Title
Dimensionality and the Stress/Growth Axis of Gene Expression in S. cerevisiae

Author
Stewart-Ornstein, Jacob

Publication Date
2012

Peer reviewed|Thesis/dissertation

Dimensionality and the Stress/Growth Axis of Gene Expression in *S. cerevisiae*

by

Jacob Stewart-Ornstein

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Dedications and Acknowledgements

None of this work would have been possible without the support of my faculty advisors Jonathan Weissman and Hana El-Samad. They provided occasionally contradictory but always informative advice and guidance. I also received substantive thoughts and reality checks from my thesis committee member Hiten Madhani, onetime rotation advisor Orion Wiener, and qualifying committee members Wallace Marshall, Wendell Lim, and Barbra Panning, and collaborator Joe DeRisi. I am indebted to David Pincus, Onn Brandman, Charles Biddle-Snead, Mike Chavalier, Chris Nelson, and all members of the Weissman and El-Samad labs for their generosity with time, reagents, and advice. My TETRAD (now TRIAD) classmates were also a source of support, excuses to have 'a' drink, eat or cook, and talk about science. In some order Colin, Jessica, Michael, Monica, Stacey, Beth, Argenta, Ellen, Ariel, Heather, Erica, Erin and Ashley all were crucial to the last five years, science wise and otherwise. My immediate family, Penni, Michael, and Noah also contributed in amorphous and difficult to specify ways.

Although I have not included it in this work I would also like to acknowledge a fruitful collaboration with Sean Summers and Andreas Milias-Argeitis and their supervisor John Lygeros on the development and deployment of light regulated tools for manipulating gene expression. Similarly, Chris Nelson and his supervisor Joe DeRisi have worked with me to obtain measurements of the affinity of Msn2 for DNA.

Dimensionality and the Stress/Growth Axis of Gene Expression in *S. cerevisiae*

Jacob Stewart-Ornstein

To thrive in different circumstances cells tightly regulate and tune gene expression to optimize their protein complement to the environment. Understanding and manipulating this regulation is an apparently intractable goal as even budding yeast has more than 6000 genes. Even from the perspective of the organism, achieving precise control over the expression of every gene would amount to an enormous and perhaps impossible burden. Instead organisms regulate genes in large groups, activating suites of genes required for stress resistance or growth on alternative carbon sources. In this work we show that by measuring gene expression in single cells in high throughput using yeast genetics and the GFP library it is possible to define expression 'regulons' which describe the functional arrangement of the budding yeast exome. From this analysis we focused on one particular axis of gene expression which regulates the transition from growth to stress resistance and is regulated by the Protein Kinase A (PKA) pathway. We show that the PKA and two of the transcription factors it regulates, Msn2/4, respond in a graded manner to a range of stresses and that demonstrate that this graded response is due to negative feedback regulation in the PKA network and unsaturated non-cooperative association of Msn2/4 with its target promoters. These features allows for co-linear activation of target genes, maintaining stoichiometry within the Msn2/4-responsive program across a wide range of conditions. In addition, we find that under different environmental conditions PKA shows distinct dynamic behaviors at the single cell level ranging from steady state activation to pseudo-oscillatory regimes. We observe that Msn2 faithfully follows these dynamics, whereas other PKA regulated transcription factors do not, suggesting a mechanism by which the PKA regulatory circuit may specifically activate subsets of regulated genes in response to different conditions.

Table of Contents

List of Figures

Chapter 1: How complicated is a single celled organism?

Or "Exploring Cellular Dimensionality"

The budding yeast genome codes for in excess of 6000 genes, which store, produce, utilize, and modify a staggering number of small molecules. Describing even the protein complement of a single yeast cell would therefore seem to require thousands of individual values situated in a 6000 dimensional space. In practice, however, genes are not expressed independently and much of the biological variation occurs along a few dimensions or axes, each of which accounts for the regulation of hundreds of genes. Initial observations of this effect was in bacteria, where DNA coding for groups of related proteins was organized into local co-transcribed units (operons), this arrangement had the effect of slaving together groups of related genes, allowing them to be expressed stoichiometrically and synchronously (Jacob and Monod, 1961). Although mechanistically distinct, similar phenomena of groups of tightly co-regulated genes were observed in eukaryotes in the regulation of alternative carbon source metabolism (Broach, 1979) and heat shock (Schedl et al., 1978). With the advent of genome wide measures of gene expression the organization of functionally related genes into co-regulated units was generalized to the point that the functionality of genes can sometimes be inferred from their patterns of expression (Gasch et al., 2000). These results raised the question—which I will explore in this chapter—of how few variables could accurately describe the state of a cell.

Particularly in budding yeast, studies of the cell cycle and stress response revealed that large groups of genes, which we will refer to as regulons, were activated and inhibited as units (Spellman et al., 1998; Gasch et al., 2000). Attempts to explain these patterns of expression structurally by defining the proximal transcription factors regulating each gene have been ongoing and increasingly successful, both in classical micro-organisms E. coli and budding yeast and more recently in mammalian systems (for example Harbison, et al, 2004; Encode Project

Consortium, 2012). Less explored, has been the second layer of regulatory architecture which coordinates groups of transcription factors into higher order units. Typically this regulation is exercised by kinases which process information from many different cellular sources and pass this information on to transcription factors to regulate gene expression. This form of post-transcriptional regulatory networks appears particularly active in micro-organisms such as budding yeast, where unlike mammalian cells most transcription factors do not appear to dramatically change in abundance in conditions where they are active.

First, we describe the framework, development, and application of a single cell approach to measuring and analyzing the covariance in the expression of different genes is described (1.1). Second, this method is applied on a large scale to the regulation of gene expression in budding yeast (1.2). Finally, extending the concept and experimental determination of dimensionality from gene expression to protein folding is explored (1.3).

1.1

Single Cell Measurements Define Regulatory Relationships

Measurements of genomes, transcriptomes, and translatomes are rapidly identifying groups of interacting genes. Determining the architecture and strength of these interactions, however, remains an unsolved problem. A common approach is to infer regulatory relationships between genes by measuring the degree of correlation in their expression over many environmental conditions. Traditionally, this correlation has been measured over time in a synchronized population using a population averaged mRNA measurement (Spellman, et al. 1998; Gasch et al. 2000; Klevecz et al, 2004). The natural extension of these methods is to use single cell reporters to measure the correlation between two genes *across a population.* Single cell resolution permits noise in protein expression to act as a probe, replacing invasive synchronizing perturbations. Further, the directionality of interactions can be inferred by measuring transmission of expression

noise between components. Establishing regulatory relationships from single cell data has been demonstrated on a limited scale in bacteria (Dunlop et al., 2008), and immune cells (Sachs et al., 2005), but has not been systematically applied and should provide a valuable complement to existing network reconstruction techniques and a useful tool for biological discovery.

The expression of a protein can be measured on a single cell level by fusion to a fluorescent protein, an approach that rarely impairs protein function in yeast (Howeson et al, 2005). The distribution of protein expression in cells across a population can be thought of as the expression "noise" of that protein. This noise can be separated into intrinsic and extrinsic components. These elements are defined by an experiment with a pair of identical promoters driving different alleles of GFP in the same cell (Elowiz et al, 2002). The variance in expression of the alleles is ascribed to intrinsic fluctuations such as transcription factor binding and unbinding, whereas the correlation is due to extrinsic factors such as transcription factor levels or the translational capacity of the cell.

The extension of the experiment described above is to place two different promoters within the same cell and measure the correlation between them. The relative expression of two proteins within a single cell will relate to the total transcriptional and translational capacity of the cell and any regulatory relationship they share. Strong positive correlation between two proteins suggests either an activating regulatory relationship or co-regulation whereas negative correlation implies repression or anti-regulation.

Synthetic genetic circuits have often been used as showcases of the applicability of engineering principles to biology (Stricker et al, 2008; Becskei et al, 2005; Hooshangi et al, 2005). The engineered nature of these circuits makes them attractive benchmarks for network reconstruction approaches and also for investigating aspects of cellular processes in isolation from the rest of the cell (Cantone et al., 2009). Unfortunately the bulk of these systems have been

designed in bacteria and many have been characterized less than completely. The development and characterization of a flexible *S. cerevisiae* based synthetic circuit was therefore necessary.

The circuit I focused on is a cascade of repressors analogous to bacterial systems developed by Van Oudenarrden and Weiss (Pedraza et al, 2005; Hooshangi et al, 2005). This network is one of the more simple motifs in biological circuitry, but is designed such that the addition of feedback loops to add complexity is possible (Figure 1A). Of particular value is the ability to perturb the system with small molecules—tetracycline, galactose, and IPTG—allowing access to a wide range of interaction strengths between genes. The construction of this network allowed for the characterization of single cell correlation measurements under circumstances where the underlying relationships were known and could be manipulated in straightforward ways.

Results 1.1

Using flow cytometry we measured the single cell expression of proteins in this circuit tagged with florescent proteins. We began by titrating in the small molecule aTC and measuring the expression of a repressor (TetR-mCherry) and its target gene (GFP) at steady state. With increasing concentrations of aTC we observe the expected monotonic increase in GFP expression (Figure 1B). In contrast the variance of GFP and covariance of GFP and mCherry show a peak (or trough in the case of covariance) at intermediate concentrations of aTC, this peak corresponds to the point where the system is maximally sensitive to variation in TetR concentration (Figure 1C). The negative correlation is expected from the relationship between a repressor and its target, and the corresponding increase in variability we observe in GFP likely represent 'noise transmission' from the upstream regulator. These results align well with work on other synthetic circuits (Pedraza et al, 2005) and provide a critical proof of concept, suggesting that single cell

data can be measured with sufficient precision to extract meaningful relationships between genes *in vivo*.

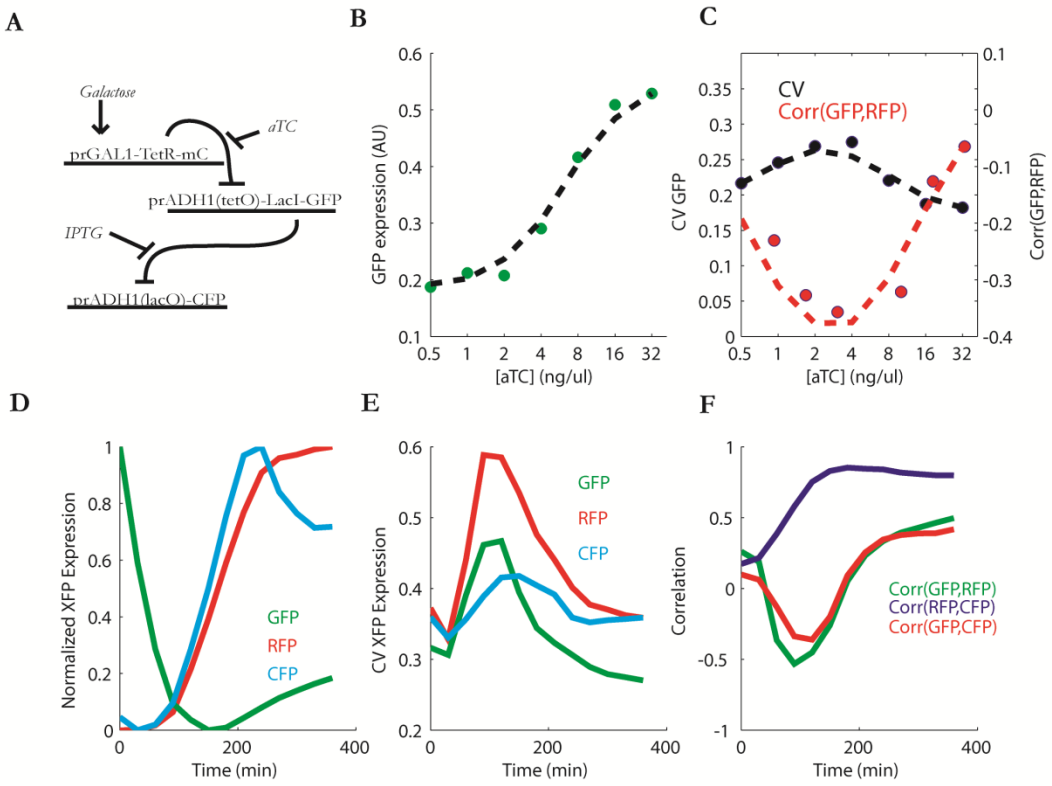


Figure 1: Design and characterization of synthetic regulatory networks constructed from orthogonal prokaryotic components and expressed in cerevisiae. (a) Schematic of the network which was constructed, briefly galactose induces the expression of TetR-mCherry that represses the expression of GFP-LacI that in turn represses the expression of CFP, the small molecules aTC and IPTG can modulate the strengths of these repressors. (b) Cell expressing the circuit from (a) show increased GFP expression as they are treated with increasing concentrations of aTC. (c) Plots showing the CV and Correlation of GFP and mCherry of the same population in (b), the maximum CV or minimum correlation occurs when the population is maximally sensitive to changes in [aTC]. (d) The circuit in (a) is activated by a switch from glucose to galactose media and cells sampled every hour and average GFP, RFP and CFP fluorescence measured, in addition CV (e) and Correlation between the species (f) were computed.

Analogous data was collected for the full three node circuit, simultaneously measuring the level of all three proteins in each cell. We took advantage of this three dimensional dataset to test the limits of information contained in pairwise steady state correlations, as opposed to higher

order interactions. If every gene were expressed independently, there would be no information within the covariance of pairs of genes. At the other extreme, if the expression of all genes were intimately connected pairwise data would miss more complex three or more gene interactions. Empirical studies using microarray data suggest that many cellular networks can be well represented by pairwise interactions (Lee et al., 2004; Lezon et al., 2006). However, a direct test of the degree to which pairwise and higher order interactions contribute to interactions in genetic networks in a single cell context is lacking.

Using the synthetic network outlined above we directly measured the contribution of pairwise and third order interactions. In general the strength of interaction between a pair of variables may be quantified by the reduction in the uncertainty about X if Y is known. This quantity is known as mutual information (I(X,Y)) and is sensitive to linear and non-linear relationships. The relative information content in binary interactions (I_2) compared to tertiary interactions (I_3) can then be approximated by the change in I(X,Y) when a third variable Z is known (McGill, 1967). Simulations of the circuit described above show that under a wide range of parameters >90% of the total information in the system can be explained with binary interactions. Empirical *in vivo* characterization of the circuit finds that binary interactions explain ~90% of the variance. These results argue that in this context pairwise measurements are sufficient to accurately define the network interactions.

Finally, to characterize if time dynamic measurements were tractable with this system we measured a 7hr timecourse of the the evolution of the system starting at time=0 with the addition of galactose. At each timepoint cells were removed for cytometry and an equal volume replaced maintaining the cell density at approximately an OD of 0.05 throughout the experiment. Plotting the expression of each florophore, its CV across the population, and the correlation between each channel we observe striking dynamic behavior with a regime in the middle of the timecourse where variability gives a great deal of information and either end of the timecourse where

correlations revert towards zero (Figure 1D, E, and F). Although we did not carry these experiments further, these results argue that a careful analysis of the covariance of time dependant systems may be informative.

Discussion 1.1

The synthetic tools we constructed provided a crucial proof of concept and a dataset with which to hone computational approaches to defining cell-to-cell covariance as a quantitative phenotype. Further, these results highlight the dynamic nature of variation, and that while cell-to-cell variation or noise is typically studied at steady state there is substantial signal in the evolution of variance and covariance in dynamic conditions.

These results also show that cell-to-cell variation can be substantial enough—even for relatively highly expressed genes which were used in these experiments—to drive distinct outputs from genetically identical cells. Other groups working with strong positive feedback loops have shown that this same variation can cause bimodal gene expression (To and Maheshri, 2010). These results suggest that either this variation must be controlled, tolerated, or exploited by cells. There are clear examples of noise minimization in decision making circuits such as those that regulate development. Elegant work on the role of the diffusible morphogen bicoid in drosophila showed that this system operates the absolute thermodynamic limit of variability (Tkacik et al, 2008). Similarly some systems appear to use variation from a steady state to encode information, particularly the temporal variation in transcription factor activity has been shown to transmit substantive information about the external environment in the calcium sensing Crz1 circuit in budding yeast (Cai et al., 2008). Less explored may be the ways in which cell may tolerate substantive variation. One possibility is that it is not so much absolute protein dose, but rather the ratio of different proteins which is important to a cellular system. If for example the bulk of cellular variation was channeled into axes in which scores of genes are co-regulated then the

stoichiometry of genes within these groups would remain constant even in the face of substantial variation.

1.2

Measuring Dimensionality with Single Cell Covariance

Gene expression in budding yeast has been studied extensively, and measured at the genomic level in many conditions (Gasch et al., 2000). From this data it is possible to define co-regulated genes and begin to explore dimensionality in gene expression (Hibbs et al., 2007). However, these expression datasets speak to the spectrum of gene expression in an organism, not how gene expression is regulated in any given state. Moreover, the states that these measurements are recorded in are not a random selection of all possible states but are a set of relatively extreme conditions which generate reliable phenotypes. To obtain an unbiased collection of gene expression measures with which to explore the dimensionality of budding yeast gene expression we turned to pairwise single cell measures of expression from florescent proteins. Taking this approach avoids the problem of condition selection at the cost of substantively lower throughput.

It is possible to use single cell measures to explore co-regulation of gene expression because even uniformly cultured, isogenic populations of cells show diversity in size, shape, cell cycle state, and gene expression. This individuality, or "noise", has been shown to affect cellular response to stimuli (Colman-Lerner et al., 2005; Volfson et al., 2006). Noise in gene expression can be caused by stochastic fluctuations in components of gene expression, global variation in cell state or expression capacity, and local variation in cell signaling pathways (Ozubdzk et al., 2002; Blake et al., 2003; Raser and O'Shea, 2004; Golding et al., 2005; Raj et al., 2005; Newman et al., 2005; Becskei et al., 2005; Taniguchi et al., 2010). Variation in specific signaling pathways has important implications for individuality, as this source of noise alone can serve to coordinate the expression of groups of related genes, perhaps resulting in phenotypes such as

stress resistance. However, the magnitude of these pathway fluctuations, and their contribution to gene expression noise across the genome, is yet to be explored.

Evidence of non-genetic individuality was reported in early studies of bacterial persistence during antibiotic treatment, λ phage burst size and bacterial chemotactic behaviour (Bigger, 1944; Delbruck, 1945; Spudich and Koshland, 1976). Such phenotypic heterogeneity arises due to stochastic fluctuations, or noise, in molecular reactions at the level of a single gene or transcript. It can also be the result of global cellular variability, or fluctuations propagated from upstream signaling pathways to which many genes could be simultaneously susceptible.

Assays for decomposing noise particular to a single gene (intrinsic, uncorrelated) and noise experienced by multiple genes (extrinsic, correlated) have been instrumental in tracing the sources of variability in gene expression. In these assays, intrinsic and extrinsic noise are measured using identical promoters which drive expression of distinguishable fluorescent proteins (FPs) in the same cell (Elowitz et al., 2002; Raser and O'Shea, 2004). Uncorrelated variation in expression of these two promoters reflects intrinsic noise resulting from stochastic fluctuations in the process of gene expression itself. In contrast, extrinsic noise is defined as the correlated variation in the expression of the two genes. While global studies in yeast and bacteria (Golding et al., 2005; Newman et al., 2006; Raj et al., 2006, Taniguchi et al., 2010) demonstrated that intrinsic noise was dominant for weakly expressed genes, the contribution of extrinsic noise to gene expression heterogeneity more generally remains unexplored. Further, we aimed to exploit this extrinsic variability to explore shared variation across signaling pathways with the aim of defining single condition measures of gene expression dimensionality and covariance.

Results 1.2

To measure gene expression in vivo in single cells we constructed reporter constructs incorporating the promoter—and therefore presumably regulatory apparatus—of specific genes

driving either YFP or RFP. The florescent signal measured in individual cells by flow cytometry or microscopy is therefore a metric of the activity of that promoter in that cell integrated over the previous few hours. The covariance in the activity of two promoters in a single cell measured by their YFP and RFP fluorescence could therefore be used to compare their degree of co-regulation, as we showed for synthetic circuits. To test this we compared the correlation in single cell expression from an exponentially growing population of two stress responsive promoters regulated by the transcription factor Msn2, and also the DBP7 promoter which is from a ribosomal biogenesis gene (Figure 2A). As expected we find a strong correlation between two copies of the HSP12 promoter (one driving YFP, one RFP), and a slightly smaller but still substantial relationship between HSP12-RFP and PGM2-YFP, and a very low correlation between HSP12-RFP and DBP7-YFP. These results argue that single cell covariance measures are a accurate approach to measuring co-regulation in natural yeast regulatory networks.

We propose that the co-variation we observe between the two stress responsive promoters is due to shared regulation by the transcription factor Msn2. To test this directly we measured the correlation between HSP12 and two other stress responsive promoters SSA1 and SSA4, as with PGM2 we observed a substantive although lower correlation. The advantage of using these promoters as opposed to prPGM2 is that both of these genes are regulated by Msn2 and other factors, so disrupting their regulation by Msn2 will not substantively alter their average expression value. We then compared the correlation of prHSP12-RFP and prSSA1 or prSSA4, and also versions of these genes with the consensus Msn2 binding sites disrupted by point mutations (prSSA1-dSTRE, prSSA4-dSTRE). We observe that removal of the Msn2 binding sites reduces the correlation with HSP12 (Figure 2B), without substantively altering their average gene expression (data not shown).

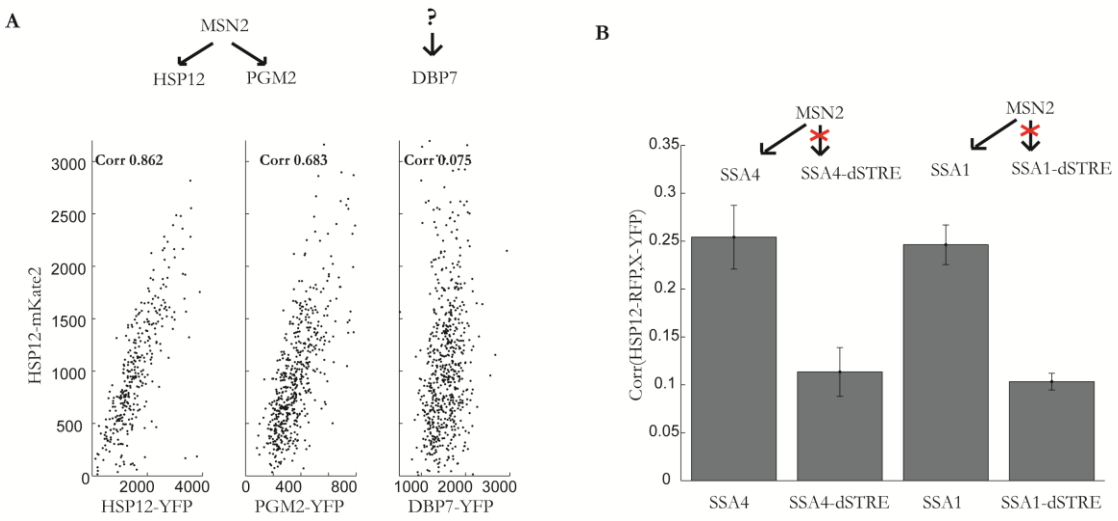


Figure 2: Single cell gene expression measures show correlation between co-regulated genes. (a)Hsp12 is regulated by Msn2, two alleles of Hsp12 expressed with distinguishable XFPs show strong single cell correlation, as does the co-regulated gene Pgm2. In contrast the ribosomal biogenesis gene Dbp7 shows only very weak correlation with Hsp12. (b) Hsp12 shows moderate correlations with Ssa1 and Ssa4, this correlation is greatly reduced in alleles of the promoters lacking Msn2 binding sites.

These results were consistent with a component of cell-to-cell variation resulting from variation in transcription factor activity which is shared across co-regulated genes. To extend this analysis and define if covariance at the single cell level could be used to define co-regulated genes more broadly we extended the above analysis to many hundreds of genes by taking advantage of the GFP library. We used a two-color strategy to measure the covariance of the most abundant ~18% genes in the GFP library (752 genes) to either Pgm2 or the "quiet" ribosomal gene Rpl17B fused to RFP (F3A). To compare covariance across all genes without bias, we defined the S-score as the covariance of two genes, normalized by the query gene's (PGM2 or RPL17B) extrinsic.

To define our S-Score metric we began by assuming that in an exponentially growing population, gene expression can be approximated as a linear process. The expression of a given gene in a single cell (b_i) is then equal to the sum of a basal expression (β_i), the activity of

upstream signaling pathways in that cell (x_i) multiplied by the gene's susceptibility to that pathway (α), and a noise term (ε_i). The following derivation will show that it is not possible to get an absolute value for α from single cell noise data, but that it is possible to measure the strength of gene *b*'s response to pathway *x* relative to a reporter gene (*a*) which is simultaneously measured in the same cell (i.e. you can find α_b/α_a). We define:

$$a_i = \beta_1 + \alpha_a x_i + \varepsilon_{1i}$$
$$b_i = \beta_2 + \alpha_b x_i + \varepsilon_{2i}$$
$$x_i = \bar{x} + \theta_i$$
$$\varepsilon_i = Norm(0, \sigma_i), i = 1,2$$
$$\theta = Norm(0, \sigma_\theta)$$
$$Corr(\varepsilon_1, \varepsilon_2) = Corr(\varepsilon_2, \theta) = Corr(\varepsilon_1, \theta) = 0$$

a_i = expression of gene *a* in cell *i*

b_i = expression of gene *b* in cell *i*

β = basal expression of a gene

x_i =activity of pathway *X* in cell *i*

α =change in a gene in response to pathway X

Under the assumption that noise terms are uncorrelated, we obtain:

$$\bar{a} = \beta_1 + \alpha_a \bar{x}$$
$$\bar{b} = \beta_2 + \alpha_b \bar{x}$$

The covariance of *a* and *b* can also be computed:

$$Cov(a,b) = \overline{(a_i - \bar{a})(b_i - \bar{b})}$$
$$Cov(a,b) = \overline{(\alpha_a \theta_i + \varepsilon_{1i})(\alpha_b \theta_i + \varepsilon_{2i})}$$
$$Cov(a,b) = \overline{\alpha_a \theta_i \varepsilon_{2i} + \alpha_b \theta_i \varepsilon_{1i} + \alpha_a \alpha_b \theta_i^{\,2} + \varepsilon_{1i} \varepsilon_{2i}}$$
$$Cov(a,b) = \alpha_a \alpha_b \sigma_x^{\,2}$$

Now, if we assume that variance in gene expression is proportional to mean expression level of the gene, then the total variance in *a* and *b* is given by

$$\sigma_a^{\,2} = \alpha_a^{\,2} \sigma_x^{\,2} + \phi \bar{a} + \gamma_a$$
$$\sigma_b^{\,2} = \alpha_b^{\,2} \sigma_x^{\,2} + \phi \bar{b} + \gamma_b$$

Where the first term represent the pathway noise from *x*, and subsequent terms represent intrinsic noise (whose variance is proportional to mean expression), and extrinsic noise from other processes unrelated to *x*. The correlation between *a* and *b* can be written as:

$$Corr(a,b) = \frac{\alpha_a \alpha_{b2} {\sigma_x}^2}{\sqrt{({\alpha_a}^2 {\sigma_x}^2 + \phi\bar{a} + \gamma_a)}\sqrt{({\alpha_b}^2 {\sigma_x}^2 + \phi\bar{b} + \gamma_b)}}$$

Note that the correlation depends on the mean expression level of *a* and *b*, indicating that as the mean expression level of the genes increase their correlation will also decrease. Correlation has a fundamental bias towards weakly expressed genes, and is therefore problematic as a measure of relatedness. Correlation does however describe in an unbiased way the proportion of variance shared by a given gene pair and is useful to compare across different genes. To achieve a measure of relatedness independent of mean expression we turn to the covariance.

$$Cov(a,a) = {\alpha_a}^2 {\sigma_x}^2$$

$$\frac{Cov(a,b)}{Cov(a,a)} = \frac{\alpha_a \alpha_b {\sigma_x}^2}{{\alpha_a}^2 {\sigma_x}^2} = \frac{\alpha_b}{\alpha_a} = G$$

$$S = G * \frac{\bar{a}}{\bar{b}}$$

From the covariance we define two metrics of relatedness, the gain (G) of a particular gene relative to a given reporter, and the sensitivity (S) of a gene to fluctuations in the reporter as a proportion of its mean.

