

Accurate Predictions of Genetic Circuit Behavior from Part Characterization and Modular Composition

Noah Davidsohn
MIT
ndavidso@mit.edu

Fusun Yaman
Raytheon BBN Technologies
fusun@bbn.com

Jacob Beal
Raytheon BBN Technologies
jakebeal@bbn.com

Yinqing Li
MIT
yinqingl@mit.edu

Ron Weiss
MIT
rweiss@mit.edu

Aaron Adler
Raytheon BBN Technologies
aadler@bbn.com

Zhen Xie
Tsinghua University
zhenxie@tsinghua.edu.cn

1. OVERVIEW

A long-standing goal of synthetic biology is to rapidly engineer new regulatory circuits from simpler regulatory elements [8, 16, 2, 7]. As the complexity of engineered circuits increases, it becomes increasingly important to utilize quantitative models to guide circuit construction effectively, but previous efforts have been hindered by lack of accuracy in predictions of circuit behavior [13, 10]. To address this shortcoming, we have developed Empirical Quantitative Incremental Prediction (EQuIP), a new method for accurate prediction of genetic regulatory network behavior. EQuIP predictions are based on a composable black-box model derived solely from empirical observations of steady-state and dynamic behavior. We have validated the precision of EQuIP predictions for six transcriptional cascades in transiently transfected mammalian cells. These cascades exhibit over 1,000-fold cell-cell variation in fluorescence, yet EQuIP’s predictions have a mean error of only 1.6-fold compared to experimental data. Such accurate predictions will foster reliable forward engineering of complex biological circuits from libraries of standardized devices.

2. EMPIRICAL PRODUCTION/LOSS MODEL

Effective construction of genetic circuits in synthetic biology involves choosing parts that combine to produce a desired behavior—a task that rapidly increases in importance and complexity for larger circuits. Current biological circuit models often explicitly encode all the biochemical reactions in the system believed to be significant (e.g., [11, 1, 9]), but their accuracy is impaired by difficult to estimate parameters and inadequate incorporation of cellular context effects. The main alternative, black-box abstractions, have been frequently explored (e.g., [7, 17, 14, 3, 12]), but until now there has been no black-box model that successfully predicted the experimental behavior of a multi-gene circuit.

Our EQuIP black box model improves on prior approaches by combining three key ingredients: an appropriate level of device abstraction, obtaining detailed and internally validated characterization data, and an effective framework for

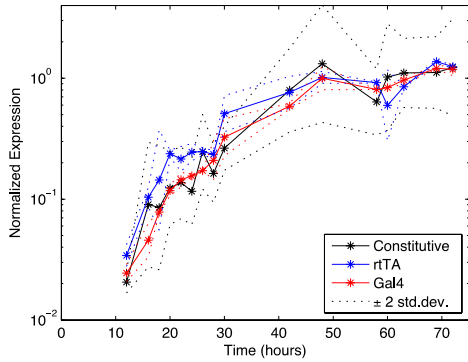
modeling circuit behavior from part characterization. In our abstraction, a device encapsulates a small set of regulatory interactions that transduce input values into corresponding outputs. For each device, we collect empirical observations of expression dynamics and device steady-state behavior within the appropriate cellular context. This information is then used to create a composable device model with commensurate units describing the relationship between input levels, the circuit (i.e., plasmid) copy number, and the corresponding production and loss rates for the output biological species. EQuIP thus abstracts away biochemical details and captures significant interactions with the cellular context without requiring either explicit or well-understood models of those interactions. Using a piecewise linear approximation of input/output transfer curves, rather than making strong biochemical assumptions (e.g., using Hill kinetics), allows us to build a model that includes the effects of interactions with the cell even when an explicit model is not available.

EQuIP’s device production and loss functions are derived from three sets of characterization data: fluorescent protein calibration and normalization, constitutive gene expression measured over time, and input/output transfer curves. The calibration and normalization data are an extended set of controls used to map each color’s fluorescence measurements to absolute MEFL units [15] and to compensate for systematic variation between data sets, following the calibration protocols that we have previously laid out in [5].

