Data and Analysis

William and Mary iGEM 2015

1 Introduction

1.1 Mathematical Representations of Noise

A biological process can be mathematically represented as a time-dependent random variable which denotes the concentration (or copy number) of an output molecule. Such a random variable has a mean level of expression, and introducing stochastic effects causes the variable to fluctuate around its mean in accordance with its underlying probability distribution. Within this framework, a variety of mathematical representations for the noise of the distribution abound in the literature. An important distinction arises between the Fano Factor, σ^2/μ, (where σ^2 denotes the variance and μ denotes the mean of the process), and the squared Coefficient of Variation σ^2/μ^2. While in our data analysis we will be following Elowitz *et al*'s choice of using the squared Coefficient of Variation to represent noise, we could also have analyzed noise strength (or the fold change of the signal's noise relative to the noise of a Poisson process with the same mean) by using the Fano Factor [**?**][**?**].

The noise $\eta^2 = \sigma^2/\mu^2$ of a process is a nondimensional quantity, and hence is difficult to interpret intuitively. We caution the reader not to focus on the specific numeric values of noise in our results, but rather to emphasize the relative differences between noise values of different promoters.

In Figure 1 we present a diagram to help develop some intuition for the effect of η^2 on a signal. The three traces are Poisson processes, and one potential interpretation for the signal is to envision the y-axis representing the number of mRNA copies produced within some time interval– in this way the processes could serve as an idealized model of transcription (take note that a Poisson process is too oversimplified to accurately model transcription). Regardless of the interpretation, however, it is clear that an increase in noise causes the signal to exhibit a much wider distribution of values within the observed timeframe, and that it is far more difficult to precisely predict what value the signal might take at a given time.

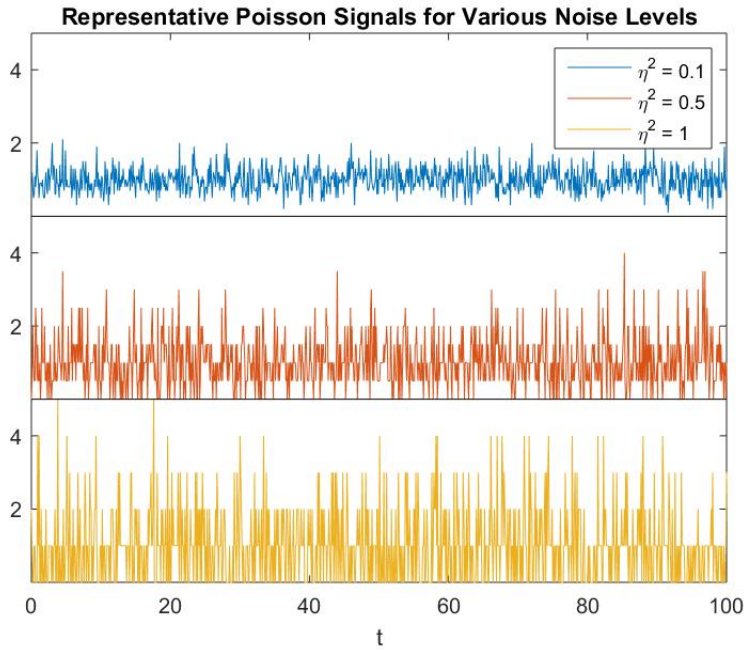

Figure 1: Representative time traces of a random signal $Y \sim \text{Poiss}(\lambda)$, normalized to its mean for various values of $\eta^2 = 1/\lambda$. y-axis represents arbitrary units.

1.2 The Dual Reporter System

In his 2002 study, Michael Elowtiz and colleagues introduced a revolutionary new concept that allows one to decompose the observed noise in a signal into its intrinsic and extrinsic components [**?**]. By placing two distinguishable copies of an identical reporter into the system, one can assume that the underlying probability distribution for the signals generated by the reporters will be the same. Under this assumption, measuring the differences between these signals will provide information about the spread of the distribution itself– this is the intrinsic noise of the process. All other noise effects present in the signal would arise from extrinsic factors. A good analogy for intrinsic noise is the correlation between two signals– in the example of transcription, if a promoter has high intrinsic noise, then the signals between two fluorescent proteins under the same reporter should be less correlated than if a promoter has low intrinsic noise (Figure 2).

