Quantitative Predictions of Gene Expression

Brian Munsky[1,#] and Gregor Neuert[2,#]

Short Abstract — **Even with continually improving biochemical understanding, it remains difficult to predict the quantitative dynamics of single-cell gene regulation. Here we discuss a comprehensive approach to identify and validate stochastic models of gene regulation with single-cell and single-molecule resolution. We apply this approach on mRNA measurements from several genes in the osmotic stress response pathway in *S. cerevisiae*, and we find a single model (among thousands of automatically-generated candidates) to balance complexity and simplicity and achieve maximal predictive capabilities. Our final model accurately predicts single-cell transcriptional dynamics for three genes in response to several combinations of environmental changes and genetic mutations.**

Keywords **— single-cell, single-molecule, stochastic gene regulation, system identification, uncertainty quantification.**

Even genetically identical cells in the same environment can exhibit wildly different behaviors due to spatial, temporal and stochastic fluctuations. Often labeled "noise," these fluctuations were previously considered a nuisance that compromised cellular responses, complicated modeling, and made predictive understanding all but impossible. However, if we examine these fluctuations more closely through the lens of new experimental and computational techniques, we actually discover a powerful information resource. Different cellular mechanisms affect these cellular fluctuations in different ways, and the resulting "fluctuation fingerprints" can help us to identify new properties of hidden cellular mechanisms. Understanding these fluctuations (or "Listening to the Noise") requires a systematic integration of single-cell measurements with stochastic analyses[1,2]. Here we show how this integration, with careful attention to complexity and parameter uncertainty, can achieve powerful predictive understanding of transcriptional pathways.

To quantitatively predict gene regulatory responses in complex networks, we utilized a comprehensive system identification and validation approach. First, we obtained quantitative measurements of single-molecule mRNA expression[3] at fast temporal resolution in individual cells. Second, we used efficient stochastic analysis approaches to analyze the discrete, time varying, and stochastic nature of mRNA transcription[4]. Third, we integrated these experimental and computational approaches within an identification framework of uncertainty quantification, experiment design, predictions and validation[5].

We applied this approach to the osmotic stress response pathway of *Saccharomyces cerevisiae* and uncovered a single semi-mechanistic gene regulatory model among several thousand automatically generated hypotheses. Looking at a single mRNA species (*STL1*) and using several rounds of parameter searches and uncertainty quantification, we inferred mechanisms and parameters that match the measured single-cell / single-molecule data yet are simple enough to avoid over-fitting and retain predictive power (Fig. 1A-B). We then extended our model to capture the quantitative effects of transcription factor (*HOT1*) over expression and chromatin modifier (*ARP8* & *GCN5*) knockouts on the *STL1* dynamics. We also extended the model to capture the dynamics of other stress response genes *CTT1* (oxidative stress) and *HSP12* (heat stress). The final identified model provided accurate quantitative predictions for the transient bimodal distributions of single-cell transcriptional response in all three genes (*STL1*, *CTT1*, *HSP12*) in response to new levels of osmotic shock and new combinations of genetic mutations (Fig 1C).

We stress that achieving quantitatively predictive understanding requires careful consideration of complexity and parameter uncertainty—adding more mechanisms and parameters may enable a more accurate fit, but can lead to losses in predictive precision. Our uncertainty quantification method represents a powerful approach to automatically choose model complexity and improve predictive understanding of inducible transcriptional networks.

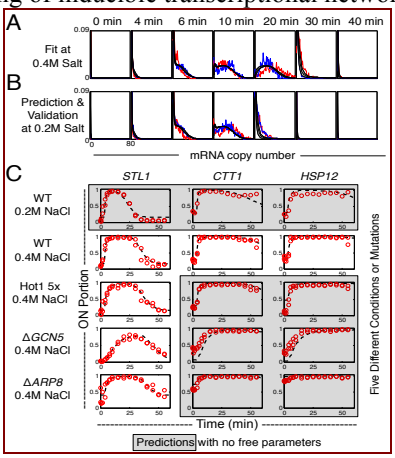

Figure 1. Predictive model identification for *STL1*, *CTT1*, and *HSP12* responses to osmotic stress. A) Model *fit* to *STL1* mRNA distributions at several time points after applying 0.4M osmotic stress. B) Model *predictions* for *STL1* distributions in 0.2M osmotic stress. C) Probability of *STL1*, *CTT1*, and *HSP12* activation during adaptation for different conditions and genetic mutations (adapted from [5]).

[1] Center for Nonlinear Studies and the Information Sciences Group, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.
[2] Molecular Physiology and Biophysics, School of Medicine, Vanderbilt University, Nashville TN 37232, USA.
[#] These authors contributed equally to this work.
E-mail: brian.munsky@gmail.com

REFERENCES

[1]Munsky, Trinh, Khammash, *Mol Syst Biol* **5**:318 (2009).
[2]Munsky, Neuert, van Oudenaarden, *Science*, **336**:6078 (2012).
[3]Raj, van den Bogaard, *et al*, *Nat. Methods* **5**:877 (2008).
[4]Munsky, Khammash, *J. Chem. Phys.* **124**:044104 (2006).
[5]Neuert, Munsky, *et al*, Science, **339**:6119 (2013).