ACS SyntheticBiology

Letter

pubs.acs.org/synthbio

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

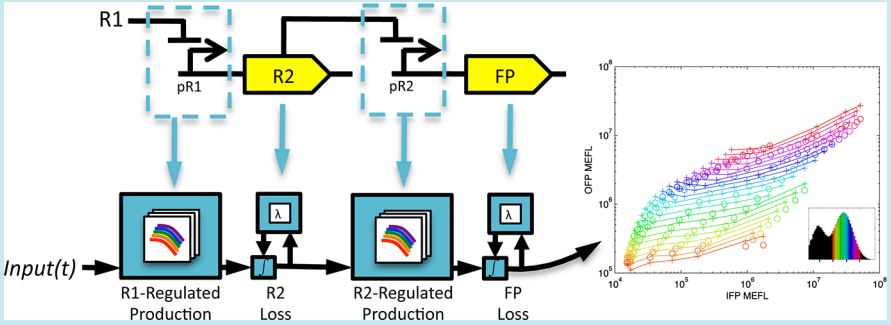
Noah Davidsohn,[†,∥] Jacob Beal,[‡,∥] Samira Kiani,[†] Aaron Adler,[‡] Fusun Yaman,[‡] Yinqing Li,[†] Zhen Xie,[§] and Ron Weiss*[,†]

[†]Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States

[‡]Raytheon BBN Technologies, 10 Moulton Street, Cambridge, Massachusetts 02138, United States

[§]Bioinformatics Division/Center for Synthetic and Systems Biology, Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China

S *Supporting Information*



ABSTRACT: A long-standing goal of synthetic biology is to rapidly engineer new regulatory circuits from simpler devices. As circuit complexity grows, it becomes increasingly important to guide design with quantitative models, but previous efforts have been hindered by lack of predictive accuracy. To address this, we developed Empirical Quantitative Incremental Prediction (EQuIP), a new method for accurate prediction of genetic regulatory network behavior from detailed characterizations of their components. In EQuIP, precisely calibrated time-series and dosage-response assays are used to construct hybrid phenotypic/mechanistic models of regulatory processes. This hybrid method ensures that model parameters match observable phenomena, using phenotypic formulation where current hypotheses about biological mechanisms do not agree closely with experimental observations. We demonstrate EQuIP's precision at predicting distributions of cell behaviors for six transcriptional cascades and three feed-forward circuits in mammalian cells. Our cascade predictions have only 1.6-fold mean error over a 261-fold mean range of fluorescence variation, owing primarily to calibrated measurements and piecewise-linear models. Predictions for three feed-forward circuits had a 2.0-fold mean error on a 333-fold mean range, further demonstrating that EQuIP can scale to more complex systems. Such accurate predictions will foster reliable forward engineering of complex biological circuits from libraries of standardized devices.

KEYWORDS: *synthetic biology, systems biology, genetic circuits*

One of the key challenges in synthetic biology is to accurately predict the behavior of novel biological systems, thereby enabling faster and more effective engineering of such systems.[1−3] This challenge is becoming a critical issue, given the growing gap between the exponential increase in length of DNA sequences that can be readily synthesized[3−9] and the much slower increase in the complexity of genetic circuits that have been demonstrated.[9−14] Accurate predictions of the behavior of genetic circuits are an important ingredient for addressing this gap. As the number of genetic elements in a circuit increases, the number of candidate designs increases exponentially. Accurate predictions help cope with this exponential explosion by dramatically reducing the number of candidate designs that must be considered. Predicting circuit behavior, however, has been extremely difficult, and without reliable predictions, even relatively simple circuits have typically required extensive and costly tuning to achieve the desired results.[9,10,12,15]

Recently, there have been major steps toward improving the accuracy of genetic circuit predictions. First, genetic elements are now being characterized using calibrated and standardized measurements.[16−19] Second, several investigations have provided means of combining families of primitive elements of a transcriptional unit, such as promoters and 5′UTRs, in order to more reliably and predictably control constitutive gene