To derive the expressions above, we made two main assumptions. First, we assumed that gene expression could be modeled as a linear process. This assumption seems to be warranted for the Msn2/4 regulated genes which respond linearly to changes in the activity of Msn2/4. Indeed when a strain containing two distinct Msn2/4 reporters is tested under many different conditions (deletion strains) the two reporters across experiments were linearly related (R^2 =0.59, data not shown).

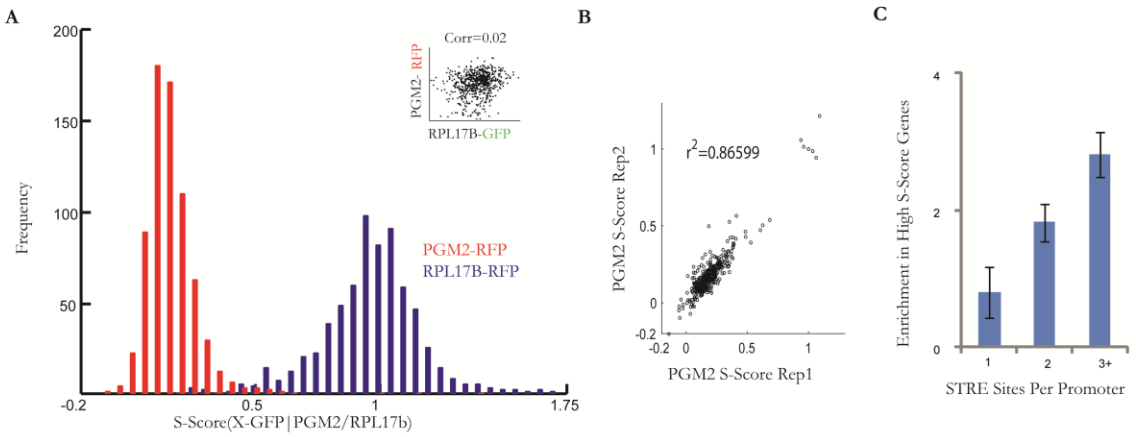


Figure 3: To measure correlations at a genome scale two unrelated query genes, Pgm2 and Rpl17B (inset) tagged with RFP were crossed to 740 GFP strains and the covariance of each pair measured. From these measurements quantitative S-scores were computed histograms of these values for the seven hundred genes are displayed (a). Measurement of replicates of these strains showed that S-score measures were fairly reproducible (b). Genes with the top 10% of S-score values for Pgm2 showed enrichment for Stress Responsive Elements (STREs), errorbars represent 95% confidence intervals from a binomial model (c).

The second assumption is that the noise terms (ε) are independent. In our experiments, these terms represent noise in measurement and also factors not explicitly accounted for in the model of gene expression such as global correlators. For example, cell size (or transcriptional and translational capacity) can uniformly affect the expression of all genes in a given cell and therefore induce interdependence. We attempted to control for this explicitly in our flow cytometry data analysis by using cell size information from the forward and side scatter parameters to remove biases generated by cell size variation. Indeed we find that on average the correlation of the ~700 genes we measured to the stress responsive gene Pgm2 was 0.0471 (+/- 0.0022), looking at ribosomal genes only this correlation drops further to 0.0113 (+/-0.0042). These results suggest that most of the noise from global fluctuations has been accounted for with our data analysis approach, validating that our normalized covariance reports on pathway specific covariation.

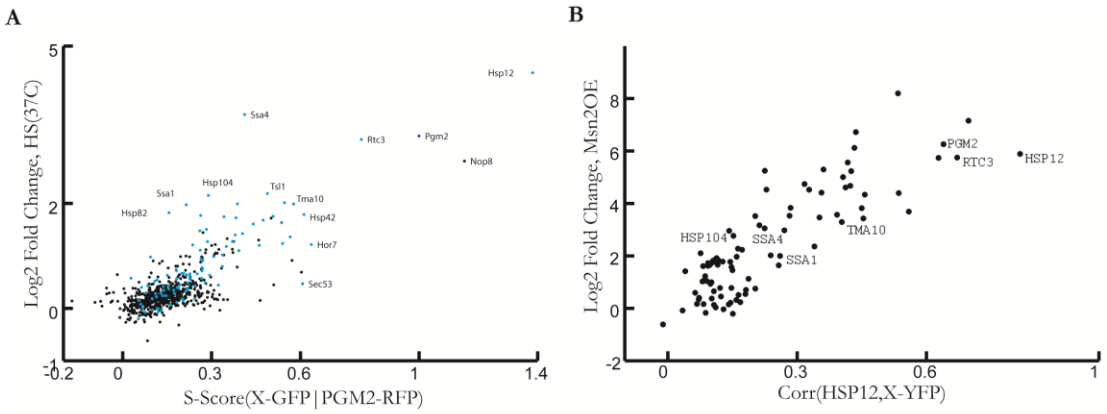


Figure 4: Single cell covariance measures at steady state are predicative of gene expression dynamics under inducing conditions. (a) S-score measures with Pgm2-RFP for 740 functionally diverse genes (Figure 3A) shows strong correlation with the change in GFP expression in those genes after heat shock. (b) Similarly, a strong relationship is observed in ninety promoters enriched for stress responsive genes expressing YFP, between steady state covariance measures and the fold-change in gene expression (measured by YFP) when a constitutively active allele of Msn2(5A) is expressed from a Gal1 promoter.

The distribution of Pgm2 S-scores was reproducible (R^2=0.87, Figure 3B) and right skewed, with the majority of genes weakly correlated with Pgm2 (Figure 3A). Furthermore, for genes in this set, the value of their Pgm2 S-Score was highly predictive of the presence of Msn2/4 binding sites in their promoters (17 of the top 20 genes have consensus 'AGGGG' sites in their promoters, Figure 3C).

The S-score metric aims to capture the local sensitivity of a gene to activity of an upstream regulator with a simple linear model. Although such a linear model is appropriate for low amplitude variability, it should be extendable to larger input perturbations when genes respond linearly to changes in activity of the transcription factors. This is plausibly the case for stress responsive genes, which have been observed to exhibit graded induction profiles in yeast and other organisms (Giorgetti et al., 2010). To test whether the low amplitude variability resulting from noise at steady-state, as captured by an S-score, is quantitatively predictive of the response of a gene to changes in the environment, we compared the PGM2 S-score of a gene with the gene's response to a mild heat shock—37C for 30min. In general, a positive correlation (0.68)

was observed between these two quantities (Fig. 4A), with the genes exhibiting highest Pgm2 S-scores enriched for Msn2/4 binding sites in their promoters. The genes that violated this relationship (including Hsp82, Ssa4, Ssa1) by responding strongly to heat shock while exhibiting moderate S-scores are targets of the heat shock transcription factor Hsf1 (Boy-Marcotte et al., 1999).

To extend and refine this analysis we measured the correlations in a small library of 90 promoters fused to YFP which is heavily biased towards stress responsive genes and expressed in strains which expressed RFP from a HSP12 promoter. For each promoter we compared the correlation to RFP expression in exponential growth with the fold change of that promoter induced by Msn2 overexpression (F4B). Response to Msn2 overexpression is a more specific metric of response to Msn2 than simply heat shock and we expect and observe a much stronger correlation between single cell steady state measures and dynamic activation of gene expression in this data set.

These results argue that noise in gene regulation has physiological implications, and show that extrinsic noise at steady-state is shared across genes in the Msn2/4 stress response pathway. The presence of this pathway-specific noise defines a "noise regulon", with the S-score representing a quantitative measure of pathway membership.

What other noise regulons might exist in a yeast cell? To address this question more globally, we selected a set of 43 mCherry tagged genes (in addition to Pgm2) exhibiting high extrinsic noise. We then measured their S-scores with a set of 180 GFP-tagged genes spanning a range of pathways, noise, and expression levels. The 44 by 180 noise matrix was then clustered by a metric of S-score distance, revealing four major blocks (Fig. 5A). GO term analysis indentified specific three of these blocks as corresponding to stress response, mitochondrial, and amino acid biosynthesis functional categories (Figure 5A,B). The fourth block, which dominates the upper left quadrant, was a more heterogeneous mix of genes and may represent generic

cellular noise. Other smaller groups could also be identified, including a group containing three out of the four histone proteins featured our dataset (HTZ1, HTA2, HTB2) in addition to TCB3 and YMR295C, two other cell cycle regulated proteins (Spellman et al., 1998).

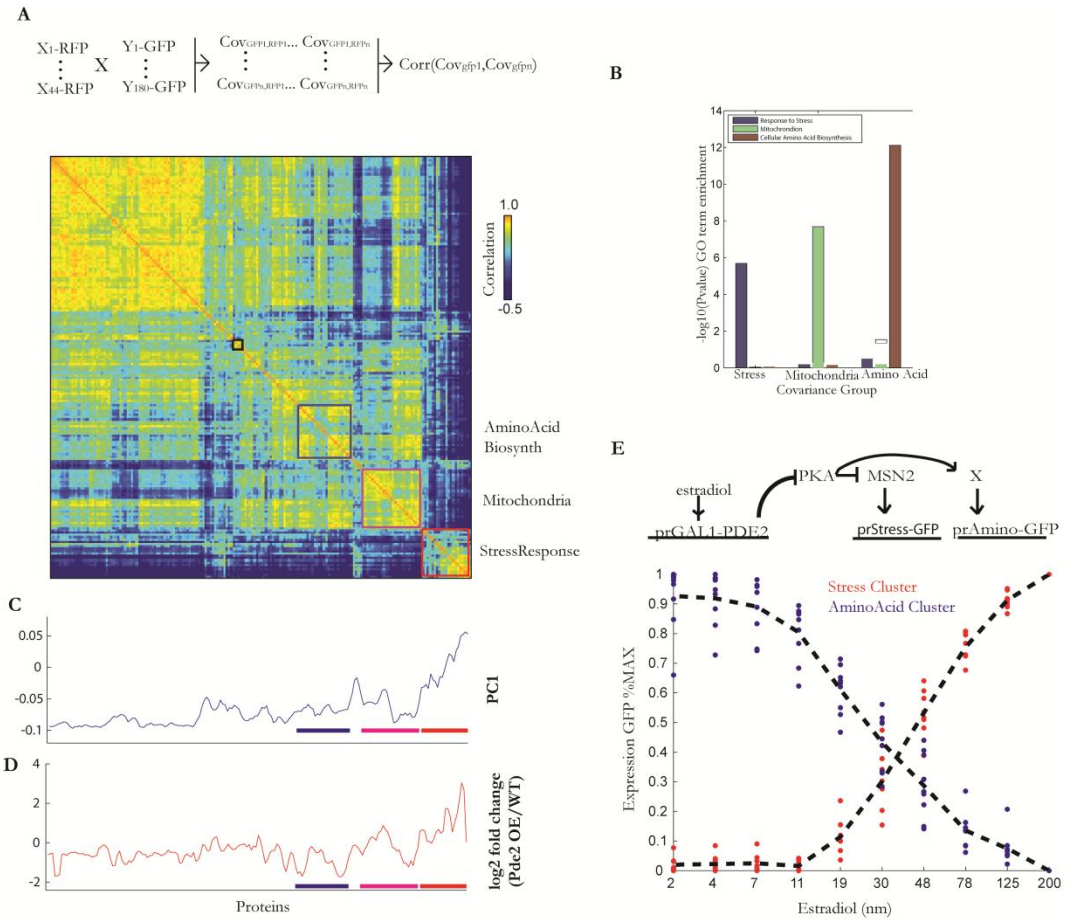


Figure 5: Noise measurements divide the genome into distinct regulons and inform an understanding of the dimensionality of a cell. (a) A map of noise correlations among 182 proteins showing distinct blocks that share patterns of covariance. Prominent among these are Amino Acid biosynthesis (blue outline), Mitochondrial (magenta outline), and Stress responsive (red outline) clusters. A smaller group of cell cycle regulated genes are also apparent (black). (b) p-values of GO term enrichments for each block identified in the covariance map. (c) A principal component analysis of the covariance dataset shows that five principle components (PCs) can describe ~80% of the observed variance. Contribution of the noise in different genes to the first PC, PC1, is plotted in the order in which they appear in (a). Values are smoothed using a sliding window of size 3. (d) Expression of 184 GFP-tagged proteins in a Pde2 over-expression background. The results are reported as a log2 of the ratio of GFP in the overexpression strain normalized to that of the wild-type. Overexpression of Pde2 was achieved using an estradiol inducible GAL1 promoter. Proteins are in the order in which they appear in *(a)*. Induction values were smoothed with a sliding window of size 3. (e) Repression of PKA by Pde2 overexpression using a titrateable estradiol regulated system results in increased expression of genes in the stress gene cluster and repressed expression of the amino acid cluster as estradiol is added and PKA is inhibited.

We examined the putative amino acid, mitochondrial, and stress responsive clusters and assessed the enrichment of transcription factors binding sites in the promoters of genes featured in these clusters. Drawing on ChIP data (Zhu et al., 2011), we determined that the amino acid biosynthesis group was enriched for the amino acid regulatory transcription factors GCN4, MET4, and ARG80/81, that the mitochondrial group was strongly enriched for members of the HAP2/3/4/5 complex which is known to regulate heme synthesis and oxidative metabolism, and that the stress responsive group was significantly enriched for Msn4 binding (data not shown). Therefore, our data reveals that in addition to generic extrinsic fluctuations resulting from variability in general cellular factors (background noise), genes across a yeast cell are subject to a small number of structured and pathway-specific extrinsic noise sources. Such noise sources induce correlated fluctuations in the expression of pathway genes, and hence could be used as a quantitative metric for pathway membership.

To explore this idea further, we took advantage of the data representation in terms of a covariance matrix, which renders it interpretable in terms of standard dimensionality reduction techniques such as Principal Component Analysis (PCA). By subjecting the covariance matrix to PCA, we can compute the number of dimensions necessary to describe the noise patterns under unperturbed conditions. PCA analysis revealed that ~80% of the variance in the data could be explained by the first 5 principal components (Figure 5C).

Further, these principal components appeared readily interpretable in terms of cellular pathways. For example, the contribution of the first principal component to each gene, plotted in the same order as the clustered data, showed a strong peak spanning the Msn2/4 stress responsive cluster and a broad negative region corresponding to the remaining genes (Figure 5C). To extended this qualitative analysis we correlated the first principal component to mRNA expression data over a range of environmental conditions (Gasch et al., 2000), the most correlated

conditions were for growth at low temperatures and on poor carbon sources (pvalue=6.60*e-3). Growth on non-glucose carbon sources is known to inactivate the PKA pathway, which also results in MSN2/4 activation (Gorner et al., 1998), consistent with the pattern we observe for the first principal component. To directly test if in activation of PKA would result in the pattern of gene expression predicted by our PCA analysis we overexpressed PDE2 which inhibits PKA activity by degrading cAMP and measured GFP expression in our set of 180 strains. Principal component one correlates with the log2 change in gene expression after overexpression of PDE2 (Pearson correlation 0.62, pvalue 4.04e-21), and shows a very similar pattern of stress and mitochondrial gene induction and repression of other types of genes (Figure 5D). These results underscore the predictive power of steady state fluctuations for genetic or environmental conditions, and also suggests that PCA derived components of this dataset have biological as well as mathematical implications.

We then asked if given our analysis of the data there are any surprising interactions, one feature that caught our eye is that if we assume the stress responsive genes are correlated by PKA activity it suggests that things that are anti-correlated to the stress response should be positively regulated by PKA activity. To test this we compared the expression of genes from our stress responsive and amnio-acid biogenesis cluster in response to a gradient of PKA activity. We generated this gradient, by overexpressing PDE2 to a range of different final concentrations by adding different quantities of estradiol and using an Estradiol regulated transcriptional induction system (Stewart-Ornstein et al., 2012) to obtain a graded degree of PKA inhibition. We preformed this experiment and normalized the expression of each gene to the maximum and minimum values (Figure 5E), intriguing we observe that the amino-acid biosynthesis gene are indeed inhibited by PKA inhibition (or to rephrase PKA exerts a positive effect on their expression), whereas the stress responsive genes rise. Interestingly when normalized in this manner genes of each type behave as a coherent block, suggesting that PKA acts on all genes in a

uniform way. These results argue that PKA, through unknown intermediates positively regulated the expression of amino acid biogenesis genes.

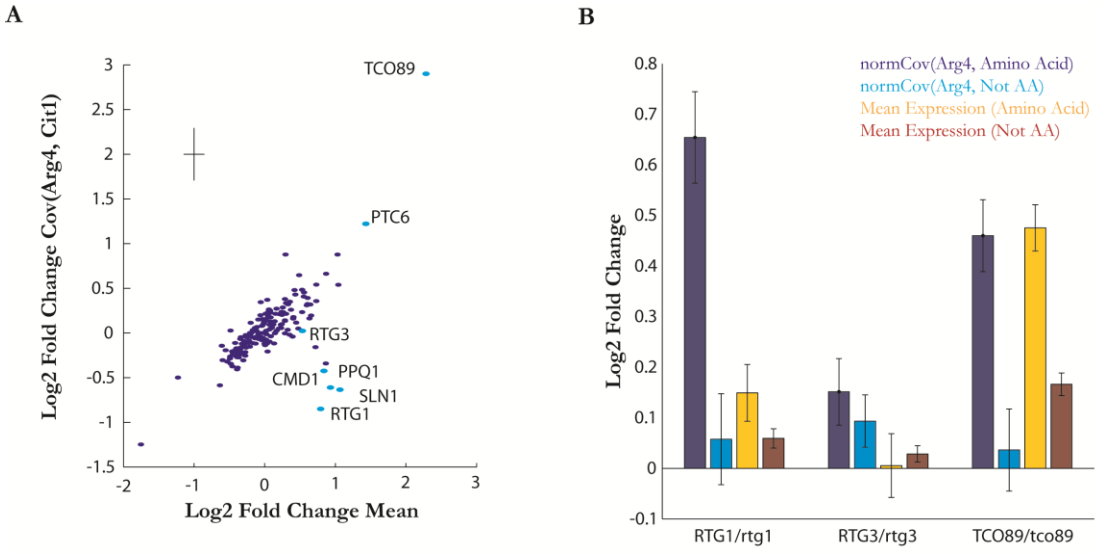


Figure 6: Covariance measurements of heterozygous strains link noise in amino acid biosynthesis to Tor signaling. (a) Covariance of the amino acid biosynthesis protein Arg4-mCherry and the mitrochondrial protein Cit1-GFP plotted against the geometric mean of Arg4-mcherry and Cit1-GFP for 188 heterozygous deletion mutants. Values are displayed as log2 fold change over WT. Strains referred to in the text are highlighted in cyan. Error bars in upper left show standard deviation of replicate measurements.(b) Change of Arg4 mean and covariance between Arg4 and members in the amino acid biosynthesis group in *RTG1/rtg1*, *RTG3/rtg3,* and *TCO89/tco89* heterozygous strains. Values are displayed as log2 fold change over WT. *RTG1/rtg1*and *TCO89/tco89* heterozygous deletions cause significant increases in covariance, while *RTG3/rtg3* has no significant effect on the covariance. *TCO89/tco89* heterozygous deletion, but not *RTG1/rtg1* heterozygous deletion, increases mean gene expression of amino acid biosynthesis genes. Error bars represent standard error of means

To explore the regulation and specifically the sources of variability in the amino acid biosynthesis genes in an unbiased way we adopted a covariance based approach where we constructed a query strain expressing Arg4-GFP and CIT2-mKate2, a pair of genes from the mitochondrial and amino acid regulons which show substantial covariance and crossed this strain to 192 genes from the deletion collection selected to focus on kinases, regulatory proteins, and transcription factors to produce a set of heterozygous strains. We measured the RFP and GFP signals in each of the resulting strains and computed the covariance and mean expression in each

case (Figure 6A). These two quantities showed a linear relationship as one would expect for a poissonian noise source. Most strikingly the *TCO89/tco89* strains showed strongly elevated variability and mean expression for both genes. Also interesting was that the mitochondrial regulatory *RTG1/rtg1* strains showed greatly reduced covariance with little change in mean expression. To see if the perturbations induced by these mutations were general we constructed 180 strains which were ARG4-RFP, X-GFP (where X is one of the strains we assayed earlier) and either WT or *RTG1/rtg1*, *TCO89/tco89*, *RTG3/rtg3*. Measuring covariance and mean of these strains we found that consistent with our screening results among the amino acid regulon *TCO89/tco89* greatly increased mean gene expression and covariance, whereas *RTG1/rtg1* resulted in increase covariance with little change in mean (Figure 6B). These results implicate the TOR signaling pathway and mitochondrial activity as drivers of cell-to-cell variability in the amino acid regulon, and suggest an experimental approach using classical screening strategies to explore the source of variability in gene expression is viable.

Discussion 1.2

In this work we define the contribution to noise from small stochastic single allele fluctuations inversely correlated to expression level, large global fluctuations, and pathway specific fluxes. Focusing on extrinsic noise our results show that global fluctuations induce a minimum level of noise in gene expression in even the most highly expressed genes, but that excursions above this limit are due to fluctuations in signaling in specific pathways. These fluctuations can be exploited to define co-regulated groups of genes through a shared noise signature.

Noise within a pathway is shared across co-regulated genes and in conjunction with libraries of GFP tagged genes can be used to define regulons and in a relatively unbiased way search for genes involved in any given processes. A noise based approach to assigning regulation to genes stands apart from other expression based methods as it requires no external perturbation.

Thus, no knowledge of when, how, or what induces the pathway of interest is necessary. Further, this method requires no assumption that cellular wiring is constant across conditions or time, indeed applying a similar covariance analysis to a network as it evolves in response to an input might allow a dynamic mapping of network connections as they occur.

Correlations induced by molecular fluctuations also proved to be a quantitative measure of the sensitivity of a given gene to its regulator. This is intriguing as it suggests equivalence between local fluctuations and global induction and implicitly requires that gene expression is graded rather than switch like, a results we will return to later.

Using a quantitative covariance measurement strategy we delineated several noise regulons operating across a yeast cell. Regulons we assigned to stress response, amino acid biosynthesis, and mitochondria were strongly enriched in corresponding GO terms. Using a genetic strategy we traced the fluctuations in the general stress response and amino acid biosynthesis to PKA and TOR regulation respectively implicating these master regulators in tuning the balance between stress and growth across a population. Consistent with the importance of these two crucial pathways the results of a PCA analysis of the covariance matrix suggested that only a handful of factors were necessary to explain the noise observed. It is, however, entirely possible that the simplicity is a result of the rapid growth conditions we observe, and might disappear or evolve if noise was measured under different conditions.

The existence of large correlated groups of genes in populations of isogenic cells suggests that correlated modules might exist to maintain useful, but controlled, diversity across an unperturbed population. Further studies will be needed to elucidate the molecular basis and quantitative role of this organization for the fitness of a yeast cell. However, the prominence of the stress protective MSN2/4 pathway as an active organizational module is consistent with a role for correlated noise in ‘bet-hedging’ strategies that are used by unicellular organisms to navigate varying environments. Similar investigations of noise regulons under a variety of environmental

conditions and stimuli will be a powerful tool for probing systematically the logic and implementation of such strategies.

1.3

Protein/Chaperone interactions

In addition to transcribing genes and translating mRNAs the protein products of these processes need to be folded and maintained in a functional conformation. If all proteins were meta-stable in a folded conformation this would require little additional cellular complexity, but increasing evidence suggests that many proteins and particularly active signaling molecules such as kinases need to be actively maintained in a productive conformational state by the concerted action of networks of chaperones. We therefore considered how the complexity of the budding yeast protein folding network could be explored using yeast genetics.

Protein misfolding caused by temperature, stress, or the genetic or chemical inhibition of chaperones can often result in aggregation of poorly folded proteins. Extensive work in *C. elegans* and *S. cerevisiae* suggests that temperature sensitive proteins, which were selected to have impaired function and presumably folding at higher temperatures, are particularly vulnerable to changes in the proteostatic environment often aggregating at permissive temperatures in the face of such perturbations (Kaganovich et al., 2008; Gidalevitz et al., 2006; Silva et al., 2011). This suggested that monitoring how a large number of different temperature sensitive mutants behave in response to different perturbations to the cellular folding environment might reveal shared folding pathways and give insight into the dimensionality and complexity of protein folding.

Results 1.3

We reasoned that if different proteins required different chaperones then these proteins would also have different sensitivities to deletion of chaperones. To measure the folding of proteins we focused on essential proteins whose folding impairment would result in a measureable growth defect, and further we used strains where one essential protein per-strain was a temperature sensitive allele. We would therefore expect a strain with a ts-protein to be particularly sensitive to perturbations that effected the folding of that particular protein. To assay these interactions in high throughput we took advantage of an existing tsLibrary of ~290 ts strains from the Hieter group (Ben-Aroya et al., 2008), using SGA technology we crossed these strains to eleven strains with deletions of core chaperones (HSP104, SSA2, HSC82, STI1, GET1, and YDJ1), overexpression of the chaperone HSP104, the prion protein URE2, or constitutively unfolded PEP4(delta-signal sequence), or deletion of the C-terminal of the core folding regulator HSF1. We then quantified the colony size of these strains grown in 24 by 16 arrays pinned onto YPED plates at 25, 30, and 35C (Figure 7A).

