Distribution and regulation of stochasticity and plasticity in *Saccharomyces cerevisiae*

R. D. Dar[1,2], D. K. Karig[1], J. F. Cooke[2], C. D. Cox[3,5], and M. L. Simpson[1,4,5,*]

[1]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Bethel Valley Road, Oak Ridge, TN 37831, USA

Departments of [2]Physics and Astronomy, [3]Civil and Environmental Engineering, and [4]Materials Science and Engineering, and the [5]Center for Environmental Biotechnology, University of Tennessee, Knoxville, TN, 37996-2010, USA

*corresponding author: simpsonml1@ornl.gov

Keywords: transcriptional plasticity, gene expression noise, gene circuit fluctuation analysis, regulatory architecture, stress response, SACCHAROMYCES CEREVISIAE

Abbreviations: transcriptional start sites (TSS), environmental stress response (ESR)

Abstract: 197 words

Lead Paragraph: 431 words

Word Count (rest of text): 3525 words

42 References

8 Figures

Abstract

Stochasticity is an inherent feature of complex systems with nanoscale structure. In such systems information is represented by small collections of elements (e.g. a few electrons on a quantum dot), and small variations in the populations of these elements may lead to big uncertainties in the information. Unfortunately, little is known about how to work within this inherently noisy environment to design robust functionality into complex nanoscale systems. Here, we look to the biological cell as an intriguing model system where evolution has mediated the trade-offs between fluctuations and function, and in particular we look at the relationships and trade-offs between stochastic and deterministic responses in the gene expression of budding yeast (*Saccharomyces cerevisiae*). We find gene regulatory arrangements that control the stochastic and deterministic components of expression, and show that genes that have evolved to respond to stimuli (stress) in the most strongly deterministic way exhibit the most noise in the absence of the stimuli. We show that this relationship is consistent with a bursty 2-state model of gene expression, and demonstrate that this regulatory motif generates the most uncertainty in gene expression when there is the greatest uncertainty in the optimal level of gene expression.

Nanoscale science, engineering, and technology aspire to usher in a new level of technology where engineered systems embody complex functionality such as self-organization, self-repair, and adaptability. The research community has aggressively pursued the science to meet these expectations, with the most significant results seen in the synthesis and characterization of nanoscale materials that exhibit extraordinary physical, chemical, or biological functionality controlled by their nanoscale structure. Yet, the highest levels of functionality are envisioned from the collective behaviors of interacting nanoscale components, and the question shifts to how these materials should be organized into ensembles that exhibit the highest levels of functionality. The best examples of high levels of functionality emerging from ensembles of nanoscale elements are found in biological cells where complex functionality such as sensing, communication, navigation, and cooperation, has evolved. As we look to cells to learn design principles that apply at such high levels of scale and complexity, we see striking differences between natural and synthetic systems. For example, most often in engineered systems fluctuations (or noise) are minimized to the point of insignificance, yet these fluctuations are an inherent and inescapable feature in cells. A typical cell may contain 50M molecules of a few thousand different types, and the fluctuations in any one molecular population could vary from negligible to dominant. The noise in any one molecular species could be made negligible just by the selection of a large population for that species, but only at the cost of moving this noise to other molecular species. That is, there is a loose sort of conservation of fluctuations that leaves the distribution of this conserved stochasticity across the functional elements as the choice left to the system. At present there are no design rules to guide how such a distribution should be made in engineered systems, so here we look to the biological cell as an intriguing model

system where evolution has mediated the trade-offs between fluctuations and function. We focus on the relationships and trade-offs between stochastic and deterministic responses in the gene expression of budding yeast (*Saccharomyces cerevisiae*). We find gene regulatory arrangements that control the stochastic and deterministic components of expression, and show that genes that have evolved to respond to stimuli (stress) in the most strongly deterministic way exhibit the most noise in the absence of the stimuli. This study provides another example of the exciting era that biological sciences have entered where extensive experimental databases enable the mapping of complex system principles, which may be the only approach to understanding the rules that enable a new design science for synthetic complex nanoscale systems.

Introduction

Advances in nanoscale fabrication, assembly and characterization are providing the tools to construct dense, distributed, and highly interconnected information processing systems[1]. However, as these synthetic systems approach such levels of density and complexity, fluctuations become a major challenge to the representation and processing of information. Stochasticity is an inherent feature at the nanoscale as information is represented by small populations of objects. Unfortunately, very little is known about how to design robust function into synthetic, dense, highly interconnected, and highly stochastic systems, and here we look to the biological cell to understand how deterministic and stochastic responses are balanced to achieve complex function.

At the extremes of density and complexity, the biological cell presents an intriguing model system where evolution has mediated the trade-offs between fluctuations and function. For

example, a typical cell may contain 50 million molecules of a few thousand different types, and the fluctuations (noise) in any one molecular population could vary from negligible to dominant. The fluctuations in any one molecular population could be made negligible just by the selection of a large population for that molecular species. However, because of the bounded, crowded environment of the cell, such a selection made in favor of one molecular species comes at the cost of moving this stochasticity to other species. Consequently, the choice made at the system level is the distribution of stochasticity across the functional elements. Although selection processes have optimized the distribution of stochasticity across elements in natural biological systems, at present there are no design strategies to guide how such a distribution should be made to enhance function in engineered systems. Conversely, over the last several years the new field of Noise Biology has started to uncover some of the fundamental principles of noise structure and function in cellular systems with the eventual goal of implementing these design principles in engineered systems.

Noise Biology focuses on the sources, processing, and functional consequences of the inherent stochastic fluctuations in the molecular populations that store and process information in individual cells[2]. In these systems there is competition between the benefits of low noise and those of low population that has produced an uneven distribution of stochasticity across the molecular components of cells. In contrast to stochastic gene expression is phenotypic plasticity, the ability of an organism with a given genotype (genetic sequence) to change its phenotype (represented here as gene expression profile) in response to changes in the environment. Plasticity is a deterministic response to a stimulus, while stochasticity is a random response. In the former, the gene network of the organism has evolved to respond with deterministic and optimized expression levels to a specific perturbation. On the other hand, stochastic gene

expression does not necessarily respond to a given perturbation. It is 'un-programmed' and perhaps provides the mechanism for an organism to hedge its bets against sudden changes in the environment by distributing diverse gene expression profiles across a cellular population. Here we examine the system-wide distribution and structure of noise in *Saccharomyces cerevisiae* (budding yeast) and focus on the relationships between gene circuit plasticity and stochasticity.

