Controlling *E. coli* gene expression noise

Kyung Hyuk Kim, Kiri Choi, Bryan Bartley, Herbert M. Sauro.

Abstract

Intracellular protein copy numbers show significant cell-to-cell variability within an isogenic population due to the random nature of biological reactions. Here we show how the variability in copy number can be controlled by perturbing gene expression. Depending on the genetic network and host, different perturbations can be applied to control variability. To understand more fully how noise propagates and behaves in biochemical networks we developed stochastic control analysis (SCA) which is a sensitivity-based analysis framework for the study of noise control. Here we apply SCA to synthetic gene expression systems encoded on plasmids that are transformed into *Escherichia coli*. The objective of the study was to show that we could differentially control the noise and mean levels of molecular concentrations in biological networks. We show that (1) dual control of transcription and translation efficiencies provides the most efficient way of noise-vs.-mean control. (2) The expressed proteins follow the gamma distribution function as found in chromosomal proteins. (3) Bursting size and frequency are strongly correlated, implying that transcription efficiency can affect transcript lifetimes and/or translation efficiency. (4) Lastly, genetic encoding in plasmids amplifies intrinsic noise of gene expression, showing that the two-promoter state model, commonly used to describe chromosomal gene expression, may need to be modified.

Keywords

Synthetic biology, gene expression noise, stochasticity, noise control, two state model, stochastic control analysis

Kyung Hyuk Kim, Kiri Choi, Bryan Bartley, and Herbert M. Sauro are with the Department of Bioengineering, University of Washington, Seattl, WA 98195 USA. The e-mail address of Kyung H. Kim is kkim@uw.edu.

Controlling *E. coli* gene expression noise

I. Introduction

CELL-TO-CELL variability in protein copy numbers within isogenic populations are typically observed in various types of cells due to underlying random biochemical reaction processes [1], [2], [3]. The variability can lead to noise-induced cellular phenotypes such as cellular differentiation [3], multiple stability [4], and either sensitivity enhancement or suppression [5], [6]. Here we investigate the ability to differentially control the noise and mean levels of gene expression in *E. coli*. The approach we use is based on stochastic control analysis (SCA)[8], [7], a body of theory we developed and reported in previously publications.

SCA is a sensitivity analysis framework, that is a direct extension of metabolic control analysis [9], [10] to the stochastic regime [8]. This approach is based on a *local* sensitivity analysis that can be applied to study first-order effects of finite-size perturbations. SCA can identify which parameters in stochastic systems – here, gene regulatory circuits – need to be varied by how much to achieve a desired control aim. This includes orthogonal control of noise levels with respect to mean levels, and simultaneous changes in noise and mean levels in the same or opposite directions for the same or different protein species. SCA can provide control efficiency and strength to identify the most effective control schemes that are experimentally relevant [7]. Here, we apply SCA experimentally to *E. coli* genetic systems and achieve differential noise control *in vivo*.

In this paper, gene circuits are encoded on plasmid backbones, which are inserted into *E. coli* MG1655. We show that encoding circuits in plasmids amplifies intrinsic circuit dynamics. In the *E. coli* transcriptome study [11], the extrinsic noise was found in many cases to completely suppress the intrinsic noise when transcription factor copy numbers were larger than 10 (refer to Fig. 2B in [11]). This means that all dynamics faster than cell doubling time such as transcription-translation processes is significantly averaged out. Therefore, to study these processes, it is necessary to use fast-responsive probes [13] or to come up with methods that amplify the intrinsic processes. Encoding genetic systems of interest in plasmids will be shown to amplify the intrinsic processes and this allows us to investigate *E. coli* gene expression with flow cytometry and fluorescence microscopy without resorting to single-molecule fluorescence microscopy [11], [12].