We began by quantifying the reproducibility of these measurements by comparing two replicates of growth at 30C across all eleven plates. These replicates showed good agreement with an R^2 of 0.927. By contrast comparing strains grown at 30C to those grown at 25 or 35C we see similar overall trends but large numbers of strains that grow better at 25C as compared to 30C, or worse at 35C as opposed to 30C (Figure 7B).

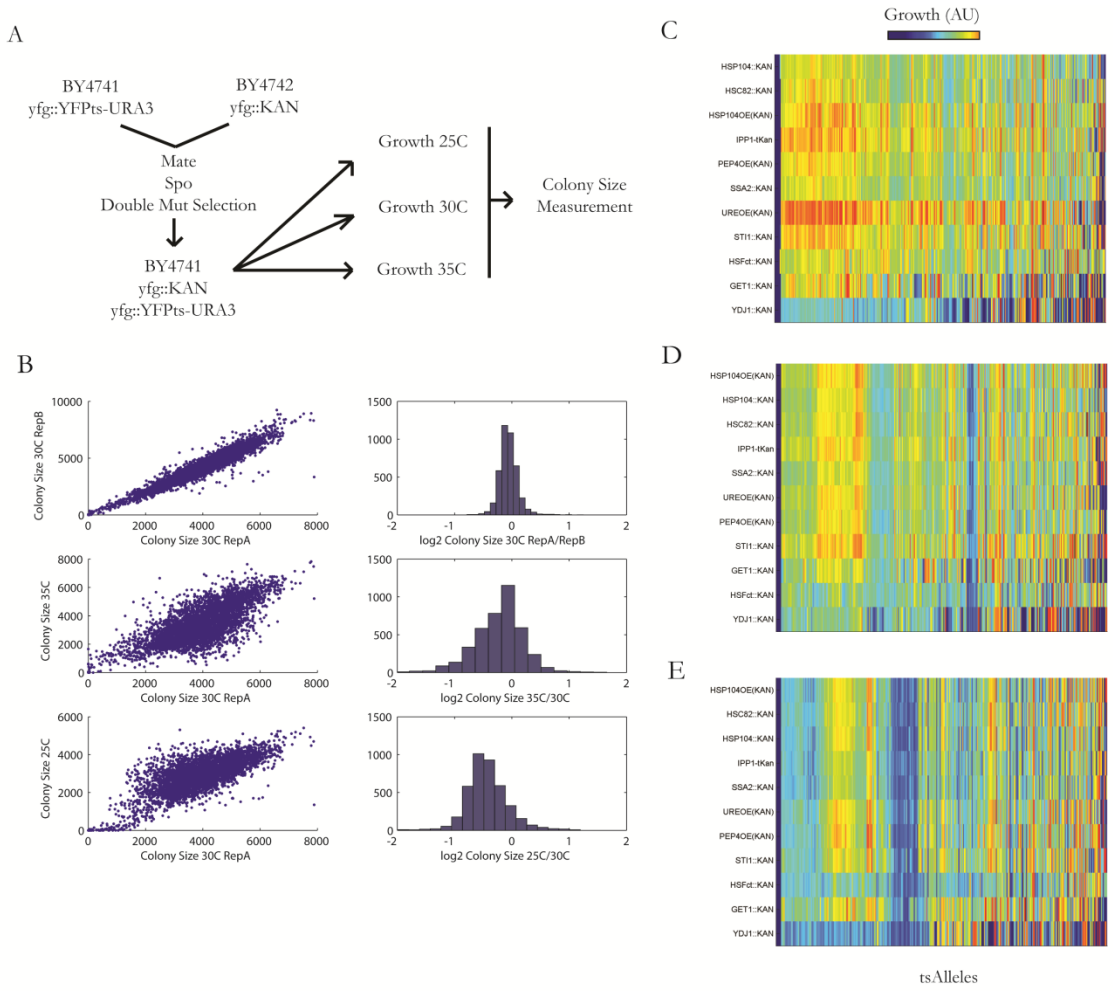


Figure 7: An attempt to identify groups of co-folded genes by sensitivity of tsAlleles to proteostatic network purturbation in an array format. (a) Mating strategy to generate tsAllele libraries, breifly MATa strains with URA3 marked tsAlleles of essential genes were crossed to MATalpha strains with a proteostatic purturbation marked with G418 resistance, these diploids were then sporulated, selected for double mutants and then grown at various temperatures in 384 colony plate format. (b) Measurement of colony size was highly reproducible between replicates, on average colony size was decreased by growth at higher temperatures (35C) and increased by low temperature(25C). (c,d,e) Clustering of data at each growth temperature of 11 (proteostatic purturbations) by 380 (tsAlleles), no obvious clustering of proteome purturbations was observed.

We then hierarchically clustered this dataset and asked if we observed like perturbations to cluster together, specifically we expected deletion of HSC82 and its co-factor STI1 to cluster together and the control perturbation (insertion of the KAN cassette into an intragenic region downstream of the gene IPP1) to cluster poorly with the other perturbations. Unfortunately, we

observed that the clustering resulted in no particularly coherent grouping of genes (Figure 7C). Moreover, comparison of the clustering obtained with the 30 and 35C datasets revealed no strong changes suggesting that patterns we do observe are not strongly temperature dependent (Figure 7D,E).

Discussion 1.3

These analyses suggested that the use of this temperature sensitive library to explore chaperone mediated protein folding was unsuccessful. The reasons for this failure were not explored further, but we suggest two possibilities. First, that the temperature sensitive alleles used in this strain are not compromised for folding per-se, but rather compromising the activity of these genes makes these strains temperature sensitive for growth. Second, that chaperones in yeast form a complex interconnected network that perturbations made by deleting a single component have extremely pleiotropic effects which are not specific for any particular folding process. Follow up work using the same approach is not recommended, but recent results from other groups suggest that functional assays of folding aggregation combined with yeast genetics may prove a powerful tool to probe proteostasis networks (Theodoraki et al., 2012). During the planning stages of this project some work was put into devising a transcriptional assay for mis-folding of a temperature sensitive allele of the Gal80 protein, whose activity can be monitored by the transcription of the GAL1 promoter, providing a flow-cytometry applicable tool for measuring mis-folding, further work in this direction could be warranted.

Chapter 1 Methods

Construction of synthetic circuits

Individual components were constructed consisting of DNA binding domains (TETR, GAL4dbd, and LACI) fused to XFPs (C-terminal for TETR and GAL4dbd, N-Terminal for LACI). These components were placed under control of the GAL1 promoter, ADH1 promoter, or these promoters modified by the insertion of TET or LAC operators [Collins, Tang]. In all cases constructs were assembled in 'single-integration vectors' which integrate at the TRP1, LEU2, or HIS3 loci.

Strains were derived from W303a, reverted to ADE+ to reduce auto-florescence, and serially transformed with linearized plasmids and then selected for on media lacking the appropriate amino acid.

Flow Cytometry Experiments

For steady state measurements cells were grown to saturation in 96-shallow well plates (Costar) and then diluted into fresh media, grown at 30C on orbital shakers (Elim) for 12hrs to an OD of ~0.5. These cells were subsequently diluted (typically 1:200) and drugs (or galactose) added as necessary and grown for 8hrs to an OD of ~0.05 before measurement. Strains with GAL1 promoters were induced by growth for >18hrs in SC+2%rafinose, followed by the addition of 2% Galactose to induce the strains, under these conditions steady state GAL1 expression was observed by 6-8hrs. Tetracyclin (aTC, Sigma Aldrich, stock 2mg/ml in DMSO) was typically applied in a log 2 dilution series from 128ng/ul to 0.5ng/ul.

Timecourse experiments were performed by growing strains to saturation in SC+2%RAF, followed by dilution (1:1000) and growth to OD0.5 (~14hrs), these cells were then spun down and washed 2X with SC(no sugar), resuspended in 10X volume of SC2%Gal and transferred to 96 well deepwell plates which were placed in a shaking incubator at 30C. 100ul samples were

collected at each timepoint and immediately run on the flow cytometer, an equal volume of SC+2%GAL was added to the culture to maintain constant volume and ~constant OD.

All cytometry measurements were made on a Becton Dickinson LSRII flow cytometer, along with an autosampler device (HTS) to collect data over a sampling time of 6-12 seconds, typically corresponding to 2000-10000 cells. GFP and YFP were excited at 488nm, and fluorescence was collected through a HQ530/30 bandpass filters (Chroma), mCherry was excited at 561(or 532) nm and fluorescence collected through 610/20 bandpass filter (Chroma), CFP was excited at 405nm and emission collected through a 465/30 bandpass filter.

Flow Cytometry Data Analysis

Flow cytometry data was first processed to remove outliers by filtering the data using a minimum covariance determinant method [Rousseeuw and Van Driessen, 1999] (~80-90% of original cells are kept). Other approaches to filtering data or calculating 'robust' measures of variance and mean (e.g. Mean Absolute Deviation or S-estimators) were explored and gave similar, though less reproducible, outcomes. In practice these approaches proved similar to gating as both remove the most extreme outliers.

Second, to correct for variations in cell size (and cell cycle), we utilized forward and side scatter information (FSC and SSC). Traditionally this is done by selecting only a subset of cells with similar size/shape (gating), and calculating means and covariances from this subset of data [Newman et al., 2005]. As the gate size becomes infinitely small (i.e. all cells within the gate have identical forward and side scatter properties), any covariance measured will be independent of these parameters. On the other hand, as the gate becomes smaller and the number of measured cells decreases errors magnify. Further gating typically discards the vast majority of the data (80-99% in some reports).

We addressed this problem in two ways. First, we viewed the measurements from a particular experiment as a population from which to draw samples. We selected a single cell

randomly and found the N cells that are most similar to it in forward and side scatter space (based on a distance metric with a determined threshold). We then computed variance and covariance from this sample of cells. Note that by taking a fixed number of nearby cells, the error in computing the covariance for each group is constant, but the degree of similarity between cells within a group is not. By repeating this procedure many times (500 was sufficient to produce sampling errors in measured covariance of <1%), robust estimates of variance and covariance were computed. The size of N was empirically set at 100, as further reductions result in negligible changes in covariance. We will refer to this procedure as multi-gating.

The second approach is to view the flow-cytometry (GFP, RFP, FSC, SSC) measures as a multi-dimensional dataset and determine using partial correlations the covariance between GFP and RFP (or variance within GFP or RFP) measurements given the correlation due to cell size as measured by side scatter. This is similar in spirit to a recent analysis of cell-to-cell variation in flow cytometry data [Rinott et al., 2011]. This assumes relatively linear and normal relationship between side scatter and fluorescence. In practice, this approach results in estimates for mean and noise that are similar to a simple linear transformation of the data by dividing by side scatter or other measures of cell size which other groups have adopted [Murphy et al., 2010;, Raser and O'Shea, 2005].

The correlation values reported in Figure 1 (and subsequent figures in the manuscript) are from results computed with partial correlations. Code to implement these analyses was written in matlab (see appendix).

Dual color strain construction and covariance measurements

For crossing to generate dual color strains, individual RFP marked strains were were constructed by PCR based homologous insertion of an RFP protein (mCherry or mKate2) at the C-Terminal end of the open reading frame with a URA3 marker immediately 3'. Reporter strains marked with mCherry (or mKate2) were crossed to strains from the GFP library (Open Biosystems), with

a selection step for diploids in SD -Ura/-His. Pure diploid populations were verified as containing no cells lacking either fluorophore using flow cytometry.

Briefly, a RFP marked strain was gown up in 5ml YPD, and GFP strains grown in 96 or 384 well plates (200 or 80ul of YPD). These cells were mixed in liquid on 96 or 384 well plates and incubated at 30C for 24hrs to allow for mating. Diploids were then selected by 1:8000 fold dilution into SD-URA-HIS using a biomech robotic system. These strains were then diluted 1:32000 using a Biomech robotic system and grown for 20-22 hrs before measurement (final OD ~0.05-0.1), the 96 or 384 plates were measured using an HTS autosampler on the LSRII cytometer (BD). Plate growth was at 30C with orbital shaking on Elim heater shakers.

Overexpression of PDE2 and MSN2

To over-express PDE2 or MSN2(5A), the appropriate sequence was amplified from the genome and placed in a vector with a GAL1 promoter, this construct was then integrated at the TRP1 locus in cells expressing an estradiol inducible system as described above. Cells were induced using concentrations of estradiol ranging between 1 and 200nm in SDC for 6hrs before measurement.

Hetrozygous Deletion Strain Construction and Analysis

Strains for HSP12 and PGM2 covariance measurements were constructed by deletion of the stated genes (IRA2, GPB, PDE2, RAS2) by homologous recombination with a NAT marker from a PGM2-mCherry MATα strain (MATα *his3 leu2 met15 ura3* can1::STE3pr-HIS3 lyp1::STE2pr-LEU2 PGM2-mCherry (URA3) YFG::NAT). This strain was then crossed to (MATa *his3 leu2 met15* HSP12-GFP(HIS3)) to construct heterozygous strains expressing PGM2-mCherry and HSP12-GFP and selected for in SD-HIS-URA.

To construct diploids hetrozygous strains, a MATα strain with tagged copies of ARG4 and CIT1 (MATα *his3 leu2 met15 ura3* can1::STE3pr-HIS3 lyp1::STE2pr-LEU2 CIT1-mCherry (URA3) ARG4-GFP (HIS3)) was crossed to deletion strains from the Yeast deletion collection

(MATa *his3 leu2 met15 YFG::*KAN). Diploids were selected for in two steps with an initial selection in YPD+G418(200ug/ml) followed by selection in SD-URA+G418(500ug/ml). A complete list of these strains is in table S4.

To study the effects of heterozygous deletions of key genes on the covariance between ARG4 and many different GFP tagged genes, we crossed an ARG4-mCherry strain with a specific gene deletion (MATα *his3 leu2 met15 ura3* can1::STE3pr-HIS3 lyp1::STE2pr-LEU2 ARG4-mCherry (URA3) YFG::NAT) to the 182 strains used in the covariance matrix analysis. Selection for diploids was in SD–URA-HIS.

These strains were inoculated into shallow 384 well plates (Fisher) and grown to saturation overnight in 80ul SD-URA/HIS. They were then diluted 1:32000 using a Biomech robotic system and grown for 22-24 hrs before measurement (final OD ~0.05-0.1), the 384 plates were measured using an HTS autosampler on the LSRII cytometer (BD). In all cases, growth was at 30C with orbital shaking. Flow cytometery data for these strains was collected, the data analyzed as described below.

tsStrain Construction

A collection of ts mutant stains was obtained from the Hieter group [], these strains were in a BY4741/2 background with the following genotype (MATa *yfg*::tsYFP(URA3) *can1*::LEU2 prMFA1-HIS3, *met15* or *lys2*). To mate to these strains we constructed the following query strains BY4742(*yfChaperone*::KAN *his3, leu2, ura3, lys2*) or alternatively BY4742(*trp1::*prADH1-YFunfoldedprotein (TRP1,KAN) *his3, leu2, ura3, lys2*). Where yfChaperone was a chaperone of interest and YFunfolded protein was a unfolded protein to be overexpressed.

Mating and selection was based on the SGA technology developed in the Boone Lab. Briefly the query strain was mated to the temperature sensitive collection on YPD (1Day 30C), diploids selected on SD-URA+KAN (1Day 30C), pinned to sporulation media and allowed to

sporulate (3Days room temperature), and pinned to SD+ Canavanine+KAN-Ura-Leu-His-Met (2Days, 30C), repined to the same media (1Day, 30C), pinned onto YPD+KAN (1Day).

tsSensitive Strain Measurement

Strains were assayed by pinning to fresh YPD plates that were then incubated at various temperatures for 1-3 days and then imaged by scanning on a document scanner (Epson). Colony sizes were analyzed with custom matlab software written by Onn Brandmann.

Chapter 1 References

Becskei A, Kaufmann B.B., Oudenaarden A (2005). Contributions of low molecular number and chromosomal positioning to stochastic gene expression. Nature Genetics **37**: 937-944.

Ben-Aroya S, Coombes C, Kwok T, O'Donnell KA, Boeke JD, Hieter P (2008). Toward a comprehensive temperature-sensitive mutant repository of the essential genes of Saccharomyces cerevisiae. Mol Cell. **30**(2):248-58.

Bigger W (1944). Treatment Of Staphylococcal Infections With Penicillin by Intermittent Sterilisation. The Lancet **244**(6320): 497-500.

Blake WJ, Kaern M, Cantor CR, Collins JJ (2003). Noise in eukaryotic gene expression. Nature **422**: 633–637.

Blake WJ, Balazsi G, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, Cantor CR, Walt DR, Collins JJ (2006). Phenotypic consequences of promoter-mediated transcriptional noise. Molecular Cell **24**:853-865.

Broach JR 1979 .Galactose regulation in Saccharomyces cerevisiae. The enzymes encoded by the GAL7, 10, 1 cluster are co-ordinately controlled and separately translated. J Mol Biol. **131**(1):41-53.

Burnett JC, Miller-Jensen K, Shah PS, Arkin AP, Schaffer DV (2009). Control of Stochastic Gene Expression by Host Factors at the HIV Promoter. PLoS Pathog **5**(1): e1000260. doi:10.1371/journal.ppat.1000260

Cantone I, Marucci M, Iorio G, Ricci MA, Belcastro B, Bansal M, Santini S, di Bernardo M, di Bernardo D, and Cosma MP. A Yeast Synthetic Network for In Vivo Assessment of Reverse-Engineering and Modeling Approaches. Cell, **137**(1): 172-181.

Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, Pesce CG, Brent R (2005). Regulated cell-to-cell variation in a cell-fate decision system. Nature **427**: 699-706.

Delbruck M (1945). The Burst Size Distribution in the Growth of Bacterial Viruses (Bacteriophages). J Bacteriol **50**(2): 131-5.

Dunlop MJ, Cox RS 3rd, Levine JH, Murray RM, Elowitz MB (2008). Regulatory activity revealed by dynamic correlations in gene expression noise. Nat Genet. **40**(12):1493-8.

Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002). Stochastic gene expression in a single cell. Science. **297**(5584):1183-6.

The ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature **489**: 57–74.

Gardner TS, Cantor CR, Collins JJ (2000). Construction of a genetic toggle switch in Escherichia coli. Nature, **403**(20): 339:342.

Gasch AP, Spellman PT, Kao CM,Carmel-Harel O, Eisen MB,Storz G, Botstein D, Brown PO (2000). Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. Molecular Biology of the Cell **11**: 4241–4257.

Gidalevitz T, Ben-Zvi A, Ho KH, Brignull HR, Morimoto RI (2006). Progressive disruption of cellular protein folding in models of polyglutamine diseases. Science, **311**(5766):1471-4.

Giorgetti L, Siggers T, Tiana G, Caprara G, Notarbartolo S, Corona T, Pasparakis M, Milani P, Bulyk ML, Natoli G (2011). Noncooperative Interactions between Transcription Factors and Clustered DNA Binding Sites Enable Graded Transcriptional Responses to Environmental Inputs. Molecular cell **37**(3): 418-428.

Golding I, Paulsson J, Zawilski SM, Cox EC (2005). Real-Time Kinetic of Gene Activity in Individual Bacteria. Cell **123**:1025-1036.

Görner W, Durchschlag E, Martinez-Pastor MT, Estruch F, Ammerer G, Hamilton B, Ruis H, Schüller C (1998). Nuclear localization of the C2H2 zinc finger protein Msn2p is regulated by stress and protein kinase A activity. Genes Dev. **12**, 586-597.

Görner W., Durchschlag E., Wolf J., Brown E.L., Ammerer G., Ruis H., Schuller C (2002). Acute glucose starvation activates the nuclear localization signal of a stress-specific yeast transcription factor. EMBO **21**: 135-144.

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KS, Lander ES, Gifford DK, Fraenkel E, and Young RA (2004). transcriptional regulatory code of a eukaryotic genome. Nature **431**: 99-104.

Holgate P (1964). Estimation for the bivariate Poisson distribution. Biometrika, **51**: 241-245.

Hooshangi S, Thiberge S, and Weiss R (2005). Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. PNAS **102**(10): 3581-3586.

Howson R, Huh WK, Ghaemmaghami S, Falvo JV, Bower K, Belle A, Dephoure N, Wykoff DD, Weissman JS, O'Shea EK (2005). Construction, verification and experimental use of two epitope-tagged collections of budding yeast strains. Comp Funct Genomics. **6**(1-2):2-16.

Huh D, Paulsson J (2011). Non-genetic heterogeneity from stochastic partitioning at cell division. Nature Genetics **43**: 95–100.

Jacob F, Monod J (1961). Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol. **3** (3): 318–56

Jacquet M, Renault G, Lallet S, DeMey J, Goldbeter A. Oscillatory nucleocytoplasmic shuttling of the general stress response transcriptional activators Msn2 and Msn4 in *Saccharomyces cerevisiae*. Journal of Cell Biology **161**(3), 2003: 497-505.

Jonikas MC, Collins SR, Denic V, Oh E, Quan EM, Schmid V, Weibezahn J, Schwappach B, Walter P, Weissman JS, and Schuldiner M (2009). Comprehensive Characterization of Genes Required for Protein Folding in the Endoplasmic Reticulum. Science **323**(5922): 1693-1697.

Kaganovich D, Kopito R, Frydman J (2008).Misfolded proteins partition between two distinct quality control compartments. Nature, **454**(7208):1088-95.

Lee HK, Hsu AK, Sajdak J, Qin J, and Pavlidis P (2004). Coexpression Analysis of Human Genes Across Many Microarray Data Sets. Genome Res. **14**: *1085-1094*

Lehner B (2010). Conflict between Noise and Plasticity in Yeast. PLoS Genet **6**(11): e1001185. doi:10.1371/journal.pgen.1001185

Lezon TR, Banavar JR, Cieplak M, Maritan A, and Fedoroff NV (2006). Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. Proc Natl Acad Sci U S A. **103**(50):19033-8.

Liu Z, Butow RA (1999). A transcriptional switch in the expression of yeast tricarboxylic acid cycle genes in response to a reduction or loss of respiratory function. Mol Cell Biol. **10**:6720-8.

Martinez-Pastor MT, Marchler G, Schuller C, Marchler-Bauer A, Ruis H, Estruch F The Saccharomyces cerevisiae zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). EMBO J **15**(9), 1996:2227-35.

McGill WJ, IRE Trans. Info. Theory **4**, 93 (1954).

das Neves RP, Jones NS, Andreu L, Gupta R, Enver T, Iborra FJ(2010). Connecting Variability in Global Transcription Rate to Mitochondrial Variability. PLoS Biol **8**(12): e1000560. doi:10.1371/journal.pbio.1000560

Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS (2005). Single-cell proteomic analysis of S. cerevisiae reveal the architecture of biological noise. Nature **441**, 840-846.

Ozbudzk EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. (2002) Regulation of noise in the expression of a single gene. Nature Genetics **31**: 69:73.

Pachkov M, Erb I, Molina N, van Nimwegen E (2007). SwissRegulon: a database of genome-wide annotations of regulatory sites. Nucleic Acids Research, **35**, Database issue D127-D131

Park H., Pontius W., Guet C.C., Marko J.F., Emonet T. and Cluzel P (2010). Interdependence of behavioural variability and response to small stimuli in bacteria. Nature **468**(7325):819-23.

Paulsson J (2004). Summing up the noise in gene networks. Nature **427**, 415–418.

Pedraza JM, van Oudenaarden A (2005). Noise Propagation in Gene Networks. Science **307**: 1965-1969.

Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biology **4**(10): e309. doi:10.1371/journal.pbio.0040309.

Raser JM, O'Shea EK (2004). Control of Stochasticity in Eukaryotic Gene Expression. Science **304**: 1811-1814.

Reinke A, Anderson S, McCaffery JM, Yates J 3rd, Aronova S, Chu S, Fairclough S, Iverson C, Wedaman KP, Powers T (2004). TOR complex 1 includes a novel component, Tco89p (YPL180w), and cooperates with Ssd1p to maintain cellular integrity in Saccharomyces cerevisiae. J Biol Chem. **279**(15):14752-62.

Rousseeuw PJ and Van Driessen K (1999). A Fast Algorithm for the Minimum Covariance Determinant Estimator, Technometrics **41**:212-223.

Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP (2005). Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. Science **308**: 523-529.

Schedl P, Artavanis-Tsakonas S, Steward R, Gehring WJ, Mirault ME, Goldschmidt-Clermont M, Moran L, Tissières A (1978). Two hybrid plasmids with D. melanogaster DNA sequences complementary to mRNA coding for the major heat shock protein. Cell. **14**(4):921-9.

Schmit AP, McEntee K (1996). Msn2p, a zinc finger DNA-Binding protein, is the transcriptional activator of the multistress response in Saccharomyces cerevisiae. PNAS **93**: 5777-5782.

Schäfer, J. and Strimmer, K. (2005a) An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics*, **21**, 754–764.

Silva MC, Fox S, Beam M, Thakkar H, Amaral MD, Morimoto RI (2011). A genetic screening strategy identifies novel regulators of the proteostasis network. PLoS Genet. **7**(12):e1002438.

Singh DK, Ku CJ, Wichaidit C, Steininger RJ 3rd, Wu LF, Altschuler SJ (2010). Patterns of basal signaling can distinguish cellular populations with different drug sensitivities. Molecular Systems Biology **6**:369.

Smith MCA, Sumner ER., Avery SV (2007). Glutathione and Gts1p drive beneficial variability in the cadmium resistances of individual yeast cells. Molecular Microbiology **66**: 699-712.

Spellman PT, Sherlock G, Zhang MQ , Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B (1998). Comprehensive Identification of Cell Cycle-regulated Genes of the Yeast Saccharomyces cerevisiae by Microarray Hybridization. Molecular Biology of the Cell **9**: 3273-3297.

Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature **459**(7245):428-32.

Springer M, Weissman JS, Kirschner MW (2010). A general lack of compensation for gene dosage in yeast. Mol Syst Biol **6**:368.

Spudich JL, Koshland DE Jr (1976). "Non-genetic individuality: chance in the single cell." Nature **262**(5568): 467-71.

Suel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB (2006). An excitable gene regulatory circuit induces transient cellular differentiation. Nature **440**: 545-550.

Swain PS, Elowitz MB, Siggia, ED (2002). Intrinsic and Extrinsic contributions to stochasticity in gene expression. PNAS **99**: 12795-12800.

Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS (2010). Quantifying E. Coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. Science **329**: 533-538.

Thattai M, van Oudenaarden A (2001). Intrinsic Noise in Gene Regulatory Networks. PNAS **98**: 8615-8619.

Thattai M, van Oudenaarden A (2004). Stochastic Gene Expression in Fluctuating Environments. Genetics **167**:523-530.

Theodoraki MA, Nillegoda NB, Saini J, Caplan AJ (2012). A network of ubiquitin ligases is important for the dynamics of misfolded protein aggregates in yeast. J Biol Chem., **287**(28):23911-22.