Noise in genetic circuits

A simple birth-death process (e.g. constitutive protein expression; Fig. 1a) illustrates the origin of intrinsic noise in genetic circuits. In this example the time evolution of the population of molecule *M*, *M*(*t*), may be modeled using (Fig. 1b)

$$\frac{dM(t)}{dt} = \alpha - \gamma M(t)$$
$$M(t) = \frac{\alpha}{\gamma}\left(1 - e^{-\gamma t}\right) \quad , \tag{1}$$

where α is the average rate of production and γ is the rate constant for decay of molecule *M* (rate of decay = $\gamma M(t)$). However, this continuous representation neglects the random timing and discrete nature (integral numbers of molecules) of molecular transitions (Fig. 1b), while an actual time evolution could follow any one of a number of possible trajectories (Fig. 1c). Here we define the noise in any individual trajectory as the difference between that trajectory and the average of all trajectories (Fig. 1c).

The magnitude of this noise is most often quantified using the square of the coefficient of variation, or CV^2, which is the variance of the noise divided by the square of the population

mean. It can be calculated by examining an ensemble of many trajectories at a single point in time, as given by flow cytometry, or from single trajectories over time, as given by fluorescence microscopy. If we assume that molecules are produced in processes governed by Poisson statistics (an assumption that tends to underestimate the magnitude of the noise) we find that

$$\sigma_M^2 \approx \langle M \rangle$$
$$CV^2 = \frac{\sigma_p^2}{\langle M \rangle^2} \approx \frac{1}{\langle M \rangle}, \quad (2)$$

where σ_M^2 and <*M*> are the variance and mean population of molecule *M*.

A particularly important question in noise biology is the extent to which stochastic fluctuations are used for a functional benefit. There are three possible consequences of noise within a specific gene circuit: (1) noise plays little or no role in function; (2) noise is detrimental to function; or (3) noise is an important functional component. Some studies have supported a noise picture consistent with the first two of these consequences. For example, Fraser *et al.* found that the production of essential and complex-forming proteins involved lower levels of noise than the production of most other genes [3], implying that natural selection has distributed the noise to populations where it will do the least harm while minimizing noise in protein populations where it may be detrimental. Conversely, biological systems have found strategies to make functional use of the noise, including the control of the swimming and tumbling periods of bacteria during chemotaxis [4], stochastically-driven phenotype variability in the response to mating pheromone in yeast [5], and fitness enhancing phenotypic individuality in microbial cultures [6-8].

Along these lines, following from an earlier theoretical study from the same lab [9] Acar *et al.* used an engineered *S. cerevisiae* strain that stochastically switched between two environmental growth phenotypes and revealed a fitness advantage when stochastic properties were matched between phenotypic stochastic switching and environmental fluctuations [10]. This result suggests that in nature, cells may tune the stochastic fluctuations in gene expression to match the stochasticity of environmental variations, and that under the right conditions a population derives an advantage by stochastic phenotype switching compared to the energy and resources required to sense and respond when challenged by a varying environment [6,9]. That is, there is a trade-off between the costs and benefits of stochastic and deterministic strategies for responding to fluctuating environments. Understanding these trade-offs requires an examination of the system-wide distribution of stochasticity and plasticity.

System-wide noise

Over the last several years there has been a significant body of work focused on quantifying stochastic fluctuations in gene expression, including measurements with bacterial [11-16], yeast [17-19], and mammalian [20-22] cells. While much of this work has focused on individual or small networks of synthetic [11-13,15-19] or natural [23] gene circuits, some studies have started to look at the distribution of noise across larger systems of gene circuits [24-26]. One of the most extensive genome-wide noise magnitude studies to date was reported by Newman *et al*. in 2006 [25]. Using both rich (YPD) and minimal (SD+) growth media, Newman *et al*. measured the noise in the populations of more than 2000 proteins in yeast using flow cytometry and found that most proteins displayed the inverse relationship between protein abundance and noise described in equation (2).

For our study of the relationship between noise and plasticity we focus on proteins that exhibit noise that exceeds the level predicted by the protein abundance. Newman *et al*. quantified this 'excess' noise as the difference (*DM*) between the measured noise and the noise expected at that protein abundance. In our current study we use the minimal medium (SD) *DM* measurements reported by Newman *et al*.

System-wide plasticity

While stochasticity is defined as variation that occurs irrespective of the presence of a stimulus, plasticity is defined with respect to variation in gene expression that occurs in concert with stimuli. We begin by considering a system consisting of genes i=1, 2….I whose transcription level (as measured by mRNA abundance) is determined in environments j=1, 2…..J. We define the *relative transcriptional response* of gene i to environment j, e_{ij}, as

$$e_{ij} \equiv \log_2\left(\frac{m_{ij}}{m_{i0}}\right) \tag{3}$$

where m_{i0} is the mRNA level in some reference (unstressed) environment. Using a definition similar to Promislow's 2005 phenotypic plasticity study [27] we define the *average transcriptional response*, E_i, and *environmental transcriptional plasticity*, Pl_i, of gene i, as follows:

$$E_i \equiv \frac{\sum_j e_{ij}}{J} \tag{4}$$

and,

$$Pl_i \equiv \log_{10}\left(\sigma^2_{e_{ij}}\right) \tag{5}$$

where $\sigma^2_{e_{ij}}$ is the variance of e_{ij} across the J environments. Pl_i then is a measure of how much gene i changes expression in response to environmental variation. Low plasticity implies that expression either remains fairly constant regardless of environmental condition, or that any significant changes in expression occur in very few different environments. Conversely, high plasticity suggests significant expression changes in response to many different environments.

