control,[48] where transfer characteristics may be uncertain. Spike frequency control, or rate coding, is also fundamental to neural computation,[49] where it may overcome noise background for signal propagation, among other functions.[50] In genetic networks, frequency-modulation regulation solves a fundamental problem, allowing cells to co-regulate proportionally a large set of target genes with a diverse range of input functions. Note also that these results represent a complementary mode of regulation compared to previous observations of calcium oscillation frequency-dependent gene expression.[51] FM-coordination connects the dynamic behavior of a single protein to the expression of large regulons, and may thus provide insight into genome-scale expression patterns. Furthermore, at the evolutionary level, FM-coordination enables promoter mutations to alter the level of expression of individual genes without disrupting their coordination. In contrast, to achieve coordinate regulation using amplitude modulation would require ‘fine-tuning’ of target promoter input functions, severely constraining promoter architecture and regulatory evolution. In light of previous observations of transcription factor pulsing in p53,[52] NFκB,[53] and SOS stress response systems,[54] and because of its potential utility in protein and metabolic, as well as transcriptional networks, we anticipate that frequency-modulated regulation may represent a general principle by which cells coordinate their response to signals.

Acknowledgements

The work done in this chapter was a collaboration with Long Cai, a postdoctoral fellow in the Elowitz lab.

Chapter 4

SIGNAL PROCESSING ACROSS THE YEAST PROTEOME

Motivation

In our previous investigation, we observed that the yeast transcription factor Crz1 exhibits frequency-modulated bursts of nuclear localization in response to its stimulus, calcium. This behavior enables proportional regulation of Crz1 target genes in response to calcium. Because this regulatory strategy is very general, it could be employed more generally by cells in order to optimally tune expression levels of groups of promoters (regulons). Moreover, cells may employ other dynamic regulatory strategies to enable other functional behaviors. Based on these considerations, we performed a systematic screen for dynamic regulatory behaviors in yeast.

Screen Methods/Results

Since we were unable to determine which proteins would be most appropriate to observe, we decided to look at as many as we could. Taking advantage of the yeast GFP library,[23] we used multi-well time lapse microscopy to systematically search for dynamic and/or heterogeneous localization behaviors. Cells were imaged in one of 4 media conditions: Standard synthetic complete or wild-type, carbon-limited, nitrogen-limited, or high salt. The screen involved two stages: First, we acquired low time-resolution (~51 min/frame) movies of each protein in each condition. These movies were inspected visually to detect heterogeneous localization patterns in one or more media conditions. This screen suggested a few candidate genes, but, because of its low time resolution, left ambiguity in many of these cases. We classified each protein based on the likelihood that we observed real heterogeneity in any one of four conditions. Interestingly, almost all protein localization patterns were independent of condition.

In the second stage, heterogeneously localized protein strains were re-analyzed at higher time-resolution movies (~3 minutes between frames). If they still showed dynamic or heterogeneous localization patterns we examined them further in several ways:

- Focal drift exclusion: Several strains showed dynamic patterns of non-nuclear localization to regions smaller than a cell nucleus. To test for possible effects due to 3-D diffusion of a localized spot, we re-acquired movies of these strains with multiple z-slices. In all cases, fluctuations in fluorescence intensity in these strains could be explained by movement of a localized sub-cellular spot, rather than fluctuations in the intensity of the spot (fraction of proteins localized).

- Cell cycle exclusion: We excluded proteins whose localization patterns were principally determined by cell-cycle phase. To investigate cell-cycle dependence, we grew up all the proteins on our list and synchronized the cell cycle with hydroxyurea. After releasing the synchrony , we added the appropriate stimuli to the media and took movies. (In some cases, because the hydroxyurea does not work perfectly, we also took very long movies where we could observe several cellular divisions and checked by eye to see whether there was any correlation between cellular division and burst occurrence). From these results, we were able to erase several proteins off our list.

The remaining proteins exhibited dynamic and/or heterogeneous localization dynamics not principally correlated with cell cycle. These proteins consisted almost entirely of transcription factors.

Based on these results, we undertook a final additional screen, which focused specifically on transcription factors. For each transcription factor, we recorded a series of movies with varying levels of factors or conditions known to induce its transcriptional activity. The results of the final screen are tabulated in table 1.

The validated dynamic/heterogeneous proteins consisted exclusively of transcription factors. Several conclusions emerged from this dataset: First, most transcription factors are constitutively nuclear. Second, except for a two cases that require special media conditions (alpha pheromone), our original proteome-wide screen identified

10 out of the 12 bursting proteins. Third, we observed distinct timescales of localization bursts. We found five rapid-bursting proteins (τ<30 min), and seven with slower dynamics (τ~30 min to 2 hour), most of which appear to be involved in metabolism in some way. Note that these slower proteins were not cell cycle correlated in phase. Fourth, some proteins exhibited heterogeneous patterns of nuclear localization that did not involve dynamic bursts. Rather, stimuli appear to control the fraction of cells which exhibit strong nuclear localization. In these "fractionally heterogeneous" proteins, the number of cells with fully localized protein increases with the level of stimulus. Finally, we identified "amplitude-modulated (AM)" transcription factors, in which the fraction of total protein localized to the nucleus increased with stimulus level. We identified no additional proteins resembling Hog1, which has been shown to adapt perfectly to osmotic stress.[24]

Localization Behavior	**Protein (Media Condition)**
I. Fast Bursting	Crz1 (Calcium), Msn2 (salt, rapamycin, calcium, glucose deprivation), Mig1 (glucose deprivation), Dot6 (wild-type), Rtg1 (wild-type)
II. Slow Bursting (no Modulation)	Nrg2 (Glucose Deprivation), Mig2 (Glucose Deprivation), Kar4 (alpha pheromone), Tos1 (alpha pheromone), Arg81 (Arginine), Leu3 (Leucine), Tea1 (Arginine)
III. Statically Heterogeneous	Dig1 (alpha pheromone), Dig2 (alpha pheromone), Pci8 (alpha pheromone), Sch9 (salt), Ino2 (lithium acetate), Sfl1 (salt)
IV. Constitutive but Amplitude	Abf1 (glucose deprivation), Cin5 (salt), Cti6 (no glucose, galactose titration), Mcm1

Modulated	(alpha pheromone), Put3 (Nitrogen deprivation) , Ste12 (alpha pheromone), Xbp1 (glucose deprivation),
V. Perfect Adaptation	Hog1 (salt)

Table 1. **Modes of Signal Processing in Yeast**. The different types of localization behavior from the transcription factor screen are summarized.

As a nice positive control, Crz1 came up in our screen; as previously stated, it exhibits FM bursts in response to calcium. Msn2 and Mig1 had also been previously identified as fast bursting proteins.[55] Msn2 is a general stress response protein that undergoes nuclear localization to a wide variety of stresses.[56,57] Our screen results suggest that Msn2 responds to several stresses similarly, with fast bursts. Mig1 is a carbon utilization transcriptional repressor[58,59] that exhibits both FM and pulse width modulation (PWM) in response to glucose deprivation; both burst duration and frequency increases as you decrease the amount of glucose. Dot6 is not a well-characterized protein, but has been implicated in transcriptional silencing.[58,59] Without any stressful stimuli, Dot6 exhibits bursts. Similarly, Rtg1, a transcription factor implicated in nitrogen stress,[58,59] bursts without any provocation.

The second class of localization behavior observed is that of the slow bursting proteins. These proteins exhibit bursts with a timescale of approximately one hour. Interestingly, all these transcription factors are involved in sensing nutrient availability but none of them show any sort of modulation. For example, Nrg2 and Mig2 are both transcriptional repressors that respond to glucose.[58,59] However, their localization appears to remain consistent across all concentrations of glucose. Similarly, Arg81, Leu3, and Tea1 respond to amino acid deprivation and Kar4 and Tos1 respond to alpha pheromone,[58,59] but the localization of all these proteins show slow bursting across various concentration of the stimulus.

At any given time during a movie of a bursting transcription factor, the heterogeneity in localization is quite striking as some cells are nuclear while others are cytoplasmic. Of course, one can imagine a far simpler mechanism to generate such heterogeneity, if only a fraction of the cells respond to the stimulus and localize. This is precisely the behavior we observe in the third class of proteins, the statically heterogeneous ones. The interesting aspect of these proteins is that their localization is modulated by stress also. As the amount of stress increases, the fraction of cells that respond via nuclear localization also increases. These proteins, Dig1, Dig2, Ino2, Pci8, Sch9, and Sfl1, respond to various stresses including alpha pheromone, salt, and lithium acetate.[58,59]

Finally, we also observe amplitude-modulated (AM) proteins. These proteins are constitutively nuclear, but their intensity increases with stress, yielding an AM phenotype, as described in Figure 4 of Chapter 3. From the movies, it appears that the intensity increase is from direct or indirect transcriptional regulation of the transcription factor and not from increased nuclear import. The transcription factors that behave in this manner are Abf1, Cin5, Cti6, Mcm1, Put3, Ste12 and Xbp1.

In other work,[24] the localization dynamics of Hog1 have been characterized as being perfectly adapted in response to salt, i.e. upon salt stress, cells respond with a quick pulse of localization before redistributing back to their basal pre-stressed cytoplasmic state. We reconfirmed Hog1 behavior but did not find any other proteins that behaved similarly in our screen.

Integration of Multiple Bursting Regulatory Systems

These data show that multiple bursting systems are active simultaneously in the yeast cell. In at least some cases, multiple burst systems co-regulate common target genes. These observations provoke the questions of (a) how multiple burst systems combine to regulate target genes and (b) what additional signal encoding and regulation capabilities are provided by such combinatorial burst systems.

We chose to investigate Msn2 and Mig1, two bursting transcription factors that both respond to carbon deprivation. This pair has several ideal features for this analysis: (a) both factors can be induced to burst in the same conditions, (b) both factors can be induced

by the same condition (glucose) allowing analysis of a combinatorial response to a single stimulus, and (c) they regulate both separate and common sets of target genes, allowing analysis of co-regulation. Interestingly, Msn2 also responds to several other stimuli, making it unique amongst the hits in our screen.

Mig1 represses genes involved in metabolism of alternate carbon sources when glucose is present.[56] At high concentrations of glucose it is constitutively nuclear while at lower concentrations, it exhibits bursting behavior. Under similar conditions, Msn2 bursts as it senses insufficient carbon in the media. In the dual-color strain, it is apparent from the traces that both proteins can burst coincidentally (Fig. 1a- at 50 min) but there do appear to be signs of an anti-correlation (Fig reference).

This is different than what we observed in a previous experiment where we measured the cross-correlation of Msn2 and Crz1 in response to calcium.[60] Here we found two largely uncorrelated bursting systems.[60] With Msn2 and Mig1, we find two timescales of correlation. There is a positive correlation between the bursts of both proteins on the timescale of an hour; however, at shorter timescales a lack of correlation becomes evident (Fig. 1b). This lack of correlation is reproducible in several low-glucose conditions, but it is difficult to quantify its effect due to its subtle nature.

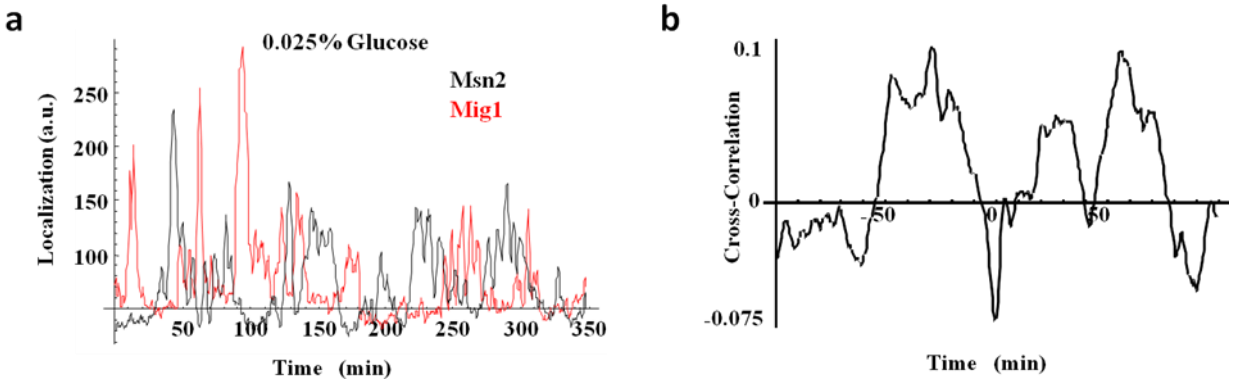


Fig. 1. **Correlations amongst Bursting Transcription Factors**. A strain tagged for both Msn2 (black) and Mig1(red) is imaged at limiting amounts of glucose (.025%). In **a**, the sample trace shows that both proteins bursts on a fast timescale. In **b**, the average cross-correlation of 27 cells shows a positive correlation with a timescale of ~ 1 hour and an anti-correlation with a timescale of ~ 1 minute.

It is unclear whether this correlation function suggests that such an interaction affects localization bursts in any meaningful way. Because there are several similarities between the two proteins: nuclear import of Msn2 and Mig1 are dephosphorylation-dependent and the cytoplasmic redistribution of these proteins is dependent upon phosphorylation by several kinases, one of which is Snf1,[58,59]any interaction between the two may simply be due to utilizing similar enzymes. Furthermore, there are several conditions in which only one of the two proteins exhibits localization bursts. In these cases, the fact that one of the proteins is not bursting does not appear to change the behavior of the other, showing that neither protein is necessary for the other to burst nor does one protein modulate the other's burst statistics.

One possible explanation for this lack of correlation is that the export machinery used by both Msn2 and Mig1to leave the nucleus can be saturated when both proteins leave the nucleus simultaneously. Although this seems unlikely given the typical load these export enzymes are known to carry, the only way to disprove this hypothesis would be to find conditions in which the lack of correlation disappears from the cross-correlation function. After screening through different carbon sources, we have been unable to find such a condition.

We now have two bursting regulatory systems with common target genes. Without understanding the molecular mechanisms responsible for generating the observed cross-correlation function, we can still generate a toy model to understand the potential consequences this can have on downstream combinatorial target genes. Suppose a target gene promoter has binding sites for an activator and a repressor; but neither protein can bind to the promoter simultaneously. If only a repressor were bound, expression would be downregulated significantly while if only an activator were bound, expression would be increased significantly. However, if both an activator and repressor were bound, the change in expression would be tempered. By inhibiting the binding of both factors, the system enables the target promoter to experience the full range of repression and activation (Fig. 2).

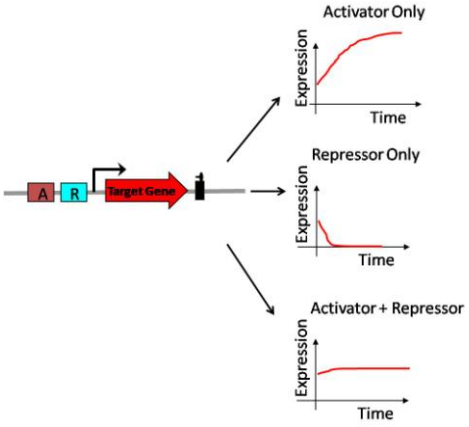


***Fig. 2*. Model of Combinatorial gene Regulation**. A target gene (red) is controlled by an activator (A) and a repressor R. If only the activator were present, expression would increase. If only the repressor were present, expression would be suppressed. If both were present, expression would not change much. By limiting the DNA occupancy on the promoter to only of the two transcription factors, the dynamic range of the combinatorial target gene is increased.

Future Directions

As with any genome-wide experiment, there are always several small-scale follow-up experiments. For the purposes of the screen, our first priority is to take higher time-resolution movies of the proteins classified as bursters to try to quantify their bursts statistics. Additionally, we would like to do more in-depth bioinformatic analysis to understand what properties, if any, distinguish bursting transcription factors from their non-bursting counterparts.

In terms of our analysis of Msn2 and Mig1, there are three important sets of experiments that we will conduct. First, we will continue taking more low glucose movies to get better statistics and quantify the lack of correlation observed at time lag 0. Next, we will screen pharmacological inhibitors of kinases and phosphatases involved in Msn2 and/or Mig1 nuclear localization to see how they affect the cross-correlation between the proteins. These inhibitors have very specific targets, enabling direct mechanistic hypotheses from any observed effects. Finally, we will experimentally verify the combinatorial target gene expression prediction (Fig. 3). Here we will create synthetic promoters with binding sites for both Mig1 and Msn2, integrate them into yeast cells, grow the cells in low glucose +/- glycerol and measure expression of the target promoters. We expect that the dynamic range of the combinatorial promoters is higher in just glucose when compared to cells grown in glucose *and* glycerol.