Tkacik G, Callan CG Jr, Bialek W (2008). Information flow and optimization in transcriptional regulation. Proc Natl Acad Sci U S A. **105**(34):12265-70.

To TL, Maheshri N (2010). Noise can induce bimodality in positive transcriptional feedback loops without bistability. Science. **327**(5969):1142-5.

Volfson D, Marciniak J, Blake WJ, Ostroff N, Tsimring LS, Hasty J (2006) Origins of extrinsic variability in eukaryotic gene expression. Nature **439**: 861–864.

Yamamoto N, Maeda Y, Ikeda A, Sakurai H (2008). Regulation of Thermotolerance by Stress-Induced Transcription Factors in Saccharomyces cerevisiae. Eukaryotic Cell **7**:783-790.

Chapter 2: What is the simplest model you can think of?

Or "The interaction of Msn2 with its promoters"

Although the identity of genes regulated by many transcription factors are increasingly well established by genomic scale datasets the nature of this qualitative and quantitative nature of this regulation is much less explored. Transcription factors are often sorted into activator and repressors but the strength of this activation and how it depends on the identity of the promoter, number of binding sites, and nature of the transcription factor activation domain are rarely interrogated. In this chapter we aim to take one particular factor, Msn2, a crucial stress responsive regulator in budding yeast, and systematically dissect the qualitative nature of its interaction with target genes (2.1) and the biophysical underpinnings of this interaction (2.2).

2.1

The Transcription Factor Msn2 Drives Linearly Related Promoters

To thrive in challenging and rapidly changing circumstances cells tightly regulate the production of cytoprotective proteins. Expression of these genes is costly and often toxic in conditions amenable to rapid growth, but essential in stressful environments. A stress responsive program must therefore balance rapid and sustained activity in the face of genuine peril with limited basal activity to avoid toxicity and waste of resources in benign environments. Although the details of these responses vary across organisms the basic strategy is highly conserved, on stress the activity of a core set of kinases—PKA, TOR, AMPK—are modulated, activating a range of transcription factors which regulate the expression of protective genes. In the budding yeast *S. cerevisiae* a wide range of stresses evoke the Environmental Stress Response (ESR) which is largely coordinated by Protein Kinase A (PKA) and realigns the cellular transcriptional program away from ribosomal and growth regulated genes and through the homologous transcription factors Msn2/4 towards stress responsive genes (Gasch et al., 2000; Slattery et al., 2007; Capaldi et al., 2008). These factors are important for survival under a range of conditions and are themselves

sufficient when activated to confer resistance to these same conditions (Stewart-Ornstein et al, 2012).

The transcription factors Msn2/4 bind to at least 200 promoters, activating the expression of chaperones, trehalose and glycogen synthase machinery, oxidative stress response, mitochondrial components, and alternate glycolytic enzymes; in short everything a cell needs to survive in stressful nutrient-poor conditions (Boy-Marcotte et al., 1999; Zhu et al., 2009; Huebert et al., 2012). The expression of such a rich transcriptional program raises questions about how a single factor coordinates all of these processes or why it would be desirable to do so. Most existing datasets on the behavior of the transcriptional targets of Msn2 has been measured under strong stress conditions and shows little obvious differentiation in timing or magnitude of different target promoters (Gasch et al., 2000; Capaldi et al., 2008], although recent work on the sensitivity of different target promoters to the duration of Msn2 activity (Hao and O'Shae., 2011; Huebert et la., 2012) suggested some differences in competence of promoters to be rapidly activated.

Recent measurement of transcription factor-DNA association using genomic techniques *in vivo* have given us a snapshot of the regulatory landscape we have a lack an understanding of how this snapshot blurs and changes with conditions (Harbison et al., 2004). Much of our understanding of the expression of genes is regulated by transcription factor activity comes from classical carbon source switches (Galactose in yeast, Lactose in bacteria) or abrupt heat shock. Less studied, although probably more common, are graded homeostatic responses such as those that are observed in the expression of amino acid biosynthesis genes in response to amino acid concentrations (Mumberg et al., 1994). The environmental stress response (ESR) in yeast shows characteristics of both a switch and a homeostat, Msn2 is activated within minutes in response to stress (Gorner et al., 1998), but gene expression regulated by this factor shows a graded response

proportional to the degree of insult (Sadeh, et al., 2012). Graded gene expression has been observed in other stress responsive systems such as NFkB, where it has been proposed to allow proportionate responses to varied stress stimuli (Giorgetti et al., 2010).

Here we took advantage of synthetic tools that allowed us to systematically measure the sensitivity of many promoters to Msn2 activity and found that Msn2 non-cooperatively and linearly associates with even high affinity targets. This pattern of binding is enabled by relatively low-affinity interactions between Msn2 and its cognate binding site and also the number of potential Msn2 binding sites in the genome relative to the number of Msn2 molecules. This low-affinity unsaturated binding results in co-linear expression of many different Msn2 targets, allowing them to be induced stoichiometricaly and suggests that Msn2 provides a proportional response to stressful conditions. These results position Msn2/4 activity as a homeostatic system capable of providing a precisely balanced response across a range of environmental conditions, in addition to its properties as an 'emergency' stress response.

Results 2.1

To define the interaction of Msn2 with promoters it regulates we required a reproducible and quantitative way to modulate Msn2 activity, initially we explored various stresses but found that these inputs gave complex time varying Msn2 activity. For example addition of 500mM KCl resulted in a relatively rapid peak of Msn2 activity, measured by nuclear localization, at ~15minutes followed by a decline (Figure 8A). To avoid these complications we developed synthetic tools to specifically increase the amount of Msn2 in the nucleus by either direct regulated overexpression of a constitutively active Msn2 (5A allele) or specific inhibition of the PKA pathway (Figure 8B). In comparison to the relatively slow and transient localization of Msn2 on KCl treatment, inhibition of cAMP production by expressing a dominant negative allele of Ras2(S24N; Poullet et al., 1995) results in a switch-like and sustained activation (Figure 8C).

The discreet and sustained activity induced by this synthetic overexpression allows for analysis of promoter activity without complex measurement and analysis of the time-dependant expression and dynamics of Msn2 activity. Further, activation of Msn2 by expression of Ras2(S24N), a second PKA inhibitor the phosphodiesterase PDE2, or Msn2(5A) results in activation of the canonical Msn2 target genes PGM2 and HSP12 several fold greater than is observed under heat shock or in mid-stationary phase (Figure 8D).

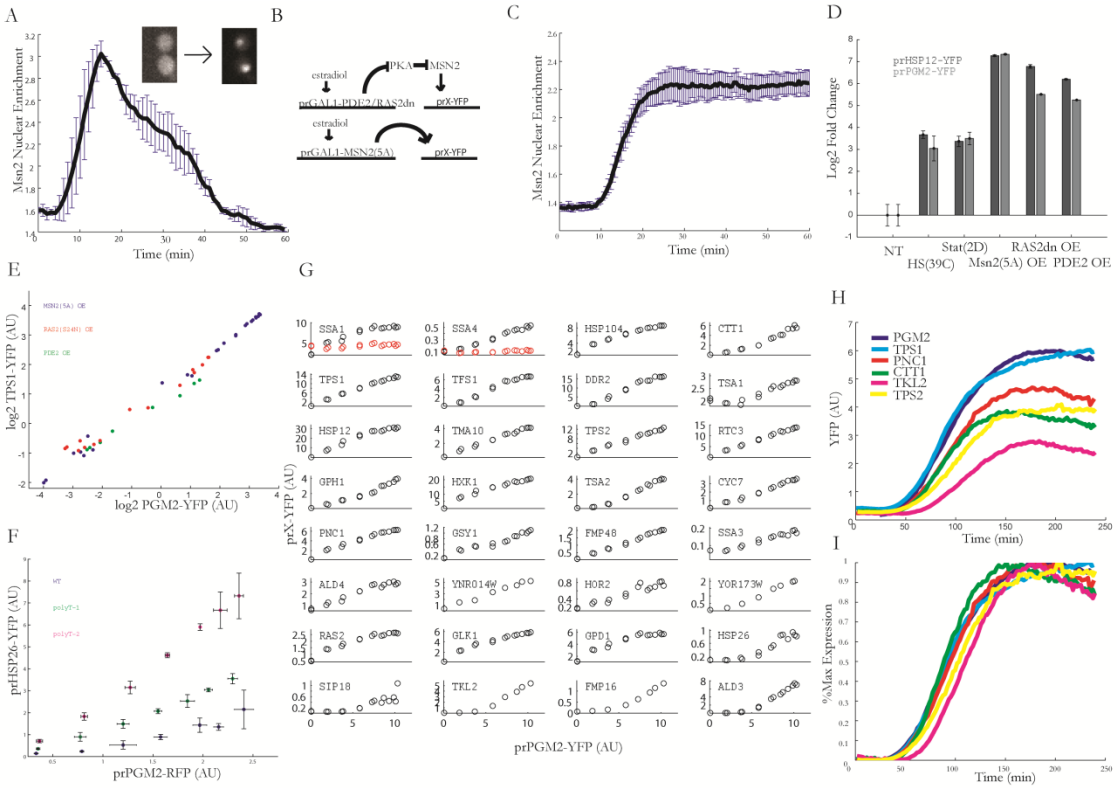


Figure 8: Msn2 activity can be precisely and comprehensively tuned *in vivo* with titratable synthetic inputs, revealing co-linear relationships in the expression of many distinct transcriptional targets. (a) Osmotic stress induced by addition of 500mM KCl in a strain expressing Msn2-mCherry results in complex dynamic Msn2 localization profile. Average nuclear enrichment of Msn2 in a population over time is plotted, errorbars represent standard deviation of six replicates. (b) Schematic representation of our synthetic approaches to modulating Msn2 activity through overexpression of a constitutively active allele or expression of PKA inhibitors Pde2 or Ras2dn. (c) At t=0 estradiol is added to a strain incorporating pGal-RAS2dn and Msn2-mCherry, average nuclear localization of Msn2 across a population is plotted as a function of time, errorbars show standard deviation of six replicates. (d)Expression of the Msn2 activity reporters prPGM2-YFP or prHSP12-YFP under conditions of heat shock (39C for 30min), or stationary phase (2Days in Synthetic media), or expression of a synthetic induction system, values are fold change of YFP florescence measured by flow cytometry over exponential phase cells, errorbars are Std. error from triplicate measurements. (e) Expression of prPGM2 or prTPS1 under induction with synthetic Msn2 activators results in a linear relationship. (f) prHSP26 shows a non-linear relationships with prPgm2, this

relationship was linearized by insertion of chromatin disrupting 12xpolyT sequences into the Hsp26 promoter. (g) the YFP expression of 32 promoters at various levels of Msn2(5A) overexpression, plotted against Pgm2-YFP. (h) Timecourse measurements of the average YFP expression in a population of cells over time expressing one of six promoters. At T=0 the expression of Msn2(5A) is induced by 200nM estradiol. (i) Data from (h) normalized to maximum expression of each promoter.

To take advantage of these tools to study the regulation of Msn2 dependant promoters we cloned a set of ~90 promoters (informatically 50 were expected to be at least partly dependant on Msn2 (Schmitt, 1996; Gasch, 2000; Capaldi, 2008), the remaining promoters were reporters for other cellular process of interest) in front of YFP in a single integration vector. We then monitored the expression of YFP monitored by flow cytometry in response to induction of Msn2 by either expression of Msn2(5A) or expression of the PKA inhibitors Ras2(S24N) or PDE2. Initial screening of these promoters by measuring their fold-change in expression on maximum induction of Ras2(S24N) and the dependence of this on Msn2/4 reduced us to 32 promoters that showed >4 fold activation by one assay and at least 50% dependence on Msn2 for this induction. For these promoters we then systematically measured their expression level (YFP florescence) at different levels of induction (Figure 8G).

To examine these results we selected a single canonical Msn2 responsive promoter, PGM2 and plotted the expression of all other promoters against this standard. Comparing Pgm2 to a second strongly Msn2 sensitive promoter TPS1 we observe that their expressions are linearly superimposable over 6log2s of activity with all three perturbations (Figure 8E). Indeed, most promoters show strong linear relationships when plotted this way despite different substantially in minimum and maximum expression (Figure 8G). These results suggest that the 'input functions' of these promoters are relatively uniform. A subset of these promoters showed a response with two regimes, a weakly responsive initial stage where the concentration of Msn2 has little effect on the promoter activity followed by a linear regime. One possible explanation for the two regime behavior of these promoters is that they are constrained by chromatin, to test this we selected one example, prHSP26 and inserted chromatin disrupting elements (12xpolyT) by quick

change mutagenesis. The promoters with either one or two polyT tracts added showed substantially more sensitivity to MSN2 and co-linearity with PGM2 (Figure 8F).

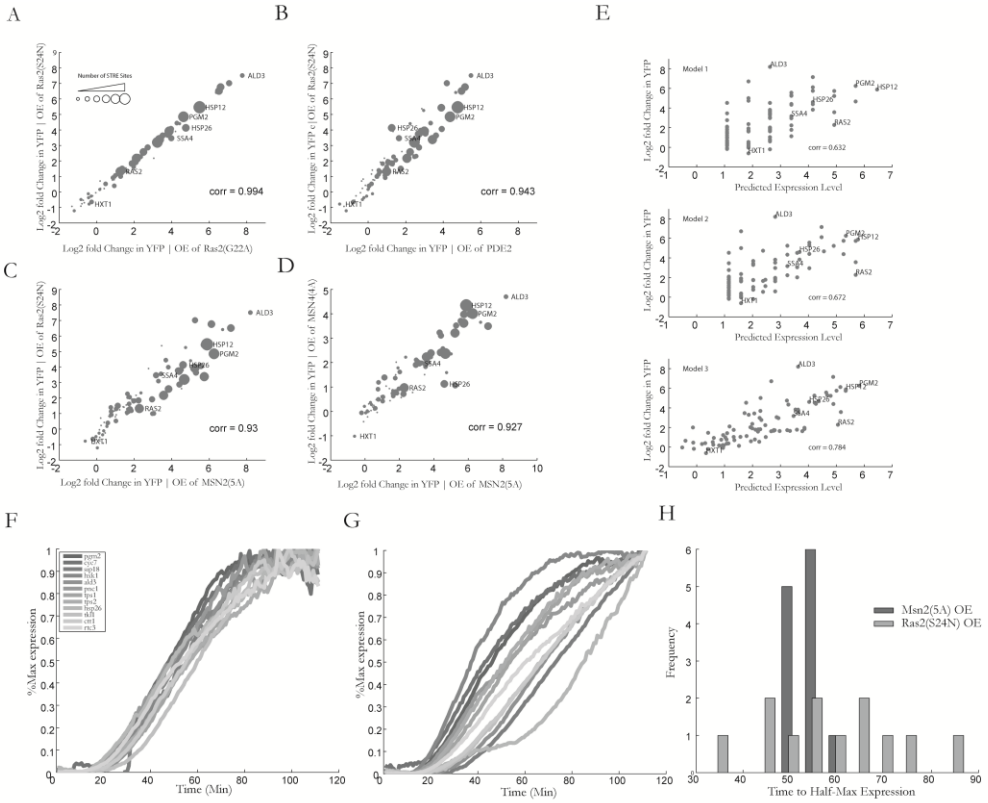


Figure 9: Multiple approaches to modulating Msn2 activity result in qualitatively similar results with minor quantitative difference, data from these assays can be predicted with reasonable accuracy by a simple model of Msn2 activity based on binding site number of each promoter. (a-d) Expression of YFP driven by 90 promoters enriched for stress response was measured in exponential growth and in response to overexpression of Msn2(5A), Msn4(4A), Ras2(S24N), Ras2(G22A), and PDE2. Log2 fold change in expression over exponentially growing cells for each perturbation is plotted. The size of each circle represents the number of Msn2 binding sites in that promoter. (e)The fold change induced by overexpression of Msn2(5A) predicted with a regression model taking into account the number of Msn2(STRE) binding sites in each promoter (model 1), the identity of those sites (the N position in the 'NAGGGG' site, Model 2), and both the identity and the genes exponential phase expression level (Model 3). Strains expressing one of twelve Msn2 sensitive promoters fused to YFP were imaged every ~2 minutes for two hours, at time zero Msn2 was activated by addition of 200nm estradiol resulting in (f)direct overexpression of Msn2(5A), (g)PKA inhibition by expression of Ras2(S24N). Average YFP expression over a population of cells as a percentage of the maximum observed in the dataset is plotted. The time to half-maximum expression was computed for each promoter (h).

To further examine this we selected a set of these promoters and by microscopy followed single cell traces over time as we overexpressed Msn2(5A). Again despite greatly different initial and maximum expression most promoters behaved similarly, inducing synchronously (Figure 8H,I). Slight, though measureable, delays were observed in those promoters such as TKL2 or HSP26 which showed a thresholded behavior in our dose response screen, suggesting that there may be a kinetic components to their regulation, consistent with recent results from the O'Shae lab (Hao and O'Shae, 2011).

To further analyze these promoters we compared the fold change on overexpression of Msn2(5A), Msn4(4A), or PKA inhibition by Ras2dn or PDE2 overexpression. Comparing two different alleles of RAS2dn (G22A and S24N; Camus et al, 1995) we observe nearly identical behavior, underscoring the reproducibility of these experiments (Figure 9A). Plotting promoter activity induced by PDE2 and RAS2dn we observe very similar behavior, but somewhat more variability compared to the two RAS2 alleles (Figure 9B). Msn2(5A) and Ras2 show highly correlated but distinct patterns of activity with outliers including HSP26, TKL2, and SIP18, all promoters that showed thresholded behavior (Figure 9C). Finally, Msn2(5A) and Msn4(4A) show similar inductions, but with Msn4(4A) in general being ~8X less active than Msn2(5A) in this assay (Figure 9D).

In all cases the number of Stress Responsive Elements (STREs, 'AGGGG'; Schmitt and McEntee, 1996) is moderately predictive of the induction by Msn2 activation (Circle size, Figure 9A-D). To quantify this we asked how predictive a model of promoter activation by Msn2 would be if we used simply the number of binding sites. We used linear regression to quantify the degree of influence the number of binding sites has on our fold change measurement, (Figure 9E-model 1) and found that even this minimal data was fairly predictive (R^2=0.4), we could improve this value slightly by taking into account the type of Msn2 binding site (the identity of 'N' in the consensus 'NAGGGG', R^2=0.45, Model 2), and could substantially improve it by

also taking in to account the basal expression of the gene (R^2=0.6147, Model 3). These results suggest that a simple additive model of Msn2 binding to promoters is potentially viable.

To compare the kinetics of promoter activation by direct transcription factor over expression and inhibition of PKA by expression of RAS2dn. We selected twelve promoters fused to YFP and measured their expression over time as Msn2(5A) or Ras2(S24N) was induced (Figure 9F,G). Consistent with earlier data, we observe that Msn2(5A) overexpression induces nearly synchronous expression of all promoters, whereas PKA inhibition results in a more heterogeneous activation (Figure 9H).

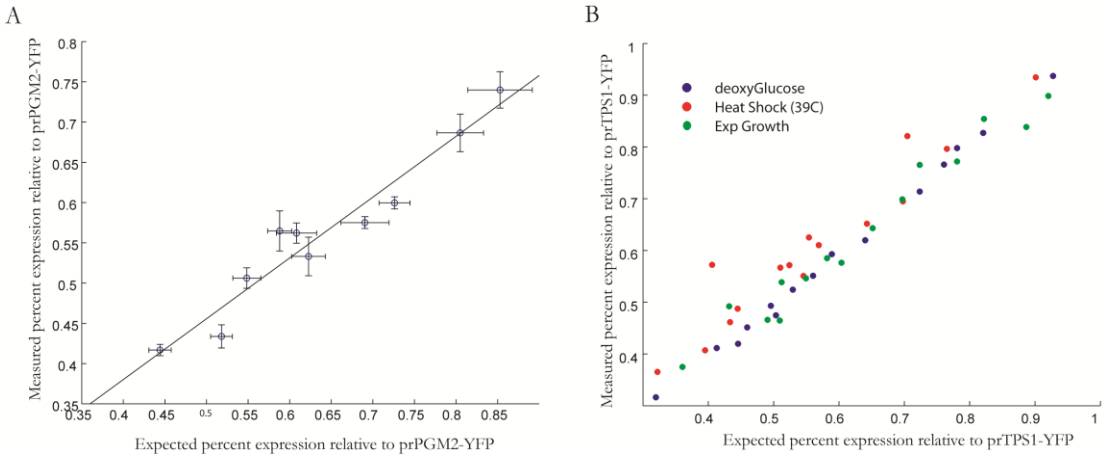


Figure 10: Mutational analysis of the PGM2 and TPS1 promoters argues for non-cooperative binding of Msn2. (a) Mutations were made to each of the five consensus STRE in the PGM2 promoter, all ten possible double mutants were also constructed, expression of each of these promoter alleles was measured. A simple multiplicative model was used to predict the expression of the double mutants from the single mutant data (x-axis), this model was compared to the actual double mutant values and a best fit line of linear model plotted. Errorbars represent standard error (n=3). (b)The same procedure was followed for the six binding sites in the TPS1 promoter, the quality of this fit was compared for exponential growth and also heat shock(39C for 30min, 60min recovery at 30C) and treatment with the toxic glucose analog deoxglucose (0.5% for 6hrs).

To explore the behavior of Msn2 as it binds to a promoter we focused on the Pgm2 promoter which has five STRE sites in the first 500bp before the start codon. We mutated each binding site is series, creating five single mutant alleles, we measured the expression of YFP in each instance. We then surmised if these binding sites are interacting independently then we

should be able to estimate the activity of double mutants from the single mutant data. To test this we constructed all possible double mutants and assayed their YFP expression, we then compared the measured expression of the doubles to the product of the constituent single mutants and obtained a strong linear relationship (Figure 10A).

We preformed the same analysis with a second promoter, TPS1 and its six STREs and obtained a similarly good correlation. Moreover this relationship was not condition sensitive, comparing measurements obtained in exponential growth with heat shock or addition of the glycolysis inhibitor deoxyglucose resulted in similarly linear relationships (Figure 10B)

Discussion 2.1

The use of synthetic biology tools to precisely set the concentrations of the active transcription factor Msn2 rather than much more pleiotropic—albeit more physiological—stress conditions that are typically used allowed us to quantitatively measure the dose response relationship between a transcription factor and many different target promoters. These tools should allow for the widespread analysis of transcription factor-promoter and more generally signaling pathway-gene expression interactions. In this work we limited ourselves to straightforward to analyze steady state data, however, initial experiments with these tools and timecourse microscopy suggest that there is a potential for more dynamic analyses.

Our measurements show a strictly graded and non-cooperative activation of gene expression by the Msn2 system, which combined with other work in mammalian and yeast systems suggests that far from being switch-like many stress responsive systems are graded and capable of precise dose dependant activation (Giorgetti et al, 2010; Sadeh et al., 2012). This should perhaps come as no surprise as many stress responsive systems are tightly embedded in negative feedback loops that insure regulated and limited expression (Wang et al., 2010; Lahav et

al., 2004), in these circumstances highly cooperative activation of downstream genes would be counter-productive.

These results suggested that the bio-physical interactions of Msn2 and its DNA binding sites generated the linear expression profiles we observe, this hypothesis is explored further in the next section.

2.2

Msn2 as a Low Affinity Transcription Factor

Eukaryotic transcription factors are complex proteins with multiple domains that bind DNA, recruit repressive or activating factors for the transcriptional machinery, and regulate these activities. Early work on these factors showed that these domains were often modular, for example swapping DNA binding domains between factors resulted in functional proteins with altered binding specificity (Wharton et al, 1984). Studies on the binding of transcription factors to promoters and how these interactions regulate gene expression in pro- and eukaryotic systems has found both cooperative and independent binding of factors without clear rules about in which context or transcription factor types these modes occur (Spitz and Furlong, 2012; Hahn and Young, 2011) . First principal arguments suggest that building switch-like binding and activation functions should be favored in decision making systems whereas more graded response might be preferred in homeostatic systems.

Based on our earlier measurements of Msn2 activity we observed that the promoters regulated by this factor behave in a graded and co-linear manner, this suggested that Msn2 might associate with these promoters in a graded non-cooperative fashion. Further, we wished to explore if this graded activity of Msn2 extended to its post-translational regulation by PKA.

The transcription factor Msn2 consists of four functional domains, a transactivation domain which spans the N-terminal 300 amino acids, a nuclear export sequence (NES) located at

the end of this domain (~AA260-300), a nuclear localization sequence (NLS) region located from ~AA570-642, and a DNA binding domain consisting of two zinc finger motifs arranged in tandem at the extreme C-terminus (Gorner, 1998; Gorner, 2002; Boy-Marcotte 2004). Additionally, the bulk of the regulation of Msn2 is provided by five consensus PKA phosphorylation sites distributed throughout (and regulating) the NES and NLS regions. To study the role of each of these regions in regulating Msn2 localization and activity we constructed alleles which compromised or altered these activities and subjected them to a battery of phenotypic tests. Our overall goal was to define the bio-physical nature of the interaction of Msn2 and its promoters and use this data to explain the observed ultra-graded nature of stress responsive promoters.

Results 2.2

Our initial focus was on how Msn2 activity responds to PKA activity, to study this we systematically mutated each of the five phosporylation sites to non-phosphorylatable alanine (S to A), expressed the allele and measured its localization by microscopy. As expected we observe a graded increase in nuclear localization as more of the PKA phospho-sites are mutated to alanine (Figure 11A,B; Gorner, 2002). We also note that this greater nuclear localization results in reduced Msn2-YFP expression, consistent with earlier reports suggesting that Msn2 that is localized to the nucleus is degraded more rapidly (Chi et al., 2004; Durchschlag et al, 2004). From our earlier data on the linearity of Msn2 promoters we suspect that nuclear Msn2 linearly relates to the amount of downstream transcription. Given this information on the degradation and linearity of this system we wrote a simple set of equations that described our expected relationship between the expression of Msn2 alleles and the expression of downstream genes (Figure 11C).

Consider a model of MSN2 gene regulation whereby MSN2 is produced in the cytoplasm (at rate *alpha*), and degraded (at rate *γ1*) and translocated into the nucleus (at rate *kin*). Nuclear MSN2 then is degraded (at the higher rate *γ2*) and can re-enter the cytoplasm (at rate *kout*). RFP is produced at a rate proportional to the amount of MSN2n and degraded (*γ3)*.

$$dMSN2c/dt = alpha + MSN2n * kout - MSN2c * kin - MSN2c * \gamma_1$$
$$dMSN2n.dt = MSN2c * kin - MSN2n * kout - MSN2n * \gamma_2$$
$$dRFP/dt = MSN2n * k - RFP * \gamma_3$$

Solving this system of equations for at steady state (dx/dt=0) we get the following for RFP

$$RFP = \frac{k}{\gamma_3}\alpha\frac{Kin}{(kout+\gamma_2)(kin+\gamma_1)} * \frac{1}{1-\frac{kin*kout}{(kout+\gamma_2)(kin+\gamma_1)}}$$

The solution results in a linear relationship between the production rate of MSN2 and RFP concentration with a slope governed by the relative rates of import/export and degradation.