For our study, we extract measures of plasticity from the report of Gasch *et al.* that describes an extensive genome-wide stress response microarray study in yeast [28]. The study employed 13 stressors that included heat shock, hydrogen peroxide, amino acid starvation, and nitrogen depletion. All of the stressors were applied at various strengths for a total of 173 measured environments. Each environment had a separate microarray containing the ~6200 genes of the yeast genome. In brief, when the budding yeast is subjected to stress it stops growth by repressing ~600 genes and reallocates its expression capacity to the induction of ~300 stress response genes. This large-scale switching between the growth and stress states was coined the environmental stress response (ESR) and is executed by highly coordinated gene regulation. Expression from non-ESR yeast genes clustered in the region defined by near zero average transcriptional response and low plasticity (~5k genes, Fig. 2a), i.e. the region where expression is little changed by environmental stimuli. The minority of genes with the highest levels of plasticity (the ~900 ESR genes) are most often found associated with either highest or lowest average transcriptional responses (Fig. 2b).

Regulatory arrangements that control noise and plasticity

To begin exploring the relationships between stochasticity and plasticity, we first consider three different gene regulatory arrangements that control expression. First we consider the

presence or absence of a TATA box (5'-TATAAA-3' or other variant), which is a core promoter motif present in ~20% of all *S. cerevisiae* genes [29] and in ~24% of human genes [30]. The TATA box is a competitive binding site for histones and transcription factors and is generally associated with greater expression noise [8,25]. Here we rely on the database provided by Basehoar *et al.* [29] to determine which yeast genes contain TATA boxes.

Next we consider the co-activation complexes involved in the coordinated regulation of the yeast stress response versus growth and housekeeping genes. TFIID is known to regulate ~90% of the measureable genome while the SAGA complex only regulates ~10% of the measurable genome [31]. Here we rely on the data reported by Huisinga *et al.* [31] to identify the co-activation complexes responsible for regulating the expression for each of the genes considered in our study.

Finally, we consider the nucleosome occupancy pattern near the transcriptional start sites (TSS) of the genes considered here. In our study we rely on the high resolution atlas of yeast nucleosome occupancy patterns reported by Lee *et al*. This study covered over 80% of the yeast genome [32] and characterized nucleosome occupancy patterns at +/- 400 bp from the TSS of each gene. We will use the four occupancy motifs (identified here as clusters 1-4; Fig. 3) Lee *et al.* generated with *k*-means clustering using the Euclidean distance metric.

Results

We used the first two regulatory elements described above to segregate the yeast genes into 4 ***main categories***: TATA-SAGA; TATAless-SAGA; TATA-TFIID; and TATAless-TFIID. Each of these were further divided into 1 of 4 ***sub-categories*** (nucleosome occupancy pattern

clusters) such that there were 16 distinct grouping of genes. An average DM and plasticity were calculated for genes where both their TATA/TATAless and SAGA/TFIID architecture were reported in the datasets. These averages did not include any genes for which the architecture was unknown or ambiguous. The averages were also calculated separately for each of the 4 nucleosome occupancy pattern clusters.

Excess noise was strongly dependent on the gene main category with the average DM of TATA-SAGA an order of magnitude greater than that for TATAless-TFIID (Fig. 4, upper). This result is not surprising as TATA box genes tend to have higher noise levels due to transcriptional bursting [8,25]. TATA-containing promoters are often associated with stress-response functions [29], and it may be that the TATA-generated DM provides a noise-mediated benefit for the response to acute environmental stress[8]. Conversely, TATA-less promoters are often associated with housekeeping functions [29] where noise may be detrimental or have little impact on function. The SAGA or TFIID classification had an even greater effect on DM than TATA, with TATAless-SAGA having higher noise than TATA-TFIID.

For the case of nucleosome occupancy clusters, cluster #1 had ~4x the excess noise as compared to clusters #3 & #4 which had similar low levels of excess noise. Using gene ontology (GO) analysis, Lee *et al* reported an enrichment in stress response genes in cluster #1 [32] consistent with the high excess noise in TATA architectures (Fig.4 upper).

Intriguingly, high noise architectures were also high plasticity architectures, as plasticity follows exactly the same pattern as DM (Fig. 4). Accordingly, there is a strong positive correlation between DM and plasticity, consistent with additional findings in recent studies [33-34]. With respect to nucleosome occupancy cluster and the 16 distinct architecture subgroups (Fig.

5), DM and plasticity followed a similar pattern for all main categories except TATAless-TFIID, with clusters 1 and 2 associated with higher DM and plasticity than clusters 3 and 4 (Fig. 5). The TATAless-TFIID genes had very low DMs that had little or no correlation with plasticity.

Discussion

On first inspection, the strong positive correlation between DM and plasticity is surprising as one might have expected a large degree of plasticity (expression varying in a deterministic way in response to environmental signals) to instead be associated with low levels of random variability (Fig. 6). This expectation would be consistent with a hypothesis where noise is used as a bet hedging strategy when the optimal expression level is unknown. However, closer inspection indicates that our results and this bet hedging hypothesis are not at odds. The TATA-SAGA genes that have the highest DM and plasticity are often associated with stress-response functions [29], and their plasticity implies that an optimal expression level is known, but only for stressful environments. In the non-stressed environment where DM was measured, the optimal expression level for stress genes is unknown, response attempts to anticipate a randomly timed environmental stress, and the high level of noise is coincident with a high level of uncertainty in the proper gene expression level.

The relationships between DM and plasticity presented here are consistent with a gene expression model dominated by 2-state transcriptional bursting [18,35-37]. Transcriptional bursting is a model of gene expression where the expression rate is controlled by switching between discrete high and low transcriptional rates. The average rate is determined by the fractional amount of time spent in each of the two states. This model is consistent with transcription controlled through protein–DNA interactions at an operator site within the gene promoter region or with the bursty expression of genes compacted in chromatin. To illustrate the model we present equations

adapted from an earlier analysis [36] with the simplifying assumptions that the low expression rate is 0, and that other than burst dynamics there is only one dominant time constant (assumed here to be protein decay time) represented by the rate constant γ_p.