II. Stochastic control analysis: Review

SCA [7], [8] is a local sensitivity analysis based on control coefficients, which are defined approximately as percentage change in a response signal (y) divided by the percentage change in a system parameter (p):

$$C_p^y = \frac{p}{y}\frac{dy}{dp} = \frac{d\log y}{d\log p}.$$

We note that the slope in the log-log plot of p vs. y corresponds to C_p^y. The response signal can be the mean or noise levels of mRNAs or proteins. The parameters can include transcription and translation efficiencies, degradation rates of mRNAs and proteins, dilution rate due to cell growth, and reaction rates of transcription-factor binding and unbinding from promoter regions, etc. Another important quantity in SCA, is the control vector, each element of which corresponds to a control coefficient for a given response signal (y):

$$\mathbf{C}_\mathbf{p}^y = (C_{p_1}^y, C_{p_2}^y, \cdots, C_{p_N}^y),$$

where N defines the number of parameters (dimension of the parameter space) that will be varied to control the value of y. In this paper, we are mostly interested in dual control of transcription and translation efficiencies, i.e., $N = 2$. One of the important properties of the control vector is that its inner-product with a parameter perturbation vector $\delta\mathbf{p}$ becomes the amount of change in the response signal δy,

$$\mathbf{C}_\mathbf{p}^y \cdot \frac{\delta\mathbf{p}}{\mathbf{p}} = \frac{\delta y}{y}.$$

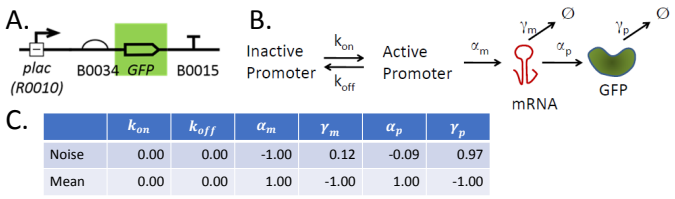
We can quantify which parameter value, and by how much it should be controlled to achieve specific control aims. For example, consider a case where the noise level of a protein needs to be reduced by 9%, while its mean level should remain the same. Here, the noise level (n) is defined by the variance divided by the squared mean value, i.e., squared coefficient of variation. Two control vectors for the noise and mean levels, $\mathbf{C}_\mathbf{p}^n$ and $\mathbf{C}_\mathbf{p}^m$, need to be computed based on a given mathematical model. System parameters need to be perturbed while satisfying

$$\frac{\delta n}{n} = \mathbf{C}_\mathbf{p}^n \cdot \frac{\delta\mathbf{p}}{\mathbf{p}} = -0.09 \quad \text{and} \quad \frac{\delta m}{m} = \mathbf{C}_\mathbf{p}^m \cdot \frac{\delta\mathbf{p}}{\mathbf{p}} = 0.$$

The perturbation vector $\delta\mathbf{p}/\mathbf{p}$ satisfying these two equations can be solved, but the solutions can be infinite. In that case, it is important to select the optimal control scheme (perturbation vector) among the possible solutions. For this, the control efficiency and strength were introduced [7]. Based on these two quantities, one can choose desired control schemes that are appropriate to systems of interest with the maximum control strength and/or efficiency.

III. SCA for a single gene expression cassette

We constructed plasmid systems that express green fluorescent protein (GFP) under *lac*-promoters in *E. coli* (Fig. 1A). The plasmid copy number in a single cell fluctuates in time because the a set of plasmids are randomly partitioned during cell division and are synthesized in a stochastic fashion. Thus, the copy number of *lac*-promoters per cell fluctuates, which will be discussed later in the two-state model. For simplicity,



C.

	k_{on}	k_{off}	α_m	γ_m	α_p	γ_p
Noise	0.00	0.00	-1.00	0.12	-0.09	0.97
Mean	0.00	0.00	1.00	-1.00	1.00	-1.00

Fig. 1. GFP expression system: (A) GFP is expressed under the *lac*-promoter (BioBrick part BBa_R0010) with a ribosome binding site (BBa_B0034). This expression cassette was placed in a low-medium copy number plasmid backbone (pGA3K3; origin of replication p15A). The 'T' symbol represents a terminator (double terminator used here to ensure transcription termination). (B) Two-promoter-state model. When the promoter is active, mRNA is transcribed with a rate constant α_m. From the transcript, GFP is translated with a rate constant α_p. mRNA and GFP degrade or are diluted with net rate constants γ_m and γ_p, respectively. (C) Control coefficients for noise and mean levels are listed. All control coefficients in the same row add up to zero, satisfying summation theorems in SCA [8]. Parameters (unit: hr^{-1}): $k_{on} = 10$, $k_{off} = 0.01$, $\alpha_m = 10$, $\gamma_m = 30$, $\alpha_p = 300$, and $\gamma_p = 1$.