Received: June 13, 2014
Published: November 4, 2014

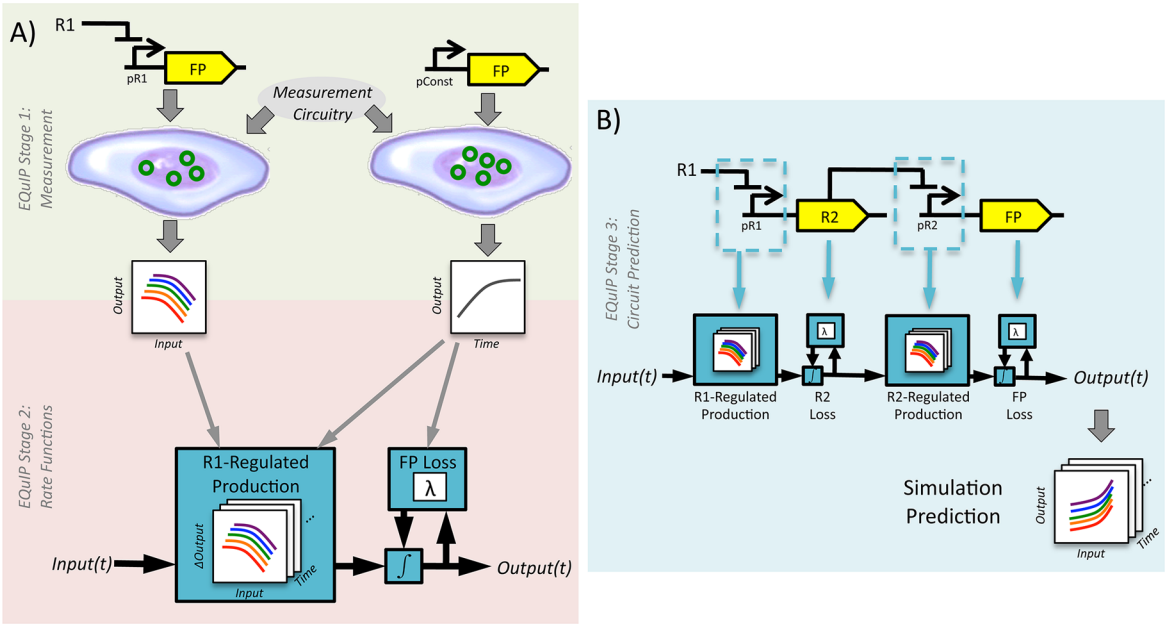


Figure 1. Stages of EQuIP from data gathering to predictions. (A) Experimental observation of the behavior of regulatory and constitutive elements in cells, where these elements are combined with additional biological circuitry for calibrated measurement (top: stage 1). Behavior of constitutive elements is measured over time. For regulatory elements, the relationship between input, circuit copy number (indicated by different colored lines), and output is measured at a single time point. Data is used to build rate functions for time-dependent regulated production and for loss of protein concentration, which can be mathematically integrated for computational simulation (bottom: stage 2). (B) The behavior of a biological circuit is predicted by linking production functions for each regulatory relation and loss functions for each relevant protein, according to circuit topology, then simulating concentrations over time according to the network of rate functions (stage 3).

expression levels.[20−23] For gene regulation, progress has been made toward accurate predictions of single self-regulating negative feedback transcriptional units,[24] which may exhibit decreased variability,[25] increased variability,[26] or oscillatory behavior,[27] depending on conditions. For multicomponent regulatory circuits, however, there is a critical need for a fundamentally new approach to prediction. Prior efforts have either focused on characterizing a complex circuit and then predicting the influence of modulating or replacing specific elements[11,27−29] or else have suffered from reduced precision when parts were first characterized and then subsequently used for predictions of a more complex circuit.

Prior methods for predicting multicomponent regulatory circuits have typically relied on explicit biochemical models,[10,28] such as Hill functions or chemical reaction networks, that depend strongly on the completeness and correctness of models of relevant cellular mechanisms. Such models are also frequently under-constrained by experimental data and thus require significant parameters to be set by heuristics or untested assumptions, rather than through direct (or indirect) experimental observations. This is a critical problem for predictions: the inherent uncertainty of an under-constrained model means that the same observation can be explained by a number of different sets of parameter values.[30] The predictive accuracy of a model is therefore impaired, because even if the model fits observations for one particular use, if the wrong parameters are chosen it is likely to fail on future predictions. In sum, although there has been much recent progress in characterization and prediction of genetic parts and circuits, even the behavior of a "simple" circuit such as a two-repressor cascade cannot generally be predicted accurately and reliably. We thus focus on the prediction of combinational genetic circuits (i.e., circuits without feedback or state), as both an important goal in its own right and as a step toward prediction of circuits with more complex dynamics such as oscillations and bistability.