Transcriptional bursting may be represented using three model parameters (Fig. 7): (1) the transcription rate in the high expression state, α; (2) the fraction of time spent in the high expression state, *O*, which we will also refer to as the 'on fraction'; and (3) the kinetics of the switching between off and on expression states, which is represented by *k* (referred to here as the burst kinetic rate), the sum of k_{ON} and k_{OFF} (see Fig. 7). As previously shown [38], with these assumptions, the autocorrelation function of the noise, $\Phi(\tau)$, is

$$\Phi(\tau) \approx \frac{\alpha O b}{\gamma_p} b e^{(-\gamma_p \tau)} + \left(\frac{\alpha O b}{\gamma_p}\right)^2 \frac{(1-O)}{Ok}\left(\frac{\gamma_p}{\left[1-\left(\frac{\gamma_p}{k}\right)^2\right]} e^{(-\gamma_p \tau)} + \frac{k}{\left[1-\left(\frac{k}{\gamma_p}\right)^2\right]} e^{-k\tau}\right), \qquad (6)$$

where *b* is the translational burst rate (average number of proteins translated from each mRNA). The average protein population, <*p*>, is

$$\langle p \rangle = \frac{\alpha O b}{\gamma_p}, \qquad (7)$$

and

$$CV = \sqrt{\frac{\Phi(0)}{\langle p \rangle^2}} \approx \sqrt{\frac{1}{\langle p \rangle} b + C_k \frac{(1-O)}{O}}. \qquad (8)$$

where C_k approaches 0 for fast bursting ($k >> \gamma_p$) and approaches 1 for slow bursting ($\gamma_p >> k$). The first term under the radical on the right, referred to as the shot-noise term[36], is most prominent at low protein population (<*p*>), high on fraction (*O*), or fast burst dynamics (low C_k). Conversely, the second term on the right, referred to as burst noise, is most prominent for high <*p*>, low O, or slow burst dynamics ($C_k \rightarrow 1$). For model clarity, exact stochastic simulation utilizing the Gillespie algorithm[39] was performed to generate time-dependant gene expression trajectories for constitutive gene expression where $O \rightarrow 1$ (shot noise dominated), and noise magnitude is low (CV = 0.1, Fig. 8a), and transcriptionally bursty gene expression where O = 0.25 (burst noise dominated), and noise magnitude is high (CV = 1.14, Fig. 8b). Transcription rate (α) was adjusted to get the same mean protein abundance for both simulations, even though their stochastic behaviors are drastically different (Fig. 8). The excess noise, *DM*, is generated by the burst noise and can be approximated as

$$DM \approx \sqrt{\frac{1}{\langle p \rangle} b + C_k \frac{(1-O)}{O}} - \sqrt{\frac{1}{\langle p \rangle} b}, \tag{9}$$

If the shot noise term is dominant, then $DM \rightarrow 0$. Conversely, if the burst noise term is dominant the maximum excess noise is

$$DM_{\max} \approx \sqrt{C_k \frac{(1-O)}{O}} \tag{10}$$

Plasticity is a measure of the variability of the protein population in response to environmental signals. The protein population,

$$\langle p \rangle = C \alpha O \tag{11}$$

(we use the constant C to represent b/γ_p that we assume does not vary) can be controlled through variation of O, α, or both, such that plasticity (Pl)

$$Pl \propto \langle p \rangle_{\max} - \langle p \rangle_{\min} = C\left((\alpha O)_{\max} - (\alpha O)_{\min}\right) \tag{12}$$

If only α is responsive to stress

$$Pl \propto CO\left(\alpha_{\max} - \alpha_{\min}\right) \tag{13}$$

and

$$DM_{\max} \approx \sqrt{C_k \frac{(1-O)}{O}} = \text{constant for all <p>} \tag{14}$$

For large plasticity, O→1, in which case

$$DM_{\max} \approx \sqrt{C_k \frac{(1-O)}{O}} \to 0 \tag{15}$$

In this case, high plasticity is found in combination with low excess noise.

Alternatively, if only O is responsive to stress

$$Pl \propto C\alpha\left(O_{\max} - O_{\min}\right) \tag{16}$$

and high plasticity is gained for low $O_{\min}$ and high $O_{\max}$, such that high excess noise

$$DM_{\max} \approx \sqrt{C_k \frac{(1-O_{\min})}{O_{\min}}} \tag{17}$$

is found in concert with high plasticity in the non-stresses environment. It is also important to note that

$$DM_{\min} \approx \sqrt{C_k \frac{(1-O_{\max})}{O_{\max}}} \to 0 \ , \tag{18}$$

indicating that the excess noise is much reduced when the gene is expressed in the stressed environment where the optimal expression level is more certain.

Our results suggest that plasticity and stochasticity are not only actively regulated in complex genetic systems, but are also co-regulated. This co-regulation is correlated with specific regulatory architectures, which likely control both through the manipulation of transcriptional bursting. At least for the stress genes it appears that the expression level is varied by controlling the on fraction of the transcriptional bursts. This arrangement provides for the most noise to be exhibited when the optimal expression level is the most uncertain, and may represent a bet hedging strategy where noise in expression is used to anticipate future stresses. Alternative systems biology methodologies for studying gene expression relationships and gene network inferences exist. Here a combined analysis of dynamic time-dependant gene expression signatures are mapped onto the genetic network [40]. Graph theory methods are then used to find significant subgraphs of well connected genetic activity and architectures [41-42]. The time-dependant kinetic analysis of stochastic and plastic gene expression is of great interest and would yield important insights as the biological system is intrinsically dynamic with transient regulation and expression patterns and environmental conditions are time-dependant as well. In the current study we confine ourselves to a steady-state gene expression picture as currently no relevant genome-wide time-dependant gene expression databases in both healthy and stressful conditions exist for the budding yeast.