To experimentally test this model we placed each allele of Msn2 tagged with YFP under the regulation of a GAL1 promoter inducible with estradiol and monitored the expression of the downstream gene prPGM2 driving RFP in response to a range of estradiol concentrations (Figure 11D). We observe the predicted linear relationship between Msn2 expression and PGM2-RFP expression with a slope that increases as more PKA sites are ablated from Msn2. These results highlight both the linear nature of the relationship between the transcription factor activity and downstream gene expression and the additive effect of each additional phosphorylation on Msn2.

These results focused on the interaction of Msn2 and its promoters, to explore the interaction between Msn2 and PKA we took a similar approach of examining the dose response of a Msn2 sensitive promoter to variation in PKA activity, and explored how this relationship changed in different alleles of Msn2. To modulate the activity of PKA in a graded and specific way we expressed the phosphodiesterase PDE2 under the control of a tritrateable promoter.

PDE2 catalytically degrades cAMP, the degradation rate of cAMP is therefore proportional to the quantity of PDE2 present in the cell. By titrating the level of PDE2 we could indeed achieve strong and graded activation of Msn2 target genes. We then compared the dose response to *msn2/msn4* deletion strains expressing alleles of MSN2 with compromised NLS activity, NES activity, or mutated PKA phosphosites. Intriguingly while each of these mutations substantially affected the basal activity of our reporter gene (prHSP12-RFP) the maximum activity was unaltered (Figure 11E).

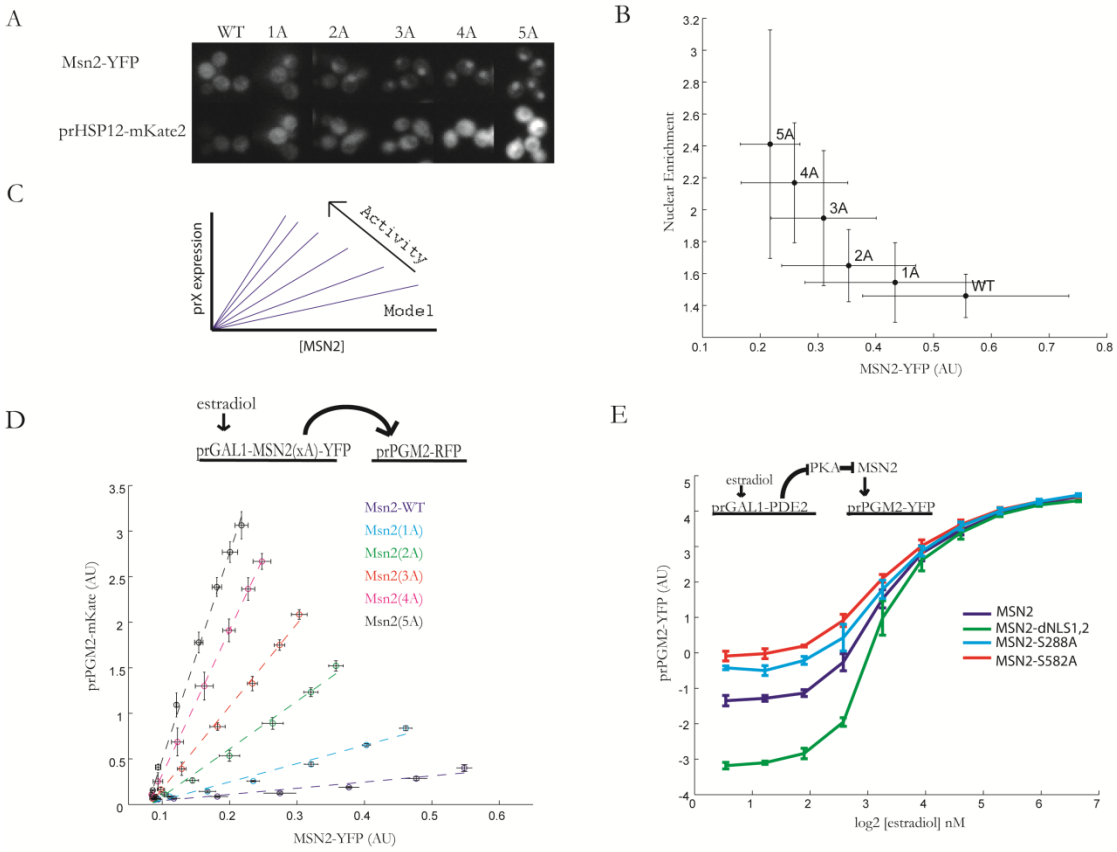


Figure 11: Sequential removal of repressive PKA phosphorylation sites results is graded increase in Msn2 nuclear localization and transcriptional activity but does not change the essentially linear character of the transcriptional regulation. (a) We constructed Msn2 alleles with serial removal of each of the five PKA phosphorylation sites resulting in six alleles (WT, 1-5A). Each of these alleles tagged with YFP was expressed in a strain also expressing a HSP12 promoter driving the expression of the RFP mKate2, the degree of nuclear localization was measured by microscopy as well as the expression of RFP. (b) Quantification of the nuclear localization and abundance of each Msn2 allele, increased nuclear localization results in lower protein likely due to degradation. Errorbars are standard deviation across single cells

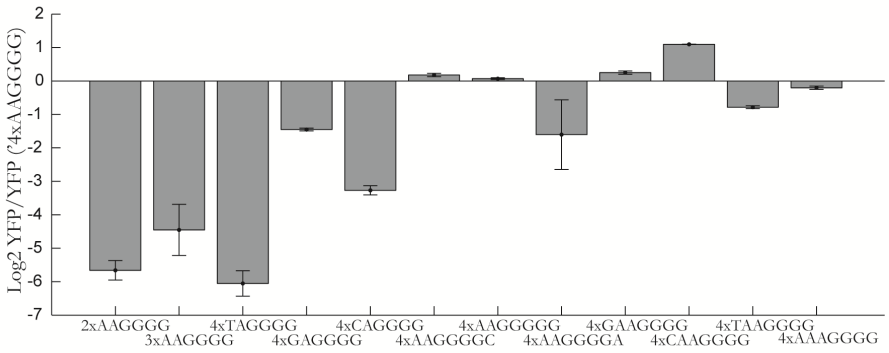
(N=20-50) (c)A model of gene expression predicts that all Msn2 alleles should show linear relationships between Msn2 expression (X-axis) and target gene expression, with a slope that increases with more active alleles. (d) Experimental measurement of the expression of the Msn2 reporter prPGM2-mKate2 in response to variable expression of each allele of Msn2 obtained using the titratable estradiol system. (e) Comparison of WT Msn2 and alleles lacking phosphosites or with a mutation to the NLS sequence in response to PKA inhibition by PDE2 overexpression. Errorbars in (d) and (e) are std. error of triplicate measurements.

We find that Msn2 regulation by PKA is, like its interaction with promoters, graded and the sum of many distinct phosphorylation events. Further, this data suggests that the maximum Msn2 transcriptional activity is independent of PKA activity, suggesting that PKA itself acts as a dynamic rheostat, modulating Msn2 activity to conform to a particular cellular state. This setpoint is influenced by PKA activity in exponential phase, and also by the number and identity of the phosphosites on Msn2. This suggests that design of transcription factors with differential sensitivities to PKA would be relatively trivial matter of modulation of the number of phosphorylation sites. It would be interesting to explore how this prediction holds across different yeast species where PKA sites have been gained or lost.

Modulation of the phosphorylation state of Msn2 does not therefore alter the inherent linearity between Msn2 activity and promoter activity, arguing that the dynamics of the system do not play a role in this linearity, but rather it is a property of the interaction between the transcription factor and DNA. To explore this process we asked first if Msn2 exhibits binding specificity beyond the 'AGGGG' sequence which is canonical. To do this in vivo we construct reporter plasmids with a crippled cyc promoter driving YFP expression and placed various 'NNAGGGGN' sequences upstream and measured the YFP expression (Figure 12A). This work defined 'AAGGGG' as the optimal MSN2 binding site, with small effects (<2 fold) induced by altering additional flanking nucleotides, this consensus is also consistent with the sequence of Msn2 zinc finger binding domain (Dreier et al., 2001; Segal et al., 1999). To see if we could find support for this sequence preference in vivo we searched the *S. cerevisiae* genome for Msn2

binding sites, curiously in vivo AAGGGG sequence makes up almost 40% of NAGGGG sequences, greatly more than would be expected by chance. Further, in promoters with multiple Msn2 binding sites the AAGGGG sequence is favored, whereas the low affinity TAGGGG sequence is disfavored (Figure 12B).

A



B

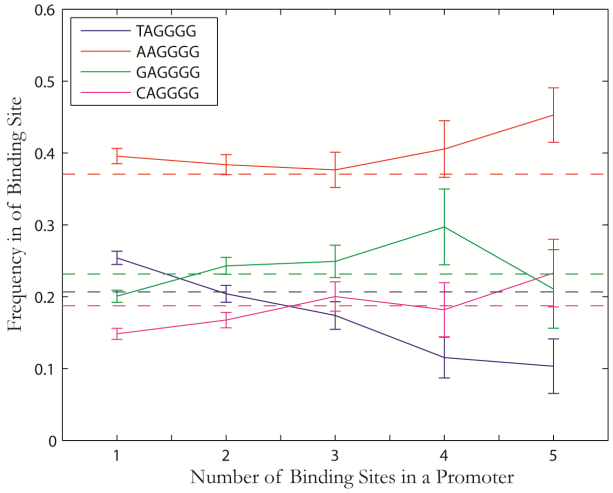


Figure 12: Msn2 binding affinity depends strongly on the identity of flanking residues, in particular we observe a strong preference for the sequence 'AAGGGG'. (a) To assess Msn2 binding preferences *in vivo* we constructed a synthetic reporter consisting of the minimal cyc1 promoter driving YFP, we than introduced four repeated STRE elements ('NNAGGGGN), modulated the N bases, and measured the resulting change in YFP expression. To compare the ability of each substitution to be activated by Msn2 we overexpressed the Msn2(5A) allele and measured the resulting YFP expression. We show this expression as fold change relative to the optimal 'AAGGGG' sequence and note that modification to

flanking positions outside the core five nucliotide motif can have up to 64 fold reductions in transcriptional activity. (b)To explore the role of flanking nucleotides in vivo we scanned the genome for all potential Msn2 binding sequences ('NAGGGG') and compared the frequency of nucleotides in the 1st position for promoters with varying numbers of binding sites. Promoters with many Msn2 sites tend to be strongly regulated and among these promoters we see a mild increase in the number of strong 'AAGGGG' sites and a decline in the number of weak 'TAGGGG' sites. Errorbars are standard error computed from bootstrapping of the genomic dataset.

These results outlined a stronger consensus site for Msn2 binding, but did not answer the basic question about the linearity of Msn2-DNA association. To understand this behavior we asked under what circumstances a transcription factor associates linearly with DNA, the basic physics of the binding of two molecules suggest that the process is linear from the perspective of one molecule when the association between the species is weak or when there is a great excess of one species over the other (Figure 13A). These two parameters can complement one and other, allowing a factor with both an excess of binding sites and weak affinity to be robustly linear (Figure 13B).

If these models are accurate than altering the affinity of Msn2 for DNA could alter the linearity we observe, to test this we constructed two alleles we predicted to have lower or higher affinity. To design these alleles we began by comparing the sequence of the MSN2 DNA binding domain to that of related cerevisiae 2xC2H2 zinc finger proteins MSN2, COM2, GIS1, RPH1, MIG1, MIG2, ADR1 (Figure 13C) The critical C2H2 residues which coordinate the zinc ion are marked in red, the linker between the two fingers in purple. The linker region is of particular note as the consensus sequence (TGEKP) is highly conserved across zinc fingers and mutations to this sequence are strongly associated with reduced affinity (Figure 13D). It is thought this linker caps the end of the recognition helix and positions the adjacent finger for effective binding to DNA (Wuttke, 1997; Laity, 2000). However Msn2 also has one additional feature which is the unusual spacing between the two histidines in the first finger (highlighted in green it makes up four residues, not the more common three we observed in Mig1 or Adr1 for example). To see if the

increased spacing of histidines could explain the non-cannonical spacing we querried a protein structure database HMMER (Finn et al., 2001) and using the MSN2 DBD(residues 648-705) as a querry downloaded all matching 2xC2H2 proteins and compared the linker structure for fingers with a HX_3H (101864) and those with a HX_4H spacing (4626) by constructing consensus seqlogs. Consistent with Msn2 divergent linker, we observe substantially less conservation of the core linker sequence in the HX_4H fingers, perhaps due to the disrupted helix which has been observed in the NMR structure of the second finger of ADR1, a HX_4H finger (Bernstein et al., 1994).

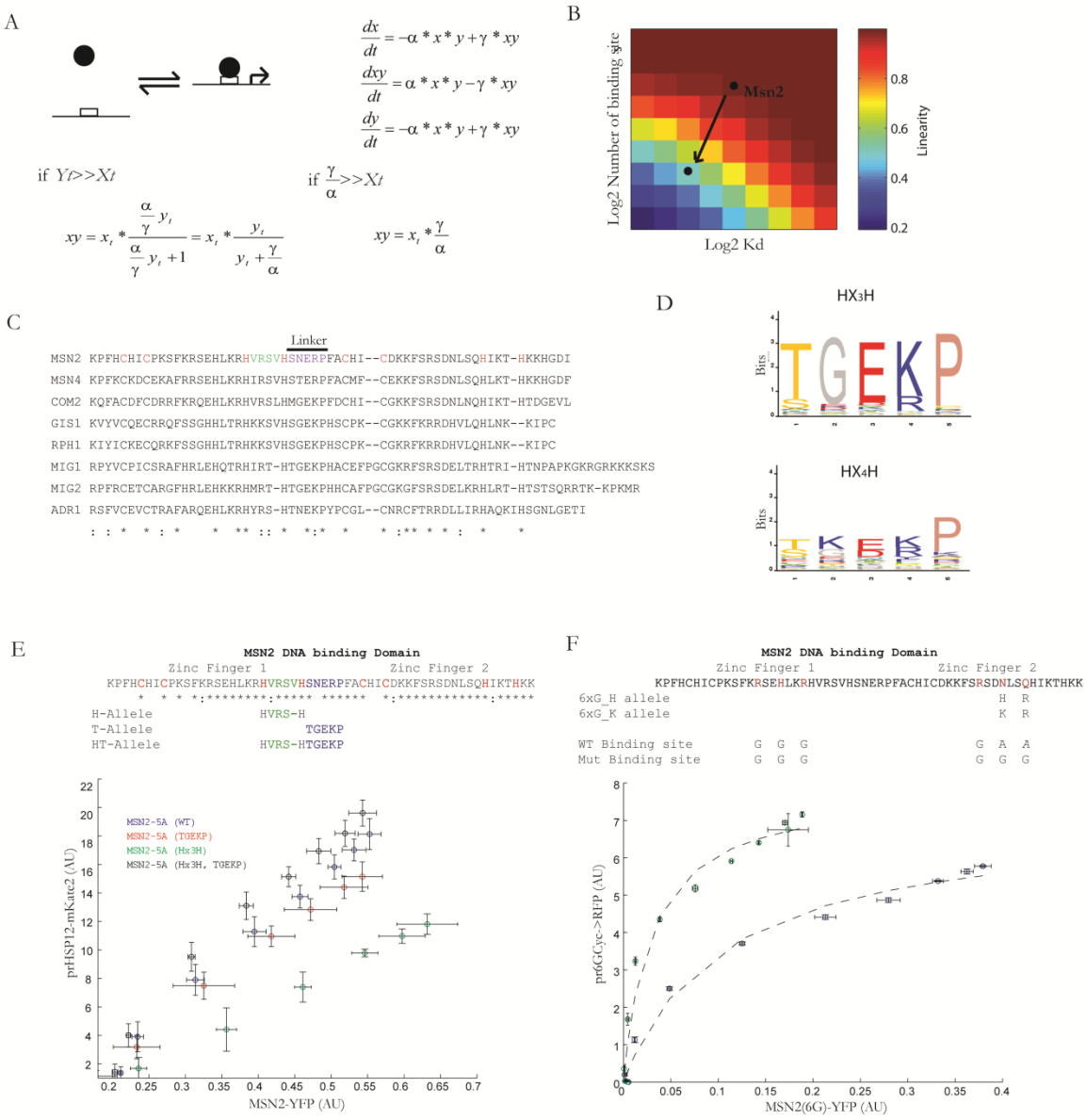


Figure 13: Linear association of Msn2 concentration and transcriptional activity can be explained by biophysical models combining the low affinity of Msn2 for its binding site and the number of potential Msn2 binding sites in the genome. (a) A simple model of transcription factor DNA association is linear if the number of binding sites greatly exceeds the number of transcription factors or if the affinity between TF and binding site it low. (b) Using the model in (a) simulations defining the region of parameter space where linear associations should be expected, Msn2 estimated position in this space is highlighted. This model predicts that the relationship between [Msn2] and transcriptional activity should be altered by increasing affinity or reducing the number of binding sites. (c) Alignment of *S. cerevisiae* proteins with homologous arrangements of zinc fingers to Msn2, crucial zinc coordinated residues are in red, the linker region between the fingers is highlighted and divergent across the proteins. (d) Alignment of linker regions from Zinc fingers with HX_3H and HX_4H arrangements show that X_4 fingers are less constrained in linker sequence. (e) Msn2 alleles with consesnsus linker arrangements have altered activity but not linearity. (f) modification of Msn2 consensus site to the less common 6xG sequence shows saturating activation of the

transcriptional reporter, with the K allele (in blue) showing ~2 fold reduced affinity compared to the H allele. Dashed lines represent fits to the [TF]/([TF]+kd). All error bars are standard error of triplicates.

Previous work that has explored the function of the linker residues suggests that perturbations should not affect DNA binding specificity, but only affinity (Wuttke et al., 1997; Laity et al., 2000, Kochoyan et al., 1991), drawing on this extensive literature we constructed three MSN2 mutants with altered linker arrangements, one construct (H allele) which converted the HX_4H spacing to a more conventional HX_3H spacing by removing the final valine (V669), the second construct (T allele) converted the linker sequence to the consensus (S671T, N672G,R674K), and finally an allele that combined these two mutations (HT). Based on previous data we suspected that the H and T alleles would by themselves reduce affinity, but in combination move the protein towards the consensus and increase the affinity without altering the binding properties. As anticipated the H and T alleles had severely and moderately reduced activation of the HSP12 gene, while the HT allele had moderately increased expression in all cases the linear relationship between gene expression and Msn2 abundance was maintained arguing that the divergent linker arrangement and reduced affinity due to that is not to blame (Figure 13E).

Another possible explanation for how Msn2 could linearly activate the expression of downstream genes is that there is not sufficient MSN2 in the cell to saturate binding at all sites simultaneously resulting in effective competition between binding sites for Msn2. There are 8450 consensus (AGGGG) binding sites in the genome, roughly half of which (4122) are present in promoter regions—loosely defined as DNA sequences 700bp upstream of a start codon—and just over one hundred Msn2 molecules. Consistent with these sequence based estimates, recent *in vivo* measurements of Msn2 binding identified 1290 associated loci—many of which contain multiple STRE sites—that associate with Msn2 during stress conditions (Heubert et al, 2012). A simple model taking these numbers into account suggests that induction of target genes would

indeed be linearly dependent on the concentration of Msn2 over a wide range (supplement, F13A).

These results are supportive of a model whereby the low-affinity competitive interactions of Msn2 with DNA create a regime where collinear induction of target genes is possible. This suggests that if the Msn2 binding site was altered to a less common sequence it might reveal more typical saturating interactions with its target promoters. To define a low occurrence motif we scanned a database of the cerevisiae promoter sequences for the number of occurrences of GNNGNN, we restricted ourselves to GNN sequences because canonical zinc fingers bind this sequence well (Segal et al., 1999). Interestingly we noticed that the single least common site 'GTCGGG' (416 occurrences) matches the consensus of the proteosomal regulator RPN4 and that the fifth least common site 'GCGGGG' matches the Mig1 consensus (Harbison et al., 2004; Zhu et al., 2009). We selected the 9^{th} most common sequence 'GGGGGG' as rare sequence (540 occurances) for which specific zinc fingers have been constructed and which no other yeast transcription factor has been reported to bind to. We made two mutations to the second Msn2 zinc finger (Q693R, and N690H or N690K) which we predicted based on the zinc finger engineering literature to alter its specificity from AAG to GGG, with high affinity for the N690H (RH) and low affinity for the N690K (RK) variants (Figure 13F; Pomerantz et al., 1995; Segal et al., 1999).

Expression of the RH or RK construct resulted in strong activation of a CYC1 promoter whose UAS was modified to contain three 6xG sequences, we observed no basal activity from these elements or cross activation of this sequence by WT MSN2. Neither the RH nor RK variants were capable of activating a promoter with the wild type MSN2 binding site (data not shown). As predicted both the RH and RK variants showed strong non-linear saturating interactions with a target promoter, which were well fit (R^2>0.95) by a simple binding model (Figure 13F). Additionally, the RK variant showed weaker binding with less maximum

expression and a two-fold lower apparent KD for the promoter (F4E). This data shows that Msn2 can be switched from a low affinity graded activator to a saturating system with two amino acid changes, suggesting that the graded nature of Msn2 binding to its promoters is a feature of the system, not a consequence of biophysical constraints.

Discussion 2.2

Our experiments show that transcription factor binding to promoters can be surprisingly straightforward and linear and that these relationships can be explained by first principal models taking affinities and molecular numbers into account. Obtaining a linear relationship does not require complex dynamic mechanisms but simply requires low affinity and low transcription factor abundances, a simpler and perhaps more broadly applicable mechanism than the previously explored linearity-through-frequency modulation of the Crz1 transcription factor (Cai et al., 2008).

We also note that this linear coupling between the transcription factor and promoter needs to be considered in the context of the PKA regulatory system which produces dynamic changes in Msn2 activity in response to stress. Low affinity interactions with DNA allow for rapid (sub-second) binding and un-binding of Msn2 to its response elements, enabling rapid dynamic control of Msn2 activity. Our experiments were performed at relatively slow timescales, purposely ignoring dynamic features to focus on the steady state regulatory relationships, if however the rate limiting step for promoter activation is transcription factor binding then the low affinity interactions of Msn2 argue that the dynamic response of the promoters will mirror the steady state. In many cases other steps may be rate limiting such as chromatin remodeling and as this would predict different genes seem to show different dynamics at the minute timescales in response to Msn2 activity as has been demonstrated (Hao et al., 2011).

There are two strong benefits for maintaining a linear response to transcription factor concentration, first co-linear activation of target genes allows for stoichiometric expression of large groups of genes without extensive promoter tuning, and second that low affinity linear interactions allow for precise tuning of promoter activity by addition of binding sites. Although speculative, the benefits of this approach are likely substantial as a handful of mutations are sufficient to alter Msn2 pattern of binding from a linear to rapidly saturating suggesting evolutionary pressure has maintained Msn2 in this linear regime. These features also potentially make the Msn2 regulon very 'evolvable' with binding site changes modifying the quantitative relationship between different genes without changing quantitative features, consistent with this stress responsive networks in fungi show rapid rewiring across evolutionary timescales (Rhind et al., 2011; Wapinski et al., 2007; Tirosh et al., 2011).

The model we derive for competition between binding sites for limiting transcription factors requires only that binding sites substantially outnumber transcription factors, a situation that is common in eukaryotic systems where relatively small and frequently degenerate binding motifs are combined with large genomes. Certainly there are ways of avoiding this competitive binding, by occluding unused sites with nucleosomes, or using co-factors or cooperative binding to increase specificity, but for monomeric transcription factors with moderate affinities these results may prove general.

Chapter 2 Methods

Yeast Strains/protocols

All yeast strains used for these experiments are derived from W303A-1 in which the *ade2* marker was reverted to ADE2+ to reduce the autoflorescence. Promoter constructs were integrated at the TRP1 locus of a HIS3+ MATa strain. Overexpression constructs were integrated into the TRP1

locus of a LEU2+ MATalpha strain which contained the estradiol inducible construct (Stewart-Ornstein et al., 2012). These strains were then mated and diploids selected in SD-leu/his media. All strains were constructed using standard yeast protocols and LioAc/PEG transformation.

Growth and fluorescence measurements by flow cytometry

For all measurements cells were grown to saturation in 96-shallow well plates (Costar) and then diluted into fresh media, grown at 30C on orbital shakers (Elim) for 12hrs to an OD of ~0.5. These cells were subsequently diluted and estradiol added as necessary and grown for 6hrs to an OD of ~0.05 before measurement. Expression of the estradiol regulated system was activated by addition of 0-200nm estradiol (stock of 1.6mM in 90/10 mixture of EtOH and DMSO (sigma)), typically applied in a log1.6 titration series.

To apply a strong heat shock cells were grown as above in 96 well shallow bottom plates and then switched to 25C for 4hrs, followed by a temperature upshift to 39C for 30min, and recovered at 25C for 1hr before measurement. Cells were grown as above, 6hrs before measurement cells were diluted 100 fold and Deoxyglucose (stock solution 20% in H2O, Sigma) was added to a final concentration of 0.5%, cells were grown at 30C for 6hrs before measurement. To approximate stationary phase cells were grown without shaking at 30C for two days, diluted 100fold into fresh media and immediately measured.

All cytometry measurements were made on a Becton Dickinson LSRII flow cytometer, along with an autosampler device (HTS) to collect data over a sampling time of 6-12 seconds, typically corresponding to 2000-10000 cells. GFP and YFP were excited at 488nm, and fluorescence was collected through a HQ530/30 bandpass filters (Chroma), mCherry and mKate2 were excited at 561 nm and fluorescence collected through 610/20 bandpass filter (Chroma) .

Microscopy and image analysis

Cells expressing Msn2-YFP or related constructs were plated in SD complete onto ConcanavalinA coated 96 well glass bottom plates, allowed to settle and then washed twice with fresh media. Samples were imaged on a Nikon Ti inverted scope with arc-lamp illumination using RFP(560/40nm excitation, 630/75nm emission, chromas) and YFP (510/10nm excitation, 542/27nm emission, semrock) filters. Images were processed and analyzed with ImageJ and Matlab.

Flow cytometry Data analysis

All data was analyzed with custom Matlab software. Raw cytometry data were filtered to remove errors due to uneven sampling and remove outliers using an MCD method (Rousseau and Van Driessen, 1999). Variability in cell size was corrected using a linear transformation from the side scatter parameter (see chapter one for detail).