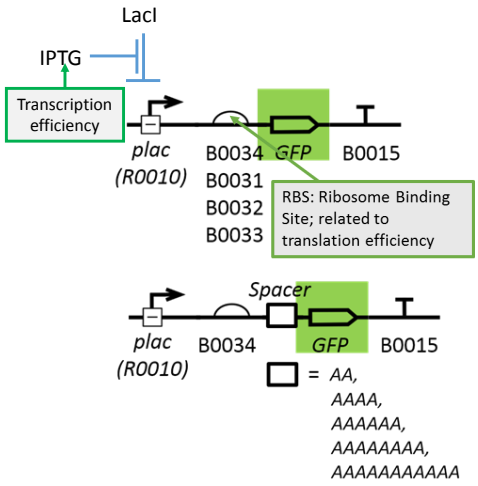


Fig. 2. Perturbations in the GFP expression systems: The GFP expression cassette is placed in the plasmid backbone pGA3K3 in *E.coli* MG1655Z1 that constitutively expresses LacI. IPTG concentrations were varied for a given complex of ribosome binding site and spacer. 9-16 fold increase in the GFP noise level can be achieved without changing its mean level.

we will assume that the plasmid copy number is tightly controlled, i.e., constant at the first level of approximation. The total number of *plac* will be the sum of the number of inactive and active *lac*-promoters (Fig. 1B), which will be set to a constant, N_p. We call this the two-state model. The plasmid backbone that we used is pGA3K3 with the replication origin, p15A ($N_p = 10 - 30$). Based on this two-state model, we computed control vectors for the mean and noise levels of GFP fluorescence as shown in Fig. 1C (refer to [7] for the control vector computation).

The computed control coefficients show that noise can be controlled efficiently by varying both α_m and γ_p; $C^n_{\alpha_m} = -1.00$ and $C^n_{\gamma_p} = 0.97$, indicating that with an increase in α_m for example by 10%, n will reduce by 10% (this is a first-order approximation, because control coefficients are defined locally). Similarly, with an increase in γ_p by 10%, n will increase by 9.7%. With similar change in α_m or γ_m, the mean level can also be efficiently controlled; $C^m_{\alpha_m} = 1$ and $C^m_{\gamma_m} = -1$, indicating that the mean level increases and decreases by 10% in respect to 10% increase of α_m and γ_m, respectively.

IV. Mean Level Control

From the computed control coefficients, the mean protein levels can be controlled without changing the noise level (with a minor change, $\sim$10 folds less than the change in the mean levels) by varying either α_p and γ_m. To confirm this theoretical prediction, we changed the translation efficiency α_p by using a library of both ribosome binding sites (RBSs) and spacer sequences as shown in Fig. 2. Among them, four different spacers – TACTAG, AAAAAA=$(A)_6$, $(A)_{10}$, and $(A)_{13}$ – that are placed between B0034 and the start codon showed distinct GFP expression levels when *plac* is fully active ([IPTG] = 1mM). Here, the introduced spacer sequences are presumed to change ribosome binding affinity, in particular, translation initiation – the limiting step for a translation rate [14], [15].

Based on our flow cytometry data, the mean level was successfully varied by using different spacer sequences as shown in Fig. 3 and Fig. 4A and C. We compared two different cases: Points A and B, and Points C and D in Fig. 3. As shown in Fig. 4A and C, the rescaled probability density functions (pdfs) were overlapped with excellent accuracy. This scale invariance confirms that the noise levels of both points are the same.

Scale invariance in the gamma distribution: Furthermore, the observed invariance implies a special property that we need to consider carefully. This invariance property is satisfied by the gamma distribution function as shown in the Materials and Method section when the bursting size is rescaled together. This implies that the difference between the system parameters of Points A and B is only the bursting size, which is consistent to our perturbation experiment. This result supports that the observed distribution function is the gamma distributions (confer to [16], [17] about claims for other types of distribution functions). To confirm this, we fit the GFP pdfs to the gamma distribution functions as shown in Fig. 5A. For the cases that the background fluorescence is well separated from GFP signals, we confirmed that the pdfs follow the gamma distributions well.