On the whole the heavily studied *S. cerevisiae* enabled us to explore genome-wide plastic and stochastic gene expression relationships in a complex and highly dimensional system. We learned that distinct regulatory architectures occupy and function in specified regions of the stochastic versus plastic gene expression space. We observed that stochasticity and plasticity are positively correlated, and specifically that stress genes with TATA-SAGA regulatory arrangements display the most noise in unstressed environmental conditions. This behavior is

consistent with a 2-state transcription burst model where the plasticity is achieved through modulation of the burst duration (on fraction), a condition that generates the greatest excess noise when the gene is expressed at its lowest level.

This study provides another example of the exciting era that biological sciences have entered. Here we have elucidated organizational principles relating two complex system traits of stochasticity and plasticity. We utilized a total of 5 genome-wide yeast studies; two gene expression studies to quantify stochasticity and plasticity; and three architecture studies for the presence of TATA boxes, co-activation complexes, and nucleosome occupancy patterns. This empirically driven approach to mapping out complex system principles may possibly be the only solution to mapping the system's complexity when dimensionality is too high to apply first principles reasoning. Furthermore, as nanoscience attempts to create *de novo* complex nanoscale systems, these studies of biological systems may be the only guides to aid in the development of a new design science.

Acknowledgements

We gratefully acknowledge fruitful discussions with our colleagues L. S. Weinberger, J. M. McCollum, and G. S. Sayler. This research was supported by the in-house research program of the Center for Nanophase Materials Sciences, which is sponsored at Oak Ridge National Laboratory by the Scientific User Facilities Division, U.S. Department of Energy.

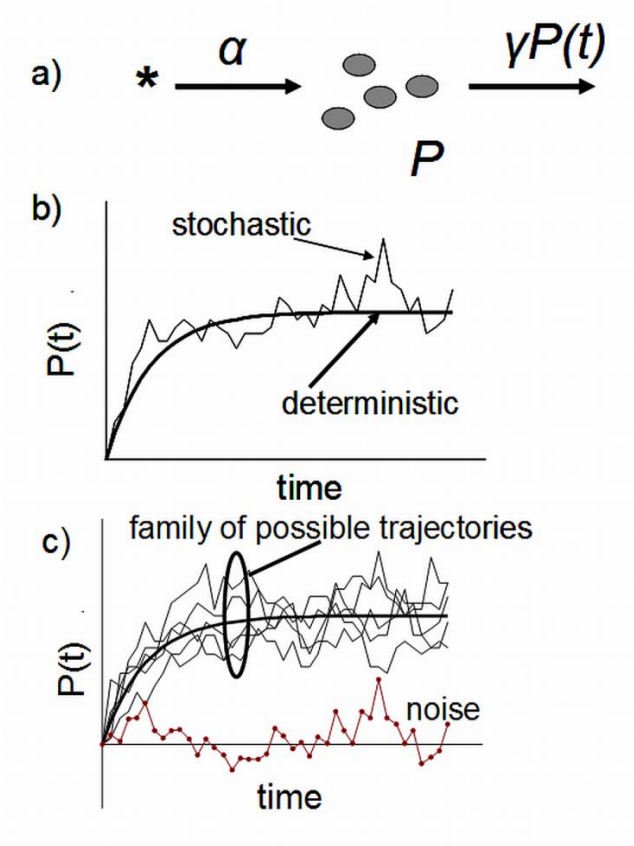
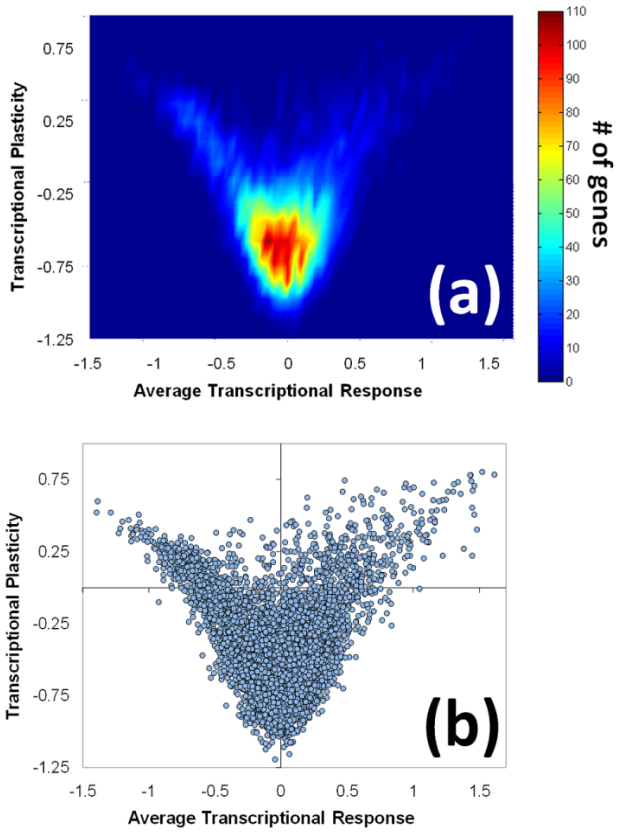
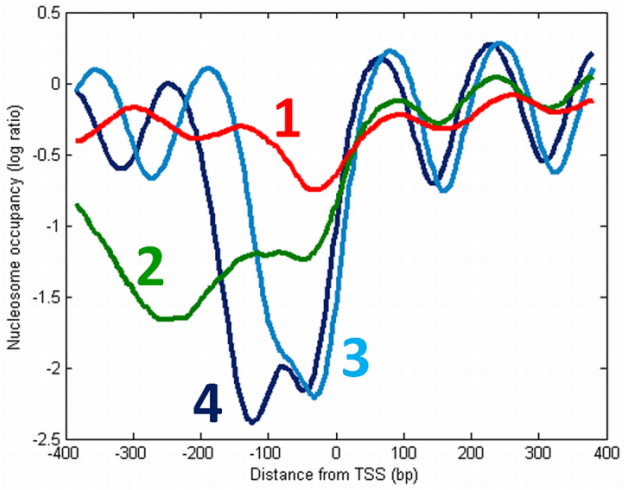