We address the current challenges of multicomponent circuit prediction with a new method, Empirical Quantitative Incremental Prediction (EQuIP), that models expression of each gene using a piecewise function of regulatory inputs, circuit copy number, and time, based strictly on high-precision experimental observations. EQuIP predicts the behavior of a biological circuit by mathematically composing, in accordance with circuit topology, these gene expression models along with models of exponential dilution and decay. As can be expected, accurate circuit predictions require accurate models. To this end, EQuIP ensures that significant observable phenomena are incorporated in each gene expression model whether or not they agree with current mechanism hypotheses and also ensures that every parameter of the model is directly grounded in experimental data. Given current limitations in the understanding of biochemical mechanisms and in the ability to determine relevant parameter values through experiment, the flexibility of piecewise approximation is highly valuable. A piecewise function can directly approximate unmodeled or poorly modeled mechanisms and can substitute simple empirical functions for mechanisms whose parameters cannot be determined from observable data. EQuIP thus combines mechanistic models (i.e., derived from the underlying molecular processes) and phenotypic models (i.e., models aiming to capture observed behavior with minimal assumptions), using mechanistic models where the underlying parameters can be adequately determined and phenotypic approximation where they cannot. This combination greatly improves the accuracy with which the behavior of biological circuits can be predicted.

dx.doi.org/10.1021/sb500263b | ACS Synth. Biol. 2015, 4, 673−681

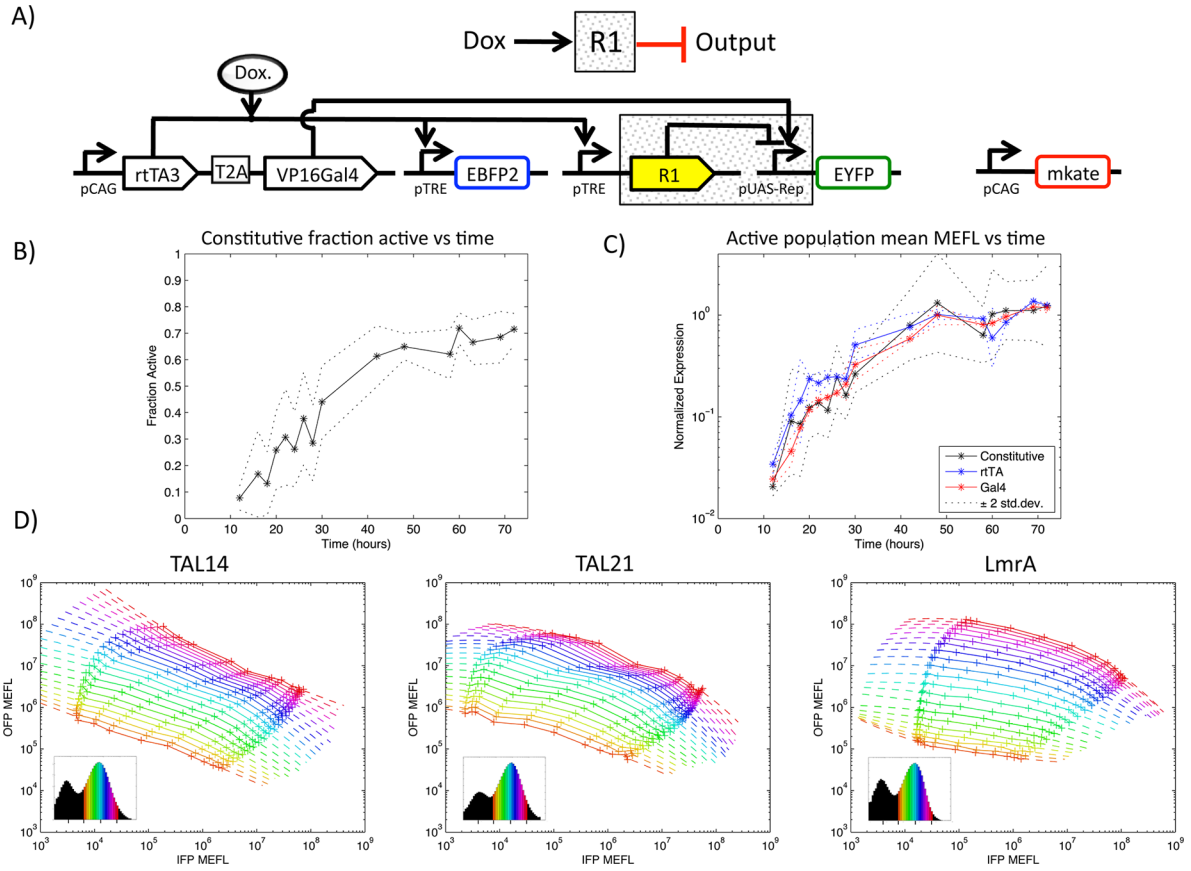


Figure 2. EQuIP characterization via time-series and dosage-response assays. (A) Biological circuit architecture for calibrated measurements, using fluorescent reporter proteins to quantify induced expression of repressor, regulated expression of output, and constitutive expression as an indicator of relative circuit copy number. EBFP2, a blue fluorescent protein, is input (IFP); EYFP, a yellow fluorescent protein, is output (OFP); mKate, a red fluorescent protein, is constitutive. (B) Time series characterization shows a linear increase in the fraction of cells constitutively expressing a fluorescent reporter, beginning a short time after transfection, until reaching saturation at approximately 70% transfection efficiency. Dotted lines show ±2 standard deviations. (C) Progression of mean fluorescence is similar for constitutive and activator-driven fluorescence, implying little impact from transcriptional activation delays. Normalized expression is computed by dividing by mean MEFL for t = 48−72. (D) Relations between input, copy number, and output for TAL14, TAL21, and LmrA: data from 12 different