Sequence Analysis

The latest S288C genome was downloaded from SGD (yeastgenome.org) and the DNA sequence was processed and examined with custom matlab scripts. Promoters were operationally defined as 700bp upstream of the start (ATG) condon of the gene.

Overexpression Constructs

To construct prGAL1 fused overexpression constructs we PCRed Msn2, Msn4, Ras2, or PDE2 from the genome and cloned these genes into a TRP1 marked single integration vector downstream of a GAL1 promoter, between Xho1 and BamH1 restriction sites. We than constructed the appropriate mutants by quickchange mutagenesis using the pfuTurbo enzyme mix (stratagene). Briefly, these were Msn2(S288A, S582A, S620A, S625A, S633A), Ras2(G22A, S24N), and Msn4(S263A, S316A, S531A, S558A).

Construction of Promoter Library

One kilobase upstream of each gene of interest was PCRed from genomic DNA with primers containing pspOMI (or Not1) and Xho1 (or Sal1) sites, digested, and cloned into a TRP1 marked single integration vector directly upstream of a yeast optimized Venus YFP. Clones were sequence verified and integrated into yeast by digestion with Nae1, transformation using a standard PEG/LioAc protocol and selection on SD-TRP plates.

Deletion of STRE elements was accomplished by quickchange mutagenesis using the pfuTurbo enzyme mix (stratagene). The binding sites AGGGG were mutated to AGaGG, a mutation which appeared to eliminate activity *in vivo* and we observed to dramatically reduce the affinity for MSN2 *in vitro*.

Insertion of polyT(12x) elements was accomplished by quickchange mutagenesis, insertions sites were chosen based on location in the promoter (250-400bp upstream of the ATG) and availability of suitable regions for design of efficient quickchange primers (16GC basepairs within a 40bp region).

Primer design was accomplished by custom matlab scripts complemented by manual editing.

Construction of synthetic promoters

We constructed a vector with the prCYC1(1-243) sequence cloned directly upstream of Venus in a TRP1 marked vector. Additionally, to reduce background expression we included the prCYC1(1000-684) sequence. In between these sequences (which lack UASs) we inserted our STRE sequences or STRE like sequences.

4xSTRE: GGGCCCCTNCATTACCCCTNCTTTACCCCTNCAAACCCCTN

Where the N nucleotide was varied

3x6C: GGGCCCATTTACCCCCCATTTACCCCCCATTTACCCCCCA

These constructs were then integrated into yeast and assayed as above.

Construction of over expression constructs

Genes of interest were PCRed from yeast genomic DNA and cloned downstream of a prGAL1 promoter in a TRP1 marked single integration plasmid. Site directed mutagenesis was preformed as necessary to construct the constitutively active alleles of MSN2 (S288A, S582A, S620A, S625A, S633A) or dominant negative alleles of RAS2 (S24N or G22A) using a standard QuickChange protocol and the pfuTurbo enzyme mix (stratagene). Constructs were sequence verified, cut with Pme1 and integrated into yeast with selection on SD-TRP plates.

Measurement of MSN2 abundance

To convert arbitrary intensity units to absolute values we expressed constructed strains where we tagged one of NUF2, ASC1, SPC42 with YFP. These proteins localize to the kineticore and spindle pole body structures and have a well described abundance in these structures and have been used previously as molecular standards. To measure these abundances we imaged cells expressing these tagged proteins and defined a linear function relating their expression to the intensity measured in the microscope [R^2>0.95]. We then applied this function to measurements of Msn2-YFP to convert measured intensity to molecular numbers.

Chapter 2 References

Bernstein BE, Hoffman RC, Horvath S, Herriott JR, Klevit RE (1994) Structure of a histidine-X4-histidine zinc finger domain: insights into ADR1-UAS1 protein-DNA recognition. Biochemistry **33**(15):4460-70.

Berry DB, Gasch AP (2008). Stress-activated genomic expression changes serve a preparative role for impending stress in yeast. Mol Biol Cell. **19**(11):4580-7.

Boy-Marcotte E, Lagniel G, Perrot M, Bussereau F, Boudsocq A, Jacquet M, Labarre J (1999). The heat shock response in yeast: differential regulations and contributions of the Msn2p/Msn4p and Hsf1p regulons. Mol Microbiol. **33**(2):274-83.

Boy-Marcotte E, Garmendia C, Garreau H, Lallet S, Mallet L, Jacquet M (2006). The transcriptional activation region of Msn2p, in Saccharomyces cerevisiae, is regulated by stress but is insensitive to the cAMP signalling pathway. Mol Genet Genomics. **275**(3):277-87.

Camus C, Hermann-Le Denmat S, Jacquet M (1995). Identification of guanine exchange factor key residues involved in exchange activity and Ras interaction. Oncogene, **11**(5):951-9.

Cai L, Dalal CK, Elowitz MB (2008). Frequency-modulated nuclear localization bursts coordinate gene regulation. Nature. **455**(7212):485-90.

Capaldi AP, Kaplan T, Liu Y, Habib N, Regev A, Friedman N, O'Shea EK (2008). Structure and function of a transcriptional network activated by the MAPK Hog1. Nat Genet. **40**(11):1300-6.

Chi Y, Huddleston MJ, Zhang X, Young RA, Annan RS, Carr SA, Deshaies RJ (2001). Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-dependent kinase. Genes Dev. **15**(9):1078-92.

Dreier B, Beerli RR, Segal DJ, Flippin JD, Barbas CF 3rd. (2001). Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. J Biol Chem. **276**(31):29466-78.

Durchschlag E, Reiter W, Ammerer G, Schüller C (2004). Nuclear localization destabilizes the stress-regulated transcription factor Msn2. J Biol Chem. **279**(53):55425-32.

Ferguson SB, Anderson ES, Harshaw RB, Thate T, Craig NL, Nelson HC (2005). Protein kinase A regulates constitutive expression of small heat-shock genes in an Msn2/4p-independent and Hsf1p-dependent manner in Saccharomyces cerevisiae. Genetics. **169**(3):1203-14.

Finn RD, Clements J, Eddy SR (2011). HMMER web server: interactive sequence similarity searching.Nucleic Acids Res. **39**(Web Server issue):W29-37.

Gasch AP, Spellman PT, Kao CM,Carmel-Harel O, Eisen MB,Storz G, Botstein D, Brown PO (2000). Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. Molecular Biology of the Cell **11**: 4241–4257.

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003). Global analysis of protein expression in yeast. Nature. **425**(6959):737-41.

Giorgetti L, Siggers T, Tiana G, Caprara G, Notarbartolo S, Corona T, Pasparakis M, Milani P, Bulyk ML, Natoli G (2010). Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. Mol Cell. **37**(3):418-28.

Görner W, Durchschlag E, Martinez-Pastor MT, Estruch F, Ammerer G, Hamilton B, Ruis H, Schüller C (1998). Nuclear localization of the C2H2 zinc finger protein Msn2p is regulated by stress and protein kinase A activity. Genes Dev. **12**, 586-597.

Görner W, Durchschlag E, Wolf J, Brown EL, Ammerer G, Ruis H, Schüller C (2002). Acute glucose starvation activates the nuclear localization signal of a stress-specific yeast transcription factor. EMBO J. **21**(1-2):135-44.

Hahn S, Young ET (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics. **189**(3):705-36.

Hao N, O'Shea EK (2011). Signal-dependent dynamics of transcription factor translocation controls gene expression. Nat Struct Mol Biol. **19**(1):31-9.

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KS, Lander ES, Gifford DK, Fraenkel E, and Young RA (2004). transcriptional regulatory code of a eukaryotic genome. Nature **431**:99-104.

Huebert DJ, Kuan PF, Keles, S, Gasch AP (2012) Dynamic changes in nucleosome occupancy are not predictive of gene expression dynamics but are linked to transcription and chromatin regulators. Mol Cell Biol. **32**(9):1645-53.

Jorgensen P, Edgington NP, Schneider BL, Rupes I, Tyers M, Futcher B (2007). The size of the nucleus increases as yeast cells grow. Mol Biol Cell. **9**:3523-32.

Kochoyan M, Keutmann HT, Weiss MA (1991). Alternating zinc fingers in the human male-associated protein ZFY: HX3H and HX4H motifs encode a local structural switch. Biochemistry. **30**(39):9396-402.

Lahav G, Rosenfeld N, Sigal A, Geva-Zatorsky N, Levine AJ, Elowitz MB, Alon U (2004). Dynamics of the p53-Mdm2 feedback loop in individual cells. Nat Genet. **36**(2):147-50.

Laity JH, Dyson HJ, Wright PE (2000). DNA-induced alpha-helix capping in conserved linker sequences is a determinant of binding affinity in Cys(2)-His(2) zinc fingers. J Mol Biol, **295**(4):719-27.

Mumberg D, Müller R, Funk M (1994). Regulatable promoters of Saccharomyces cerevisiae: comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Res. **22**(25):5767-8.

Pomerantz JL, Sharp PA, Pabo CO (1995). Structure-based design of transcription factors. Science. **267**(5194):93-6.

Poullet P, Créchet JB, Bernardi A, Parmeggiani A (1995). Properties of the catalytic domain of sdc25p, a yeast GDP/GTP exchange factor of Ras proteins. Complexation with wild-type Ras2p, [S24N]Ras2p and [R80D, N81D]Ras2p. Eur J Biochem, **227**(1-2):537-44.

Rhind N, Chen Z, Yassour M, Thompson DA, Haas BJ, Habib N, Wapinski I, Roy S, Lin MF, Heiman DI, Young SK, Furuya K, Guo Y, Pidoux A, Chen HM, Robbertse B, Goldberg JM, Aoki K, Bayne EH, Berlin AM, Desjardins CA, Dobbs E, Dukaj L, Fan L, FitzGerald MG, French C, Gujja S, Hansen K, Keifenheim D, Levin JZ, Mosher RA, Müller CA, Pfiffner J, Priest M, Russ C, Smialowska A, Swoboda P, Sykes SM, Vaughn M, Vengrova S, Yoder R, Zeng Q, Allshire R, Baulcombe D, Birren BW, Brown W, Ekwall K, Kellis M, Leatherwood J, Levin H, Margalit H, Martienssen R, Nieduszynski CA, Spatafora JW, Friedman N, Dalgaard JZ, Baumann P, Niki H, Regev A, Nusbaum C (2008). Comparative functional genomics of the fission yeasts. Science, **332**(6032):930-6.

Sadeh A, Movshovich N, Volokh M, Gheber L, Aharoni A (2012). Fine-tuning of the Msn2/4-mediated yeast stress responses as revealed by systematic deletion of Msn2/4 partners. Mol Biol Cell. **22**(17):3127-38.

Segal D.J., Dreier B, Beerli RR, Barbas III CF (1999). Toward controlling gene expression at will, selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. Proc. Natl Acad. Sci. USA, **96**;2758–2763.

Spitz F and Furlong EEM (2012). Transcription factors: from enhancer binding to developmental control. Nature Reviews Genetics **13**, 613-626

Slattery MG, Liko D, Heideman W (2007). Protein kinase A, TOR, and glucose transport control the response to nutrient repletion in Saccharomyces cerevisiae. Eukaryot Cell. **7**(2):358-67.

Stewart-Ornstein J, Weissman JS, El-Samad H (2012). Cellular noise regulons underlie fluctuations in Saccharomyces cerevisiae. Mol Cell. **45**(4):483-93.

Tirosh I, Wong KH, Barkai N, Struhl K (2011). Extensive divergence of yeast stress responses through transitions between induced and constitutive activation. Proc Natl Acad Sci U S A. **108**(40):16693-8.

Wang X, Xu H, Ha SW, Ju D, Xie Y (2010). Proteasomal degradation of Rpn4 in Saccharomyces cerevisiae is critical for cell viability under stressed conditions. Genetics. **184**(2):335-42.

Wapinski I, Pfeffer A, Friedman N, Regev A (2007). Natural history and evolutionary principles of gene duplication in fungi. Nature, **449**(7158):54-61.

Wharton RP, Brown EL, Ptashne M (1984). Substituting an alpha-helix switches the sequence-specific DNA interactions of a repressor. Cell. **38**(2):361-9.

Wuttke DS, Foster MP, Case DA, Gottesfeld JM, Wright PE (1997). Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. J Mol Biol. **273**(1):183-206.

Zhu C, Byers KJ, McCord RP, Shi Z, Berger MF, Newburger DE, Saulrieta K, Smith Z, Shah MV, Radhakrishnan M, Philippakis AA, Hu Y, De Masi F, Pacek M, Rolfs A, Murthy T, Labaer J, Bulyk ML (2009). High-resolution DNA-binding specificity analysis of yeast transcription factors.Genome Res. **19**(4):556-66.

Chapter 3 : Guns versus Butter Or "Protein Kinase A as the cellular swing vote"

Optimal growth requires cells to measure the environment, process this information, and alter their gene expression profile as appropriate. For many situations this alteration amounts to changing the balance of ribosomal production and other growth related protein expression and stress responsive genes. In budding yeast this balance between ribosomal (Butter) and stress (Guns) expression is regulated by Protein Kinase A. In this chapter we will explore the dynamics and regulation of this pathway (3.1), the influence of the TOR pathway (3.2), and mechanisms by which a single kinase can coordinate many distinct cellular processes (3.3).

3.1

The PKA network and Regulation of MSN2

In the budding yeast *S. cerevisiae* a wide range of stresses evoke the Environmental Stress Response (ESR) which is largely coordinated by Protein Kinase A (PKA) and realigns the cellular transcriptional program away from ribosomal and growth regulated genes and through the homologous transcription factors Msn2/4 towards stress responsive genes (Broach, 2012). Msn2 regulated by phosphorylation on at least five sites by PKA which regulate this localization in response to carbon source and stress (Gorner et at., 1998). Time lapse imaging of Msn2 sub-cellular localization revealed highly dynamic behavior, even in exponential growth, with bursts of nuclear localization occurring at apparently stochastic intervals (Jacquet et al., 2003). These periods of localization have been linked to variation in the PKA pathway, which has been proposed to have oscillatory behavior (Garmendia-Torres et al, 2007; Bodvard et al., 2011). Consistent with these dynamics, years of work on the PKA pathway has uncovered many nested feedback loops regulating this pathway which could potentially lead to oscillatory or perhaps excitable states (Toda et al., 1985; Nikawa et al., 1987; Colombo et al., 2004).

Knowledge of the rough network structure has not, however, proved sufficient to construct a predictive model of PKA activity. Attempts to do so have generated models that are

in qualitative agreement with observed Msn2 behavior (Garmendia-Torres et al, 2007; Gonze et al., 2008), but these models have not proved sufficient to describe our data in any detail or predict the phenotypes of deletion mutants. We therefore have attempted in this work to define the functional behavior of Msn2 and therefore PKA in exponentially growing wild type cells and also in strains with mutations to various components of the PKA network. In each case we measured the magnitude and auto-correlation function of Msn2 localization. To interpret shifts in autocorrelation functions we took advantage of work on how the presence and removal of feedback loops leads to shifts in intrinsic noise signatures of networks (Cox et al, 2008). In particular deletion of the small G-protein Ras2 (and its regulators) shows a distinct shift towards longer autocorrelation times, consistent with its well established role in negative feedback in the PKA network (Colombo et al., 2004).

Although it is possible to obtain substantive information from observational methods, the PKA network is designed to be environmentally responsive so it is unlikely that exponentially growing cells excite all modes of the system. To extend our analysis we developed the use of a photoactivable adenylcyclase which allows for real time regulation of cAMP and the PKA pathway (Stierl et al., 2011; Ryu et al., 2010). Using this system as a realtime probe we could explore the frequency response of the regulatory network and the role of different protein components in regulating this response. In particular we defined the role of the catalytic subunit TPK2 in feedback regulation of the PKA network in situations of high PKA activity; in contrast we find that RAS2 regulation serves to allow the cell to recover from very low PKA states.

Results 3.1

In a population of exponentially growing cells the transcription factor Msn2 is localized to the cytoplasm in the majority of cells, a small subset of cells however show largely nuclear localization. Time lapse imaging of this population shows that Msn2 localization is a dynamic

phenomenon with ‘bursts’ of 1-5minutes of nuclear localization occurring apparently spontaneously in individuals cells (Figure 14A). We studied the nature of this localization by analyzing the single cell trajectories of hundreds of individual cells growing exponentially in rich media.

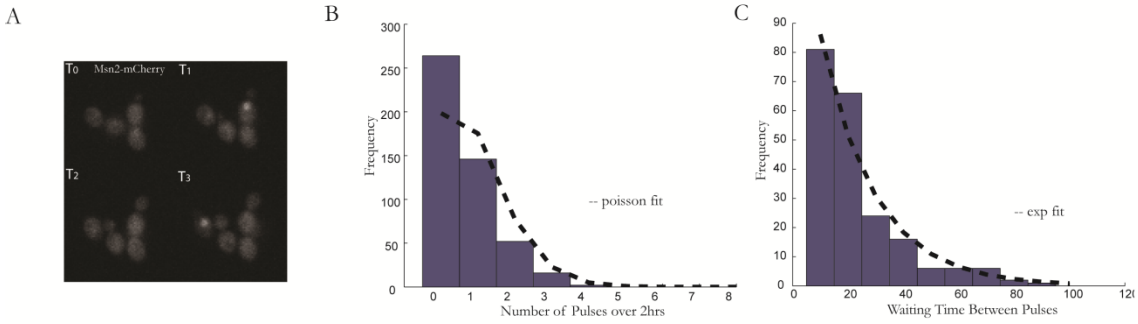


Figure 14: The transcription factor translocates into the nucleus in a subset of cells in an exponentially growing population. (a) Four images taken ten minutes apart of a group of yeast cells expressing Msn2(C649S)-mCherry. Localization of Msn2 in single cells was measured over time, with 2 minute time resolution. Over the course of three hours Msn2 was observed to enter the nucleus at least once in ~60% of cells. We obtained single cell traces of Msn2 localization and computed the number of bursts of localization per cell (b) and also for cells with more than one burst of localization the waiting time between events (c). The resulting histograms were fit to count statistics with exponential waiting times.

We compiled number of bursts of nuclear localization per cell and also, for cells which experienced more than one burst, the length of time between bursts. We then asked if these statistics were in agreement with a simple first order count statistics model (poisson) with exponential waiting times between subsequent events (Figure 14B,C). Although there was rough agreement between a poisson model and the data we observed, Msn2 burst statistics differed from a simple first order exponential process in two ways. First, more cells than would be expected showed no bursts of Msn2 localization over the two hour period they were observed. Second, the waiting time between events showed a spike at short timescales (5-10 minutes), suggesting that once a cell shows a burst of Msn2 localization it is more likely to experience a second event soon after.

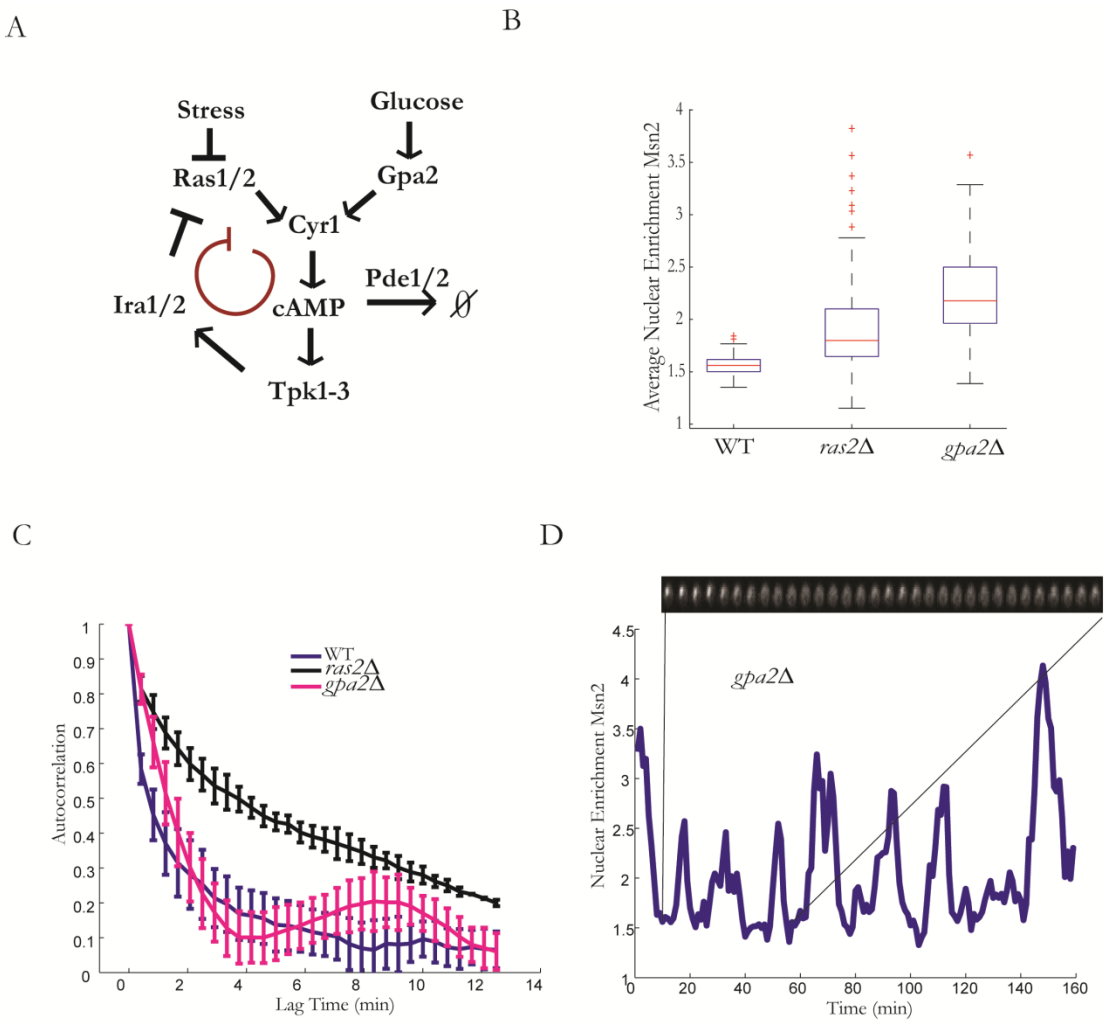


Figure 15: The PKA network regulates Msn2 localization. (a) A schematic of the connectivity of the PKA network components. (b) Average nuclear enrichment over time of Msn2-mCherry in wild type cells or *gpa2Δ* or *ras2Δ* strains. (c) Autocorrelation functions of Msn2-mCherry localization in wild type or mutant strains show that the two different G-proteins govern different aspects of Msn2 dynamics. (d) The trace of Msn2-mCherry nuclear localization in a single cell from a *gpa2Δ* strain shows periodic activation of Msn2.

The localization of Msn2 is regulated by the PKA pathway, the catalytic protein kinase subunit of this network is regulated by the concentration of the small molecule and second messanger cAMP, which in turn is regulated by a complex multi-component network whose structure is conserved throughout fungi and whose components are conserved in all eukaryotes. The basic structure of this network has been elucidated for decades (Figure 15A). To test the

hypothesis that the connectivity of this network and particularly the topologies of the core feedback loops contribute to the dynamics of Msn2 regulation we assayed the dynamics of Msn2 in strains with either of the two small G-proteins which serve as inputs into the adenyl cyclases (Nakafuku et al., 1988). Deletion of either RAS2 or GPA2 results in increased average nuclear localization of Msn2 (Figure 15B), but have substantially different effects on the dynamics of Msn2 localization. Deletion of RAS2 greatly slows the localization dynamics of Msn2 as shown by greatly increased autocorrelation time (Figure 15C). The increased autocorrelation time is consistent with theoretical models of the consequences of removing a negative feedback loop from a noisy system (Cox et al., 2008). In contrast GPA2 shows similar falloff of autocorrelation as wild type, the one change being the appearance of a small peak at 9 minutes suggestive of somewhat oscillatory behavior in this strain. We can see evidence of this behavior in a subset of individual cells which show periodic busts of nuclear localization (Figure 15D).

These results show that Msn2 dynamics are regulated by the PKA pathway and hint that the feedback heavy regulatory architecture generates complex pseudo-oscillatory phenotypes. To explore these behaviors experimentally we took two approaches. The first was to extend our analysis of the nuclear localization of Msn2 in genetic backgrounds with different PKA components deleted. This analysis is ongoing and is complicated to some extent by the strong growth phenotypes displayed by deletion of many components of the PKA pathway. Our second approach was to develop high time resolution input tools to modulate PKA activity. To accomplish this we utilized a recently characterized photo-activateable adenyl cyclase (PBAC; Stierl, 2011; Ryu, 2010), expression of a codon optimized version of this construct in cerevisiae resulted in PKA activation which was dependent on illumination with blue light. The light regulation of this protein emerges from its BLUF domain, which is a common prokaryotic light sensing module. When we express a PBAC construct with a mutation to the Tyrosine-7 residue

(Y7F), which is invariant across related BLUF domain and critical to the photo-catalyzed conformational change, we observe no response to illumination (Mathes, 2012).