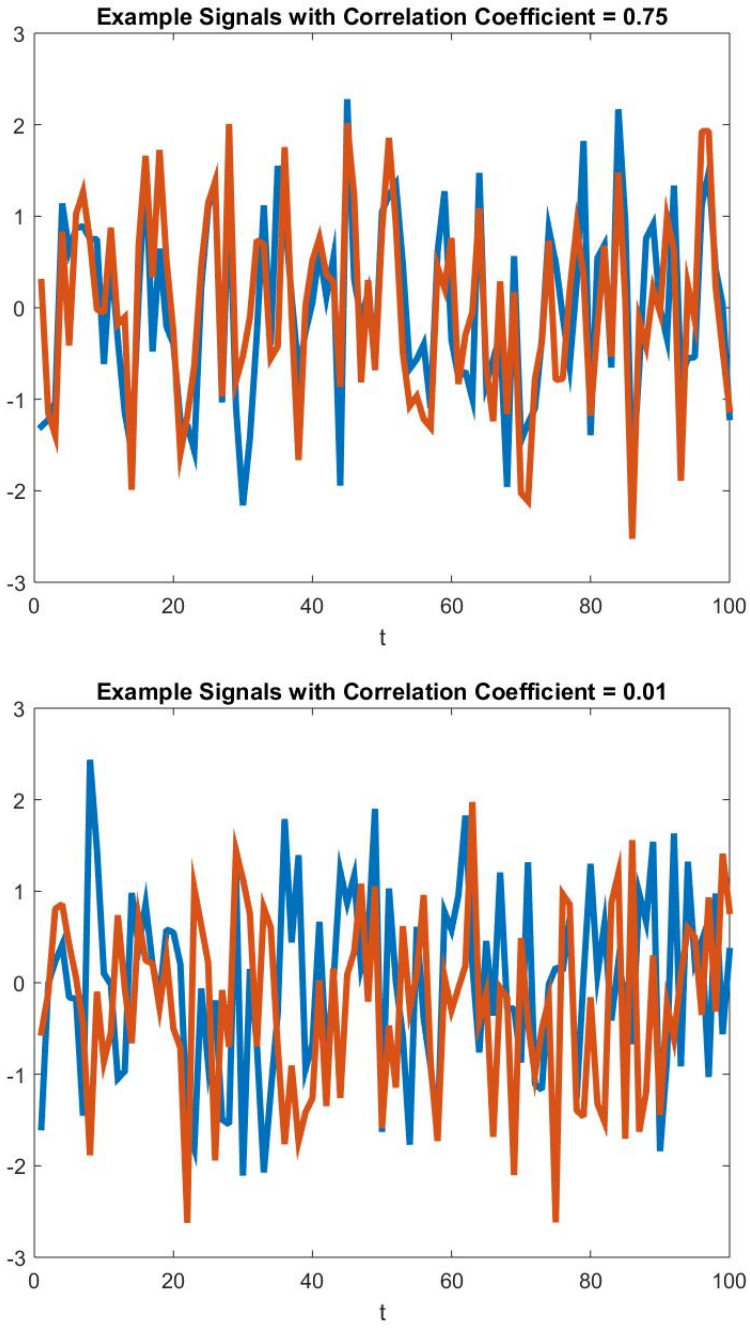

Figure 2: Two identical, independent, normally distributed random signals with correlation coefficients 0.75 and 0.01. y-axis represents arbitrary units. Within a cell, the copy number of two fluorescent proteins under two copies of the same low-noise / high-noise promoter might look like the top / bottom figure.

The final noise decomposition equation, as presented in [**?**], is

$$\eta_{int}^2 = \frac{\langle (c-y)^2 \rangle}{2\langle c \rangle \langle y \rangle}; \quad \eta_{ext}^2 = \frac{\langle cy \rangle - \langle c \rangle \langle y \rangle}{\langle c \rangle \langle y \rangle}; \quad \eta_{tot}^2 = \frac{\langle c^2 + y^2 \rangle - 2\langle c \rangle \langle y \rangle}{2\langle c \rangle \langle y \rangle} \tag{1}$$

Where c and y represent observed values from the two reporter signals, and $\langle \cdot \rangle$ denotes the sample mean of the values. Notice that $\eta_{int}^2 + \eta_{ext}^2 = \eta_{tot}^2$. See [**?**] for the derivation of (1).

2 Data Analysis Methods

Because not every cell on the slide was fluorescent, we thresholded our cells by finding the dimmest cyan-fluorescent cell in each image and removing any cell whose cyan level was below this value. We did the same for yellow fluorescence. We also removed any cells who displayed the oversaturation value (4095 au) for either channel.

In order to accurately compare the fluorescence values between CFP and YFP, we then divided each fluorescence value by the area of the cell to unbias the readout from larger cells. Next we divided all fluorescence values for a channel by the sample mean of the channel to normalize the data. At this stage we performed a two-sample Kolmogorov-Smirnov Test ($\alpha = 0.01$) between the normalized CFP values and normalized YFP values collected from three fields of view for a biological replicate. We failed to reject the null hypothesis for any of our three strains, implying that for a given promoter, the CFP data and YFP data are sampled from the same underlying distribution. This suffices to justify the assumption that the CFP and YFP genes are identically regulated by the cell and hence represent two distinguishable variants of the same reporter.