inducer dosages is segmented into subpopulations by constitutive fluorescence (plus marks) and grouped by subpopulation across dosages (colored lines). Insets show histograms of constitutive fluorescent protein expression used for segmenting the subpopulations; only colored bins have sufficient samples and separation from untransfected cells and are therefore included in the input/output curves. Extrapolation beyond the range measured in each transfer curve experiment is shown with dashed lines (see Supporting Information Section 6).

To provide circuit predictions, EQuIP operates in three main stages (Figure 1). In the first stage (Figure 1a top), we measure gene expression of constitutive and regulated elements using calibrated flow cytometry assays, with all measurements converted to equivalent standardized units. These measurements are taken at various combinations of time and regulatory inputs sufficient to precisely characterize the expression dynamics of all relevant circuit components. In the second stage, these measurements are used to compute two sets of rate functions (Figure 1a bottom). A regulated production function is a mapping from regulatory input (*e.g.*, concentration of a transcriptional repressor), circuit copy number, and time, to the gene production rate. A loss function specifies the rate at which a molecule's concentration decreases due to dilution or degradation. We can combine these two functions to describe the time evolution of a regulated gene product. In the third stage of EQuIP, the rate functions for multiple elements are combined to simulate the time evolution of more complex regulatory circuits (Figure 1b). Time evolution simulations are carried out by composing the rate functions according to the circuit topology and computing the integral with respect to time for various combinations of input and circuit copy number. As currently formulated, EQuIP can be applied to combinational circuits with relatively strong expression and low cross-interference, in conditions similar to those under which the devices were characterized. In the remainder of the paper, we present details of EQuIP and then demonstrate that its circuit simulations accurately predict experimental observations. In particular, we characterize three regulatory relations (transcriptional repressors TAL14, TAL21, and LmrA, each acting on a corresponding promoter) in mammalian HEK293 cells and use those characterizations to precisely predict the behavior of all six two-repressor cascades that can be made from these repressor/promoter pairs, as well as three feed-forward circuits constructed from the same elements. We conclude by evaluating the contribution of the various components of EQuIP to its precision in predicting composite circuit behavior.

RESULTS AND DISCUSSION

The goal of EQuIP is to predict the behavior of regulatory circuits, which we test in this paper through prediction of two-