Discussion

Examination of dynamic localization behavior reveals that yeast cells can process signals in a variety of ways, including modulating the properties of localization bursts, the intensity

of localization events, or the fraction of cells that localize in response to a stimulus. Interestingly, the majority of transcription factors in yeast appear constitutively nuclear. Clearly, in all these cases, signal processing still occurs but cannot be observed through our assay. Perhaps bursting is occurring on promoters as transcription factors are binding and unbinding to DNA binding sites. This is what one would expect in bacteria where nuclei do not exist, but such dynamic regulation can still occur. Emerging optical techniques such as super-resolution may enable characterization of such phenomena.

In-depth analysis of two bursting transcription factors, Msn2 and Mig1, reveals an atypical cross-correlation function in which a mild correlation exists on a long timescale of ~1 hour but is damped at a time-lag of 0 minutes. This lack of correlation is reproducible in several media conditions but disappears in the presence of glycerol. We predict that this lack of correlation enables a higher dynamic range in downstream combinatorial target promoters and plan to experimentally verify this prediction. Such analysis will yield fundamental insight into combinatorial gene regulation as we connect environmental input (carbon source) to transcription factor interaction (Msn2 and Mig1) to understand combinatorial expression. Not only can environmental inputs influence expression affect the interplay between transcription factors, changing the mode of regulation cells employ to control regulons of combinatorial target genes.

Acknowledgements

The entire project was once again a collaboration with postdoc Long Cai. The screen was done with the help of a SURF student Kasra Rahbar. Technician Michele Fontes is helping with the construction of synthetic combinatorial Msn2 and Mig1 target genes.

Appendix A

EPIGENETIC VARIABILITY: YEAST PRIONS

Being interested in different types of heterogeneity in cells, we were intrigued by yeast prions, proteins that change into a self-perpetuating form.[61] The most notable yeast prion is that of Sup35, a translation termination factor (Figure 1).[62] In the prion state, Sup35 can no longer carry out its function with high fidelity and as a result, stop codons are skipped, generating long proteins that may have trouble folding appropriately. The variability that results from this state manifests itself morphologically in genetic background-dependent fashion.[61] Furthermore, this state is reversible chemically, as excess amounts of guanidium hydrochloride revert cells back to the wt non-prion state.[61] We were interested in determining the impact the prion state had on gene expression in single cells, specifically on gene expression noise.

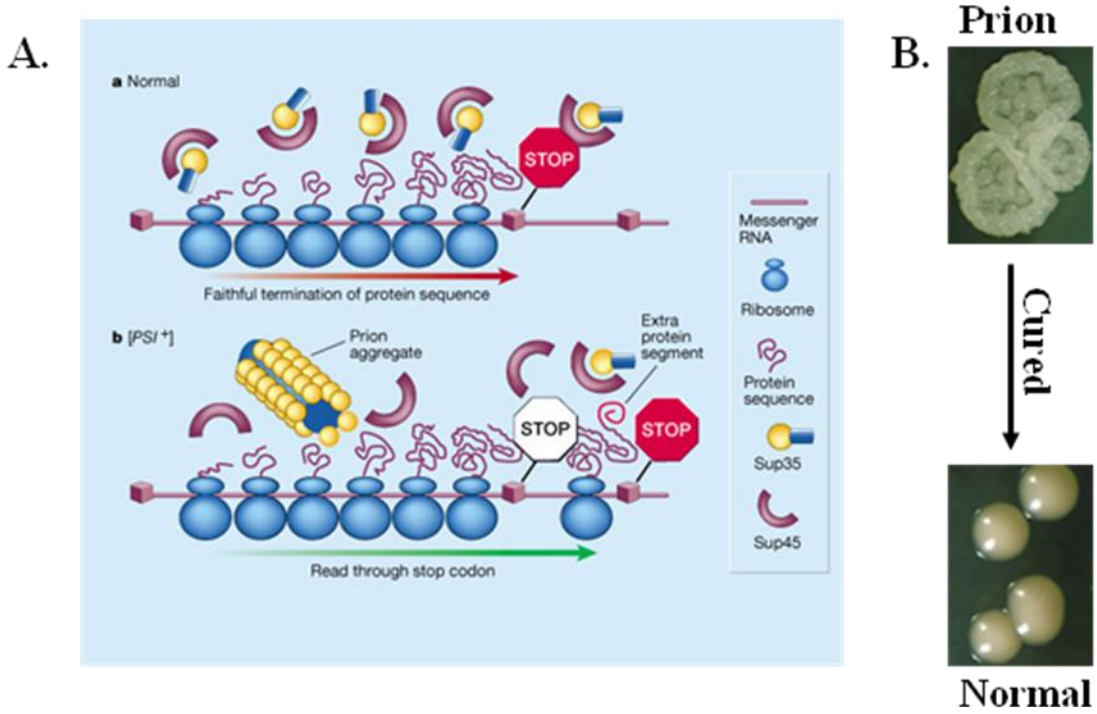


Fig. 1. **The Sup35 protein can self-aggregate and form a reversible prion state**. **a.** This figure, adapted from Partidge and Barton,[58] depicts how Sup35 normally forms a translation termination complex with Sup45, enabling translation to appropriately terminate at stop codons. However, in the prion state, Sup35 aggregates such that the appropriate translation termination complex can no longer form. As a result, stop codons are skipped, creating longer, potentially misfolded proteins. **b.** This state yields morphological changes in yeast strains and is reversible through the addition of excess amounts of guanidium hydrochloride.

We mated in six different reporters into a prion strain and measured gene expression in a flow cytometer, both in prion and non-prion states. The measurements were taken as described in Appendix B, specifically in Figure 13. In two cases, we found large gene expression changes in the prion state and were very encouraged by the results (Figure 2). We thought it might be worthwhile to measure gene expression on a larger scale and began doing bioinformatics to see if we could generate any predictions about the genes we expected to change expression in the prion state.

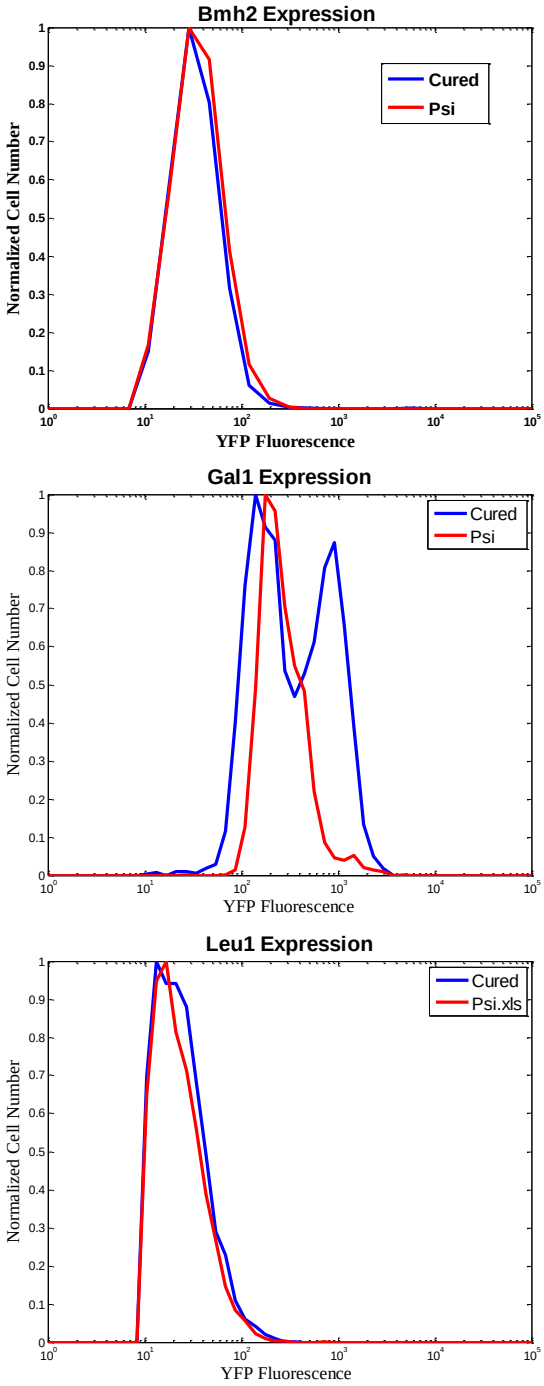
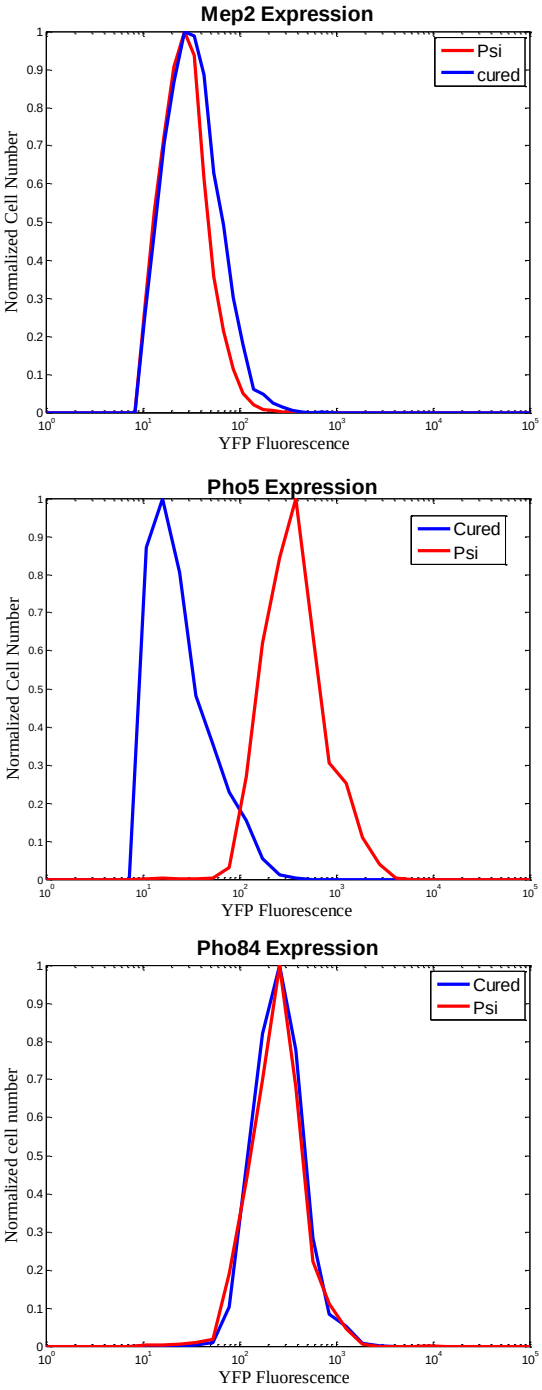
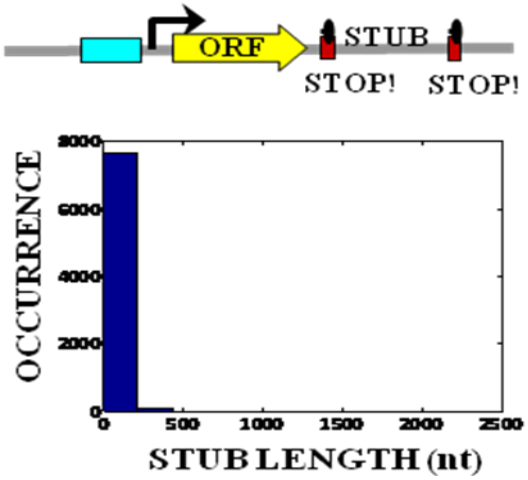


Fig. 2. **Prion and normal expression patterns.** The normalized expression histograms of six different genes are plotted in prion (or Psi) and normal yeast cells. YFP reporters were mated into a prion strain and subsequently cured with guanidium hydrochloride. YFP expression was measured through flow cytometry and cells are normalized by their maximum value. Amongst the six genes, 4 of them show no discernible differential expression across the two epigenetic states. However, normal Gal1 expression is bimodal while the Psi state shows a single peak in expression. Furthermore, Pho5 shows a 40-fold increase in expression when in the Psi state. These results indicate that there certainly are some expression changes in the prion state.



Given the function of the prion state is to generate longer than average proteins, we decided to calculate the extra length of the protein, what we term the stub length. Suppose in the prion state every protein is extended until it meets the next in-frame stop codon. If

this is true, we can calculate the lengths of these "stubs" for every protein. Presumably, proteins with longer stubs, or proteins that interact with proteins with longer stubs, are more likely to have differential gene expression in the prion state. We plotted stub lengths across the yeast proteome (Figure 3) and found most stubs to be incredibly short with an average stub length of 50 nucleotides. Given that we expect these stubs to be functional, we then looked for sequence conservation of the stubs across strains of *Saccharomyces cerevisiae*. Unfortunately, due to high sequence conservation within species, it is very difficult to determine whether 50 nucleotide stubs are more or less conserved than proteins themselves.



Fig, 3. **Stub length in yeast.** Stubs are defined as the length between the original stop codon and the next in-frame stop codon in a protein. In yeast, most stubs are extremely short and on average are only about 50 nucleotides long.

As we were unable to determine whether our gene expression results were simply due to random chance from our choice of genes or whether they were indicative of a broader pattern of gene expression differentiation in an epigenetic state, we decided to stop pursuing this project. It is, however, worth noting, that recently, several other proteins, including general transcriptional activators and repressors, have been identified as forming prions in yeast.[58] Perhaps their cumulative effects are influencing gene expression. Either way, prions remain an intriguing source of variability.

Appendix B

SUPPLEMENTARY MATERIALS: CRZ1 LOCALIZATION

Methods

Strain construction: All GFP strains were obtained from the GFP C-terminal protein fusion library (Invitrogen).[23] All other strains were constructed similarly to the library: Polymerase chain reaction (PCR) was used to amplify cassettes with yeast codon-optimized fluorescent proteins tagged to an auxotrophic marker.[63] These amplified cassettes were made with primers containing 40 bp homology towards the C-terminus of the protein of interest. The cassettes were then transformed into the w303a lab strain of yeast using standard protocols.[64] Transformants were selected using standard yeast media.[65] Plasmid pGW845, a vector containing the FRET cameleon pair under the control of the ADH1 promoter, was provided by Jochen Stadler[29] and p30, the $Crz1^{high}$ mutant with altered calcineurin affinity, was provided by Martha Cyert.[31] Both were transformed directly into a yeast strain containing Crz1-mCherry (Fig. 3b). K. Cunningham provided a Crz1 overexpression plasmid, pLE66,[46] which was transformed into each of 83 Crz1 target strains from the GFP library. A BamHI fragment of yeast codon-optimized Venus-YFP, cut from pKT103,[63] was ligated into pAMS342, pAMS363 and pAMS366, and thus contained provided 1x, 2x, and 4x CDRE's respectively.[36] This, too, was provided by Martha Cyert.

Media: Yeast were grown in synthetic complete or the appropriate drop-out media made using low-fluorescence yeast nitrogen, adapted from previous work.[63] This media is yeast nitrogen base without riboflavin, folic acid, or calcium chloride: 2 g/l $(NH_4)_2Cl_2$, 1 g/l KH_2PO_4, 0.5 g/l $MgCl_2$, 0.05 g/l NaCl, 0.5 mg/l H_3BO_4, 0.04 mg/l $CuSO_4$, 0.1 mg/l KI, 0.2 mg/l $FeCl_3$, 0.4 mg/l $MnCl_2$, 0.2 mg/l Na_2MoO_4, 0.4 mg/l $ZnSO_4$, 2 µg/l biotin, 0.4 mg/l calcium pantothenate, 2 mg/l inositol, 0.4 mg/l niacin, 0.2 mg/l PABA, 0.4 mg/l pyridoxine HCl, 0.4 mg/l thiamine, 20 g/l dextrose.

Time-lapse microscopy: Cells were attached to glass bottom dishes (Willco) functionalized with Concanavalin-A (Sigma). Fluorescence images were taken at room temperature on an Olympus IX81 with the ZDC autofocus option and an Andor Ikon (DU-934) camera. $CaCl_2$ was added during movies to nominal concentrations using a 2X stock solution. Images were acquired at 10 second intervals for single color movies and at 30 second intervals for two-color and FRET movies. Automation was controlled by Andor IQ software. Two-color movies of Crz1 and downstream target genes were taken at 30° to ensure fast maturation of the Venus fluorophore.