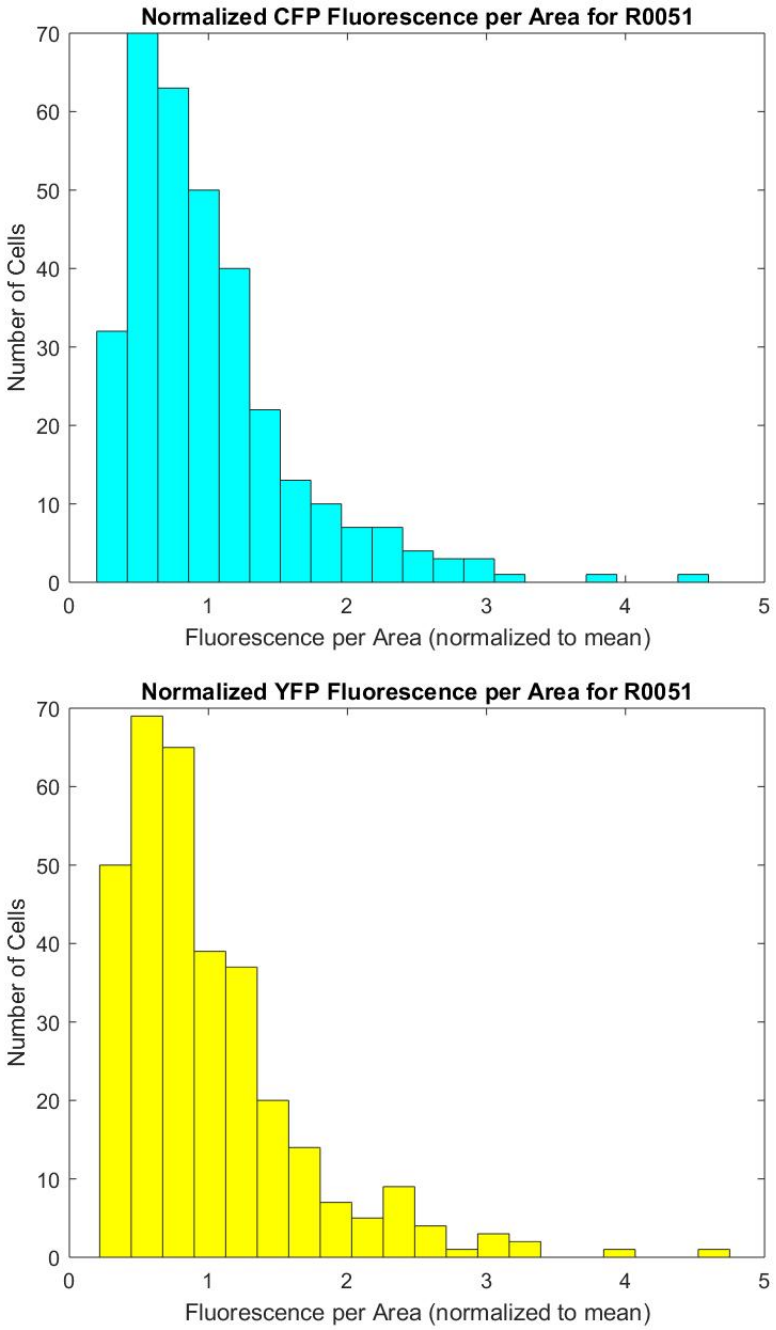

Figure 3: Representative histogram for data from the R0011 promoter. $n = 327$ cells were analyzed after thresholding.

3 Results and Interpretation

We obtained fluorescence images for dual-integrated CFP and YFP under three different promoters from the iGEM Registry: R0010, R0011, and R0051. These are the lacI-repressed promoter, the lacI-repressed lambda pL hybrid promoter, and the lambda cl-repressed promoter. All three are constitutive promoters that were expressed in the absence of their repressors. In particular, R0011 was specifically synthesized for strong transcription.

The noise values, expressed in Table 1, reveal that the intrinsic noise of R0011 is an order of magnitude lower than those of R0010 and R0051. Although all three promoters are simply labeled as 'strong' promoters on the Registry, such a discrepancy highlights a significant difference in the spread of the underlying probability distribution governing transcription under the R0011 promoter (see Figure 5 for an idealized example for a generic random process with these noise values).

Promoter	η^2_{int}	η^2_{ext}	η^2_{tot}
R0010	0.0707	0.4383	0.5089
R0011	0.0040	0.4005	0.4045
R0051	0.0869	0.4553	0.5422

Table 1: Measured noise values based on fluorescence readings from dual-integrated CFP and YFP based on $n = 1$ sample per promoter.