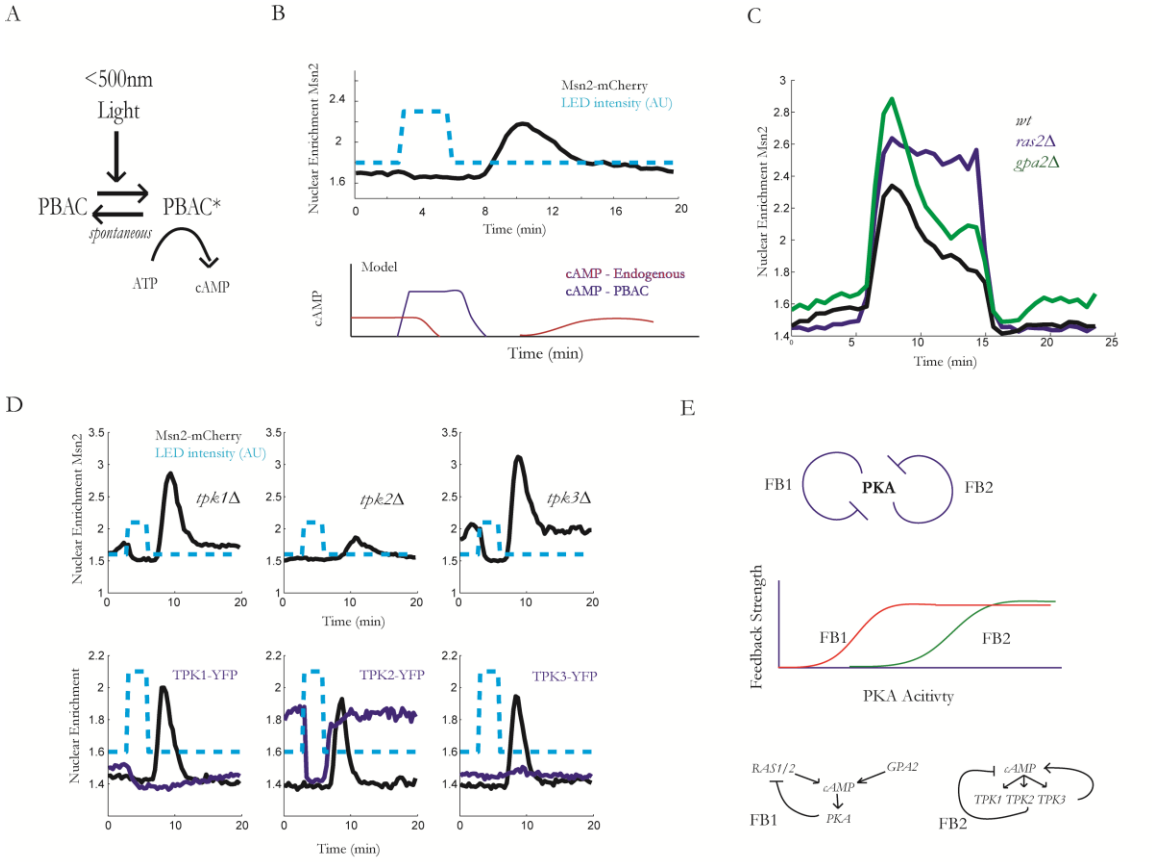


Figure 16: Photoactivatable cAMP production enables high time resolution measurements of PKA signaling. (a) The Photoactivatable Bacterial Adenyl Cyclase (PBAC), switches from an inactive to an active state on illumination with blue light. (b) A pulse of illumination from a 465nm LED results in a subsequent pulse of Msn2-mCherry nuclear localization, we ascribe this behavior to strong negative feedback loops embedded in PKA which drive down the endogenous production of cAMP when it is externally applied. (c) A pulse experiment in strains lacking key PKA components results in distinct outcomes, *gpa2Δ* strains show increased amplitude whereas *ras2Δ* strains show a conversion to s square wave-like profile. (d) Strains lacking the three redundant catalytic PKA subunits TPK1/2/3 show distinct phenotypes in response to a light pulse, with *tpk2Δ* strains showing greatly attenuated nuclear enrichment of Msn2 (upper panels). Monitoring the localization of TPK1/2/3-YFP fusions in response to a light pulse shows distinct dynamics, with TPK2-YFP showing rapid nuclear export on activation of PBAC and similarly rapid re-entry when the LED illumination is switched off, TPK1 and TPK3 in contrast show much more muted behavior (bottom panels). (e) Models of PKA regulation, two negative feedback loops seem to contribute to the regulation of PKA with distinct domains of activity (upper). As previously established RAS1/2 is involved in a negative feedback loop with cAMP while GPA2 is not. Our data suggests that TPK2 strongly inhibits cAMP, whereas TPK3 appears to have the opposite effect.

The PBAC protein, which incorporates an FAD molecule as a chromophore, is activated by a conformational change induced by excitation with <500nm light and then spontaneously reverts to the inactive state with a half life of ~10s (Figure 16A; Stierl, 2011; Ryu, 2010). To use this system to probe the dynamics of PKA we introduced the protein under a medium strength and very low noise NOP7 promoter into a strain which also expresses Msn2(C649S)-mCherry – this protein is incapable of binding DNA but displays normal translocation dynamics and therefore serves as an effective readout of PKA activity. Illumination of these cells with blue light resulted in MSN2 becoming entirely cytoplasmic, however when we turned off the illumination we observed Msn2 to rapidly (~1 minute) translocate into the nucleus (Figure 16B). These results suggested that strong feedbacks embedded within the PKA architecture were responding to the exogenous supply of cAMP by reducing the cellular production through CYR1, and that these feedbacks had slower kinetics than the shutoff of the PBAC cAMP source, this results in a transient dip in PKA activity immediately following a pulse of blue light (Figure 16B, bottom panel).

The major feedback loop in the PKA pathway is through the small G-protein RAS2 (Figure 15A), to test if this feedback loop was driving the behavior we observed we compared the response of our Msn2-mCherry reporter to pulses of light in strains where either RAS2 or GPA2 were deleted. The *gpa2Δ* strain showed substantially increased amplitude in response to a light pulse, but was kinetically indistinguishable from wild type. The *ras2Δ* strain in contrast showed both increased amplitude and a different pulse shape, converting a peak into a square wave, suggesting a role for Ras2 in adaption to low PKA conditions (Figure 16C). Surprisingly the rough behavior of the system and the presumed strong negative feedback that drove that behavior remained the same, suggesting that Ras2 does not play a role in adaption to high cAMP state. These results suggest that there are at least two negative feedback loops in the pathway which

operate in opposite regimes, one restores homeostasis from low PKA state one from the high (Figure 16E).

To identify components of the PKA pathway which contribute to this putative negative feedback we preformed the same pulse response experiment on strains with deletions of the non-essential components of the PKA pathway. Intriguingly deletion of TPK3, one of the three catalytic subunits of PKA, resulted in increased amplitude in response to a light pulse, whereas deletion of its homolog TPK2 resulted in greatly reduced response (Figure 16D). This argued that TPK2 might play a crucial role in the negative feedback we observe, consistent with this TPK2 has been implicated in phosphorylation of CDC25, the guanine nucleotide exchange factor for RAS2. The TPK proteins and the regulatory subunit BCY1 have been reported to change their localization on exposure to alternate carbon sources, we therefore wondered if their might be a dynamic component to this behavior in the context of our experiments. We therefore tagged BCY1 and each of TPK1/2/3 with YFP and monitored their behavior in response to a light pulse. Intriguingly, while TPK1/2/3 were all partly nuclear in resting conditions on exposure to blue light and activation of PBAC TPK1 and TPK2 showed rapid translocation out of the nucleus into the cytoplasm, whereas TPK3 and BCY1 showed no appreciable change in localization (Figure 16D and data not shown). In the case of TPK2 this localization rapidly reverted when the blue light illumination was turned off, whereas TPK1 show a more gradual recovery of nuclear signal. Interestingly both entry and exit of Tpk2-YFP from the nucleus was substantively faster than Msn2, particularly in the case of re-entry into the nucleus where Tpk2-YFP preceded Msn2-mCherry into the nucleus by ~30s.

These results implicate TPK2 (and TPK1 to a lesser extent) in the strong negative feedback we observe. The nuclear export of TPK1/2 is also interesting in that it occurs at in the same manner as Msn2, in that more PKA activity results in reduced nuclear localization, but with a very different offset. Basal levels of PKA activity observed in resting cells grown in glucose

are sufficient to keep Msn2 entirely cytoplasmic, whereas under these conditions TPK2 is entirely nuclear. This suggests that different PKA regulated events are precisely tuned to occur at different levels of PKA activity, allowing the RAS2 feedback loop to dominate at low levels of PKA activity, whereas the TPK1/2 feedback loop becomes relevant in high PKA situations.

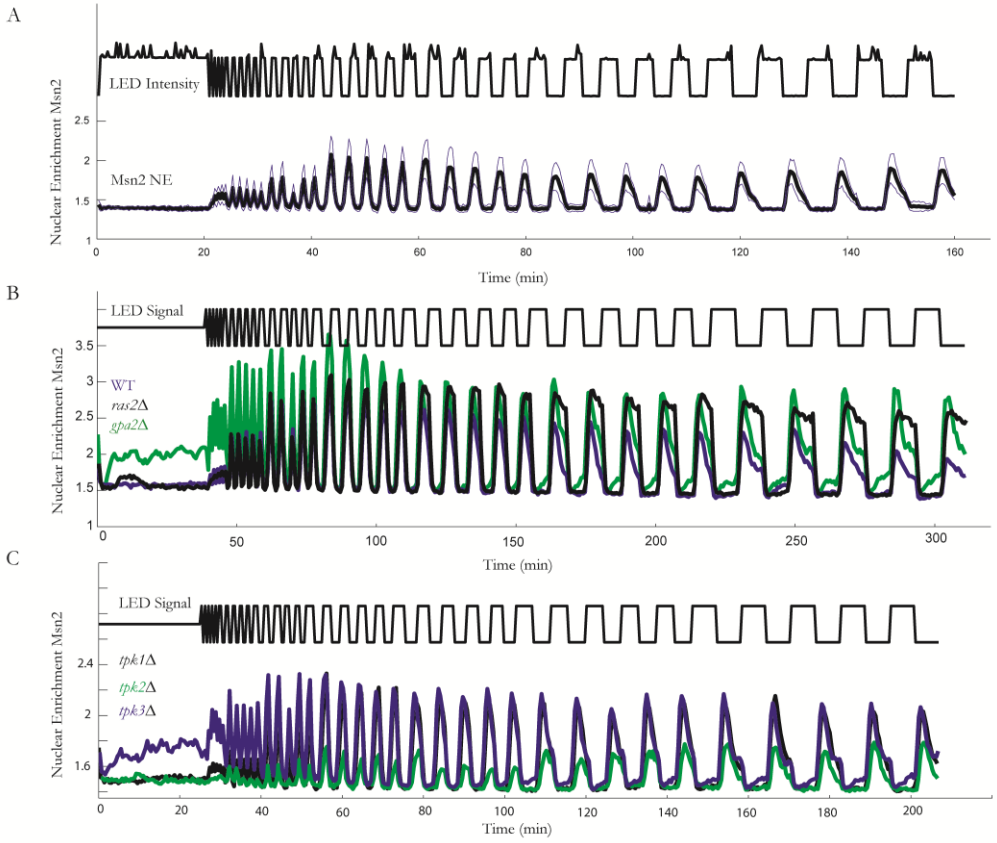


Figure 17: Frequency sweeps of Msn2-mCherry nuclear localization using a photoactivatable adenyl cyclase (PBAC). Cell were placed on a microscope slide where they could be simultaneously imaged and excited with strobed illumination with a 465nm LED light source, the frequency of this square wave was increased over time and the response of Msn2 measured. (a)Response of a wild type strain, black is the mean nuclear enrichment across a population of cells, blue lines represent one standard deviation from triplicate measurements. (b) same experiment in WT, *ras2Δ* and *gpa2Δ* strains, note increased enrichment in the *gpa2Δ* background at high frequencies and altered response profiles in the *ras2Δ* strains at low frequencies. (c) Frequency response of *tpk1/2/3* deleted strains, note increased response of *tpk3Δ* and weakened response of *tpk2Δ*.

As the cellular activity of PKA is very dynamic, with a range of different behaviors depending on the environmental condition we wondered how perturbations to the feedback loops

we identified would alter the overall behavior of the PKA circuit. To explore this we sought to define the frequency response of the PKA circuit to an oscillating signal in different genetic backgrounds. We therefore applied a set of light pulses of decreasing frequency to cells expressing PBAC and measured the amplitude of the nuclear enrichment of Msn2-mCherry at a time resolution of 20s. The nuclear enrichment faithfully followed the light pulses, with Msn2 showing strong periodic localization with amplitude that rose as the period of the light input increased (Figure 17A). We then preformed the same experiment in strains with deletions of PKA regulator genes which we identified as playing a role in the system dynamics. Deletion of GPA2 and RAS2 increased the amplitude with which the system responded at all wave lengths, with RAS2 also altering the output profile somewhat (Figure 17B). In contrast the TPKs had dramatic effects on the frequency output, with the *tpk2Δ* strain showing greatly reduced amplitude at all frequencies and the *tpk3* strain increased amplitude at high frequencies (Figure 17C). These results emphasize the critical and differential role of each of these subunits in determining the dynamic properties of the PKA circuit.

Discussion 3.1

Microorganisms continuously adapt their metabolic processes to the environment to optimize growth rate, this requires robust sensing and transduction of environmental signals. In budding yeast the cAMP/PKA circuit manages much of the environmental adaption, particularly in response to stressful conditions it activates the environmental stress response to produce an array of cyto-protective proteins and metabolites. However, PKA does not work as a simple rheostat, but rather encodes its regulation in the dynamic pattern of its activity. In conditions of rapid growth there occurs brief periods of low PKA activity which we can observe as pulses of Msn2 nuclear localization. These pulses appear to occur stochastically and have approximately exponential waiting times and Poisson statistics, albeit somewhat overdispersed.

Analysis of the pattern of spontaneous Msn2 bursts in wild type and strains lacking components of the PKA pathway revealed that both the average residency of Msn2 in the nucleus and its dynamics could be substantially altered. Strains missing the small G-protein RAS2 showed increased average localization, but also greatly increased auto-correlation time, consistent with the role of Ras2 in a negative feedback loop which regulates the PKA pathway. In constast deletion of a second G-protein and adenyl cyclase regulator Gpa2 showed similar autocorrelation to the wild type, with the appearance of a second weak mode. This suggested that in strains lacking Gpa2 some oscillatory behavior is observed in the system with a period of ~9min, examination of individual cells showed localization patterns consistent with this. These results are consistent with proposed oscillations in the PKA pathway due to the abundance of negative feedback loops (Garmendia-Torres et al., 2007), but suggests that this behavior only occurs under quite specific circumstances.

To explore the regulation of PKA experimentally we have developed a light regulated input system capable of generating cAMP signals at high resolution. When combined with automated microscopy and the Msn2 nuclear localization reporter we can simultaneously perturb and monitor PKA activity with a resolution of ~30s. This tool was applied to study the nature of feedback loops regulating PKA activity, using this approach we identified two distinct feedback loops which operate at different regimes of PKA activity. One loop which is dependent on RAS2 allows the cell to recover from low PKA states, physiologically cells lacking RAS2 show substantially slower Msn2 dynamics, with longer excursions from steady state. In contrast the catalytic subunit TPK2 is required for the cell to recover from a high PKA state but is largely inactive in physiological conditions.

Further, using this optogenetic tool it is possible to explore the frequency response of the PKA system as a whole. We found that comparing wild type and *ras2Δ* or *gpa2Δ* strains, either of the deletions increased the amplitude of the Msn2 response without greatly altering the

frequency sensitivity. The response of the system at low frequencies was, however, substantially enhanced in *tpk3Δ* strains and reduced in *tpk2Δ*, consistent with the latter's role in negative feedback regulation.

Overall, these results position PKA as a complex dynamic system which is activity encoding environmental signals into dynamic signaling. PKA accomplishes this encoding through the use of multiple feedback loops which are staggered to respond to different ranges of PKA activity. Although our data only conclusively hints at negative feedback loops, the pulse like behavior of the system suggests that there may be an as yet unknown positive feedback loop in this system. Alternatively, highly cooperative regulation of the small G-protein RAS1/2 might provide sufficient delays to allow for excitability in the absence of a positive feedback.

3.2

PKA regulation of MSN2, TOR regulation of GAT1/GLN3, a genome screen

In mammals and budding yeast two of the major cellular regulatory kinases which govern growth rate are PKA and TOR (Broach, 2012). Nominally PKA responds to carbon source stress and TOR to nitrogen starvation, but in budding yeast these pathways seem of have been adopted as general stress regulators responsive to many different stimuli (eg. Petkova et al., 2010). Moreover these pathways co-regulate many different cellular processes including autophogy, ribosomal biosynthesis, and the environmental stress response (Slattery et al., 2008; Jorgensen et al., 2004; Pedruzzi et al., 2003; Stephan et al, 2009). This apparent overlap in function is complicated by the interlinked nature of these pathways. The complexities of this interaction have resulted in a score of apparently contradictory papers about their relationship (Ramachandran and Herman, 2011; Soulard et al, 2010; Stephan et al., 2009; Zurita-Martinez and

Cardenas, 2005). To understand the regulation of the PKA network we therefore had to define the dependence and degree of interconnectivity to the TOR pathway.

The most consistent feature of the literature is that inhibition of TOR (with small molecule inhibitors) results in inhibition of PKA (Soulard et al., 2010; Huber et al., 2009). We recapitulate this result and expand on it, showing that in *S. cerevisiae* inhibition of TOR with rapamycin results in a rapid and transient inhibition of PKA. To further explore the interactions between these pathways we preformed a genome wide deletion screen using quantitative florescent reporters of each pathway. This screen showed that under many circumstances these two pathways behave effectively independently, the one striking exception was that both pathways show strong inhibition in strains which are defective in mitochondrial regulation through the retrograde response.

Results 3.2

To measure PKA activity in real time at in single cells we took advantage of the nuclear localization of Msn2, when PKA activity drops Msn2 becomes localized to the nucleus. By Expressing an Msn2 allele tagged with mCherry we could observe PKA activity in single cells on addition of the small molecule TOR inhibitor rapamycin. Using this approach we observed a sharp increase in Msn2 nuclear localization with 2 minutes of rapamycin addition in most cells in the population (Figure 18A). This localization on rapamaycin addition was not observed in strains where PKA was constitutively active by Ras2(g19v) expression arguing that TOR is acting through PKA to activate Msn2 (data not shown). Msn2 localization peaked around ten minutes and then dropped precipitously. After these initial events individual cells became de-synchronized and showed frequent bursts of nuclear localized Msn2-mCherry. Averaging across the whole population of cells there was also a second broader peak localization around 30 minutes after addition of rapamycin, although this localization was less apparent in individual

cells (Figure 18B). These results are worth contrasting to nuclear localization of Msn2 induced by other stress conditions such as osmotic stress, where individual cells show fairly synchronous behavior (Figure 8A).

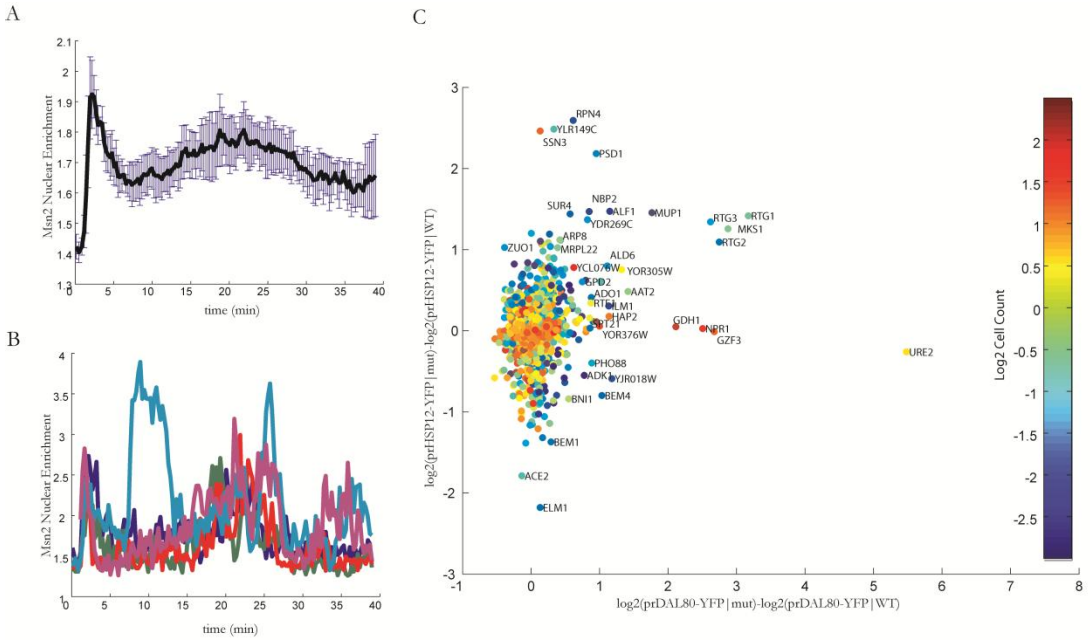


Figure 18: Inhibition of TOR results in a transient drop in PKA activity, at longer timescales PKA and TOR are largely independent. (a) Msn2 localization averaged over a field of cells, at time zero 1um of Rapamycin was added to the cells. Errorbars are standard deviation of six biological replicates. (b) The Msn2 nuclear localization traces of four single cells (differentiated by color) from the experiment described in (a), note the high variability especially after the initial peak of localization (~2 minutes). (c) A genome wide screen of the gene expression of the PKA reporter prHSP12-mKate2 and TOR reporter prDAL80-YFP in response to deletion of non-essential genes. Each dot represents expression of each reporter in a strain lacking a given gene, X- and Y-axes are in log2 fold change over the median value in the dataset, genes are colored according to their cell number (a proxy for growth rate).

This data suggested that TOR inhibition results in the inhibition of PKA, but that this inhibition is transient and PKA returns to something approaching untreated activity. However, at longer timescales rapamycin clearly still exerts an inhibitory effect on PKA activity resulting in short transient bursts of Msn2 localization reminiscent of what is observed in conditions of low glucose.

To explore the interaction of PKA and TOR more completely we designed a strain expressing florescent reporters of PKA and TOR activity. To report on PKA activity we elected to use prHSP12, a promoter that is responsive to MSN2, a transcription factor which is principally regulated by PKA activity. To measure TOR activity we used a previously validated reporter prDAL80 (Neklesa and Davis, 2009), a promoter that is bound and regulated by the GATA transcription factors GLN3 and GAT1. These promoters were cloned in front of YFP (DAL80) and RFP(HSP12) and integrated into a SGA strain which was then mated to the deletion collection. The expression of RFP and YFP was then measured in each strain by flow cytometry which allowed us to measure HSP12 and DAL80 activity and estimate cellular growth rates based on the cell numbers measured (Figure 18C).

We observed substantial variation in HSP12 activity (ranging across ~6log2s), with deletion of ~200 genes altering its expression significantly. These genes included some components of the PKA network (GPA2, IRA2, PDE2, GPB1), chromatin modifiers (IES4/5, RSC2), and proteosome associated factors (RPN4, SSN3). It is worth noting that some of these genes such as the chromatin modifiers and proteosomal factors may actually be exerting their effect through MSN2 directly rather than PKA somewhat complicating the interpretation. In contrast DAL80 had relatively few genes which substantially altered its expression. The most striking was URE2, which is a negative regulation of the transcription factors GAT1/GLN3 (Courchesne and Magasanik, 1988), and resulted in a massive increase in DAL80 signaling. Otherwise two groups of genes increased DAL80 activity, GDH1, NPR1, and GZF3 which are all related to nitrogen metabolism or signaling. The second group of was more interesting as it also increased expression from HSP12, it consisted of all four validated components of the retrograde signaling pathway (RTG1/2/3, MSK1), which transduces signals from the mitochondria to the nucleus when the mitochondria are experiencing stress (Liao and Butow 1993; Tate et al. 2002; Ferreira et al., 2005).

Additional experiments were run comparing the effects of PKA inhibition (by PDE2 overexpression) or TOR inhibition (Rapamycin addition). This data revealed quantitative changes in the relationship between individual genes and each pathway but did not quantitatively change the overall picture and was not analyzed further.

Discussion 3.2

These results argue that inhibition of Tor results in temporary loss of PKA activity. However, PKA activity rapidly recovers arguing that TOR can be best thought of as an input to the PKA pathway rather than a dominant regulator. At longer timescales which were accessible with our genetic screen it appears that PKA and TOR can be treated as relatively distinct pathways responsive to different genetic perturbations. In general we observe PKA to be much more sensitive to genetic perturbations than TOR which shows dramatic phenotypes of only a handful of deletions. Finally, we find that the mitochondrial retrograde signaling pathway is a potentially crucial link between the two signaling pathways. This is interesting given the role of mitochondria in carbon and nitrogen metabolism. However, the phenotype induced by the retrograde components does not appear to be mitochondrial dysfunction per se, as deletion of core components of the mitochondria did not result in substantive TOR phenotypes and had only a moderate effect on PKA. This suggests that the retrograde pathway may be playing a crucial and as yet unexplored *signaling* role in the regulation of the TOR and PKA pathways.

3.3

Multiple Processes are Regulated by PKA

Observing Msn2 tagged with YFP in individual cells we observed spontaneous ‘bursts’ of activity. These behaviors have been observed in other transcription factors as well, particularly the calcium sensitive transcription factor CRZ1 (Jacquet et al., 2003; Cai et al., 2008). We

wondered if these apparently 'stochastic' bursts of activity are intrinsic to the individual factors, or if there are more general patterns of variation that impact multiple transcription factors simultaneously. A small number of 'burst generators' each hooked to many transcription factors would avoid the presumably high costs of constructing such a system many times. Moreover, in the case of stress responses large numbers of factors become activated simultaneously to adjust different aspects of the cellular metabolism. For example in response to all stressful conditions budding yeast turn down ribosomal production and up environmental stress responsive genes, but these processes are managed by at least four distinct ribosomal transcription factors and at least two stress responsive factors (Broach, 2012).

To study the underlying regulatory networks which result in the 'bursty' regulation of transcription factors we first screened the GFP library collection of factors for those which underwent visible translocation under stress, this screen defined a dozen potential factors. We complemented this set with a handful of reporters of cellular processes derived from the literature. We than measured the cross-correlation of the nuclear localization of these factors and Msn2. These measurements revealed that six factors, Msn4, Dot6, Tod6, Stb3, and Crz1 showed substantive co-activation with Msn2 arguing for co-regulation by PKA. Interestingly this behavior was only partial, with each factor showing distinct frequencies of localization, in general Msn2 showed the highest frequency of bursts of nuclear localization with other factors entering the nucleus in a subset of these events. A potential model for this behavior in that PKA acts as a central pacemaker for these factors, which is then gated by other signals.

To explore this further we utilized our photoactivable cycalse, and explored the frequency response of these transcription factors in response to light. While some factors such as Dot6 and Tod6 showed strong susceptibility to our synthetic PKA input, others were resistant, supporting a model of gated activation. To test this we selected one factor, Stb3 which has been proposed to be regulated by TOR activity, indeed addition of rapamycin sensitized Stb3 to light

based regulation, arguing that this factor is maintained in the cytoplasm by both PKA and TOR activity with OR gate type logic.

Results 3.3

We began by finding transcription factors which might show dynamic co-localization by finding those factors which show dynamics in the first place. To do this we took advantage of the GFP library, we selected ~280 transcription factors from this collection and imaged them in unstressed conditions and after the addition of 0.5mM KCL. Additionally we combed the literature to define transcription factors or other proteins which had been described to show dynamic shifts of localization. This approach focused on localization, which we can measure in real time, rather than activity which is generally very difficult to measure dynamically.