V. Noise Level Control

As discussed above, the noise level can be efficiently controlled by varying α_m and γ_p. However, when these parameters are changed, the mean level also changes with the same fold difference but in the opposite direction; for example, in Fig. 1C, $C^n_{\alpha_m}$ and $C^m_{\alpha_m}$ are -1.00 and 1.00, meaning that when α_m is increased by x%, the noise level decreases by x%, while the mean level increases by x%. Thus, to change the noise level without changing the mean level, we must vary at least two different parameters simultaneously. Since the mean level can be controlled almost independently of the noise level by changing α_p, we will vary both α_p and α_m to compensate

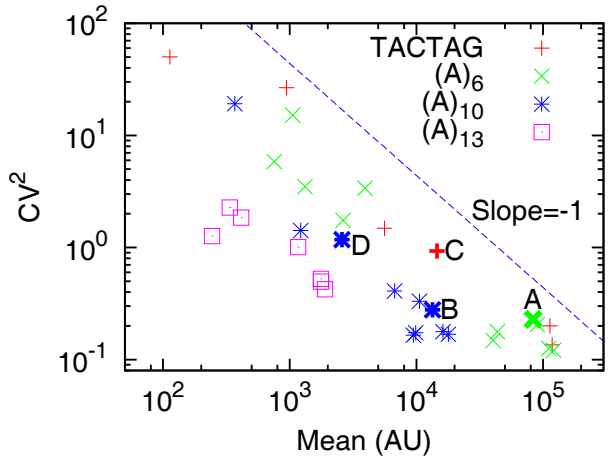


Fig. 3. Scaling relationship between noise and mean levels: Four different cases of spacer sequences between the ribosome binding site BBa_B0034 and the start codon are shown. The same symbol represents the same spacer with different [IPTG]. The noise levels (squared coefficient of variation) are inversely proportional to the mean level. For comparison, a line with a slope -1 is drawn. The contribution to the noise level by background fluorescence signals was taken out (Materials and Methods).

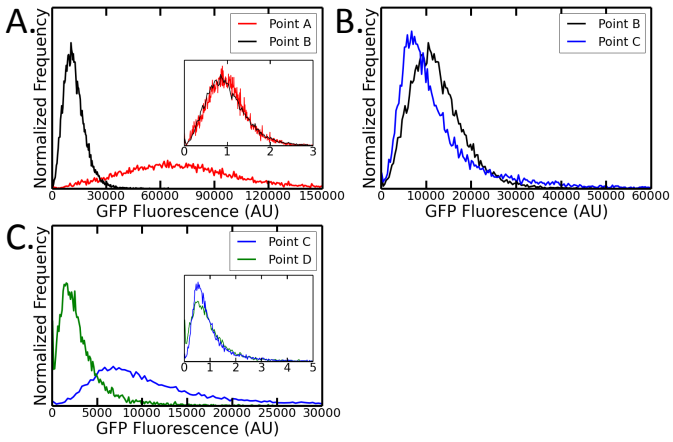


Fig. 4. Probability density functions (pdfs) of GFP fluorescence signals measured from a flow cytometer: (A) Orthogonal mean level control: Points A and B in Fig. 3. In the inset plot, both the pdfs were re-scaled by the mean values of their respective GFP fluorescence signals, so that the transformed pdfs are centered around one. (B) Orthogonal noise level control: Points B and C. Both the [IPTG] and the spacer sequences were varied. (C) Orthogonal mean level control: Points C and D. The inset plot shows the re-scaled pdfs. In this figure, background noise was not taken out.

for the change. The reason that we did not choose to vary γ_p is that this parameter is highly dependent on cell growth rate, rather than protein degradation in *E. coli*; GFP lifetime is much longer than the cell doubling time ~ 1 hr in M9 media. As shown in Fig. 3, the noise level can be controlled by varying both the parameters and $9-16$ fold change in the noise level can be achieved without changing the mean level.