Flow cytometry: Yeast cells were grown overnight in 96-well plates and diluted. $CaCl_2$ was added to nominal concentrations with a 2x stock solution. Cells were regrown in the calcium for 4-6 hours, diluted 5-fold, and gene expression was measured using a Beckman Coulter Cell Lab Quanta SC MPL. To identify cells, all particles were triggered and gated during acquisition. As shown in Figure S13, particles were triggered if their Coulter volume measurements were within a specified range. Either 5000 particles or 30 µl sample volume were acquired. These particles were then excited by a 22mW 488nm laser and side-scatter and fluorescence were measured. A polygon gate was drawn on the Coulter volume/Side-scatter two-parameter plot to identify cells. If more than 500 cells were found, mean fluorescence was calculated. Quanta Photomultiplier tube settings were 3.5 for Side-scatter, 8.00 for GFP measurements, and from 4-6 for synthetic promoter-YFP measurements.

Image analysis: Fluorescence cell images were segmented using a Hough transformation algorithm in Matlab, provided by Sharad Ramanathan.[66] Localization score was determined by the difference between the mean intensity of the 5 brightest pixels in the cell and mean intensity of the rest of the pixels in the cell. Bursts were identified by thresholding traces at >1 standard deviations above background noise, estimated from the lowest 20% of values. Subsequent data analysis used the resulting traces with standard routines in Matlab and Mathematica.

Characterization of Crz1-GFP.

Crz1-GFP abundance. Total abundance of Crz1-GFP remained approximately constant during the course of the movies (Fig. S1). Thus, the dynamic features we observe are based entirely on changes in localization.

Crz1-GFP is functional. We verified that the induction of target gene Cmk2 was similar between wild-type Crz1 and Crz1-mCherry strains, suggesting that the fluorescent protein fusion has minimal effect on the transcriptional activities of Crz1. Furthermore, in cells under calcium stress, addition of FK506, a Calcineurin inhibitor, suppresses Crz1-GFP localization bursts (Fig. S8), suggesting that the bursting behavior of Crz1 localization depends on Calcineurin activity. Finally, we grew strains containing either (a) wild-type Crz1, (b) Crz1-GFP, or (c) ΔCrz1 on rich media with 0.8M NaCl and 4mM $MnCl_2$.[31] We found that Crz1-GFP complemented the growth defects found in Crz1Δ. These results suggest that growth and stress response are not compromised by the C-terminal GFP protein fusion.

Characterization of Crz1 bursts.

Distribution of burst number. The total number of bursts per single cell trace is not Poisson distributed (Fig. S4), as determined by the ratio of variance and mean. However, the two distributions of isolated bursts and burst clusters were each individually Poisson distributed, as shown by the close match of their cumulative distribution to that of the Poisson distribution and the low Kolmogorov-Smirnov (KS) scores, whereas the total bursts are less well-fit by Poisson distributions and have higher KS scores (Fig. S4).

Relative frequencies of burst types. Single-cell localization trajectories were thresholded using the mean and standard deviation of fluorescence level in the cytoplasmic state, which is estimated from the lowest 20% of values in a trace. The threshold is the mean plus or more standard deviation of intensity fluctuation in the cytoplasmic state. Errorbars of Fig. 2a indicates the error in estimating the frequency of bursts due to ±30% variations in threshold level. Fluorescence values below threshold are set to the mean intensity of the trace to avoid spurious correlations from low-amplitude fluctuations in the absence of bursting. The autocorrelation function was computed from the thresholded traces and provided an unbiased estimate of the size of the isolated bursts and clusters, which

corresponded to the two time scale of the exponential decay in the fit (Fig. 2d). The relative frequencies of the two types of bursts can be estimated from the relative weights of the two exponential components of the autocorrelation function (Fig. 2d). At concentrations of calcium below 100mM, the average correlation function can be well-described by a single exponential, while at higher calcium levels, two exponentials are needed. In the latter case, the faster of the two timescales matched that obtained in the single exponential fit at low $[Ca^{2+}]$, suggesting that this time-scale corresponds to that of individual bursts. The slower time scale at higher $[Ca^{2+}]$ appears to occur when there is an increase in the frequency of the clustered bursts. Both the value of the slow time scale and its relative weight in the bi-exponential fit increase with $[Ca^{2+}]$, indicating that both the frequency and duration of cluster bursts increases with $[Ca^{2+}]$. This does not affect FM coordination, as the individual localization bursts that make up the clustered bursts retain the same profile across calcium concentrations. Thus, the non-zero portion of the $[Crz1]_{nuc}$ histogram remains the same shape, satisfying Eq. 5 below.

Amplitude variation. In addition to the variability in burst durations, we also found variability in burst heights within a single cell. This variability is unlikely to be due to imaging differences between cells, such as fluctuations in the position of the nucleus with respect to the imaging plane of the microscope, because there is little variability in peak intensity of the initial burst. This is usually one of the strongest localization events in a trajectory, but would be expected to exhibit similar amplitude variability during the initial response if variation in imaging conditions were responsible for amplitude variations. With the characteristic burst duration of ~120 seconds and an acquisition interval of 15 seconds, sampling is sufficiently frequent to capture burst profiles.

Localization score definition. The localization score is not normalized on a 0 to 1 scale because it is difficult to ascertain when full nuclear localization occurs, which is complicated by cytoplasmic autofluorescence and variability in the expression levels of Crz1-GFP in individual cells. Consequently, for simplicity, localization score is shown as extracted from raw data, without normalization.

Crz1 mutants. As shown in Fig. S8 and discussed in the main text, the localization burst frequency is modulated by Crz1 mutations with different affinities for Calcineurin.[31] GFP-

Crz1^{high} exhibits a much higher burst frequency than Crz1-mCherry, while Crz1-mCherry is comparable in burst frequency to a mutant with slightly lower affinity for Calcineurin, GFP-Crz1^{low} (Fig. S8 a,b). It seems that GFP-Crz1^{wt} has a slightly higher burst frequency than Crz1^{wt}-mCherry (Fig S8 c,d), but the effect is much weaker than that of the difference between GFP-Crz1^{high} and Crz1^{wt}-mCherry. In all three cases, it is exclusively the frequency of bursts that is affected, and never the burst amplitude. FK506 affected burst dynamics in a manner similar to the calcineurin affinity mutations. Frequency modulation of Crz1 bursts is further supported by results showing that expression from 1x, 2x and 4x CDRE promoters remains proportional (coordinated) with respect to FK506 concentration at 100mM Calcium (Fig. S8g).

Cell cycle dependence and daughter cells

To monitor the cell cycle dependence of the Crz1 bursts, we observed Crz1-GFP and Whi5-mCherry, a cell cycle marker, simultaneously As described previously, Whi5 is nuclear localized during late G1 phase and relocalizes again at end of mitosis in both the mother and daughter cells,[67] thus serving as a high resolution cell cycle phase indicator. When we cross-correlated the localization events between Whi5 and Crz1 (Fig. S5), we observed only a weak correlation (<10%) between Whi5 and Crz1 activities, indicating that Crz1 bursts do not occur preferentially at particular points in the cell cycle, especially not around the end of mitosis. In addition, we found little correlation (<10%) of the Crz1 bursting events between mother cells and daughter cells, suggesting that after division, Crz1 activity becomes completely independent between the mother and daughter.

FRET measurement of calcium fluctuations

Calcium oscillations have been observed in many mammalian systems, but have not to our knowledge been previously observed in yeast.[29] We measured yellow Cameleon YC2.12,[68,29] developed by R. Tsien and colleagues, and Crz1-mCherry simultaneously in the same cell. We observed both isolated calcium spikes occurring on the timescale ~20-60 seconds and also more prolonged spike trains (Fig. S6). Using 2-second acquisition intervals, we found that the majority of spikes occur on timescales longer than 30 seconds, indicating that our sampling rate is sufficient to capture most calcium spikes (Fig. S7). Every calcium spike coincides with a Crz1 localization event, but not vice versa. However,

due to low signal-to-noise ratio of the FRET reporter, we are unable to determine if low amplitude calcium fluctuations contribute to localization bursting activities. When we looked at the calcium spike dynamics as a function of calcium concentration, we found that the number of calcium spikes per trace does not follow the Crz1 burst frequency curves (Fig. S7c).

Msn2 Bursts

Msn2-GFP exhibits nuclear localization bursts under calcium stress (Fig. S9). After the addition of 200mM [Ca^{2+}], Msn2-GFP localized in a synchronized manner across a population and then exited the nucleus in ~10 minutes. Sustained bursting of Msn2 occurred in a manner similar to Crz1. The burst statistics of Msn2 (Fig. S9) shows that the nuclear residence time of Msn2-GFP is approximately 1-2 minutes, similar to that of Crz1. Furthermore, Msn2 also exhibits clustered bursts. However, other than the initial burst in response to calcium addition, the localization events were only weakly correlated between Msn2 and Crz1. As Msn2 bursts occur more frequently under calcium stress than those of Crz1, for each Msn2 burst, there is an 18.4±2.0% chance of a Crz1 burst occurring at the same time, whereas for each Crz1 burst, there is a 29.7±3.5% chance of coinciding with an Msn2 burst. In contrast, if the two proteins were localized entirely independent, the coincidence probability would have been 14.0±1.2% and 22.6±1.8% respectively (error bar determined by bootstrap). This is consistent with the weak but non-zero cross-correlation between Crz1 and Msn2, showing that while two types of bursts are predominantly independent of each other, there is a statistically significant correlation. Note that Crz1 initial localization was observed at 750mM, but not at 300mM of NaCl. Under these conditions, Msn2 shuttles in and out of the nucleus, indicating that Msn2 is also capable of nuclear localization bursts without being driven by calcium spikes.

Cmk2 in Crz1Δ strain

Cmk2-Venus does not induce in the presence of extracellular calcium in a Crz1Δ strain. The lack of induction is consistent with previous microarray results, where Cmk2 and other Crz1 target genes were not induced in the presence of the calcineurin inhibitor FK506.[22,32] We repeated this experiment at the single-cell level to select appropriate target genes (see below). The Cmk2 FK506 data is specifically plotted in Fig. S15a.

Crz1 Localization, Target Promoter Cross Correlation Analysis

The cross-correlation between the Cmk2-Venus production rate and the Crz1-localization trajectories shows a positive correlation with a peak amplitude of 0.2 and a time-delay of 15 minutes corresponding to the maturation time of the Venus fluorophore. This does *not* mean that only 20% of transcriptional noise is due to localization dynamics. Consider a scenario in which maturation time profile is a square pulse lasting 10 times longer than the localization bursts. Assuming the localization bursts are essentially delta functions, the cross-correlation follows the same profile as the maturation time profile. However, the peak amplitude of the correlation is not 1. By definition, the cross-correlation is normalized by the standard deviation of each component. Consequently, the correlation amplitude is the square root of the duration of the maturation, $1/\sqrt{10}$. Thus, the 0.2 cross-correlation between the Cmk2 promoter activity and the Crz1 localization trace should be re-normalized by the maturation time distribution, resulting in a higher correlation coefficient (>50%) between the two dynamic variables. Incomplete correlation could also result from the observation that not all Crz1 localization bursts result in corresponding transcriptional bursts (Fig. S11). The probabilistic nature of transcriptional initiation would further lower the correlation amplitude, again by the square root of the average number of Crz1 localization bursts that occur per expression event.

Maturation time of the Venus fluorophore.

We measured the maturation time of the Venus fluorophore at 30^{o}C on a spectrophotometer (TPI). A 2 ml culture of Hsp12-Venus yeast cells in SC media was induced with 400mM NaCl for 7 minutes and cycloheximide (Sigma) was added to a final concentration of 100 ug/ml to stop translation. Fluorescence traces are shown in Fig. S12. Maturation time was calculated from the derivative of the fluorescence trace, starting from the point of cycloheximide addition, producing a mean value of 15 min.

FM- Coordination

Consider two Crz1 target promoters, A and B, whose normalized rates of expression are Hill functions of nuclear Crz1 concentration, denoted x:

(1) $R_i(x) = \frac{x^{n_i}}{x^{n_i} + K_i^{n_i}}$.

Each promoter is thus characterized by half-maximal activation level K_i and Hill coefficient n_i ($i = A$ or B). Here we show that in FM-coordination, the mean promoter expression levels, P_i, are regulated proportionally: $P_A = kP_B$, where *k* is a constant, at all calcium levels, regardless of K_i and n_i. In the AM model, on the other hand, P_A and P_B in general depend differently on calcium concentration.

To proceed, consider the histogram of x at different calcium levels, $h(x, \mathrm{Ca})$ (Fig. 4). In the AM model, the center of the histogram varies with calcium concentration. In the FM model, on the other hand, the histogram has two components: a peak at high *x* due to bursts, and a region at low *x* in the resting state between bursts. Only the relative amplitude of the two peaks, but not their positions, varies with calcium. More generally, if the histogram is not bimodal, it is only important that the shape of the nuclear localization histogram remain the same while its amplitude change relative to the "off" cytoplasmic state. We assume that only the higher peak contributes to expression. As a result,

(2) $h_{FM}(x, Ca) = f(Ca)g(x)$,

where $f(Ca)$ is the burst frequency as a function of extracellular calcium, and $g(x)$ is the histogram during bursts.

The mean expression level depends on the product, or overlap, between the rate of expression and the histogram of nuclear Crz1:

(3) $\overline{P_i(\mathrm{Ca})} = \int R_i(x)h(x, \mathrm{Ca})dx$.

In AM regulation, the expression ratio of the two promoters, denoted r_{AM}, may depend in a different way on calcium and hence their ratio is calcium-dependent:

(4) $r_{AM} = \frac{P_A(Ca)}{P_B(Ca)} = \frac{\int R_A(x)h_{AM}(x,\mathrm{Ca})dx}{\int R_B(x)h_{AM}(x,\mathrm{Ca})dx}$.

In FM regulation, on the other hand, the calcium-dependence appears only in the burst frequency, and factors out of the expression ratio, producing a constant expression ratio, r_{FM}:

(5) $r_{FM} = \frac{f(Ca)\int R_A(x)g(x)dx}{f(Ca)\int R_B(x)g(x)dx} = \mathrm{const}$.

The result is easiest to see in the extreme limit, when $\mathrm{Crz1}_{\mathrm{nuc}}$ saturates both target promoters during each localization burst ($x \gg K_A, K_B$). In this case, P_A and P_B are each proportional to the fraction of time Crz1 spends in the nucleus, and hence proportional to

each other and to the frequency, f, of Crz1 bursts. Hence their expression ratio P_A/P_B remains fixed as extracellular calcium varies.

Frequency modulation of Crz1 bursts is not the only mechanism to ensure coordinate regulation of downstream targets. FliCR requires the shape of the localization histogram be independent of calcium (Eq. 2), but this could be achieved, for example, by varying the duration of localization bursts, rather than the frequency, similar to Pulse Width Modulation (PWM) in engineering. Furthermore, the clustering of bursts does not affect FliCR, as the cluster is simply an aggregate of individual bursts. Finally, we note that the argument does not depend on the Hill function itself, and FM coordination can work for more complex promoter response functions.

Flow Cytometry Selection Criterion

Of 163 genes previously shown to be Crz1-dependent,[22] 83 had GFP-fusions available.[23] Of these, 19 were excluded because of low-induced dynamic range and the resulting difficulty of distinguishing calcium-dependent induction from cellular autofluorescence. We tested the Crz1-dependence of the remaining genes by comparing their response to high concentrations of extracellular calcium with and without FK506, a calcineurin inhibitor (Fig. S15). The induction of 40 genes was completely suppressed by FK506, suggesting their expression is predominantly controlled by Crz1. We measured the expression level of each of these 40 genes under 8 concentrations of extracellular calcium (0, 25, 50, 100, 150, 200, 300, 400 mM) using flow cytometry. In the main text, we only show traces up to 300mM for consistency with other experiments. We repeated the experiments measuring the calcium induction of these genes on four separate days. The runs with low cell density (<500 cells per run) were discarded. The remaining runs were normalized according to the following procedure: Each induction curve was fit by least squares (Mathematica) to the CDRE induction curve with two free parameters: offset (minimum expression) and gain (induction range). All normalized curves thus extended from 0 at 0mM [Ca^{2+}] to 1 at 400mM [Ca^{2+}]. For each gene, the means and error bars of the normalized expression levels were computed from replicate experiments on days with sufficient cell counts.

Out of the 40 genes that passed the FK506 test, 34 showed induction (Fig. S14a) curves that closely match those of burst frequency and CDRE dependence on calcium. The remaining six (Fig. S14b) showed different induction profiles. Note that other regulatory mechanisms on the target promoter, such as indirect regulation, can interfere with coordination.