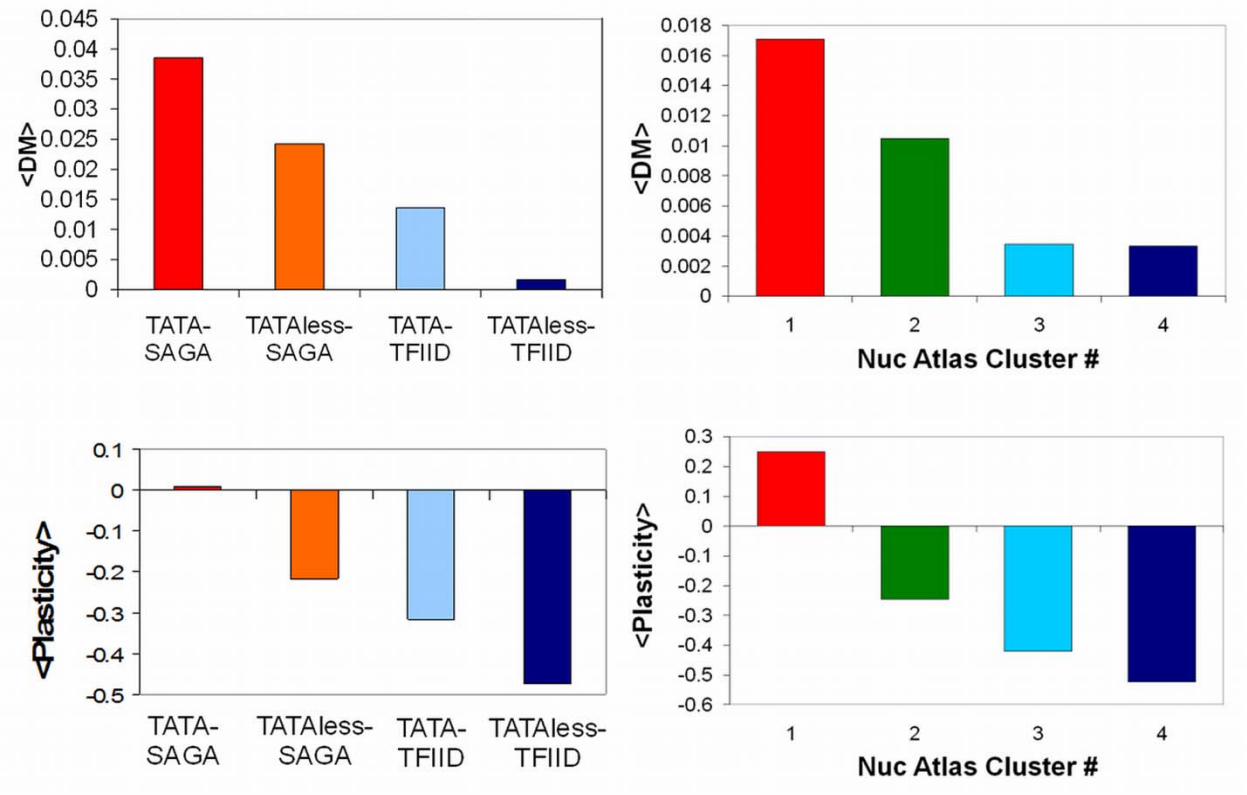
Figure 1: Noise in molecular populations. **(a)** A simple birth-death molecular process and **(b)** the deterministic and stochastic rise to steady-state molecular population level (α/γ). **(c)** A family of possible stochastic trajectories for the birth-death process. The smooth curve represents the average of all possible stochastic trajectories. The noise for any of the possible trajectories is found from the difference between the trajectory and the average of all trajectories.



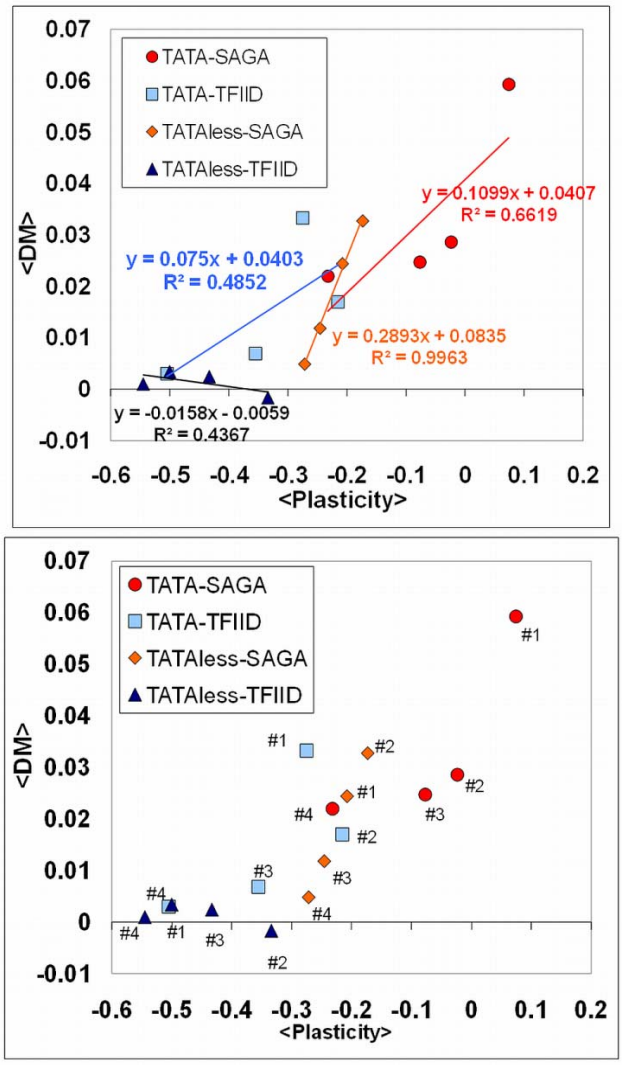
<u>Figure 2:</u> Dependence of environmental transcriptional plasticity on average transcriptional response across the yeast genome. (a) Heat map of transcriptional plasticity space. ~5k genes (5/6ths of the yeast genome) are transcriptionally invariant under stressful conditions and located at low transcriptional plasticity and close to zero transcriptional response. **(b)** Scatter plot reveals that genes with high plasticity are associated with larger response.