The reason that noise levels could be changed by varying the translation efficiency is that translation events occur in a bursting fashion. This is beause multiple ribosomes can bind

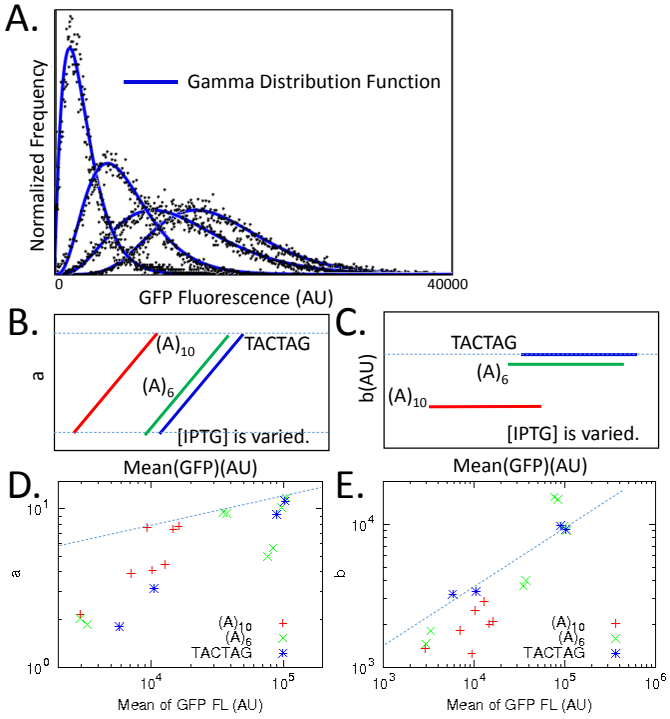


Fig. 5. Gamma distribution function: Bursting size b and bursting frequency a were estimated. (A) $(A)_{10}$ cases for different [IPTG]. The gamma distribution function fits well to the observed pdfs with a minor deviation when [IPTG] is low. (B and C) Ideal case that varying IPTG does not change translation efficiency and different spacer sequences do not affect transcription efficiency. (D and E) Estimated a and b from our flow cytometry data. In this figure, background fluorescence was not taken out.

to single mRNAs before the mRNAs degrade. This allows multiple proteins can be synthesized from a single transcript. These bursting events lead to long-tail histograms. Figure 4B shows that a longer tail in the GFP pdf can be generated by using stronger translation efficiency (Point B $\rightarrow$ Point C).

VI. Scaling Relationship Between the Noise and Mean Levels

Figure 3 shows that the noise level is inversely proportional to the mean level:

$$n = \frac{c}{m},$$

where c is a constant. It is known that the two-state model satisfies this inverse relationship [18], [11], [19]. With the plasmid copy number fixed at any positive integer values, the constant c is shown to be independent of N_p (refer to its derivation in the Supplementary Note of [18]). This indicates that the plasmid copy number N_p should not shift the plot for the noise vs. mean levels. However, we observed this shift to the right, indicating the value of c is somehow increased. Thus, the observed shift implies that the two-state model may need to be modified by taking into account the stochastic fluctuations in N_p. Alternatively, the observed shift could be due to increased bursting size; ribonuclease activity may be saturated, leading to increased lifetime of transcript.

We need to further investigate whether the proposed model is incomplete or whether there is any non-trivial biological correlation between parameter values such as N_p and b.

The observed shift indicates that the intrinsic gene expression processes were amplified and their dynamics can be observed out of the sea of strong extrinsic noise. Otherwise, the intrinsic noise would have been buried by the extrinsic noise as observed for typical *E. coli* transcription factors with their copy numbers higher than $\sim$10 [11], [20] (confer to the study on yeast transcriptome [18], where intrinsic noise is strong enough to be observed for most transcription factors because the extrinsic noise level is lower than *E. coli* due to its longer doubling time). The minimum noise level that we observed in the GFP signals is slightly larger than 0.1, which is consistent with the *E. coli* transcriptome study [11].

Bursting frequency a and bursting size b were estimated from our flow cytometry data by fitting the gamma distributions to the observed GFP pdfs (Materials and Methods). Figures 5B and C show the ideal case that varying IPTG does not change translation efficiency and different spacer sequences do not affect transcription efficiency. Our observed a and b values show the similar trend, however, with strong correlation between the two (compare the blue dotted lines in Fig. 5B and D, and C and E).

VII. Conclusions

We showed that encoding gene circuits in plasmids amplifies intrinsic circuit dynamics, so that intrinsic gene expression in *E. coli* can be observed with flow cytometry without resorting to single-molecule microscopy. In addition, stochastic control analysis was applied to identify efficient ways to control noise and mean levels of gene expression, showing that SCA can be applied to gene regulatory networks and other stochastic biological systems.