We observed 24 genes whose expression under calcium stress was not completely suppressed with FK506 (>10% induction is independent of calcineurin). In general, the induction curves for these genes depart from the Crz1 burst frequency curve (Fig. S14c). This is most likely due to contributions that are calcium-dependent, but not Crz1-dependent. Despite this, Crz1 still plays a significant role in the induction of these genes, as reflected in the similarity of some of these induction curves to the CDRE curves. The difference in target gene sets was due to the fact that our FK506 selection criteria were more stringent than that used in the previous microarray study, that had identified the larger set of target genes initially.

Finally, we examined genes that were independent of Crz1, but are induced in the presence calcium. As expected, they showed different induction profiles compared to pure Crz1 target genes (Fig. S18)

Over-expression of Crz1 rules out "fine-tuned" promoter hypothesis

In principle, coordination could be achieved by 'fine-tuning' of promoter input functions: If all target promoters had input functions that were exactly proportional to each other across $Crz1_{nuc}$ levels, then the target genes would be trivially coordinated. This possibility can be ruled out by considering the response of promoters to changes in the total Crz1 concentration in the cell (Fig. S16).

The 1x, 2x and 4x CDRE promoters exhibit different absolute induction values, but maintain the same normalized induction profiles. Given that all three constructs are driven by the same Cyc1 promoter and differ exclusively in the number of CDRE binding sites,[36] it is unlikely that the increase in expression level occurs without changes in the effective affinity of the promoters. To prove that this is indeed the case, we performed the same induction experiment while overexpressing Crz1.[46] $Crz1^{over}$ also exhibited localization bursts in the presence of calcium (Fig. S16b). Synthetic promoters induced with the same

induction profile at wild-type levels Crz1 and when Crz1 was overexpressed (see below). However, their absolute expression levels differed (Fig. S16). When Crz1 is overexpressed, the histogram of $[Crz1]_{nuc}$ during a burst shifts to a higher level, resulting in higher overall expression from the promoters. If all three synthetic promoters have the same response function to Crz1, then the induction ratios between the $Crz1^{over}$ and $Crz1^{wt}$ would be identical for all three constructs. However, if the response functions differ among the promoters, the fold change of the overlaps between the $Crz1_{nuc}$ histogram and the response function would be different for the different promoters in overexpressed vs. wild-type Crz1 abundances. This is what we observe experimentally (Fig. S16 c,d,e,f): different synthetic promoters depend differently on total Crz1 concentration. The observed induction is 7-fold for 1x and 2-fold for 4x promoters.

Furthermore, note that the larger fold change observed for the 1x promoter may be explained by the following argument. Since the 1x CDRE promoter presumably has a lower affinity to Crz1 than the 4x promoter, the input function for the 1x promoter saturates at a higher concentration of Crz1 than the 4x function. As a result, the wild-type concentration of Crz1 may already be close to saturating the 4x promoter, but not for the 1x promoter, possibly accounting for the larger fold of induction of the 1x promoter.

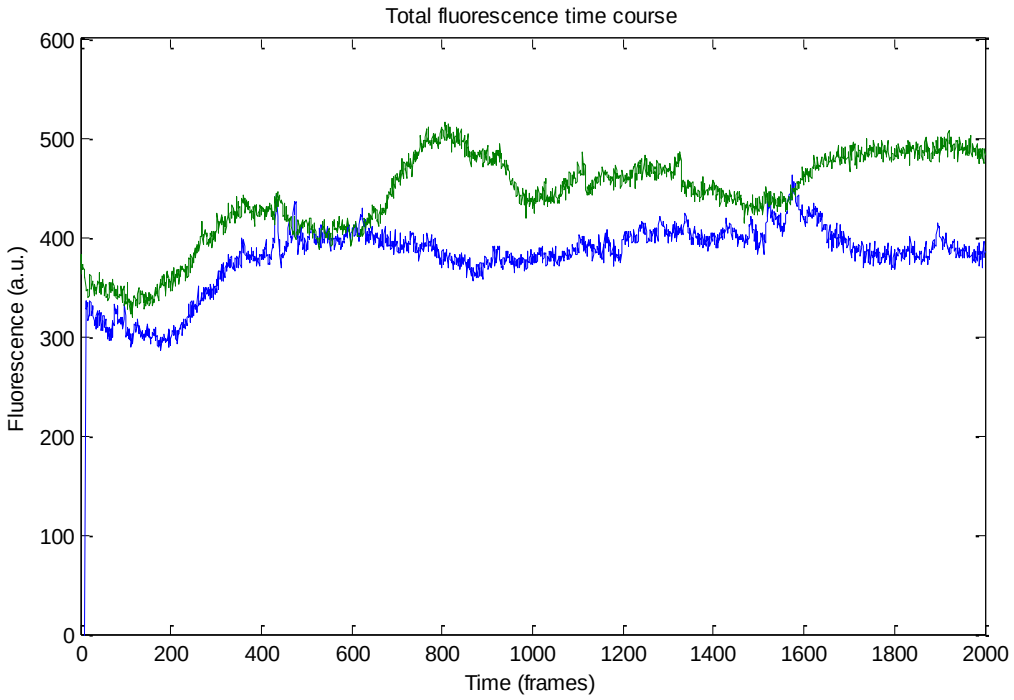


Fig. 1. Total Crz1-GFP fluorescence in two cells over the course of a 6 hour movie. 150mM calcium was added at frame 20. Interval between each frame is 10 seconds.

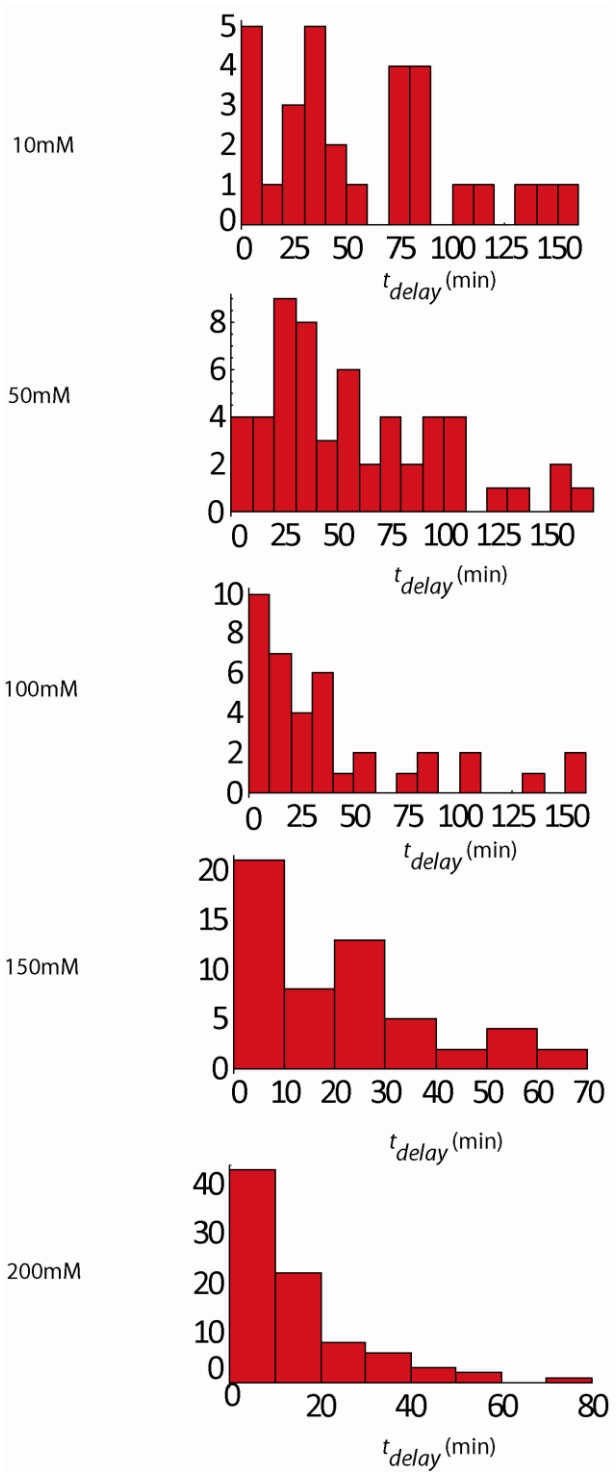


Fig. 2. Histogram of delay times between addition of calcium and the initial Crz1 localization burst, τ_{delay}.

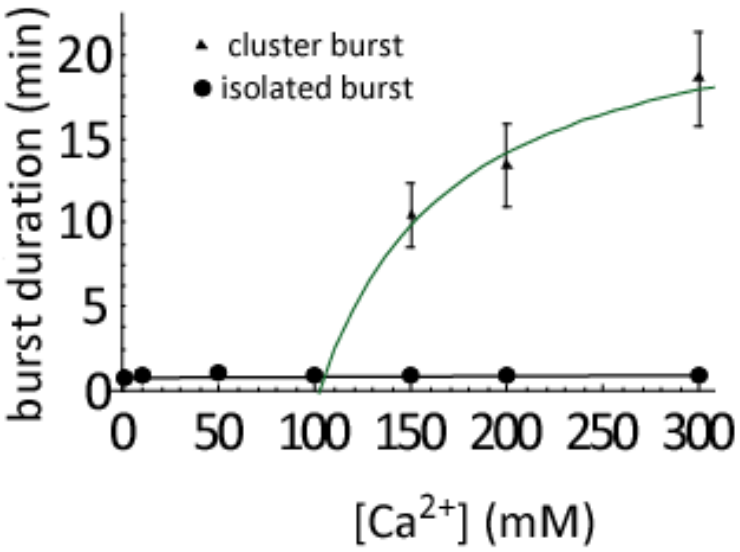


Fig. 3. Timescales, τ_{burst} and $\tau_{cluster}$, of isolated (black), and clustered (green) bursts, respectively, at varying calcium concentration. Values were determined by a fit of the autocorrelation function to $A_1 e^{-\left(\frac{t}{\tau_{burst}}\right)} + A_2 e^{-\left(\frac{t}{\tau_{cluster}}\right)}$. Error bars represent sampling errors were calculated using bootstrap.

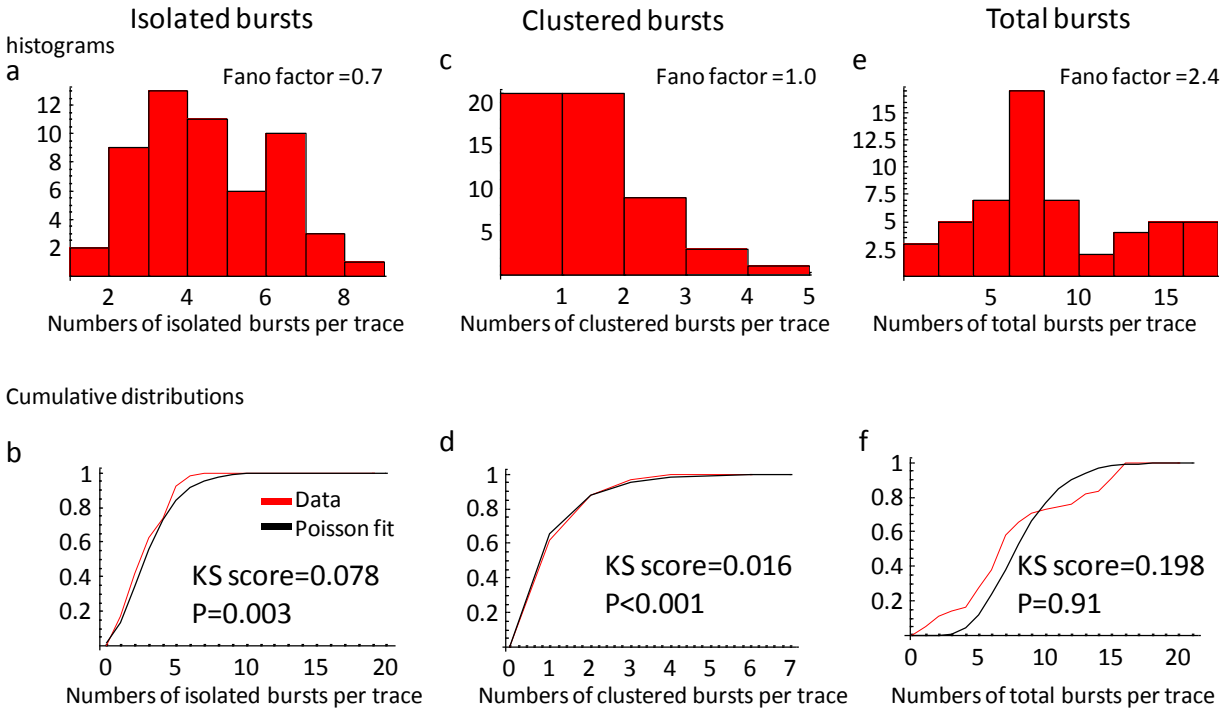


Fig. 4. Distributions of number of bursts per trace. Histograms and cumulative distributions of numbers of bursts from a Crz1-GFP movie at 150mM Calcium. Total number of burst per trace is not well-fit by a Poisson distribution, but the numbers of isolated and clustered bursts each fits well to Poisson distribution. Fano factor is the ratio between the variance and the mean of the distribution. The statistics is determined by the Kolmogorov-Smirnov test with the null hypothesis that the data distribution and Poisson distribution are dissimilar.

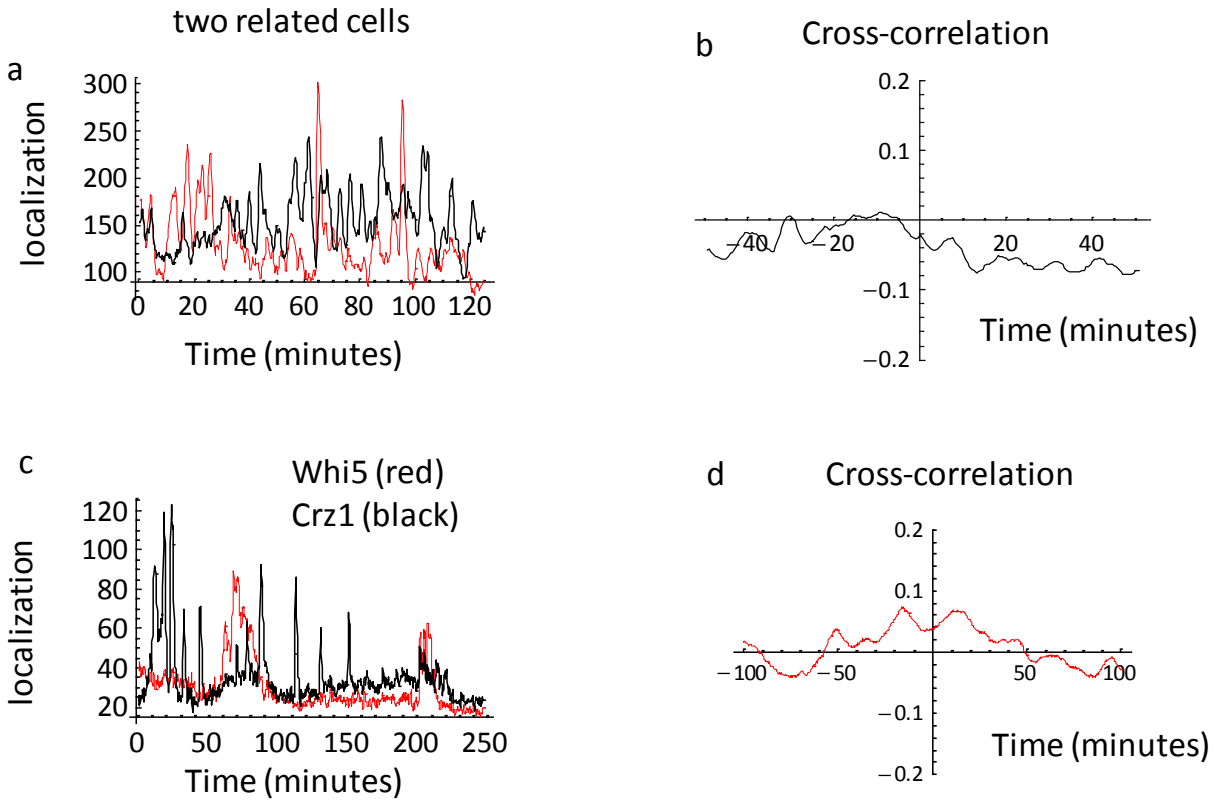


Fig. 5. Cell cycle and daughter cell analysis. **a.** A time-trace of a mother cell and daughter cell (black and red) and their **b.** cross-correlation. **c.** A time-trace of a cell containing cell cycle reporter Whi5-GFP (red) and Crz1-mCherry (black) and their **d.** cross-correlation.