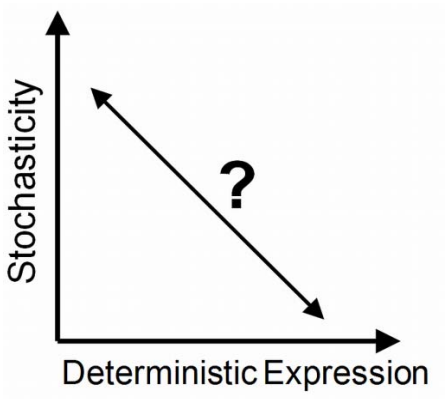
<u>Figure 3:</u> Primary nucleosome occupancy patterns reported by Lee et al [32]. Clusters 1-4 are labeled next to their respective curve in the plot and were calculated using a K-means clustering algorithm. Of note, at least half the genome has patterns 3 and 4 which have a vacant region just upstream of the transcription start site (TSS). Reproduced with permission from Lee *et al* [32].



<u>Figure 4:</u> Excess noise and plasticity are related and strongly dependent on gene regulatory architecture. (upper) Mean excess noise for the 4 main regulatory categories (TATA/TATAless, SAGA/TFIID) and 4 main nucleosome occupancy patterns. **(lower)** Mean plasticity for the 4 main regulatory categories and 4 main nucleosome occupancy patterns. High plasticity architectures share high excess noise and low plasticity architectures share low excess noise.



<u>Figure 5:</u> Excess noise and plasticity are positively correlated across 16 distinct groupings of genes. (upper) Three of the main gene categories have a strong correlation between stochasticity and plasticity. **(lower)** Labeling of sub-categories including the 4 nucleosome occupancy patterns of figure 3. The occupancy patterns conserve the same order along the linear relationship in each gene category except TATAless-TFIID where the trend is flat and TATAless-SAGA where patterns #1 and #2 are switched.



<u>Figure 6:</u> Expected inverse relationship between variability and deterministic expression. In this scenario high variability would be counter-productive to strong deterministic gene expression.

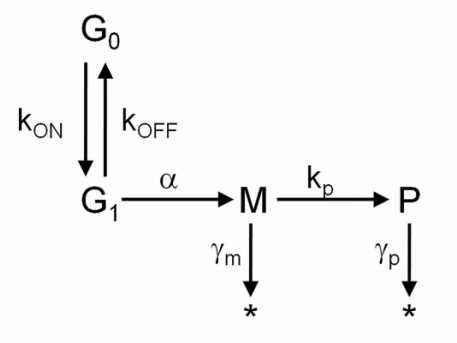
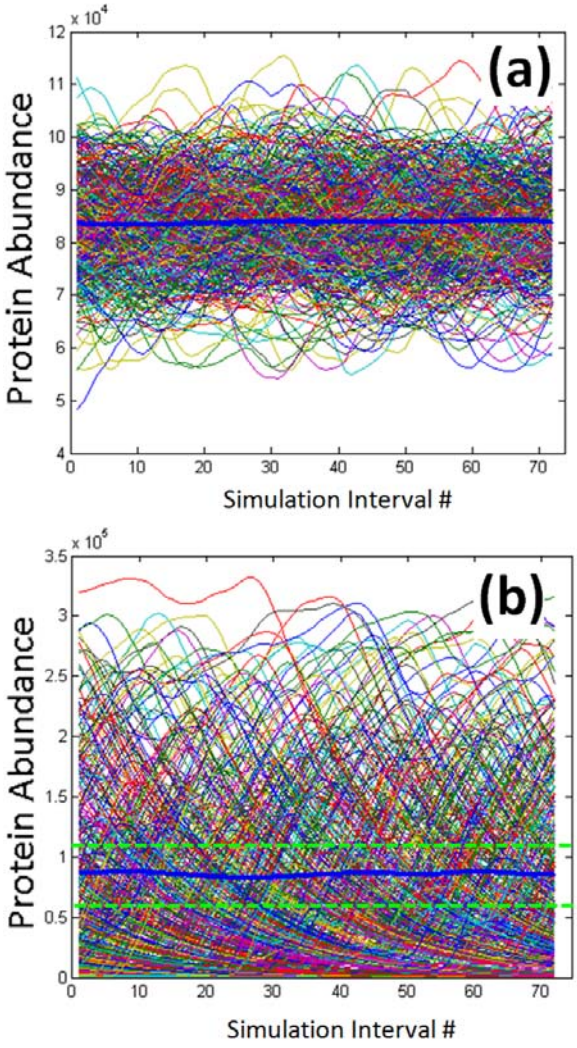


Figure 7: Diagram of the 2-state model. The gene transitions between active (G1; transcription rate =α) and inactive (G0; transcription rate =0) states. The fraction of time spent in state G1, *O*, is given by $O = \frac{k_{on}}{k_{on} + k_{off}}$.



<u>Figure 8:</u> Simulated gene expression trajectories of the 2-state model. (**a**) Constitutive gene expression with O → 1, i.e. gene constantly expressed in the elevated transcription state resulting in relatively low noise magnitude (CV = 0.1). (**b**) Transcriptionally bursty gene expression with O = 0.25 and relatively high noise magnitude (CV = 1.14). Green dashed boundaries represent the abundance range from (a) superimposed on (b). For each simulation hundreds of different expression trajectories were simulated for 12 hours. Transcription rate (α) was adjusted to yield the same average protein abundance in both simulations (bold blue line at ~83k proteins). For the simulation in (a), $\alpha = 0.0101$ sec^{-1} , and for the simulation in (b), $\alpha = 0.0403$ sec^{-1} . The rest of the parameters (other than O) were identical: translation rate = 0.783916 sec^{-1}, mRNA decay rate = 0.00078 sec^{-1}, protein decay rate = 0.00011 sec^{-1} .

