Systematic Identification of Signal-Activated Stochastic Gene Regulation

Gregor Neuert [1,2,#], Brian Munsky[3,#], Rui Zhen Tan[1,2], Mustafa Khammash[4,*] and Alexander van Oudenaarden[1,2,*]

Short Abstract — **Despite the large amount of biochemical information for many signal transduction pathways and gene regulatory mechanisms, it remains difficult to predict the quantitative dynamics of these systems at the single-cell level. Here we introduce a comprehensive approach to identify and validate gene regulation models by integrating single-cell/single-molecule experiments with discrete stochastic analyses. Applying this approach to the osmotic stress response pathway in *S. cerevisiae,* we identified the best predictive gene regulatory model among several thousands of automatically-generated candidates. This model captures several previously unknown system properties and can accurately predict the dynamics of the system in response to diverse environmental and genetic perturbations.**

Keywords **—single-cell/single-molecule, systems biology, stochastic gene regulation, system identification, prediction**

Understanding and predicting how cells sense and respond to their environments is a key goal of systems biology. Although new experimental and computational methodologies have elucidated many signal transduction pathways and gene regulation mechanisms, this goal remains elusive particularly when it comes to making quantitative predictions for the phenotypic diversity of single cell dynamics [1,2]. To better understand and predict gene regulatory responses for these complex networks, we propose a comprehensive system identification and validation approach. First, we developed a quantitative assay to measure single-molecule expression of endogenous mRNA [3] at fast temporal resolution in individual cells. Second, we developed an efficient and flexible computational approach that captures the discrete, time varying, and stochastic nature of transcriptional regulation [4,5]. Third, we integrated these experimental and computational approaches within a novel hierarchical system identification framework, which involves several clearly defined rounds of analysis, prediction, experiment design, and validation.

We applied this approach to the osmotic stress response pathway of *Saccharomyces cerevisiae* and uncovered the best gene regulatory model among several thousand automatically generated model hypotheses—each with its own assumptions for regulatory mechanisms. Using parameter identification and cross-validation analyses, we select a final model (both mechanisms and parameters) that is complex enough to match the observed single-cell/single-molecule data but simple enough to avoid over-fitting and thereby retain predictive power. The identified model describes several novel dynamical features in transcriptional regulation, including multi-step activation, low-pass filtering, kinetic proofreading and signal-modulated duration of gene expression. Furthermore, we found that kinetic proofreading, mRNA degradation and transcription dynamics are robust to environmental and/or genetic conditions.

In addition, the selected model provides accurate quantitative predictions at diverse experimental perturbations, which extend well beyond the training dataset. In fact, when used to predict responses in new conditions, our selected model out-performs any of those discarded during the identification procedure, thus validating our model selection approach. In particular, we are able to predict not only the full bimodal distributions of mRNA at different times and conditions, but also the active transcription state of the gene—a new experimental feature that was never considered in our model identification procedure.

This approach represents a powerful new methodology in the field of quantitative transcriptional regulation, which is not specific to any gene, pathway or organism, and which may lead to new insight into inducible transcriptional networks in organisms ranging from yeast to human.

[1]Departments of Physics and Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
[2]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
[3]Center for Nonlinear Studies and the Information Sciences Group, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.
[4]Center for Control, Dynamical Systems and Computation and The Department of Mechanical Engineering, University of California, Santa Barbara, CA 93106, USA.
[#]These authors contributed equally to this work.
[*]Co-senior authors.
E-mail: gneuert@mit.edu

References

[1] J. T. Mettetal, D. Muzzey, C. Gomez-Uribe, A. van Oudenaarden,*Science* **319**, 482 (2008).
[2] D. Muzzey, C. A. Gomez-Uribe, J. T. Mettetal, A. van Oudenaarden, *Cell* **138**, 160 (2009).
[3] A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, *Nat. Methods* **5**, 877 (2008).
[4] B. Munsky, M. Khammash, *J. Chem. Phys.* **124**, 044104 (2006).
[5] B. Munsky, B. Trinh, M. Khammash, *Mol Syst Biol* **5**(2009).