*Chapter 2*

# EFFECTS OF CIS-REGULATORY MUTATIONS ON NOISE

## Background

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

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

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

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

**Motivation**

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

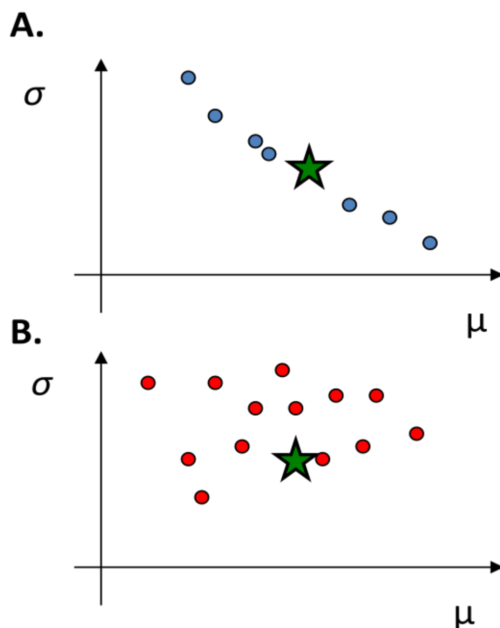


***Fig. 1.*** **Potential Mutagenesis Results.** A) If all mutants in the promoter mutant library exhibit a scaling effect between μ and σ, that means mutations affect both parameters in a correlated fashion. However, if the mutants are uncorrelated as depicted in B) then the parameters are uncorrelated.

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

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

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

**Methods/Results**

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

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

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

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

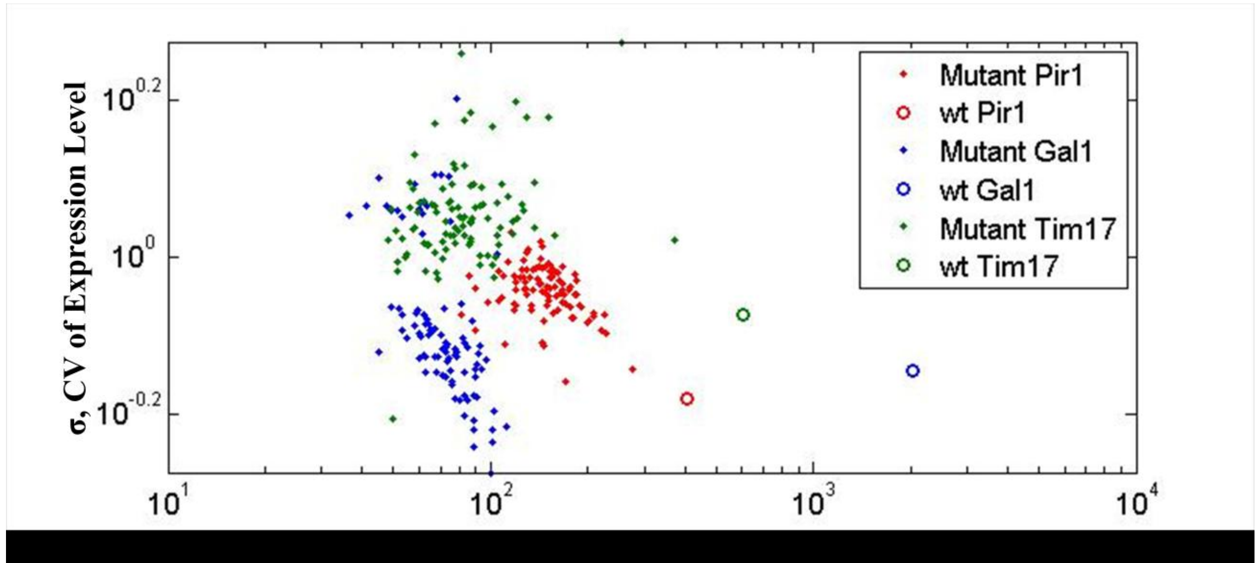


***Fig. 2.*** **Promoter Mutagenesis Results.** After 24 hours of growth in a mutagenesis cell line, three promoter mutant libraries show some intriguing results. Mutants of Pir1 seem to scale while mutants of Tim17 are fairly uncorrelated. Meanwhile, Gal1 mutants fall into 2 categories, those that are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

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

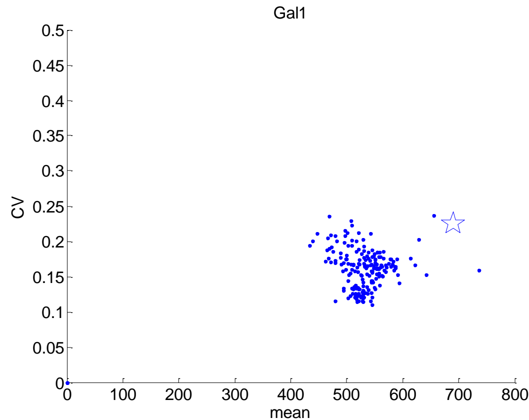


***Fig. 3.*** **Mild Promoter Mutagenesis Results.** After 12 hours of growth in a mutagenesis cell line, the Gal1 mutant promoters (blue dots) are remarkably similar to that of the wild-type promoter (blue star). In fact, the figure is displayed on a linear plot to accentuate the minor differences.