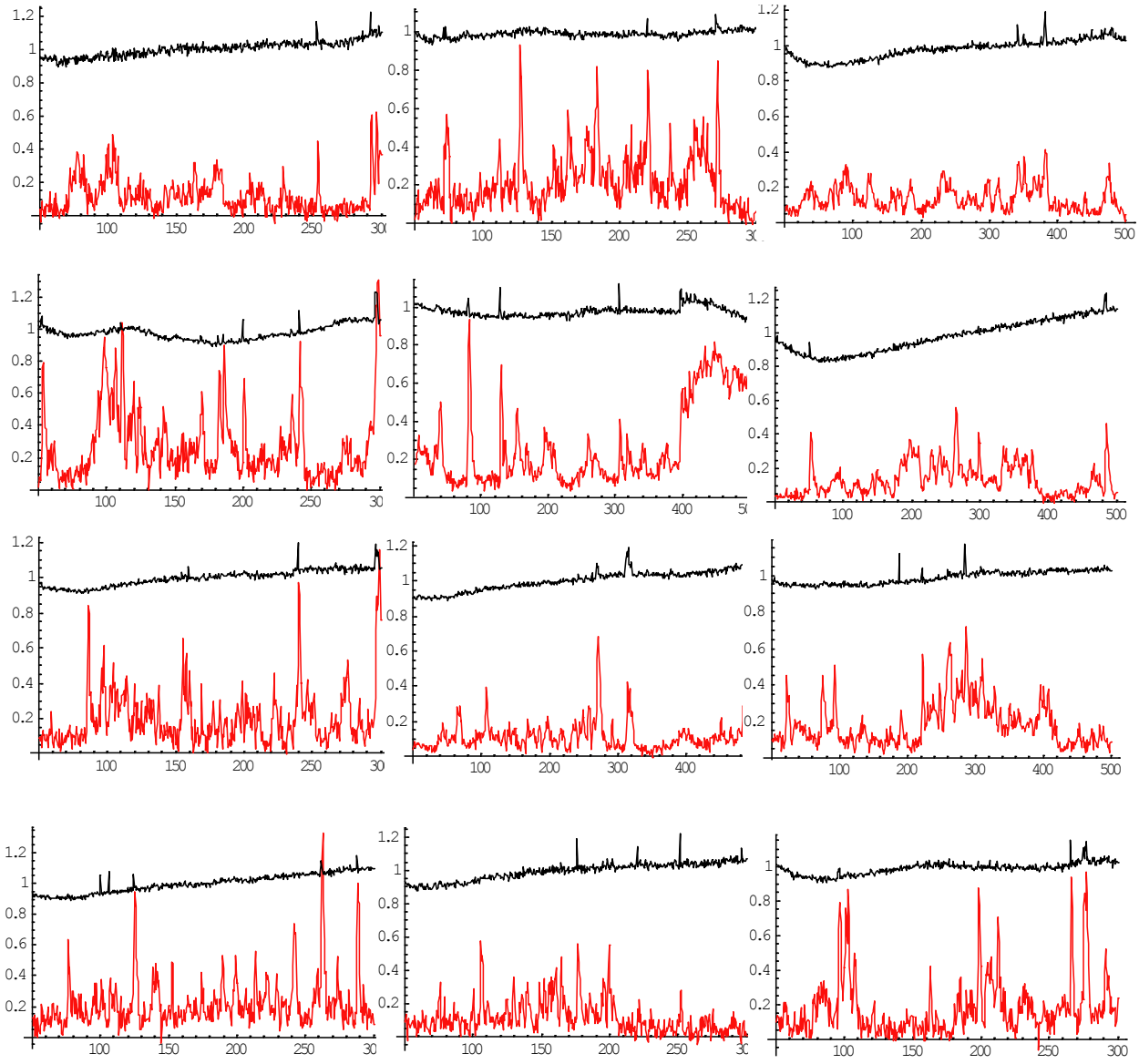


Fig. 6. Crz1 localization (red) and FRET ratio (black) traces in single cells at 200 mM extracellular Ca^{2+}. Time is on the x-axis in frames, separated by 30 seconds.

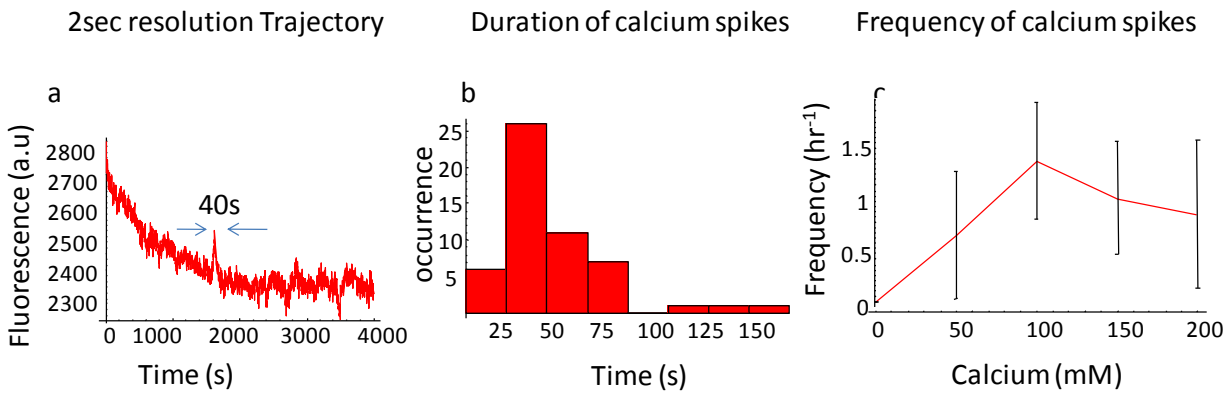
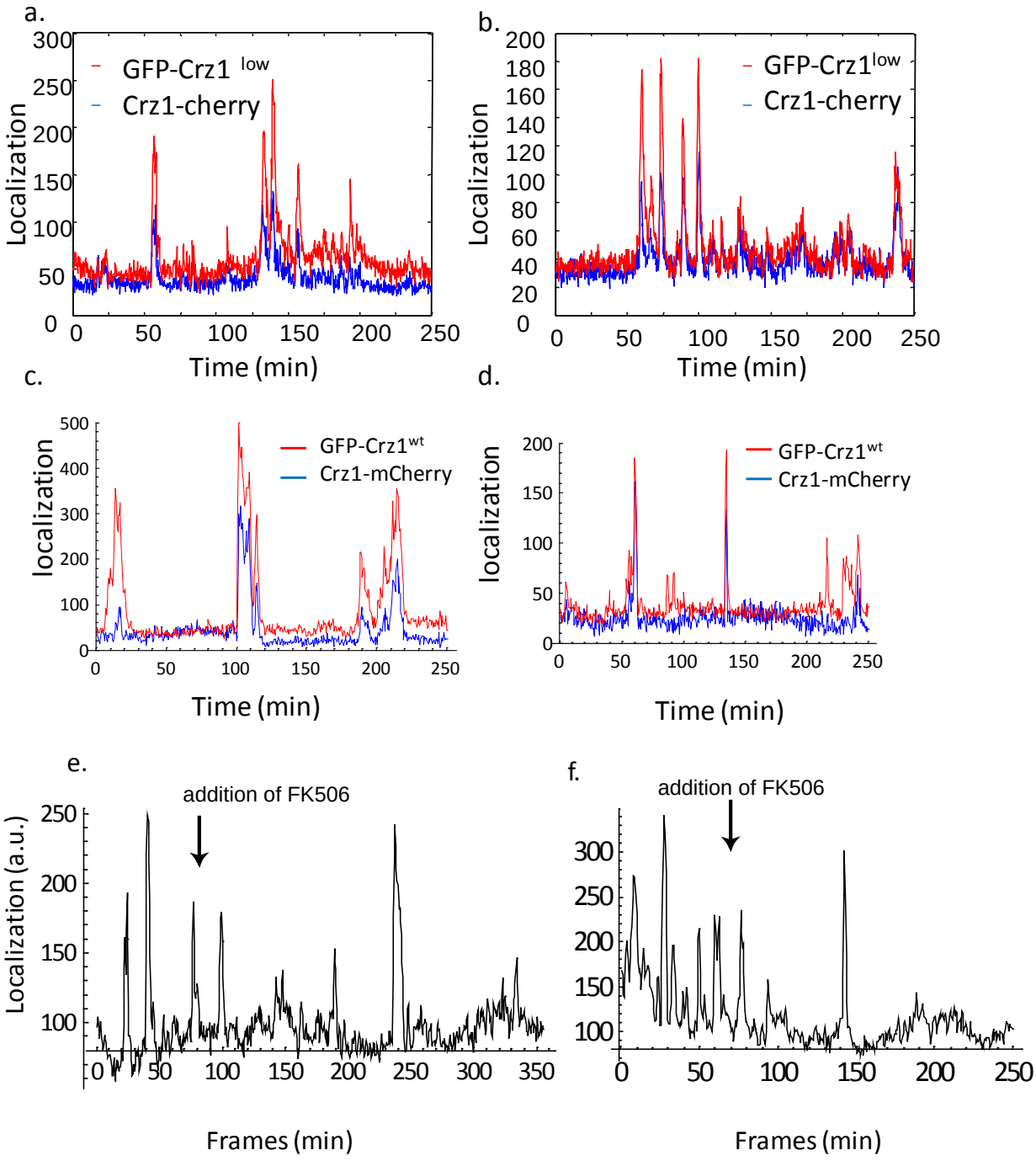


Fig. 7. **a.** Typical FRET trace acquired with a 2 second frame rate showing a calcium spike of ~40 second duration. **b.** Distribution of calcium spike durations from movies like the one shown in (a). Note that mean duration is ~38 seconds, and the distribution is peaked around this value. **c.** Frequency of calcium spikes as a function of extracellular calcium concentration. y-axis units are number of spikes per hour. Error bars denote standard deviation in the number of bursts in one trace.

Figure 8


a.
GFP-Crz1 low
Crz1-cherry
Localization
Time (min)
b.
GFP-Crz1low
Crz1-cherry
Localization
Time (min)
c.
GFP-Crz1wt
Crz1-mCherry
localization
Time (min)
d.
GFP-Crz1wt
Crz1-mCherry
localization
Time (min)
e.
addition of FK506
Localization (a.u.)
Frames (min)
f.
addition of FK506
Frames (min)

g

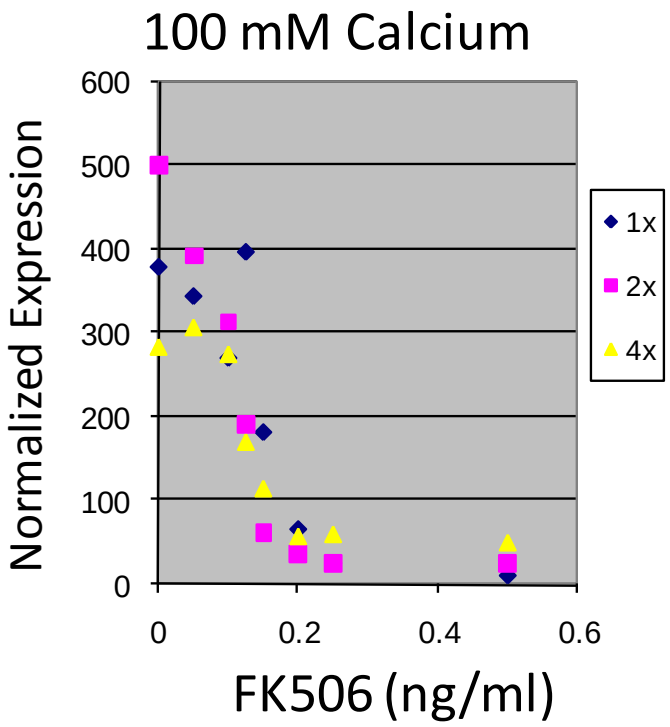


Fig. 8 (continued). **a,b.** Sample traces of Crz1-mCherry and GFP-Crz1^{low} in the same cell. Note that the amplitude differences between the two colors are due to differing fluorophore brightness. **c,d.** Sample traces of Crz1-mCherry and GFP-Crz1^{wt} in the same cell **e,f.** Sample traces of calcium-induced Crz1-GFP after the addition of sub-saturating amounts of FK506. Note that spikes of similar amplitude and duration still occur. **G**. Flow cytometry analysis of the three synthetic promoters induced with 100 mM Calcium and varying amounts of FK506. Note that the three curves collapse onto each other, consistent with frequency-modulation, rather than amplitude modulation, of bursts by FK506.

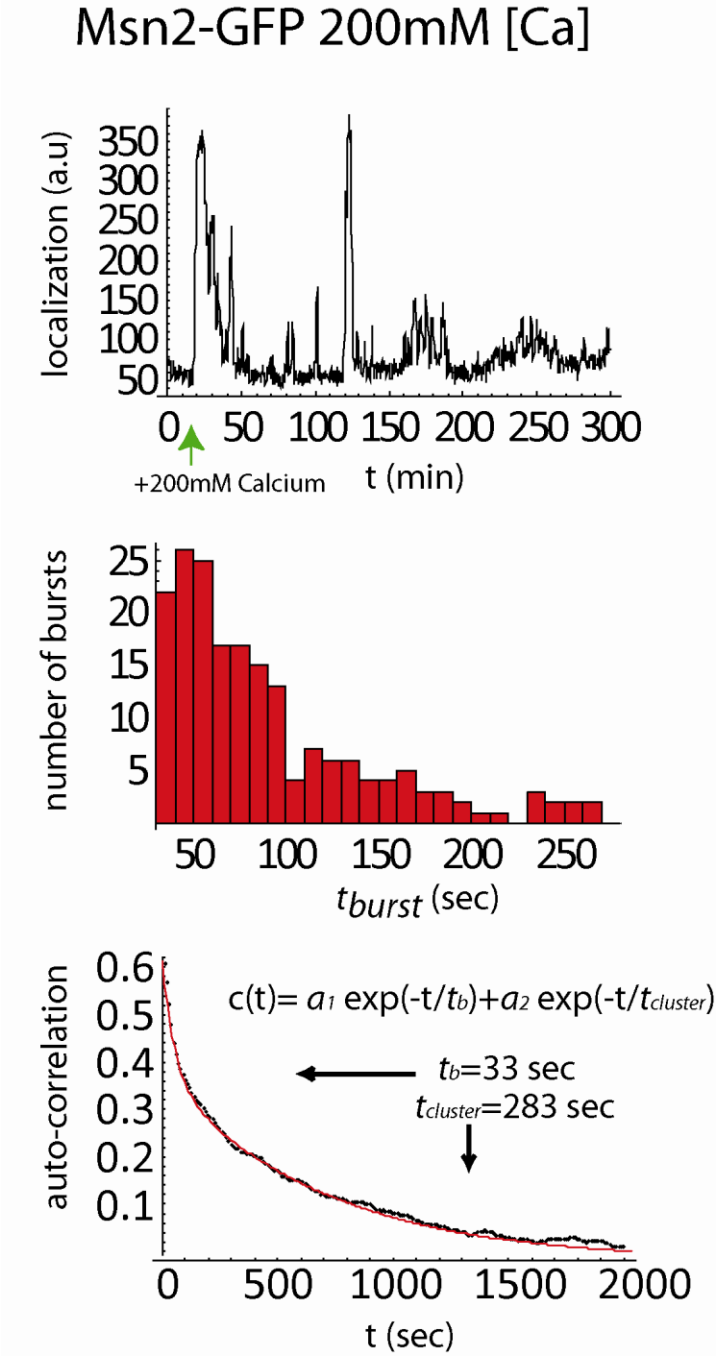


Fig. 9. Msn2-GFP bursts in presence of C=calcium. The histogram of Msn2 burst duration is similar to that of Crz1, with the majority of the bursts lasting between 0.5 to 2 minutes. Autocorrelation shows two timescales of exponential decay representing isolated bursts and clustered bursts. Data is shown in black with two exponential fit shown in red.