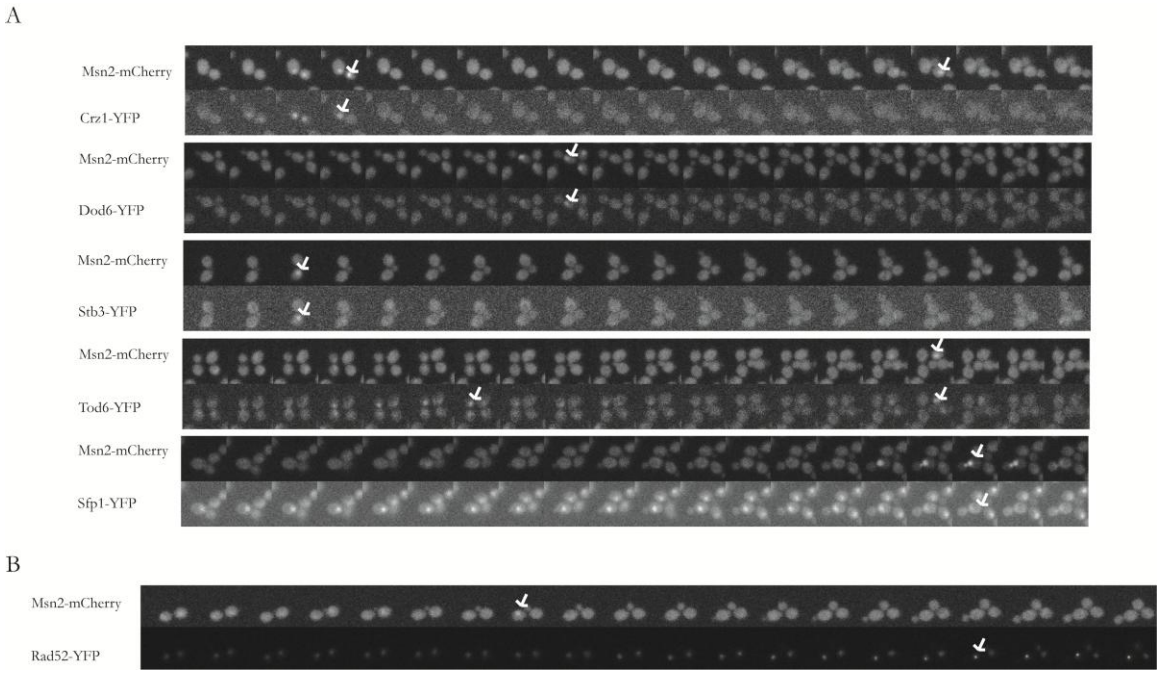


Figure 19: Monitoring nuclear co-localization in unperturbed cultures reveals shared signaling pathways regulate many different transcription factors. (a) Cell expressing Msn2-mCherry and X-YFP were maintained in exponential growth in rich media and imaged for 2hrs every minute, localization events are highlighted with arrows. Panels are montages of sets of cells, with four frames between each image. (b) Msn2 nuclear localization events do not coincide with DNA damage as assessed by Rad52-YFP foci.

Using this combination of approaches we defined ~a dozen factors which show a degree of localization shifts under some conditions, and eight (RAD52, MSN2, MSN4, DOT6, TOD6, SFP1, STB3, and CRZ1) which show rapid dynamics under rich growth conditions. Analysis of these seven factors (with the exception of RAD52) showed that they each had three or more consensus PKA phosphosites and have all been previously suggested to be regulated by PKA (Kafadar and Cyert, 2004; Huber et al., 2011; Lippman and Broach, 2009; Jorgensen et al., 2004). To explore the possible co-regulation of these factors we constructed strains with MSN2-mCherry and one of the other seven factors tagged with YFP. We imaged these strains in exponential growth at a time resolution of one minute for two hours. The images were then analyzed and single cell traces extracted. From these traces we then looked for co-occurrence of 'bursts' of nuclear localization. For the six transcription factors the five stress responsive factors (Msn4, Dot6, Tod6, Stb3, and Crz1) show strong co-localization with Msn2. The sixth factor the pro-growth ribosomal regulator SFP1 showed a weak tendency to exit from the nucleus when Msn2 was active (Figure 19A).

Closer analysis of the data revealed a more interesting pattern where localization of Msn2 to the nucleus occurred with reasonably high frequency (in perhaps 5% of cells at a single time point), whereas localization of Stb3, Dot6, and Crz1 was very infrequent, occurring in <1% of all cells, Tod6 localizations exceeded Msn2 occurring in almost 20% of cells and were also of fairly long durations compared to the other factors. The general pattern was therefore that localization of Dot6, Stb3, or Crz1 was typically accompanied by Msn2 nuclear localization, whereas Msn2 and Tod6 nuclear localization often occurred without localization of a second factor. Rad52, which forms foci on DNA damage was something of an outlier from these results, formation of transient Rad52 foci, which occur in ~10% of cells as they transit through S-Phase (Alvaro et al, 2007) showed no substantive correlation with Msn2 nuclear localization (Figure 19B). Long lived Rad52 foci (>15min), perhaps representing more catastrophic forms of damage, showed

some qualitative association with Msn2 localization, but these events occurred infrequently and we lacked the statistical power to explore this more definitively.

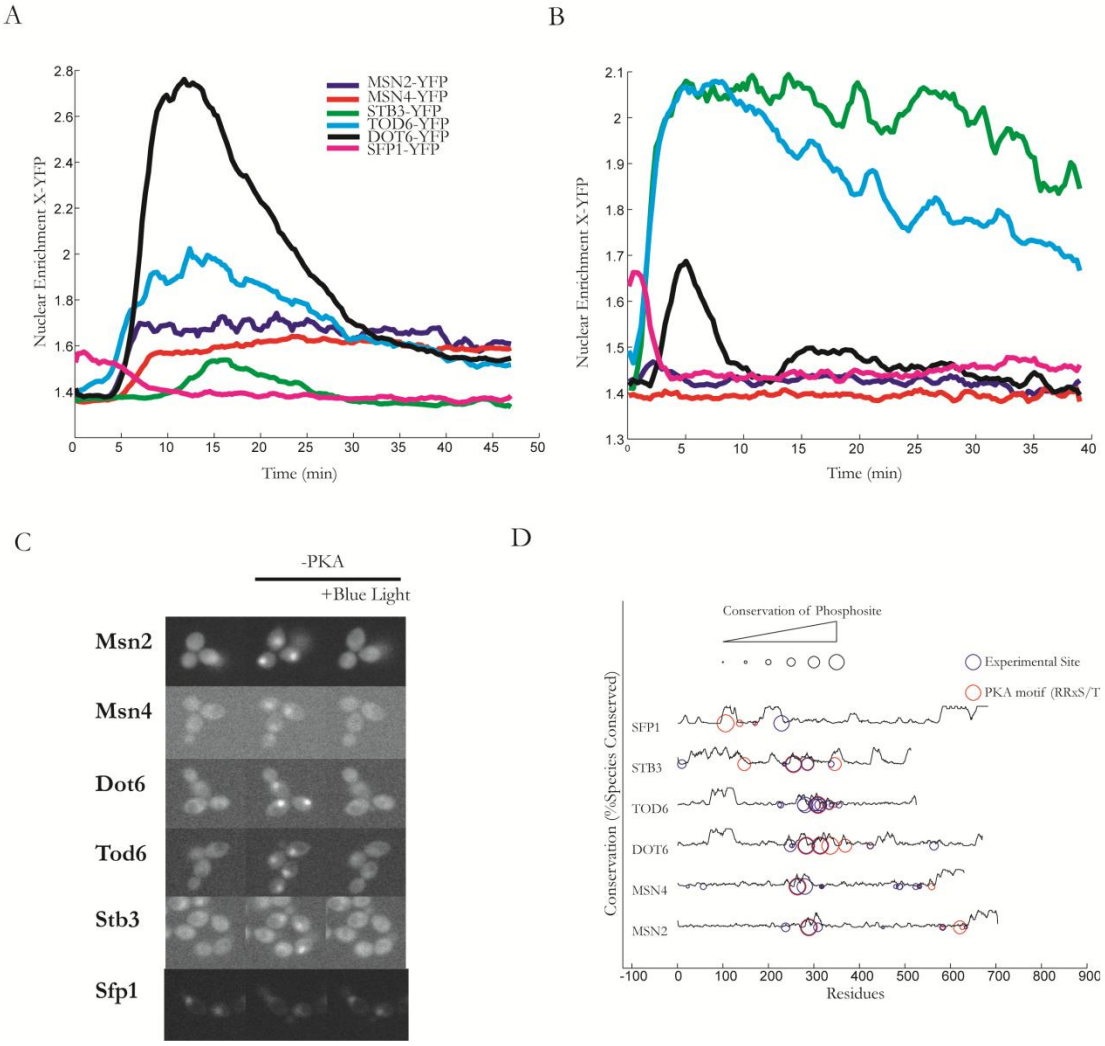


Figure 20: Different transcription factors respond with very distinct kinetics to PKA and TOR perturbations. (a) Populations of cells expressing a transcription factor tagged with YFP were grown into exponential phase and imaged after induction of RAS2(S24N) expression resulting in PKA inhibition, traces represent average nuclear enrichment across the population. (b) Cells were treated with 1um rapamycin at T=0, traces represent average nuclear enrichment across the population. (c) Images of cells expressing transcription factors tagged with YFP without treatment, with PKA inhibition by overexpression of RAS2(S24N), and after a rescue of this inhibition by activation of the PBAC construct by illuminated by blue light. (d) Regulatory sites are conserved in PKA regulated transcription factors. Conservation of each transcription factor in *K. lactis*, *C. Glabrata*, *S. Castelii*, *S. Bayanus* is plotted by residue, degree of conservation is given by height of the line (smoothed over a window of 8), conservation of phosphosites is given by circle size.

As all of the transcription factors factors have been suggested to be at least partly regulated by PKA, we therefore asked if the co-localization we observe could be caused by PKA. To test this we measured the localization of Msn2, Msn4, Stb3, Dot6, Tod6, and Sfp1 as we induced the expression of a RAS2dn allele—whose expression rapidly and completely inhibits the PKA pathway—and monitored the localization of each of these factors over time. We noted although all these factors responded to PKA inhibition they did so with markedly different kinetics and overall dynamics (Figure 20A). Similarly we measured the localization of these factors in response to the TOR inhibitor rapamycin, which again showed a strong response from all factors with distinct dynamics (Figure 20B). To show that the localization we observed is a response to PKA activity we tested that the effect of the RAS2dn expression could be reversed by activation of the photoactivable PBAC kinase (Figure 20C), which in all cases it could. Further, each factor has at least one highly conserved PKA phosphosite (Figure 20D).

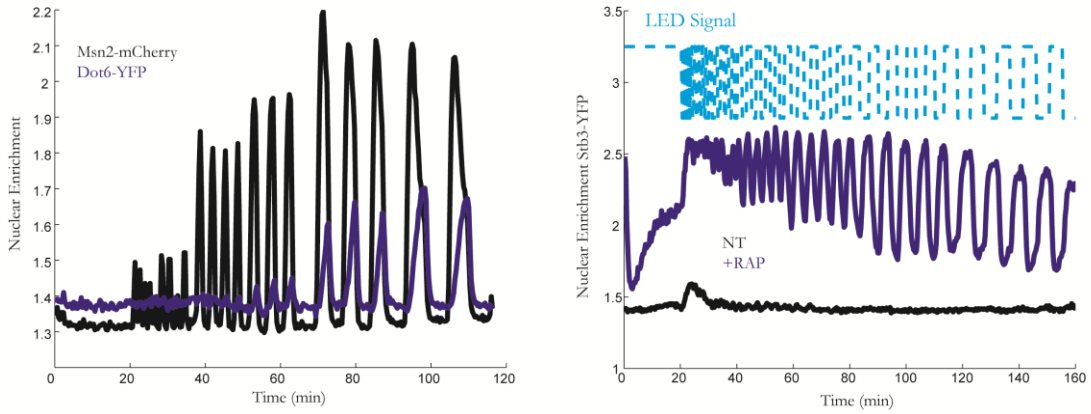


Figure 21: Transcription factors vary in their frequency response to PKA perturbation (achieved by photoactivation of an adenylate cyclase) and also the requirement for secondary signals. (a) cells expressing Msn2-mCherry, Dot6-YFP, and the PBAC construct were stimulated with a range of frequencies of light stimulation. Nuclear localization of Msn2 (black) responded robustly to relatively high frequencies wheras Dot6 nuclear localization (dark blue) did not occur until lower frequencies were reached. (b)Stb3-YFP showed little response to PKA activation in untreated cells (black line), but upon treatment with rapamycin the localization of Stb3-YFP became robustly sensitivity to PKA (dark blue).

Finally to explore the dynamic response of these factors to PKA stimulation we preformed a frequency response experiment with the blue light system. While some proteins such as Dot6 and Tod6 show moderate responsiveness to PKA stimulation, albeit (and intriguingly) with different frequency responses (Figure 21A), others such as STB3 showed a muted response to PKA stimulation (Figure 21B). Considering that these same factors do not show identical pattern of nuclear localization as Msn2, we wondered if in addition to PKA the localization of these factors is regulated by an additional regulatory signal. In each case we have a good candidate signal in that Stb3 has been proposed to be regulated by TOR activity, and Crz1 has been shown to be regulated by Calcinurin. Intriguingly when treated with rapamycin Stb3 showed substantive regulation by PKA activity arguing that this factor is indeed regulated by multiple signals, namely PKA and TOR (Figure 21C).

Discussion 3.3

Regulation of localization is a common strategy for controlling the activity of transcription factors, because this regulation can be measured in live cells with microscopy observation of transcription factor localization provides a real-time window into the activity of cellular signaling pathways. Recent work in many systems has found that far from simple on/off switches Transcription factors behave stochastically and dynamically in individual cells. Here we show with time resolved two feature measurement that much of the variability we observe in transcription factors is not intrinsic to those factors, but rather represent the activity of one or more pathways upstream. These results are reminiscent of our results in Chapter One, where a substantive fraction of variation in gene expression resulted from upstream regulation which was shared across many genes. Here we see the PKA pathway in addition to its specific role in coordinates the dynamic behavior of many different cellular factors.

These results position PKA as a central dynamic hub of the cell, potentially providing a pacemaker for diverse cellular processes. Further, by combining sensitivity to PKA and to other signaling pathways a transcription factor can produce burst like behavior as required. For example, we find that while Msn2 appears to respond fully and faithfully to PKA dynamics, whereas Stb3 shows PKA sensitive behavior that is 'gated' by TOR activity. Alternatively, a second way of gating transcription factor regulation is to give different factors distinct frequency responses, for instance Msn2 responds rapidly to PKA kinetics whereas Dot6 lags and requires longer durations of PKA activity to become active.

The approach we develop here combines long term non-invasive imaging with specific high temporal resolution perturbations to study the structure of a complex dynamic system. We expect this data to be amendable to reasonable throughput acquisition and analysis and to provide a potentially fruitful ground for modeling of the structure of PKA and its interactions with other pathways.

Chapter 3 Methods

Plasmid and Strain Construction

Overexpression of the tagged transcription factors were constructed by PCRing the genes from genomic DNA, and fusing it to the N-terminus of a yeast optimized mCherry gene, this construct was then cloned in front of an ADH1 promoter in the TRP1 marked single integration vector. The Ras2 overexpression construct was built by PCRing RAS2 from the genome, cloning it in front of a GAL1 promoter in a TRP1 marked single integration vector, and using site directed mutagenesis to obtain the dominant negative S24N allele.

All strains were derived from W303a, reverted to ADE+ to reduce auto-florescence, PBAC and TF-mCherry plasmids were serially transformed into this background using standard LioAc/PEG protocol.

The mKate2/YFP fusions strains were constructed by PCR mediated genomic tagging using a Venus-HIS3 or mKate2–URA3 plasmid as a template. Gene deletions were obtained by PCR mediated genomic disruption using a NAT or G418 cassette, and were verified by colony PCR.

The RAS2 construct was transformed into a W303(MATalpha) strain, which had a Gal4-estradiol construct integrated into the LEU2 locus and was protrophic for URA3. This strain was then mated to the W303a strain expressing the PBAC construct and TF-YFP, diploids were selected using SD-URA/-HIS media.

The photactivatable cyclase was codon optimized for cerevisiae, synthesized by IDT, and cloned in front of a strong constitutive NOP7 promoter in a single integration vector with a LEU2 marker. The Y7F mutation was accomplished by site directed mutagenesis. The Msn2 reporter was constructed by PCRing MSN2 from genomic DNA, and fusing it to the N-terminus of a yeast optimized mCherry gene, this construct was then cloned in front of an ADH1 promoter in the TRP1 marked single integration vector. The C649S mutation which eliminated DNA binding was accomplished by site directed mutagenesis.

SGA Strain Construction

A reporter strain was constructed based on the SGA technology developed in the Boone Lab []. This strain incorporated reporters for Msn2 activity (prHSP12-mKate2) and Gln3/Gat1 activity (prDAL80-YFP). Additionally an estradiol inducible construct was integrated and PDE2 was placed under control of a Galactose inducible promoter (prGAL1-PDE2). The final querry strain genotype was S288C MATalpha *can1*::prSTE2-His5 *lyp1*::prSTE3-LEU2 *ura3*::GEM, prGAL1-PDE2 (NAT) *trp1*::prDAL80-YFP (TRP1,HIS3) *met15 his3*::prHSP12-mKate2 (Ura3).

This strain was crossed to the deletion collection using modified SGA protocols. Briefly the query strain was mated to the 14 deletion collection plates on YPD (1Day 30C), diploids selected on SD-URA+KAN (1Day 30C), pinned to sporulation media and allowed to sporulate

(3Days room temperature), and pinned to SD+SAEC+Canavanine+KAN+NAT-Ura-Leu-His (2Days, 30C), repined to the same media (1Day, 30C), pinned onto YPD+KAN+NAT (1Day).

SGA Strain Measurement

Strains were pinned from plates into shallow 384 well plates with 60ul SDC, grown for 24hrs to saturation at 30C, diluted 20000 fold and grown for 14hrs at 30C with shaking. Strains were measured in media on a BD LSRII cytometer. Measurement and data processing was as described earlier.

Microscopy

Cells were grown into exponential phase in SD complete and plated onto ConcanavalinA coated 96 well glass bottom plates, allowed to settle and then washed twice with fresh media. Samples were placed in the microscope environmental chamber set at 30C and allowed to equilibrate for 30minutes. Imaging was on a Nikon Ti inverted scope with arc-lamp illumination using RFP(560/40 nm excitation, 630/75 nm emission, chromas) and YFP (510/10 nm excitation, 542/27 nm emission, semrock) filters. Images were processed and analyzed with custom Matlab scripts (see appendix).

To activate the PBAC construct the cells were illuminated with a 465nM LED source regulated by an LED driver (MightEX). Image acquisition and LED illumination intensity was coordinated by custom matlab software interfacing with micromanager microscope control software (Edelstein, et al., 2010).

Where noted the Ras2 allele was induced by the addition of 200nm estadiol (1.6mM stock in EtOH, Sigma) to the imaging medium. Rapamycin (Santa Cruz Biotech, 1mM stock in EtOH) was added to a final concentration of 1uM.

Conservation Analysis

To analyze the conservation of phosphosites in transcription factors we used a derived set of orthologous factors (Wapinski et al, 2007), aligned their sequences using clustalW (Larkin et al., 2007) , and superimposed phosphosites that were computationally identified as consensus PKA sites (RRXS/T) or were derived from a compendium of global phosphoproteomic mapping (Amoutzias et al., 2010).

Chapter 3 References

Amoutzias GD, He Y, Gordon J, Mossialos D, Oliver SG, Van de Peer Y (2010). Posttranslational regulation impacts the fate of duplicated genes. Proc Natl Acad Sci U S A. 107(7):2967-71.

Bodvard K, Wrangborg D, Tapani S, Logg K, Sliwa P, Blomberg A, Kvarnström M, Käll M (2011). Continuous light exposure causes cumulative stress that affects the localization oscillation dynamics of the transcription factor Msn2p.

Biochim Biophys Acta. 1813(2):358-66.

Broach JR (2012). Nutritional control of growth and development in yeast. Genetics. 192(1):73-105.

Colombo S, Ronchetti D, Thevelein JM, Winderickx J, Martegani E (2004). Activation state of the Ras2 protein and glucose-induced signaling in Saccharomyces cerevisiae. J Biol Chem. 279(45):46715-22.

Courchesne WE, Magasanik B. Regulation of nitrogen assimilation in Saccharomyces cerevisiae: roles of the URE2 and GLN3 genes (1988). J Bacteriol. 170(2):708-13.

Cox CD, McCollum JM, Allen MS, Dar RD, Simpson ML (2008). Using noise to probe and characterize gene circuits. Proc Natl Acad Sci U S A. 105(31):10809-14.

Edelstein A, Amodaj N, Hoover K, Vale R, and Stuurman N (2010). Computer Control of Microscopes Using ?Manager. Current Protocols in Molecular Biology 14.20.1-14.20.17

Ferreira Júnior JR, Spírek M, Liu Z, Butow RA (2005). Interaction between Rtg2p and Mks1p in the regulation of the RTG pathway of Saccharomyces cerevisiae. Gene. 354:2-8.

Gasch AP, Spellman PT, Kao CM,Carmel-Harel O, Eisen MB,Storz G, Botstein D, Brown PO (2000). Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. Molecular Biology of the Cell **11**: 4241–4257.

Garmendia-Torres C, Goldbeter A, Jacquet M (2007). Nucleocytoplasmic oscillations of the yeast transcription factor Msn2: evidence for periodic PKA activation. Curr Biol. 17(12):1044-9.

Gonze D, Jacquet M, Goldbeter A (2008). Stochastic modelling of nucleocytoplasmic oscillations of the transcription factor Msn2 in yeast. J R Soc Interface. 5 Suppl 1:S95-109.

Huber A, Bodenmiller B, Uotila A, Stahl M, Wanka S, Gerrits B, Aebersold R, Loewith R (2009). Characterization of the rapamycin-sensitive phosphoproteome reveals that Sch9 is a central coordinator of protein synthesis. Genes Dev.23(16):1929-43.

Huber A, French SL, Tekotte H, Yerlikaya S, Stahl M, Perepelkina MP, Tyers M, Rougemont J, Beyer AL, Loewith R (2011). Sch9 regulates ribosome biogenesis via Stb3, Dot6 and Tod6 and the histone deacetylase complex RPD3L. EMBO J. 30(15):3052-64.

Jacquet M, Renault G, Lallet S, De Mey J, Goldbeter A (2003). Oscillatory nucleocytoplasmic shuttling of the general stress response transcriptional activators Msn2 and Msn4 in Saccharomyces cerevisiae. J Cell Biol. 161(3):497-505.

Jorgensen P, Rupes I, Sharom JR, Schneper L, Broach JR, Tyers M (2004). A dynamic transcriptional network communicates growth potential to ribosome synthesis and critical cell size. Genes Dev, 18(20):2491-505

Kafadar KA, Cyert MS (2004). Integration of stress responses: modulation of calcineurin signaling in Saccharomyces cerevisiae by protein kinase A. Eukaryot Cell. 3(5):1147-53.

Liao X, Butow RA (1993). RTG1 and RTG2: two yeast genes required for a novel path of communication from mitochondria to the nucleus. Cell,72(1):61-71.

Lippman SI, Broach JR (2009). Protein kinase A and TORC1 activate genes for ribosomal biogenesis by inactivating repressors encoded by Dot6 and its homolog Tod6. Proc Natl Acad Sci U S A. 106(47):19928-33.

Mathes T, van Stokkum IH, Stierl M, Kennis JT. (2012) Redox modulation of flavin and tyrosine determines photoinduced proton-coupled electron transfer and photoactivation of BLUF photoreceptors. J Biol Chem. 287(38):31725-38.

Nakafuku M, Obara T, Kaibuchi K, Miyajima I, Miyajima A, Itoh H, Nakamura S, Arai K, Matsumoto K, Kaziro Y (1988). Isolation of a second yeast Saccharomyces cerevisiae gene (GPA2) coding for guanine nucleotide-binding regulatory protein: studies on its structure and possible functions. Proc Natl Acad Sci U S A. 85(5):1374-8.

Neklesa TK, Davis RW (2009). A genome-wide screen for regulators of TORC1 in response to amino acid starvation reveals a conserved Npr2/3 complex. PLoS Genet. 5(6):e1000515.

Nikawa J, Cameron S, Toda T, Ferguson KM, Wigler M (1987). Rigorous feedback control of cAMP levels in Saccharomyces cerevisiae. Genes Dev. 1(9):931-7. Erratum in: Genes Dev 1(10):1351.

Pedruzzi I, Dubouloz F, Cameroni E, Wanke V, Roosen J, Winderickx J, De Virgilio C (2003). TOR and PKA signaling pathways converge on the protein kinase Rim15 to control entry into G0. Mol Cell. 12(6):1607-13.

Petkova MI, Pujol-Carrion N, Arroyo J, García-Cantalejo J, Angeles de la Torre-Ruiz M (2010). Mtl1 is required to activate general stress response through Tor1 and Ras2 inhibition under conditions of glucose starvation and oxidative stress. J Biol Chem. 285(25):19521-31.

Ramachandran V, Herman PK (2011). Antagonistic interactions between the cAMP-dependent protein kinase and Tor signaling pathways modulate cell growth in Saccharomyces cerevisiae. Genetics 187(2):441-54.

Ryu MH, Moskvin OV, Siltberg-Liberles J, Gomelsky M. (2010). Natural and engineered photoactivated nucleotidyl cyclases for optogenetic applications. J Biol Chem.285(53):41501-8.

Schwarzel M, Oertner TG, Nagel G, Hegemann P. (2011) Light modulation of cellular cAMP by a small bacterial photoactivated adenylyl cyclase, bPAC, of the soil bacterium Beggiatoa. J Biol Chem. 286(2):1181-8.

Slattery MG, Liko D, Heideman W (2008). Protein kinase A, TOR, and glucose transport control the response to nutrient repletion in Saccharomyces cerevisiae. Eukaryot Cell, 7(2):358-67

Soulard A, Cremonesi A, Moes S, Schütz F, Jenö P, Hall MN (2010). The rapamycin-sensitive phosphoproteome reveals that TOR controls protein kinase A toward some but not all substrates. Mol Biol Cell. 21(19):3475-86.

Stephan JS, Yeh YY, Ramachandran V, Deminoff SJ, Herman PK (2009). The Tor and PKA signaling pathways independently target the Atg1/Atg13 protein kinase complex to control autophagy. Proc Natl Acad Sci U S A. 106(40):17049-54.

Stierl M, Stumpf P, Udwari D, Gueta R, Hagedorn R, Losi A, Gärtner W, Petereit L, Efetova M, Tate JJ, Cox KH, Rai R, Cooper TG (2002). Mks1p is required for negative regulation of retrograde gene expression in Saccharomyces cerevisiae but does not affect nitrogen catabolite repression-sensitive gene expression. J Biol Chem. 2002 Jun 7;277(23):20477-82.

Toda T, Uno I, Ishikawa T, Powers S, Kataoka T, Broek D, Cameron S, Broach J, Matsumoto K, Wigler M (1985). In yeast, RAS proteins are controlling elements of adenylate cyclase. Cell. 40(1):27-36.

The End

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity. I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Jarod Grant-Baker NOV. 29TH 2012

Author Signature Date

(This page must be signed and dated by the author