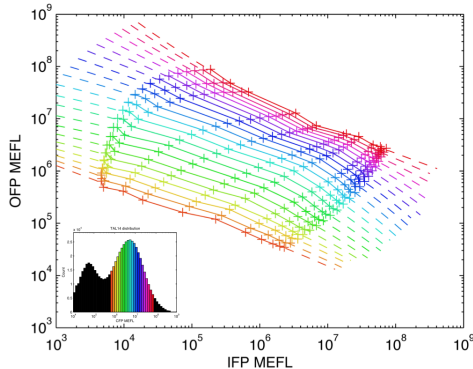
Fundamental production and loss dynamics for the system are determined from measurement of the time course of constitutive and transcriptional activator driven expression. Data from a 72-hour experiment to determine these dynamics is shown in Figure 1(a). Our analysis of this model gives a mean behavior of cell division every 20 hours through 47 hours of expression, convolved across a variable initial delay of up to one division time. From this, we determine a loss function based on exponential dilution from cell growth (expected to dominate, given stable proteins) and a relative production function uniquely determined on an inverse model of the dilution of protein and plasmids over time.

This constitutive production model is transformed into a model of regulated production with the aid of an input/output transfer curve: a snapshot of the relationship between in-

Work partially sponsored by DARPA I2O under contract HR0011-10-C-0168; the views and conclusions contained in this document are those of the authors and not DARPA or the U.S. Government.
IWBDA ’13 London, UK



(a) Expression Time Series



(b) TAL14 Transfer Curve

Figure 1: EQuIP uses two sets of input data: expression time series (a) are used to quantify incremental fluorescence production and loss rates, and input/output transfer curves (b) are used to quantify regulatory relationships (dashed lines are projection; inset shows binning by constitutive fluorescence).

put and output concentrations at a particular time. We determine this relation using a device characterization circuit with three fluorescent reporter proteins that approximate input concentration (IFP), output concentration (OFP), and number of circuit copies in a cell (CFP). We estimate the relative circuit copy number for an individual cell based on the intensity of a fluorescent protein (CFP) constitutively expressed from one of the co-transfected plasmids [6]. To account for behavior variance due to variance plasmid copy number, we split the flow cytometry characterization of each sample into logarithmically partitioned subpopulations (“bins”) based on CFP fluorescence. The mean expression level of each fluorescent protein is then computed independently for each subpopulation [4]. This allows us to diminish the impact of noise associated with single cell data while still capturing the differential behavior due to circuit copy number.

3. EXPERIMENTAL VALIDATION

To validate EQuIP, we gathered transfer curves for three hybrid repressor devices: TAL14, TAL21, and LmrA. Each experiment was performed with a co-transfection, following the characterization method in [5]: triplicate samples of

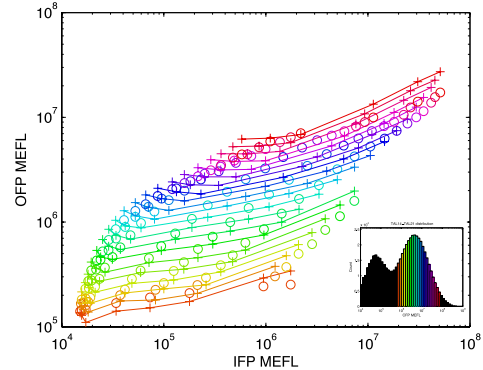


Figure 2: Experimental testing of two element cascades, such as this TAL14-TAL21 cascade, shows a good agreement between EQuIP predictions (circles) and observed experimental data (lines and crosses).

each circuit were induced with a series of Dox concentrations covering the full range of rtTA3 activation, and fluorescence was measured using flow cytometry at $t = 72$ hours post-transfection. Figure 1(b) shows the transfer curve for TAL14, segmented by circuit copy number, with the segmentation indicated on the inset histogram of constitutive expression.