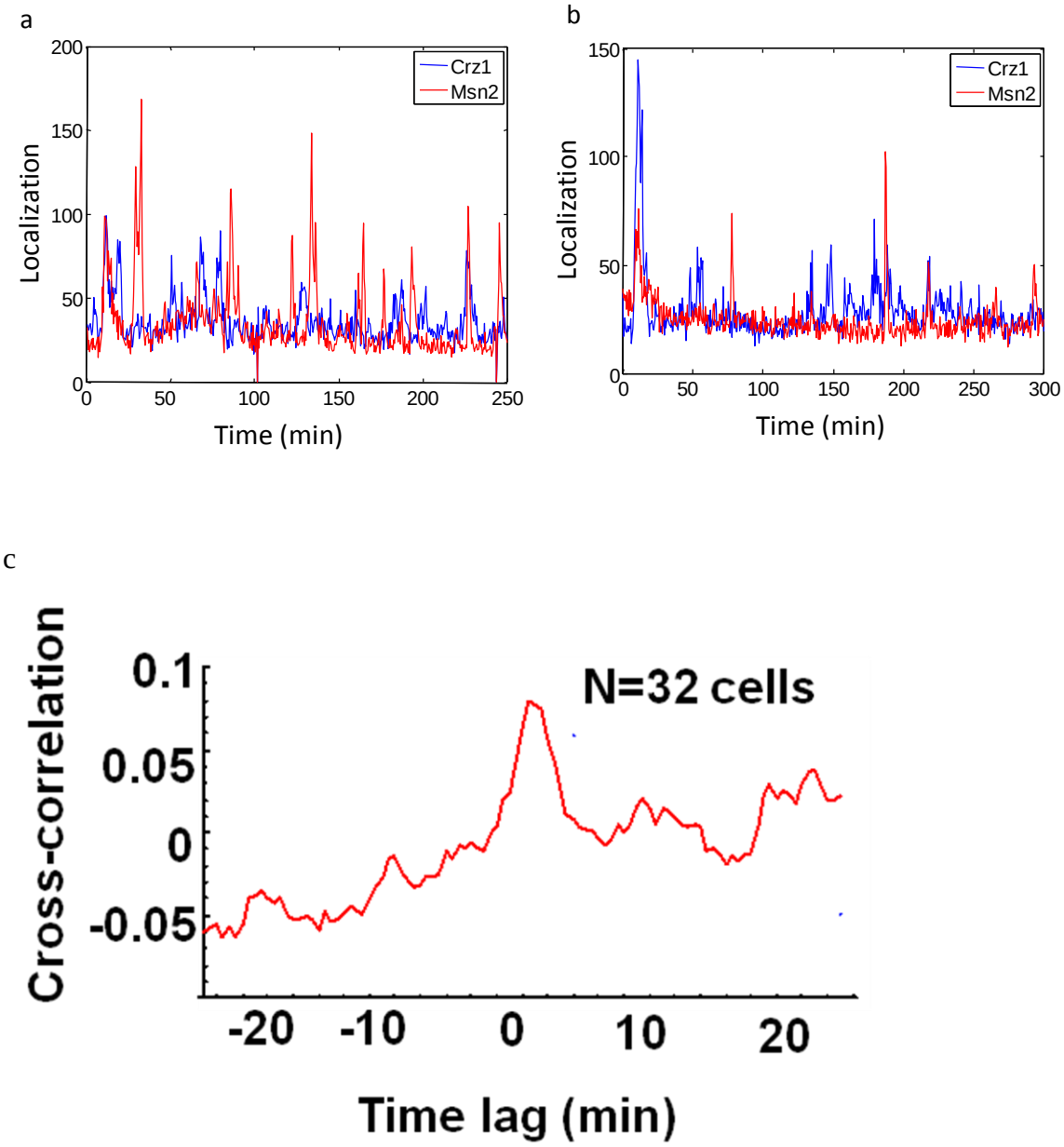


Fig. 10. **a,b.** Sample traces of cells containing Msn2-GFP and Crz1-mCherry. **c.** The cross-correlation of Msn2-GFP and Crz1-cherry at 150mM calcium. Note weak correlation (<0.1) peak at 0 time lag.

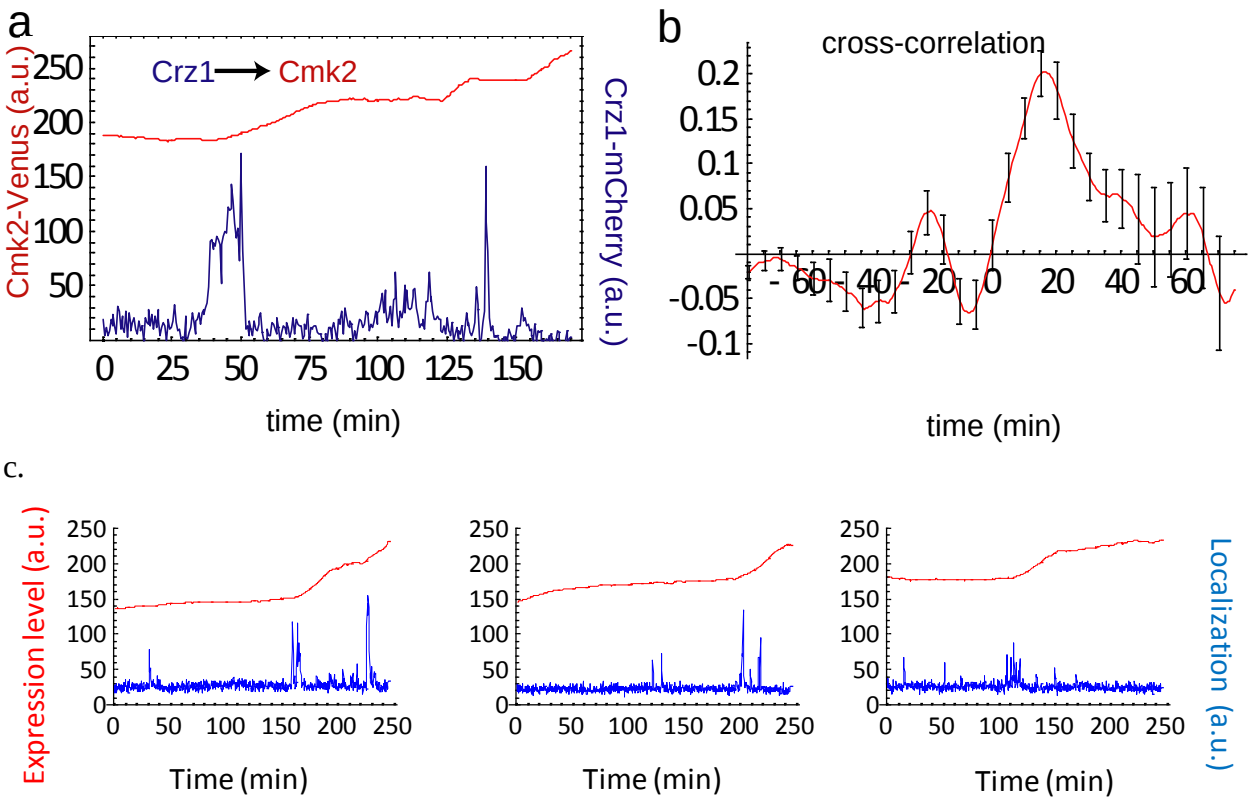


Figure 11. Propagation of Frequency-Modulated localization bursts. **a.** Sample trace of Crz1 localization (blue) and Cmk2 fluorescence (red). **b.** Cross-correlation of Crz1 localization events with Cmk2 expression. **c.** Sample traces of Crz1 localization and 2x CDRE expression.

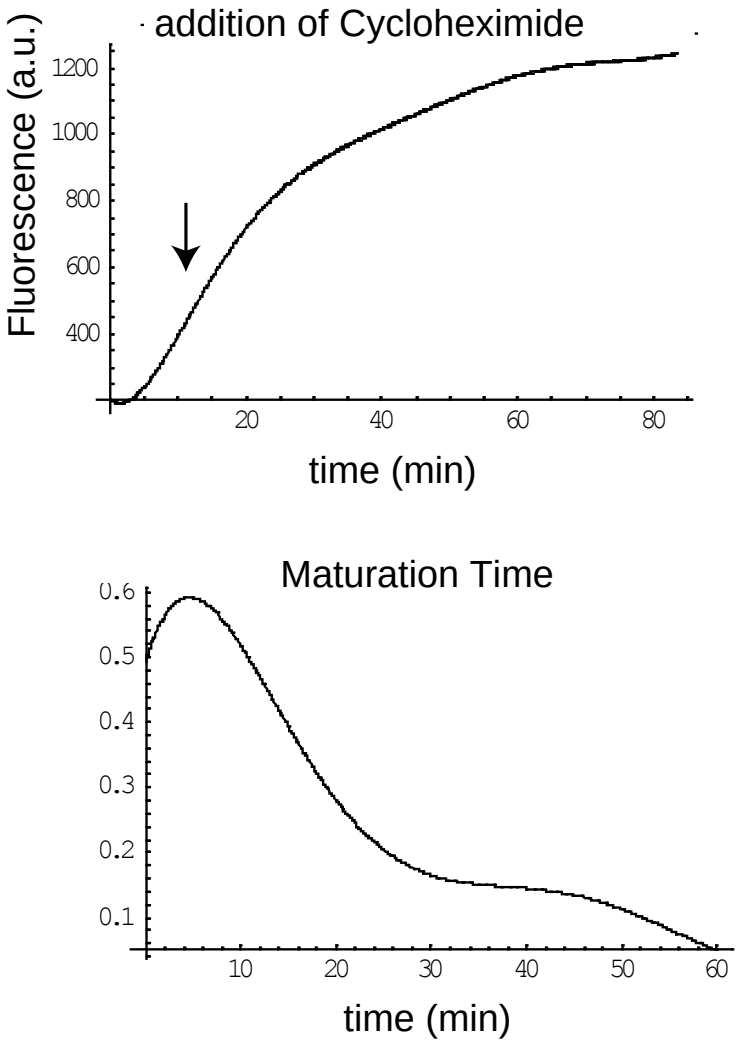


Fig. 12. Measurement of maturation time of HSP12-Venus at 30°C. Bottom figure is the time derivative of the top figure, beginning with the time point corresponding to the addition of cycloheximide. A population of cells was measured in a cuvette with a spectrophotometer.

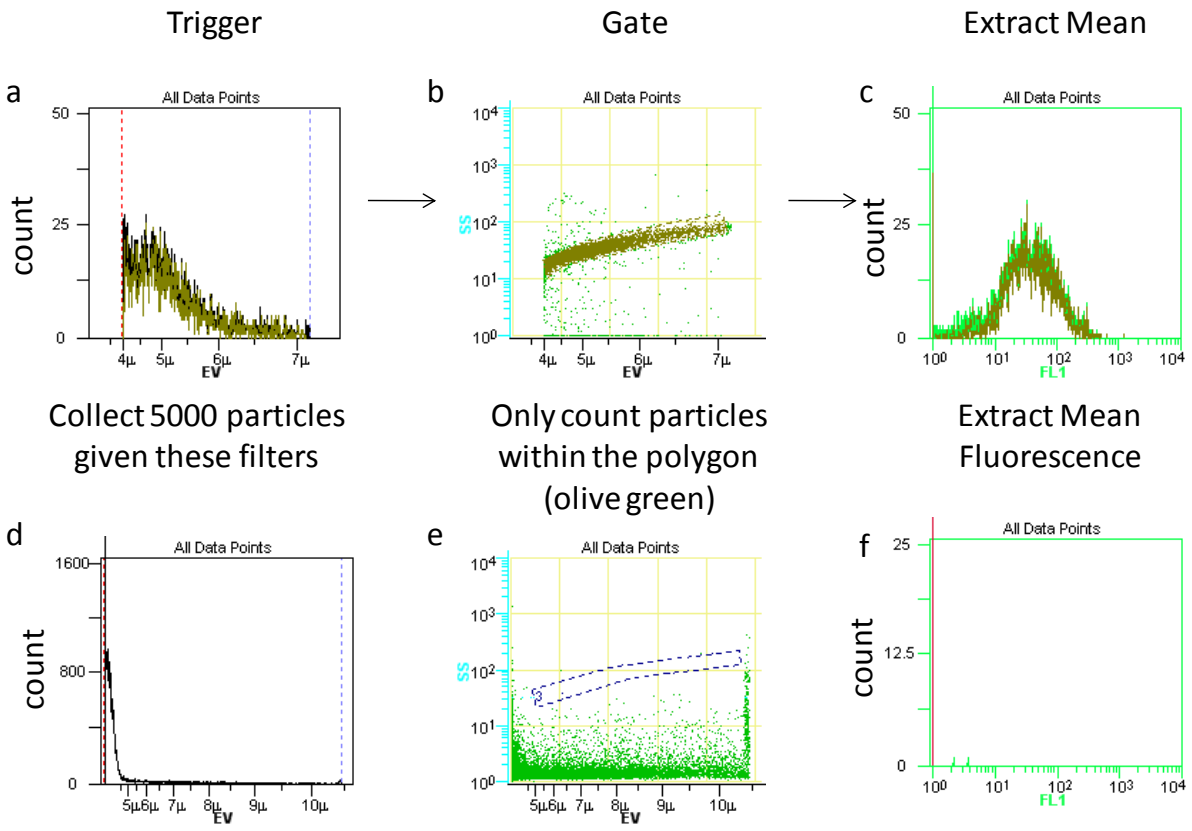
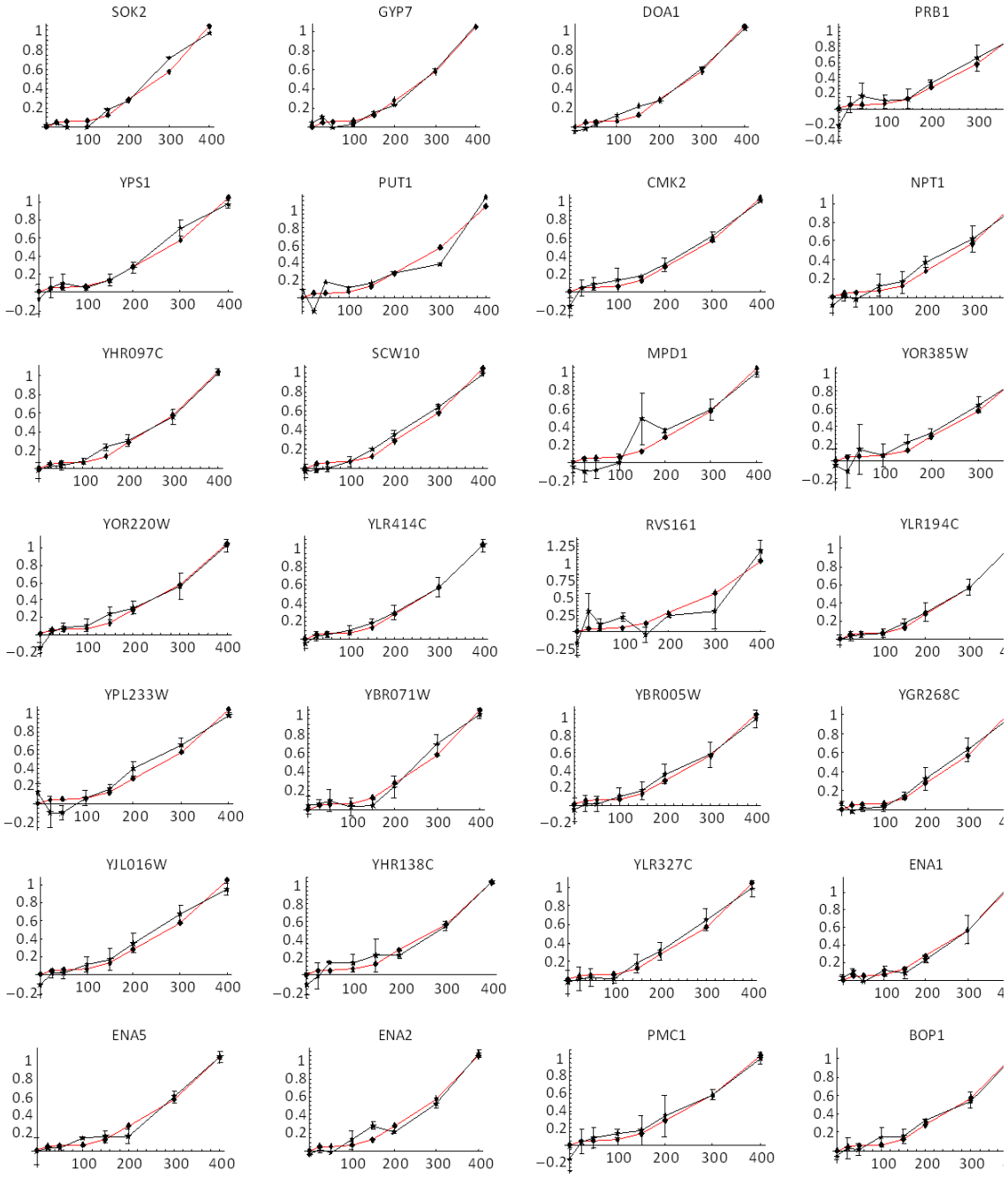


Fig. 13. Flow cytometry methods. a-c. Acquisition, trigger and gating of yeast cells on the Beckman Cell Lab Quanta Flow cytometer. EV refers to Coulter volume, or capacitance, measurement of particles. SS refers to side-scatter. **d-f.** Acquisition, trigger and gating of bleach control (no cells).