## Discussion and Future Experiments

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

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

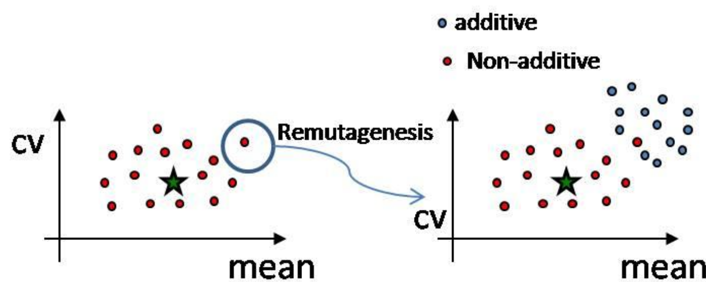


***Fig. 4.*** **Epistatic Affect of Mutations.** From our initial dataset we can pick a mutant clone with high mean and/or high noise and subject it to a second round of mutangenesis. If the mutations are additive, the mutant clone will continue increasing in mean and/ or noise (blue). If they are not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

To answer this fundamental question, we can select individual mutants with high expression level or high noise from the mutant libraries. We will subject these individual clones to additional rounds of random mutagenesis (Fig. 4). Alternatively, we will use site-directed mutagenesis to combine specific mutations with known effects (as determined from the dataset in Fig. 2). The resulting clones will be analyzed by flow cytometry so that epistasis in both mean expression level and noise phenotypes can be determined

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

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

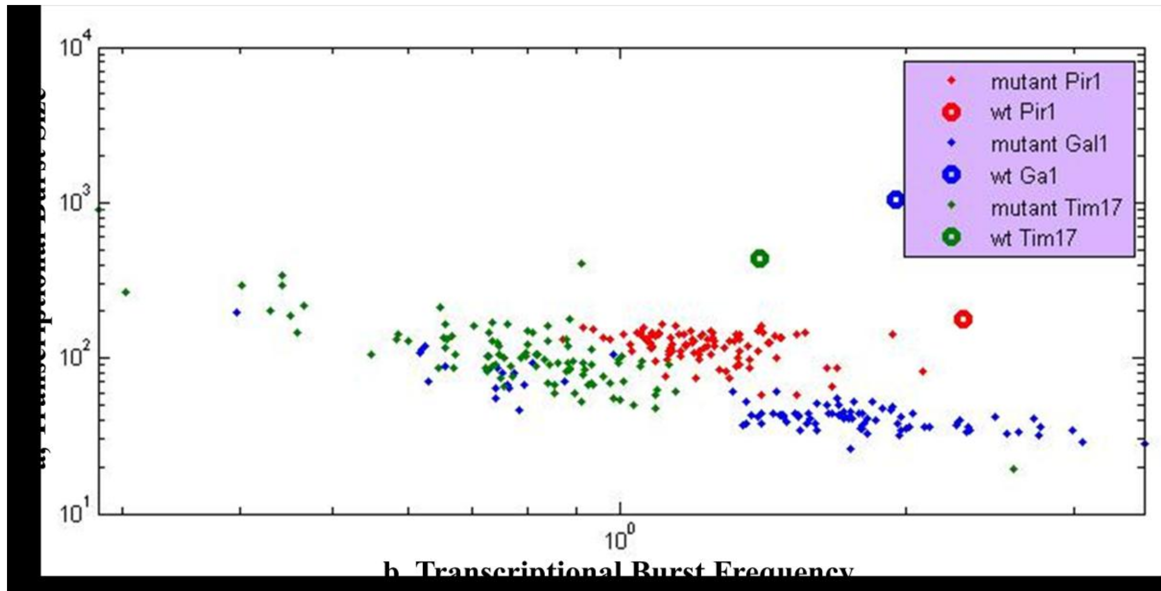


***Fig. 5.*** **Transformed Promoter Mutagenesis Results.** After 24 hours of growth in a mutagenesis cell line, the promoter mutant library data shows interesting mechanistic results. Mutants of Pir1 have a lower burst frequency than wild-type promoters which most Gal1 mutants have a burst size smaller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

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

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

**Acknowledgements**

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