dx.doi.org/10.1021/sb500263b | *ACS Synth. Biol.* 2015, 4, 673−681

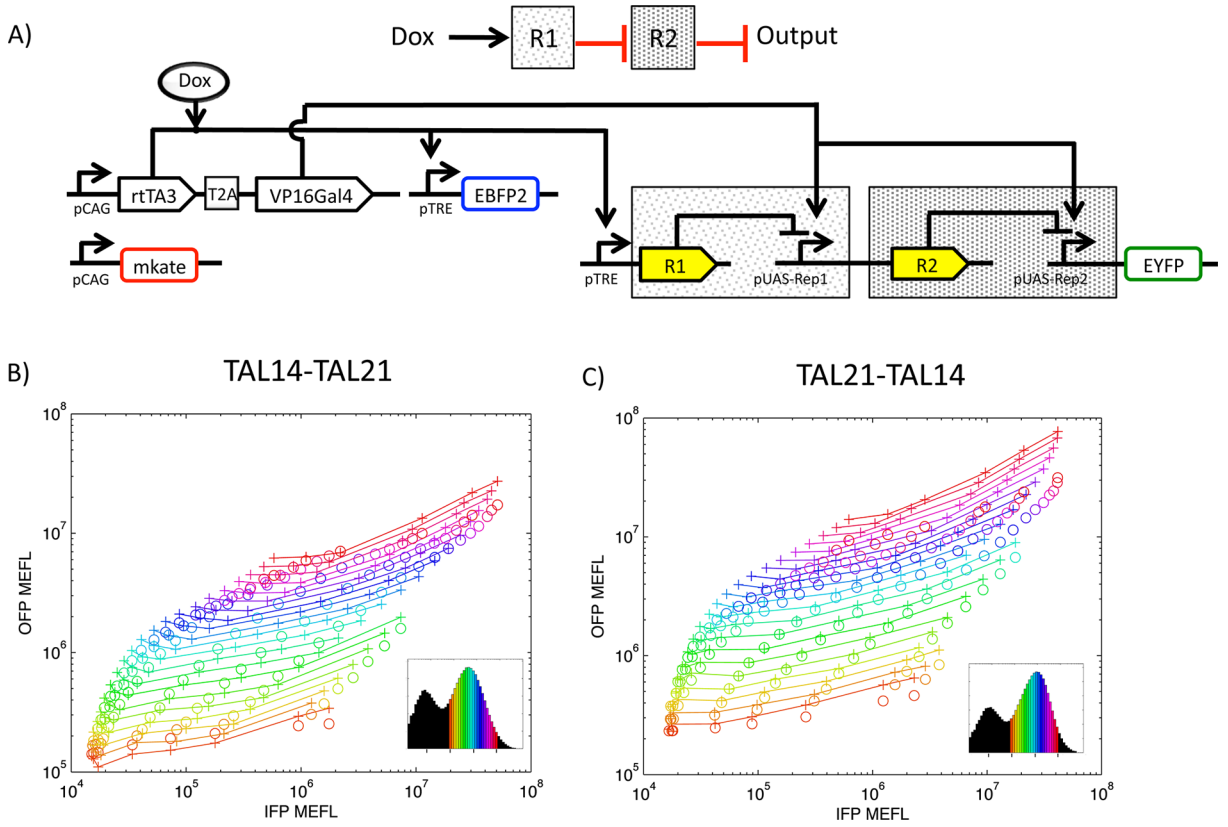


Figure 3. EQuIP produces close agreement between computational predictions and experimental data for two-stage cascades. (A) Biological circuit architecture for repressor cascades using TAL14, TAL21, and LmrA (repressor choice indicated as R1 and R2). Fluorescent reporters quantify input, output, and relative circuit copy number as in Figure 2. (B, C) Plots compare input/output relations for predictions (circles) and experimental data (pluses) for two of the cascades (others shown in Supporting Information Section 6). The predictions include all points that use extrapolation for less than 10% of the simulation steps. Insets as in Figure 2D.

repressor transcriptional cascades and feed-forward circuits. We model a transcriptional circuit using two types of functions (regulated production and loss), each taking current concentrations as input and yielding rates of concentration change as output. A transcriptional circuit, such as a two-repressor cascade, is then simulated by integrating a network of regulated production and loss functions, as described above. To restrict the scope of the problem, we consider only transient transfections of combinational circuits comprising orthogonal regulatory elements (see Supporting Information Section 2) and use each repressor/promoter pair at most once in any given circuit.

The first stage of EQuIP is to gather experimental data characterizing the regulated production dynamics for each repressor/promoter pair and the loss dynamics for each protein (Figure 1A, top). We characterize these dynamics with two experiments (Figure 2): a time series assay, which provides the mechanistic components of both production and loss models, and a dosage-response assay measured at a single time point, which provides the phenotypic components of the production model. To obtain precise and commensurate units in our models, we apply the TASBE protocol for calibrated flow cytometry, which allows us to use Molecules of Equivalent Fluorescein (MEFL)[31] as a consistent proxy unit for protein concentration (see Supporting Information Section 3).

All time-series and dosage-response characterization assays use circuits built with the same template (Figure 2A). The purpose of this circuit topology is to measure the behavior of a repressor/promoter pair at various levels of repressor concentration. We chose to regulate concentration of repressor by doxycycline/rtTA induction[32,33] as indicated by EBFP2. The concentration of the output gene product is indicated by EYFP, and constitutive mKate serves as a transfection marker and an indicator of relative circuit copy number.[34] All promoters that we characterize in this paper are hybrid promoters that also require Gal4 activation (see Supporting Information Section 2 for discussion of this modular approach to mammalian promoter design).