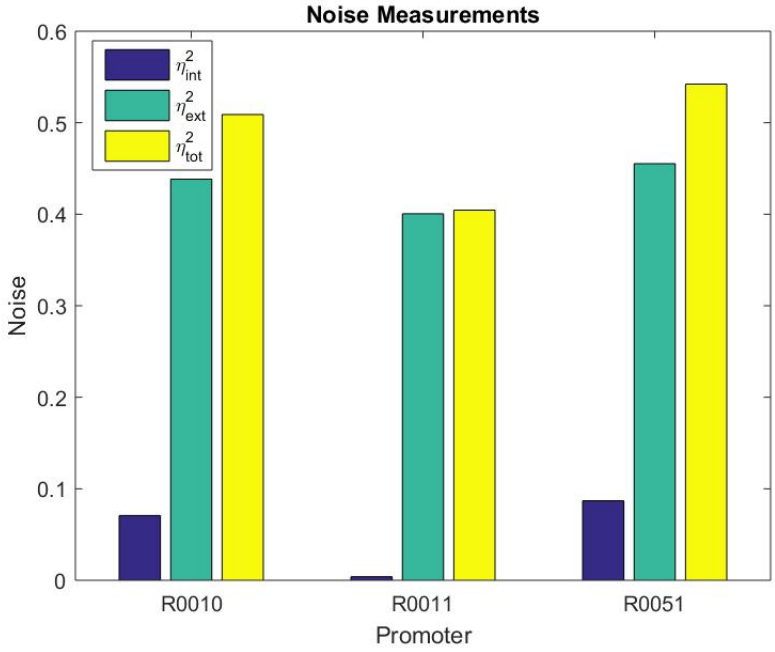

Figure 4: Intrinsic, Extrinsic, and Total Noise for each Measured Promoter. Sample size $n = 1$ for each. Values correspond to Table 1.

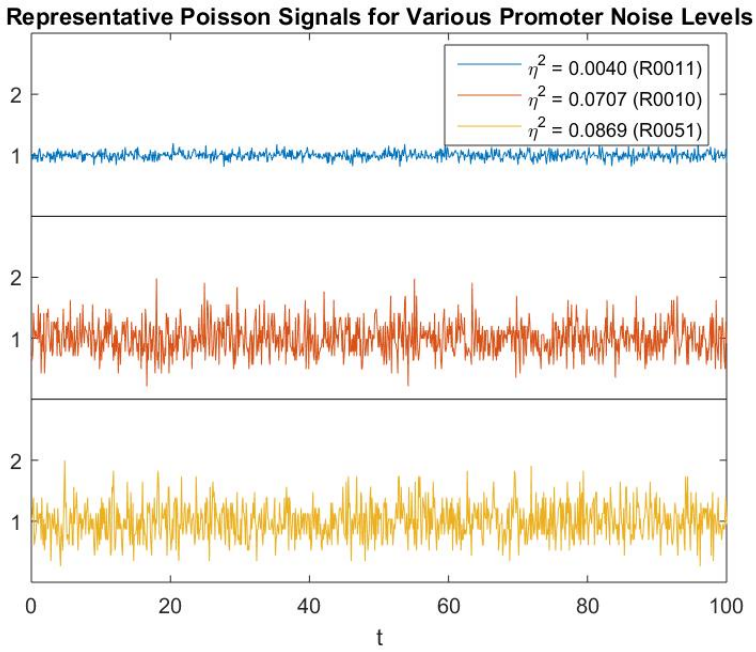

Figure 5: Time traces of random Poisson signals $Y \sim \text{Poiss}(1/\eta_{int}^2)$, normalized to their means, for the η_{int}^2 values that corresponded to those of R0011, R0010, and R0051. y-axis represents arbitrary units.

As synthetic biology moves forward and we design ever more complicated genetic networks, we will soon lose the luxury of being able to ignore the impact of stochasticity on the effectiveness of our constructs. We hope that by conducting this preliminary exploration into the measurement of the intrinsic noisiness of promoters, we convey to iGEM and the synthetic biology community the necessity of rigorous and accurate knowledge of the stochastic properties of each promoter.

References

[1] Arriaga, E.A. Determining biological noise via single cell analysis. (2009) *Anal Bioanal Chem* **393**, 73-80.

[2] Jones, D.L., Brewster, R.C., and Phillips, R. Promoter architecture dictates cell-to-cell variability in gene expression. (2014) *Science* **346**, 1533-1536.

[3] Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. Stochastic Gene Expression in a Single Cell. (2002) *Science* **297**, 1183-1186.

[4] Swain, P.S., Elowitz, M.B., and Siggia, E.D. Intrinsic and extrinsic contributions to stochasticity in gene expression. (2002) *PNAS* **99**, 12795-12800.