Materials and Methods

A. GFP expression circuits and strains

All genetic components used in this manuscript are BioBrick parts, from which genetic circuits were constructed by using the Gibson assembly method [21]. The constructed circuits were integrated into a low-to-medium copy number plasmid pGA3K3 with a Kanamycin resistance gene and *Escherichia coli* MG1655 Z1 was transformed with the plasmids. The strain (lacI^q) constitutively overexpresses LacI from its chromosome.

B. Cell Growth and Flow Cytometry Measurements

E. coli strains were grown to OD600$\sim$0.2 in 2mL Luria-Bertani (LB) media (Becton Dickinson) with kanamycin 50μg/mL at 37°C and 300rpm in a shaker. The cultures were diluted 1:200 into 200 μL prewarmed fresh M9 media (Teknova 2M1990) in 96-well plates (Costar 3904) with kananmycin 50μg/mL. 12 different IPTG concentrations (0mM, 0.02$\sim$1mM) were used for each well and grown to OD600=0.3-0.4 in a shaker (37°C, 300rpm). For the flow cytometry measuremnts, the grown cultures were diluted 1:4 in 1xPBS. A Sony Biotechnology ec800 flow cytometer was used with a 525nm filter and a 488nm excitation laser for GFP fluorescence. 100,000 events were collected for each sample and gated by using a 2-D normal distribution (Bioconductor flowCore norm2filter function with scale.factor=1) [22] within the R software environment as well as by using python package FlowCytometryTools (http://gorelab.bitbucket.org/flowcytometrytools/#). To prevent well-well contamination we executed a Medium Flush cycle after each sample well. When computing the mean and noise levels of GFP signals, background fluorescence was taken care of by using the mean and noise levels for the case without IPTG for each different gene circuit.

C. Noise Level Correction

The mean level was corrected with a simple subtraction. The noise level was corrected by using the property that the observed variance (Variance$_o$) is the sum of the GFP variance (Variance$_g$) and the background signal variance (Variance$_b$) under the assumption that the GFP signals are statistically independent of the background signals. More precisely, the noise level of GFP signals, defined by the square coefficient of variation, can be obtained by

$$\mathrm{CV}^2 = \frac{\mathrm{Variance}_o - \mathrm{Variance}_b}{(\mathrm{Mean}_o - \mathrm{Mean}_b)^2}.$$

where the subscripts o and b denote observed and background signals, respectively.

D. Nonlinear Regression

The gamma distribution function was used to fit our flow cytometry data. Protein copy number N_{pr} can be converted to fluorescence signal intensity x: $N_{pr} = c_s x$ with c_s a scaling constant. The gamma distribution function can be rescaled:

$$\begin{aligned} p(N_{pr}; a, b) &= p(c_s x; a, b) = \frac{(c_s x)^{a-1} e^{-c_s x/b}}{\Gamma(a)\, b^a} \\ &= c_s^{-1} \frac{x^{a-1} e^{-x/(b/c_s)}}{\Gamma(a)\, (b/c_s)^a} \\ &= c_s^{-1} p(x; a, b/c_s). \end{aligned}$$

Here, Γ is a gamma function,

$$a \equiv \frac{\alpha_m}{\gamma_p} \times [\text{Active Promoter Copy Number}]$$

is the number of mRNA produced per cell doubling time, called bursting frequency, and

$$b \equiv \frac{\alpha_p}{\gamma_m}$$

is the number of proteins produced during the mRNA lifetime, called bursting size. Therefore, the fluorescence intensity should also follow the gamma distribution if its corresponding copy number follows the gamma distribution, with the bursting size rescaled with c_s. Nonlinear regression was carried by using the Scipy curve_fit function (http://www.scipy.org/), which employs the Levenberg-Marquardt algorithm for the least squares fitting to estimate a and b.

ACKNOWLEDGMENT

This work was supported by the National Science Foundations (NSF MCB 1158573).

REFERENCES

[1] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators." *Nature*, vol. 403, no. 6767, pp. 335–338, Jan. 2000.

[2] A. Sanchez, S. Choubey, and J. Kondev, "Regulation of noise in gene expression." *Annu. Rev. Biophys.*, vol. 42, pp. 469–91, Jan. 2013.