For the time series experiment, we measured constitutive and transcriptional activator driven expression for 72 h post-transfection. We found that the fraction of cells with observable (*i.e.*, above autofluorescence) constitutive expression of a fluorescent protein increases linearly over time, beginning some short time after transfection, and finally saturates at approximately 70% of cells (Figure 2b). This observation is consistent with typical lipofection protocols and a model whereby plasmids enter the nucleus during mitosis (per the standard lipofection hypothesis[35]) in an unsynchronized population of mammalian cells. Given the expected stability of our fluorescent and repressor proteins, this mechanism plus a constant rate of constitutive production can be used to create a quantitative model of transfection and fluorescent protein production. Fitting against both the percentage of expressing cells and mean constitutive mKate in expressing cells (Figure 2C) gives a mean initial delay of 25 h and cell division on average every 20 h (which correlates well with independent hemocytometer measurements as well; for detailed discussion of growth rate measurements, see Supporting Information

dx.doi.org/10.1021/sb500263b | *ACS Synth. Biol.* 2015, 4, 673−681

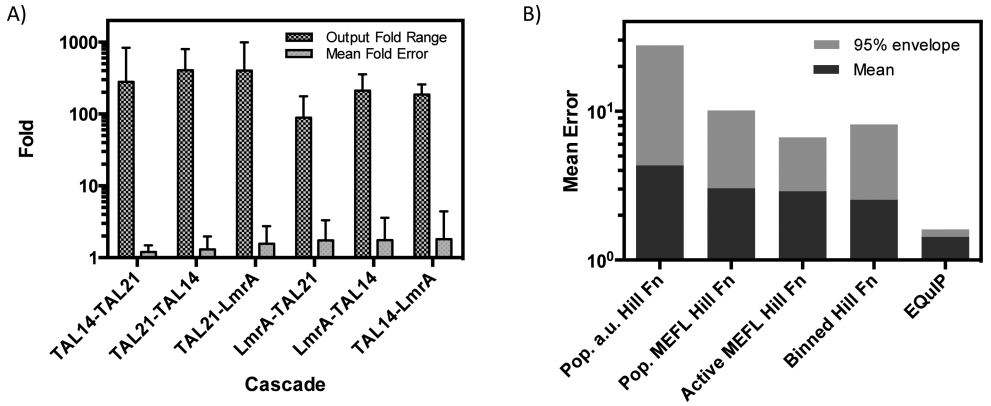


Figure 4. Comparison of the precision of EQuIP predictions for two-stage cascades. (A) Ratio of highest and lowest output means for all induction/copy-number subpopulations for given cascades (*e.g.*, the highest and lowest plus symbols in Figure 3B and C) and respective mean prediction errors. (B) Prediction errors for Hill function models fit to partial implementations of EQuIP vs the full EQuIP method, showing the improvement in both accuracy (mean error) and precision (95% envelope).

Section 4). We also note that expression of a fluorescent protein activated by constitutive rtTA3 and VP16Gal4 has no significant time lag compared to expression of mKate from a constitutive promoter, implying that, for this circuit, transcriptional activation delays are not significant and may be omitted from our production models. In contrast, even a short lag until repressors accumulate to the levels required for them to have significant impact on gene expression may result in strong transient expression from their respective promoters, and therefore, we model the time course of repressor accumulation.

From this experiment, we derive three mechanistic elements of our models. First, the cell division time provides a good approximation of the loss rate, since the proteins used are expected to be relatively stable. Second, the cell division time is also used to create an inverse function that takes an observed output expression and calculates the production rate over time that would produce that output expression (Supporting Information Section 5). This inverse function will be used to create the production rate function of a transcriptional unit from dosage-response data. Finally, simulations of gene expression take into account the mean initial delay in determining the length of time to simulate (Supporting Information Section 4).

We characterize each repressor device with a dosage-response experiment using the same characterization circuit (Figure 2A) as the time series experiment. For these experiments, we measure output as a function of input at a single point in time. We characterized the regulatory relationship between three transcriptional repressors (TAL14, TAL21, and LmrA; see Supporting Information Section 2) and a corresponding promoter for each, 72 h after transfection (Figure 2D). The observed relationships between input and output fluorescence were strongly affected by the relative number of circuit copies in the cells. Currently, there is not sufficient understanding of the underlying biological processes to create well-constrained models, based entirely on mechanistic principles, that accurately match the experimentally observed input/output relationships in Figure 2D. Instead, we estimate output gene expression for a given input level and relative copy number phenotypically by piecewise interpolation or extrapolation of the observed outputs (lines in Figure 2D). This is then transformed into a hybrid phenotypic/mechanistic model of regulated gene production using the inverse function derived from the time-series experiment (Supporting Information Section 5). Hence, this regulated production model retains every feature of the experimentally observed behavior.