REFERENCES

1. Doktycz, M.J. & Simpson, M.L. Nano-enabled synthetic biology. *Molecular Systems Biology* **3**, - (2007).
2. Simpson, M.L. *et al.* Noise in biological circuits. *Wiley Interdisciplinary Reviews-Nanomedicine and Nanobiotechnology* **1**, 214-225 (2009).
3. Fraser, H.B., Hirsh, A.E., Giaever, G., Kumm, J. & Eisen, M.B. Noise minimization in eukaryotic gene expression. *Plos Biology* **2**, 834-838 (2004).
4. Spudich, J.L. & Koshland, D.E. Non-Genetic Individuality - Chance in Single Cell. *Nature* **262**, 467-471 (1976).
5. Colman-Lerner, A. *et al.* Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437**, 699-706 (2005).
6. Kussell, E. & Leibler, S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075-2078 (2005).
7. Avery, S.V. Microbial cell individuality and the underlying sources of heterogeneity. *Nature Reviews Microbiology* **4**, 577-587 (2006).
8. Blake, W.J. *et al.* Phenotypic consequences of promoter-mediated transcriptional noise. *Molecular Cell* **24**, 853-865 (2006).
9. Thattai, M. & van Oudenaarden, A. Stochastic gene expression in fluctuating environments. *Genetics* **167**, 523-530 (2004).
10. Acar, M., Mettetal, J.T. & van Oudenaarden, A. Stochastic switching as a survival strategy in fluctuating environments. *Nature Genetics* **40**, 471-475 (2008).
11. Austin, D.W. *et al.* Gene network shaping of inherent noise spectra. *Nature* **439**, 608-611 (2006).
12. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590-593 (2000).
13. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. Stochastic gene expression in a single cell. *Science* **297**, 1183-1186 (2002).
14. Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genetics* **31**, 69-73 (2002).
15. Pedraza, J.M. & van Oudenaarden, A. Noise Propagation in Gene Networks. *Science* **307**, 1965-1969 (2005).
16. Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S. & Elowitz, M.B. Gene Regulation at the Single-Cell Level. *Science* **307**, 1962-1965 (2005).
17. Blake, W.J., Kaern, M., Cantor, C.R. & Collins, J.J. Noise in eukaryotic gene expression. *Nature* **422**, 633-637 (2003).
18. Raser, J.M. & O'Shea, E.K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811-1814 (2004).
19. Volfson, D. *et al.* Origins of extrinsic variability in eukaryotic gene expression. *Nature* **439**, 861-864 (2006).
20. Sigal, A. *et al.* Dynamic proteomics in individual human cells uncovers widespread cell-cycle dependence of nuclear proteins. *Nature Methods* **3**, 525-531 (2006).
21. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444**, 643-646 (2006).
22. Weinberger, L.S., Dar, R.D. & Simpson, M.L. Transient-mediated fate determination in a transcriptional circuit of HIV. *Nat Genet* **40**, 466-70 (2008).

23. Suel, G.M., Kulkarni, R.P., Dworkin, J., Garcia-Ojalvo, J. & Elowitz, M.B. Tunability and noise dependence in differentiation dynamics. *Science* **315**, 1716-1719 (2007).
24. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nat Genet* **38**, 636-643 (2006).
25. Newman, J.R.S. *et al.* Single-cell proteomic analysis of S-cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840-846 (2006).
26. Stern, S., Dror, T., Stolovicki, E., Brenner, N. & Braun, E. Genome-wide transcriptional plasticity underlies cellular adaptation to novel challenge. *Mol Syst Biol* **3**(2007).
27. Promislow, D. A regulatory network analysis of phenotypic plasticity in yeast. *American Naturalist* **165**, 515-523 (2005).
28. Gasch, A.P. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell* **11**, 4241-4257 (2000).
29. Basehoar, A.D., Zanton, S.J. & Pugh, B.F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699-709 (2004).
30. Yang, C.H., Bolotin, E., Jiang, T., Sladek, F.M. & Martinez, E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **389**, 52-65 (2007).
31. Huisinga, K.L. & Pugh, B.F. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. *Molecular Cell* **13**, 573-585 (2004).
32. Lee, W. *et al.* A high- resolution atlas of nucleosome occupancy in yeast. *Nature Genetics* **39**, 1235-1244 (2007).
33. Landry, C.R., Lemos, B., Rifkin, S.A., Dickinson, W.J. & Hartl, D.L. Genetic properties influencing the evolvability of gene expression. *Science* **317**, 118-121 (2007).
34. Tirosh, I. & Barkai, N. Two strategies for gene regulation by promoter nucleosomes. *Genome Research* **18**, 1084-1091 (2008).
35. Kepler, T.B. & Elston, T.C. Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophysical Journal* **81**, 3116-3136 (2001).
36. Simpson, M.L., Cox, C.D. & Sayler, G.S. Frequency domain chemical Langevin analysis of stochasticity in gene transcriptional regulation. *Journal of Theoretical Biology* **229**, 383-394 (2004).
37. Singh, A., Razooky, B., Cox, C.D., Simpson, M.L. & Weinberger, L.S. Transcriptional Bursting from the HIV-1 Promoter Is a Significant Source of Stochastic Noise in HIV-1 Gene Expression. *Biophysical Journal* **98**, L32-L34 (2010).
38. Cox, C.D., McCollum, J.M., Allen, M.S., Dar, R.D. & Simpson, M.L. Using noise to probe and characterize gene circuits. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 10809-10814 (2008).
39. Gillespie, D.T. Exact Stochastic Simulation of Coupled Chemical-Reactions. *Journal of Physical Chemistry* **81**, 2340-2361 (1977).
40. Albert, R. Network inference, analysis, and modeling in systems biology. *Plant Cell* **19**, 3327-3338 (2007).
41. Zivkovic, J., Tadic, B., Wick, N. & Thurner, S. Statistical indicators of collective behavior and functional clusters in gene networks of yeast. *European Physical Journal B* **50**, 255-258 (2006).
42. Zivkovic, J., Mitrovic, M. & Tadic, B. Correlation Patterns in Gene Expressions along the Cell Cycle of Yeast. *Complex Networks* **207**, 23-34 225 (2009).