Figure 14a.


SOK2
GYP7
DOA1
PRB1
YPS1
PUT1
CMK2
NPT1
YHR097C
SCW10
MPD1
YOR385W
YOR220W
YLR414C
RVS161
YLR194C
YPL233W
YBR071W
YBR005W
YGR268C
YJL016W
YHR138C
YLR327C
ENA1
ENA5
ENA2
PMC1
BOP1

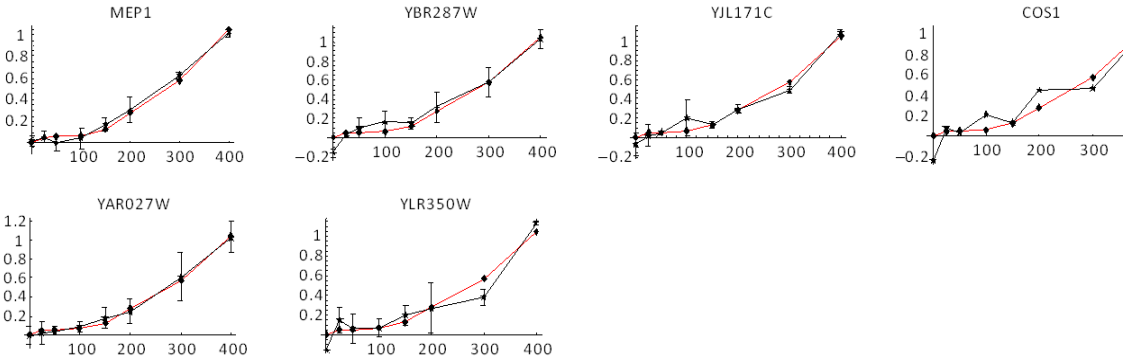


Figure 14b.

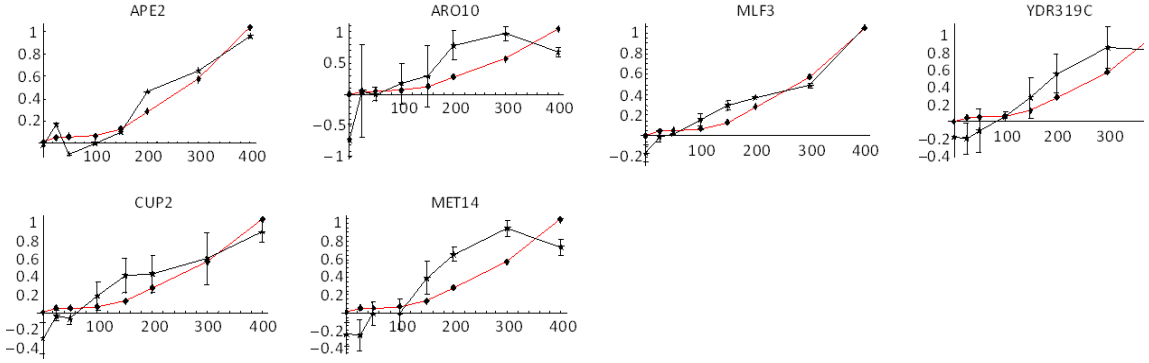


Figure 14 (continued). Flow cytometry induction curves of Crz1 target genes. a,b. Data is plotted in black, with normalized expression levels as a function of calcium concentration, with the fitted synthetic promoter induction curves in red. Error bars are standard errors from repeated measurements. (a) Based on these data, 34 genes fit the synthetic curve very well. (b) APE2, ARO10, MLF3, YDR319C, CUP2, MET14 do not seem to be coordinated, perhaps because of indirect interactions, or errors with our FK506 tests.

c.

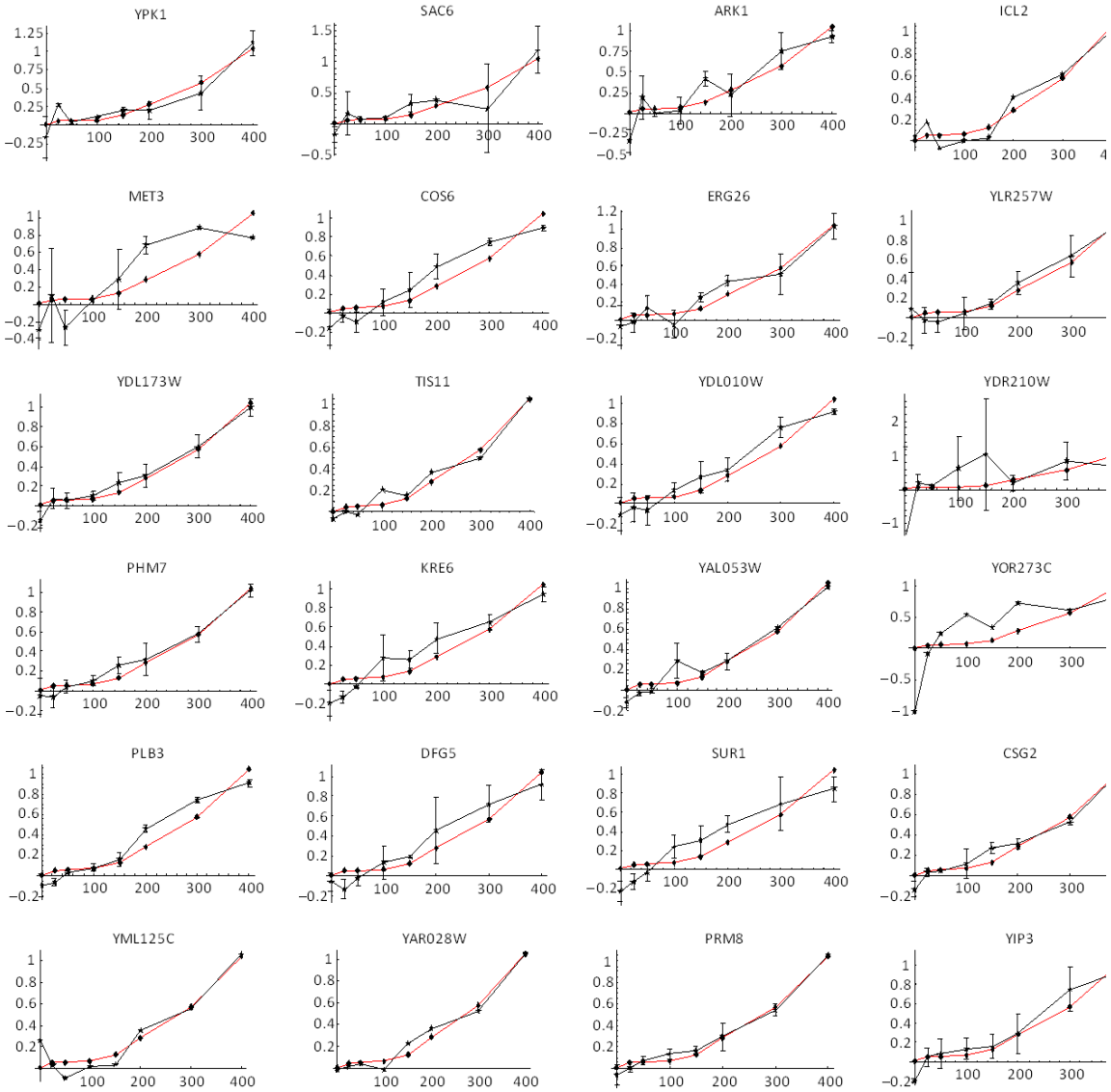


Figure 14 Flow cytometry induction curves of 24 genes that did NOT pass the FK506 test. c. Note that even though these genes did not pass the stringent FK506-dependence test (Fig. S15), which requires >90% dependence on Crz1, they can still exhibit significant Crz1-dependence, and in fact many of them also fit the CDRE induction curve quite well. For example, in the case of Ark1 (Fig. S15), Calcineurin-dependent activation accounts for >50% of the induction in calcium. Crz1 still contributes a significant fraction of its induction, despite additional contributions to its induction.

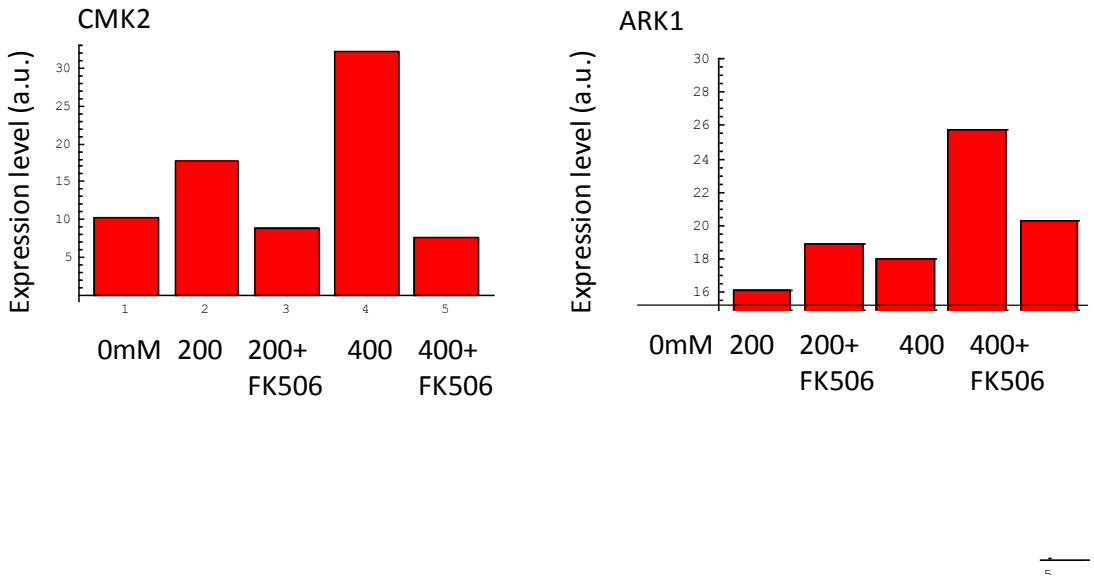
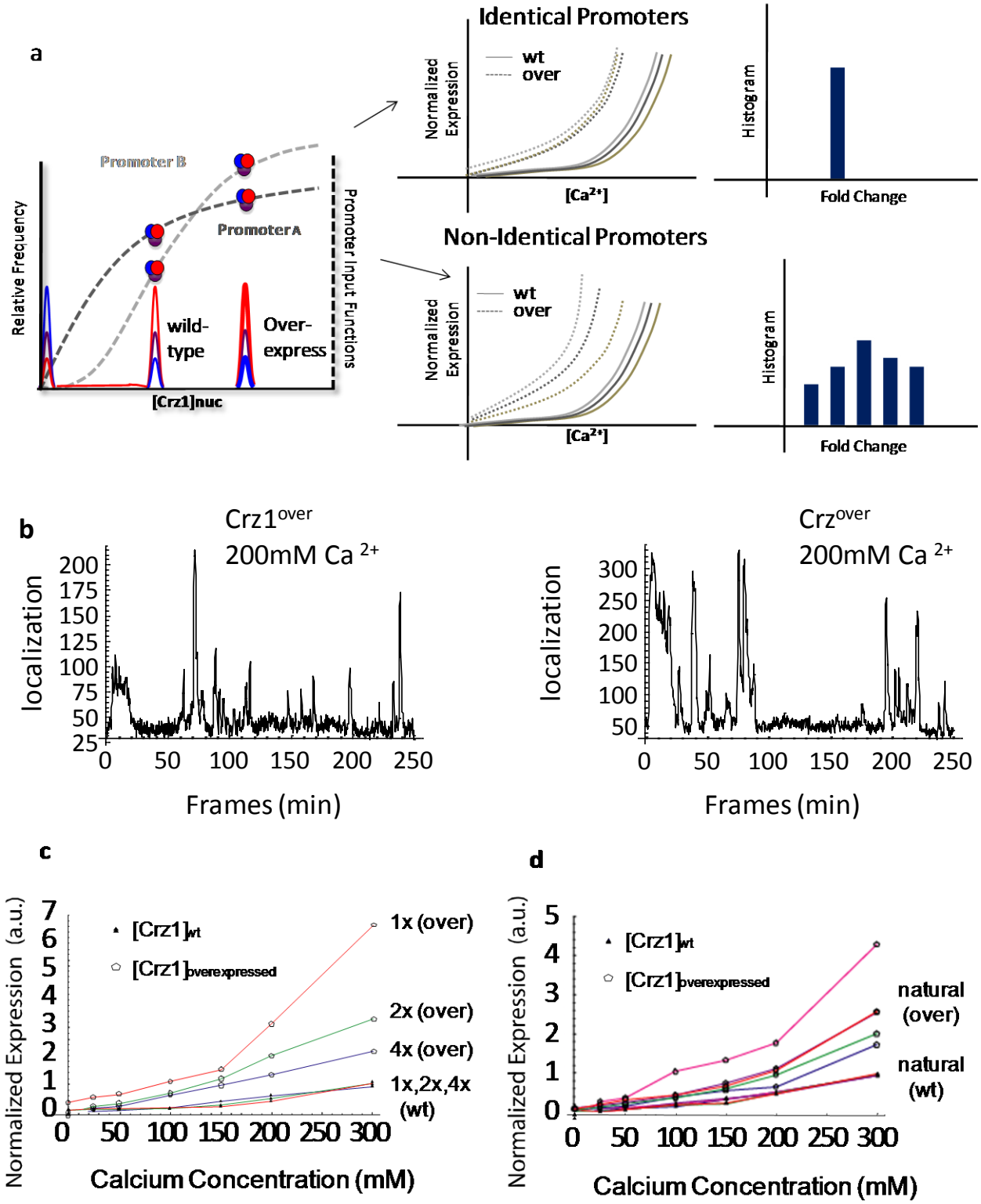


Figure 15. Selection Criterion for target genes. All 83 genes were grown in 0 mM calcium, 200 mM calcium and 400 mM calcium. Furthermore, cells grown in 200 and 400 mM calcium were also treated with saturating amounts of FK506 (1ng/ml), a Calcineurin inhibitor. In the case of Cmk2, negligible induction by calcium is observed in the presence of FK506, indicating that Cmk2 calcium induction is dependent on the Calcineurin/Crz1 pathway. Thus Cmk2 is a target gene that is strongly dependent on Crz1, whereas Ark1 is not: the 400mM Calcium + FK506 data point is not repressed all the way back to basal levels. Twenty-four genes with behavior similar to Ark1 were not considered pure Crz1 targets and were eliminated from the list of target genes.