To validate EQuIP, we constructed all two-repressor cascades comprising TAL14, TAL21, and LmrA following the architecture shown in Figure 3A. Figure 3B and C illustrates the 72 h input/output predictions vs experimental data for the TAL14-TAL21 and TAL21-TAL14 cascades, respectively. The experimentally observed output levels for different combinations of input and copy number have a wide range: across the six cascades, there is a 261-fold geometric mean difference between the highest and lowest subpopulation output means (*e.g.*, the ratio of highest and lowest plus symbols in Figure 3B and C). The mean error of predicted versus observed output across all input/copy-number combinations is only 1.6-fold for all six transcriptional cascades (Figure 4A). Predicting the output across many subpopulations also provides a prediction of the distribution of output expression for the overall population. The accuracy with which EQuIP predicts population mean and variation is even better than for individual subpopulations: the mean error of predicted versus experimentally observed output across all cascades and inductions is 1.4-fold for both population mean and population standard deviation (Supporting Information Section 6).

With such accuracy, EQuIP may guide circuit design and debugging. For example, EQuIP correctly predicts which combinations of repressors are best matched to provide the greatest differential expression between fully induced and uninduced states in the cascades. Specifically, EQuIP predicts that TAL14-TAL21 and TAL21-TAL14 cascades will have significantly stronger gain than all cascades involving LmrA (due to TAL21 and TAL14 having a better match in their dynamic ranges) and that TAL21-TAL14 will have approximately twice the gain of TAL14-TAL21, and these predictions are borne out by our experiments (Supporting Information Section 6).

We further evaluate the contribution to accurate prediction of different aspects of the EQuIP method. The first two stages of EQuIP (Figure 1A) consist of a sequence of data gathering and processing steps to produce the model for each device. Figure 4B evaluates the relative contribution of each step in this sequence to the final high prediction accuracy by comparing with a typical prior method—ordinary differential equations using Hill equations fit to device characterization data—applied to the data produced by each step in the data processing