[3] G. Balázsi, A. van Oudenaarden, and J. J. Collins, "Cellular decision making and biological noise: from microbes to mammals." *Cell*, vol. 144, no. 6, pp. 910–925, Mar. 2011.

[4] S.-L. To and N. Maheshri, "Noise Can Induce Bimodality in," *Science*, vol. 327, no. February, pp. 1142–1146, 2010.

[5] K. H. Kim, H. Qian, and H. M. Sauro, "Nonlinear biochemical signal processing via noise propagation," *J. Chem. Phys.*, vol. 139, p. 144108, 2013.

[6] J. Paulsson, O. G. Berg, and M. Ehrenberg, "Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation." *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 13, pp. 7148–7153, Jun. 2000.

[7] K. H. Kim and H. M. Sauro, "Adjusting Phenotypes by Noise Control," *PLoS Comput. Biol.*, vol. 8, no. 1, p. e1002344, Jan. 2012.

[8] ——, "Sensitivity summation theorems for stochastic biochemical reaction systems." *Math. Biosci.*, vol. 226, no. 2, pp. 109–119, Aug. 2010.

[9] D. A. Fell, "Metabolic control analysis: a survey of its theoretical and experimental development." *Biochem. J.*, vol. 286, pp. 313–330, 1992.

[10] H. Kacser and J. A. Burns, "The control of flux." *Biochem. Soc. Trans.*, vol. 23, pp. 341—-366, 1995.

[11] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, "Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells." *Science*, vol. 329, no. 5991, pp. 533–538, Jul. 2010.

[12] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, "Regulation of noise in the expression of a single gene." *Nature Genetics*, vol. 31, pp. 69–73, 2002.

[13] A. Sanchez and I. Golding, "Genetic determinants and cellular constraints in noisy gene expression." *Science*, vol. 342, no. 6163, pp. 1188–93, Dec. 2013.

[14] H. M. Salis, E. a. Mirsky, and C. a. Voigt, "Automated design of synthetic ribosome binding sites to control protein expression." *Nat. Biotechnol.*, vol. 27, no. 10, pp. 946–50, Oct. 2009.

[15] R. G. Egbert and E. Klavins, "Fine-tuning gene networks using simple sequence repeats." *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 42, pp. 16 817–22, Oct. 2012.

[16] H. Salman, N. Brenner, C.-k. Tung, N. Elyahu, E. Stolovicki, L. Moore, A. Libchaber, and E. Braun, "Universal Protein Fluctuations in Populations of Microorganisms," *Phys. Rev. Lett.*, vol. 108, no. 23, p. 238105, Jun. 2012.

[17] S. Ghosh, K. Sureka, B. Ghosh, I. Bose, J. Basu, and M. Kundu, "Phenotypic heterogeneity in mycobacterial stringent response." *BMC Syst. Biol.*, vol. 5, no. 1, p. 18, Jan. 2011.

[18] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai, "Noise in protein expression scales with natural protein abundance." *Nat. Genet.*, vol. 38, no. 6, pp. 636–643, Jun. 2006.

[19] V. Shahrezaei and P. S. Swain, "Analytical distributions for stochastic gene expression." *Proc Natl Acad Sci U S A*, vol. 105, no. 45, pp. 17 256–17 261, Nov. 2008. [Online]. Available: http://dx.doi.org/10.1073/pnas.0803850105

[20] O. K. Silander, N. Nikolic, A. Zaslaver, A. Bren, I. Kikoin, U. Alon, and M. Ackermann, "A genome-wide analysis of promoter-mediated phenotypic noise in Escherichia coli." *PLoS Genet.*, vol. 8, no. 1, p. e1002443, Jan. 2012.

[21] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. a. Hutchison, and H. O. Smith, "Enzymatic assembly of DNA molecules up to several hundred kilobases." *Nat. Methods*, vol. 6, no. 5, pp. 343–5, May 2009.

[22] F. Hahne, N. LeMeur, R. R. Brinkman, B. Ellis, P. Haaland, D. Sarkar, J. Spidlen, E. Strain, and R. Gentleman, "flowCore: a Bioconductor package for high throughput flow cytometry." *BMC Bioinformatics*, vol. 10, p. 106, Jan. 2009.