Figure 16


a
Promoter B
Promoter A
Relative Frequency
Promoter Input Functions
wild-type
Over-express
[Crz1]nuc
Identical Promoters
Normalized Expression
wt
over
[Ca2+]
Histogram
Fold Change
Non-Identical Promoters
b
Crz1over
200mM Ca 2+
localization
Frames (min)
Crzover
200mM Ca 2+
c
[Crz1]wt
[Crz1]overexpressed
1x (over)
2x (over)
4x (over)
1x,2x,4x (wt)
Normalized Expression (a.u.)
Calcium Concentration (mM)
d
natural (over)
natural (wt)

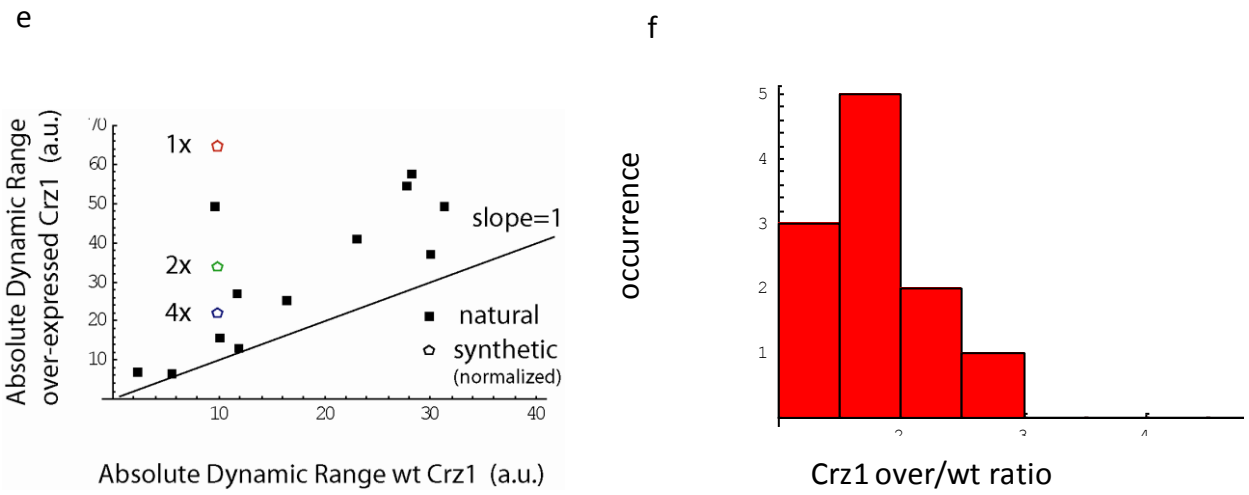


Fig. 16. Increasing total Crz1 expression level can exclude identical promoter input functions as an explanation for coordinated regulation. a. Schematic: Increased total nuclear Crz1 concentrations should cause proportional increases in the expression of promoters with identical input functions, causing them to exhibit identical fold changes between wild-type and overexpressed levels of Crz1. However, if the promoters have different input functions, this fold change should vary, producing a distribution of fold changes (top right**). b.** In a strain that overexpresses Crz1, Crz1-GFP continues to burst as shown in these sample trajectories. **c,d.** Expression profiles at wild-type and overexpressed levels of Crz1 for the synthetic and natural promoters. Each promoter activity is normalized by their activity at wild-type Crz1 levels. The diversity of fold changes plotted here shows that target promoters are coordinated despite input function differences (see text). **e.** To more clearly see the differences in expression ratio at the two Crz1 levels, we plot the expression level of each promoter. If the promoters had identical input functions the resulting points would all lie on a diagonal line; note instead the scatter in ratios. **f.** The histogram of ratio of the absolute dynamic range in overexpression vs. wild-type. If promoters have identical input functions, the histogram would have only a single peak.

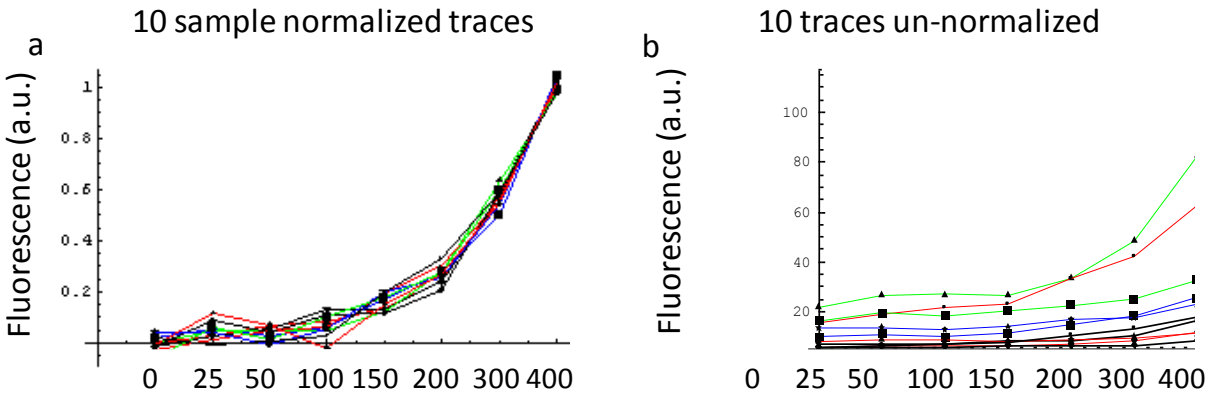


Fig. 17. Dynamic range of Crz1 target genes. a. Calcium induction curves, each normalized to its maximum, shown for 10 natural target genes. **b.** The induction curves for the same 10 genes are shown without normalization. Note the range of the absolute expression levels of Crz1 target genes.

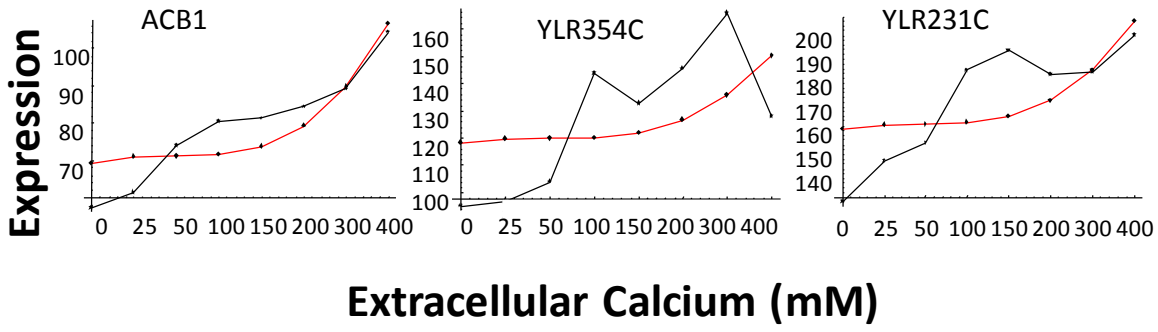


Fig. 18. Crz1 non-targets. To ensure that coordinated expression was Crz1-target-specific, we measured the expression of genes which are known to be induced by calcium but not in a Crz1 dependent manner. Their expressions do not match the synthetic CDRE induction curve (red).

BIBLIOGRAPHY

1 Bar-Even, A. et al., Noise in protein expression scales with natural protein abundance. *Nat Genet* **38** (6), 636 (2006).

2 Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S., Stochastic gene expression in a single cell. *Science* **297** (5584), 1183 (2002).

3 Ghaemmaghami, S. et al., Global analysis of protein expression in yeast. *Nature* **425** (6959), 737 (2003).

4 Sigal, A. et al., Variability and memory of protein levels in human cells. *Nature* **444** (7119), 643 (2006).

5 Raser, J. M. and O'Shea, E. K., Control of stochasticity in eukaryotic gene expression. *Science* **304** (5678), 1811 (2004).

6 Falconer, D. S. and Mackay, Trudy F. C., *Introduction to quantitative genetics*, 4th ed. (Longman, Harlow, England ; New York, 1996).

7 Clayton, G.A., Robertson, A., An experimental check on quantitative genetical theory. II The long-term effects of selection. *Journal of Genetics* **55**, 152 (1957); Clayton, G.A., Robertson, A., An experimental check on quantitative genetical theory. Short-term responses to selection. *Journal of Genetics* **55**, 131 (1957).

8 Grant, P.R, Natural Selection of Darwin's Finches. *Scientific American* **265**, 82 (1991).

9 Elena, S. F. and Lenski, R. E., Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* **4** (6), 457 (2003).

10 Hill, W. G. and Zhang, X. S., Effects on phenotypic variability of directional selection arising through genetic differences in residual variability. *Genet. Res.* **83** (2), 121 (2004).

11 Sato, K., Ito, Y., Yomo, T., and Kaneko, K., On the relation between fluctuation and response in biological systems. *Proc. Natl. Acad. Sci. U. S. A.* **100** (24), 14086 (2003).

12 Ito, Y., Toyota, H., Kaneko, K., and Yomo, T., How selection affects phenotypic fluctuation. *Mol. Syst. Biol.* **5**, 7 (2009).

13 Cai, L., Friedman, N., and Xie, X. S., Stochastic protein expression in individual cells at the single molecule level. *Nature* **440** (7082), 358 (2006).

14 Baetz, K. and Kaern, M., Predictable trends in protein noise. *Nat Genet* **38** (6), 610 (2006).

15 Newman, J. R. et al., Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441** (7095), 840 (2006).

16 Greener, A., Callahan, M., Jerpseth, B., An efficient random mutagenesis technique using an *E. coli* mutator strain. *Molecular Biotechnology* **7** (2), 189 (1997).

17 Friedman, N., Cai, L., and Xie, X. S., Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys Rev Lett* **97** (16), 168302 (2006).

18 Cyert, M. S., Regulation of nuclear localization during signaling. *J Biol Chem* **276** (24), 20805 (2001).

[19] Estruch, F., Stress-controlled transcription factors, stress-induced genes and stress tolerance in budding yeast. *FEMS Microbiol Rev* **24** (4), 469 (2000).

[20] Kyriakis, J. M., The integration of signaling by multiprotein complexes containing Raf kinases. *Biochim Biophys Acta* **1773** (8), 1238 (2007).

[21] Stathopoulos-Gerontides, A., Guo, J. J., and Cyert, M. S., Yeast calcineurin regulates nuclear localization of the Crz1p transcription factor through dephosphorylation. *Genes Dev* **13** (7), 798 (1999).

[22] Yoshimoto, H. et al., Genome-wide analysis of gene expression regulated by the calcineurin/Crz1p signaling pathway in Saccharomyces cerevisiae. *J Biol Chem* **277** (34), 31079 (2002).

[23] Huh, W. K. et al., Global analysis of protein localization in budding yeast. *Nature* **425** (6959), 686 (2003).

[24] Mettetal, J. T., Muzzey, D., Gomez-Uribe, C., and van Oudenaarden, A., The frequency dependence of osmo-adaptation in Saccharomyces cerevisiae. *Science* **319** (5862), 482 (2008).

[25] Hersen, P., McClean, M. N., Mahadevan, L., and Ramanathan, S., Signal processing by the HOG MAP kinase pathway. *Proc Natl Acad Sci U S A* **105** (20), 7165 (2008).

[26] Suel, G. M. et al., Tunability and noise dependence in differentiation dynamics. *Science* **315** (5819), 1716 (2007).

[27] Di Talia, S. et al., The effects of molecular noise and size control on variability in the budding yeast cell cycle. *Nature* **448** (7156), 947 (2007).

[28] Fewtrell, C., Ca2+ oscillations in non-excitable cells. *Annu Rev Physiol* **55**, 427 (1993).

[29] Wiesenberger, G. et al., Mg2+ deprivation elicits rapid Ca2+ uptake and activates Ca2+/calcineurin signaling in Saccharomyces cerevisiae. *Eukaryot Cell* **6** (4), 592 (2007).

[30] Boustany, L. M. and Cyert, M. S., Calcineurin-dependent regulation of Crz1p nuclear export requires Msn5p and a conserved calcineurin docking site. *Genes Dev* **16** (5), 608 (2002).

[31] Roy, J., Li, H., Hogan, P. G., and Cyert, M. S., A conserved docking site modulates substrate affinity for calcineurin, signaling output, and in vivo function. *Mol Cell* **25** (6), 889 (2007).

[32] Breuder, T. et al., Calcineurin is essential in cyclosporin A- and FK506-sensitive yeast strains. *Proc Natl Acad Sci U S A* **91** (12), 5372 (1994).

[33] Jacquet, M. et al., Oscillatory nucleocytoplasmic shuttling of the general stress response transcriptional activators Msn2 and Msn4 in Saccharomyces cerevisiae. *J Cell Biol* **161** (3), 497 (2003).

[34] Garmendia-Torres, C., Goldbeter, A., and Jacquet, M., Nucleocytoplasmic oscillations of the yeast transcription factor Msn2: evidence for periodic PKA activation. *Curr Biol* **17** (12), 1044 (2007).

[35] Medvedik, O., Lamming, D. W., Kim, K. D., and Sinclair, D. A., MSN2 and MSN4 Link Calorie Restriction and TOR to Sirtuin-Mediated Lifespan Extension in Saccharomyces cerevisiae. *PLoS Biol* **5** (10), e261 (2007).

[36] Stathopoulos, A. M. and Cyert, M. S., Calcineurin acts through the CRZ1/TCN1-encoded transcription factor to regulate gene expression in yeast. *Genes Dev* **11** (24), 3432 (1997).

[37] Golding, I., Paulsson, J., Zawilski, S. M., and Cox, E. C., Real-time kinetics of gene activity in individual bacteria. *Cell* **123** (6), 1025 (2005).

[38] Raj, A. et al., Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4** (10), e309 (2006).

[39] Rodriguez, A. J., Shenoy, S. M., Singer, R. H., and Condeelis, J., Visualization of mRNA translation in living cells. *J Cell Biol* **175** (1), 67 (2006).

[40] Kaern, M., Elston, T. C., Blake, W. J., and Collins, J. J., Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* **6** (6), 451 (2005).

[41] Kaufmann, B. B. and van Oudenaarden, A., Stochastic gene expression: from single molecules to the proteome. *Curr Opin Genet Dev* **17** (2), 107 (2007).

[42] Maheshri, N. and O'Shea, E. K., Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu Rev Biophys Biomol Struct* **36**, 413 (2007).

[43] Ozbudak, E. M. et al., Regulation of noise in the expression of a single gene. *Nat Genet* **31** (1), 69 (2002).

[44] Yu, J. et al., Probing gene expression in live cells, one protein molecule at a time. *Science* **311** (5767), 1600 (2006).

[45] Rosenfeld, N. et al., Gene regulation at the single-cell level. *Science* **307** (5717), 1962 (2005).

[46] Matheos, D. P., Kingsbury, T. J., Ahsan, U. S., and Cunningham, K. W., Tcn1p/Crz1p, a calcineurin-dependent transcription factor that differentially regulates gene expression in Saccharomyces cerevisiae. *Genes Dev* **11** (24), 3445 (1997).

[47] Armstrong, E. H., A method of reducing disturbances in radio signaling by a system of frequency modulation. *P Ire* **24** (5), 689 (1936).

[48] Song, G. B., Buck, N. V., and Agrawal, B. N., Spacecraft vibration reduction using pulse-width pulse-frequency modulated input shaper. *J Guid Control Dynam* **22** (3), 433 (1999).

[49] Adrian, E. D. and Zotterman, Y., The impulses produced by sensory nerve-endings: Part II. The response of a Single End-Organ. *J Physiol* **61** (2), 151 (1926).

[50] Sarpeshkar, R., Analog versus digital: extrapolating from electronics to neurobiology. *Neural Comput* **10** (7), 1601 (1998).

[51] Dolmetsch, R. E., Xu, K., and Lewis, R. S., Calcium oscillations increase the efficiency and specificity of gene expression. *Nature* **392** (6679), 933 (1998).

[52] Geva-Zatorsky, N. et al., Oscillations and variability in the p53 system. *Mol Syst Biol* **2**, 2006 0033 (2006).

[53] Nelson, D. E. et al., Oscillations in NF-kappaB signaling control the dynamics of gene expression. *Science* **306** (5696), 704 (2004).

[54] Friedman, N. et al., Precise temporal modulation in the response of the SOS DNA repair network in individual bacteria. *PLoS Biol* **3** (7), e238 (2005).

[55] Cai, L., Dalal, C. K., and Elowitz, M. B., Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455** (7212), 485 (2008).

[56] Estruch, F. and Carlson, M., Two homologous zinc finger genes identified by multicopy suppression in a SNF1 protein kinase mutant of Saccharomyces cerevisiae. *Mol Cell Biol* **13** (7), 3872 (1993).

[57] Schmitt, A. P. and McEntee, K., Msn2p, a zinc finger DNA-binding protein, is the transcriptional activator of the multistress response in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **93** (12), 5777 (1996); Gasch, A. P. et al., Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11** (12), 4241 (2000).

[58] Costanzo, M. C. et al., The Yeast Proteome Database (YPD) and Caenorhabditis elegans Proteome Database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. *Nucleic Acids Res.* **28** (1), 73 (2000).

[59] Issel-Tarver, L. et al., in *Guide to Yeast Genetics and Molecular and Cell Biology, Pt B* (Academic Press Inc, San Diego, 2002), Vol. 350, pp. 329.

[60] Cai, L., Dalal, C. K., and Elowitz, M. B., Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455** (7212), 485 (2008).

[61] True, H. L. and Lindquist, S. L., A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407** (6803), 477 (2000).

[62] Partridge, L. and Barton, N. H., Evolving evolvability. *Nature* **407** (6803), 457 (2000).

[63] Sheff, M. A. and Thorn, K. S., Optimized cassettes for fluorescent protein tagging in Saccharomyces cerevisiae. *Yeast* **21** (8), 661 (2004).

[64] Gietz, R. D. and Woods, R. A., Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350**, 87 (2002).

[65] Sherman, F., Getting started with yeast. *Methods Enzymol* **350**, 3 (2002).

[66] Nachman, I., Regev, A., and Ramanathan, S., Dissecting timing variability in yeast meiosis. *Cell* **131** (3), 544 (2007).

[67] Costanzo, M. et al., CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. *Cell* **117** (7), 899 (2004).

[68] Palmer, A. E. and Tsien, R. Y., Measuring calcium signaling using genetically targetable fluorescent indicators. *Nat Protoc* **1** (3), 1057 (2006).