We then used these data to predict the behavior of all two-repressor cascades comprising TAL14, TAL21, and LmrA, and compared these predictions to experimental results. The observed mean output levels for different input/plasmid combinations have more than a 1,000-fold range for each of the six cascades across the different cell sub-populations; over this wide range, the mean error of predicted versus observed output is only 1.6-fold for all six transcriptional cascades measured. Figure 2 shows one example of these six: the 72 hour input/output predictions vs. experimental data for the TAL14-TAL21 cascade. EQuIP is also effective at predicting which combinations of repressors are best matched to provide the greatest differential expression between fully-induced and un-induced states: although the induction predictions from EQuIP are conservative (they under predict the gain), they correctly identify the TAL14-TAL21 and TAL21-TAL14 cascades as having significantly stronger gain than all cascades involving LmrA.

4. FUTURE DIRECTIONS

The ability to make accurate predictions without detailed biochemical models of genetic regulatory devices makes EQuIP a powerful abstraction, as it only uses readily available single-cell fluorescence data with minimal assumptions or indirect inference. As may be expected, predictions are less accurate for extreme expression levels and circuit copy numbers. Anticipated future work includes improving models and gathering additional data to allow prediction of these more extreme cases, validation of EQuIP on more complex circuits, application of predictions to improve regulatory protein design, and adaptation of EQuIP to other organisms. On the basis of such expansion, we anticipate that EQuIP may become a critical building block in a new era of exponential growth in our ability to engineer biological systems.

5. REFERENCES

- [1] C. M. Agapakis and P. A. Silver. Synthetic biology: exploring and exploiting genetic modularity through the design of novel biological networks. *Molecular BioSystems*, 704, 2009.
- [2] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss. Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol*, 2006.
- [3] G. Batt, B. Yordanov, R. Weiss, and C. Belta. Robustness analysis and tuning of synthetic gene networks. *Bioinformatics*, 23:2415–2422, 2007.
- [4] J. Beal, R. Weiss, D. Densmore, A. Adler, E. Appleton, J. Babb, S. Bhatia, N. Davidsohn, T. Haddock, J. Loyall, et al. An end-to-end workflow for engineering of biological networks from high-level specifications. *ACS Synthetic Biology*, 1(8):317–331, 2012.
- [5] J. Beal, R. Weiss, F. Yaman, N. Davidsohn, and A. Adler. A method for fast, high-precision characterization of synthetic biology devices. Technical Report MIT-CSAIL-TR-2012-008, MIT, April 2012. Technical Report: MIT-CSAIL-TR-2012-008 <http://hdl.handle.net/1721.1/69973>.
- [6] L. Bleris et al. Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Molecular Systems Biology*, 7:519, 2011.
- [7] B. Canton, A. Labno, and D. Endy. Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology*, 26(7):787–793, 2008.
- [8] T. Ellis, X. Wang, and J. Collins. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nature biotechnology*, 27(5):465–471, 2009.
- [9] S. Hooshangi and R. Weiss. The effect of negative feedback on noise propagation in transcriptional gene networks. *Chaos*, 16, 2006.
- [10] R. Kwok. Five hard truths for synthetic biology. *Nature News*, 463:288–290, 2010.
- [11] T. S. Moon, C. Lou, A. Tamsir, B. C. Stanton, and C. A. Voigt. Genetic programs constructed from layered logic gates in single cells. *Nature*, 491:249–253, 2012.
- [12] V. K. Mutalik et al. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods*, 10:354–360, 2013.
- [13] P. E. M. Purnick and R. Weiss. The second wave of synthetic biology: from modules to systems. *Nature Reviews Molecular Cell Biology*, 10(6):410–422, 2009.
- [14] N. Rosenfeld, J. Young, U. Alon, P. Swain, and M. Elowitz. Accurate prediction of gene feedback circuit behavior from component properties. *Molecular Systems Biology*, 13(3):143, November 2007.
- [15] Measuring molecules of equivalent fluorescein (mefl), pe (mepe) and rpe-cy5 (mepcy) using sphero rainbow calibration particles. Technical Report SpheroTechnical Notes: STN-9, Rev C 071398, SpheroTech, October 2001.
- [16] B. Wang, R. Kitney, N. Joly, and M. Buck. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications*, 508:465–471, 2009.
- [17] R. Weiss and S. Basu. The device physics of cellular logic gates. In *NSC1 The First Workshop of NonSilicon Computing*, pages 54–61, 2002.