dx.doi.org/10.1021/sb500263b | *ACS Synth. Biol.* 2015, 4, 673−681

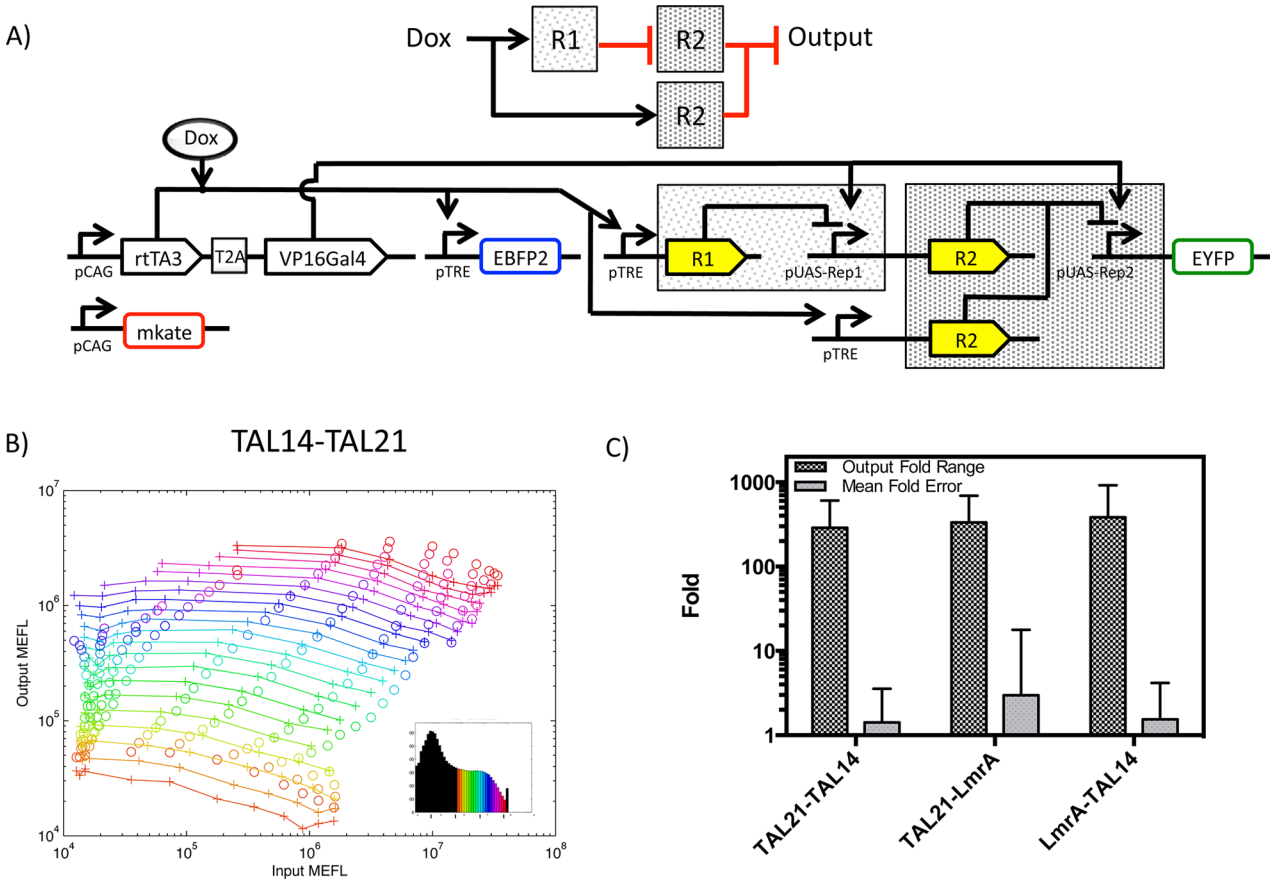


Figure 5. EQuIP produces close agreement between computational predictions and experimental data for feed-forward circuits. (A) Biological circuit architecture for feed-forward circuits using TAL14, TAL21, and LmrA (repressor choice indicated as R1 and R2). Fluorescent reporters quantify input, output, and relative circuit copy number as in Figure 2. (B) Plot compares input/output relations for predictions (circles) and experimental data (pluses) for one of the circuits (others shown in Supporting Information Section 6). The predictions include all points that use extrapolation for less than 10% of the simulation steps. Insets as in Figure 2D. (C) Ratio of highest and lowest output means for all induction/copy-number subpopulations for given feed-forward circuits (*e.g.*, the highest and lowest plus symbols in B) and respective mean prediction errors.

sequence. As described in Supporting Information Sections 3, 10, and 11, EQuIP first converts flow cytometry data to calibrated MEFL, then uses the constitutive marker to separate the subpopulation of successfully transfected cells, segments into bins by constitutive fluorescence level, then finally builds a piecewise model using the means of each bin. We thus created models generated from (1) population means in arbitrary units (our baseline model), (2) population means in calibrated MEFL, (3) population mean MEFL for transfected cells only, and (4) per-bin MEFL for transfected cells only. Each model is an ODE using Hill equations parametrized by curve fit against the observed data for each of the three repressors. We then compare the accuracy of EQuIP and these four models in predicting population means across the full range of doxycycline inductions of the various cascades. The results in Figure 4B indicate that the most important contributions to EQuIP's improved precision versus the baseline model come from calibration of measurement units and modeling with piecewise functions (full details in Supporting Information Section 7). The intermediate steps are prerequisites for piecewise models but do not appear to markedly improve prediction quality on their own. In all cases, the inaccuracies in prediction appear to derive primarily from the insufficient constraints that the experimental data provides for fitting Hill equations. Thus, although the fit often appears good (as shown in Supporting Information Section 7), this may not accurately represent the true system and may not correlate well with the predictive accuracy. Indeed, all three intermediate models show no statistically significant differences in performance (Supporting Information Section 7).

Finally, to validate the generalizability of the EQuIP method, we tested its efficacy in predicting the behavior of a more complex feed-forward circuit. This circuit, shown in Figure 5A, is similar to the cascade except that it adds a second path for repressing the output directly through Dox induction of the second repressor. We applied EQuIP to predict the behavior of all six possible feed-forward circuits 72 h post-transfection. We then constructed the three with the greatest variety of predicted behaviors and observed them experimentally. As before, we find that the predictions have a high mean accuracy, as illustrated by the example comparison of predicted and observed expression levels in Figure 5B. Across all three feed-forward circuits, there is an overall mean error of only 2.0-fold across a 333-fold geometric mean difference between the highest and lowest subpopulation output means (Figure 5C). Such a small degradation in accuracy in comparison to the cascades is expected for a more complicated circuit and is an indicator that EQuIP is likely to scale to even more complex circuits.

For synthetic biology to become a full-fledged engineering discipline, it must be possible to accurately predict the behavior of novel biological circuits from that of their constituent parts. Our results with EQuIP on six transcriptional cascades and

dx.doi.org/10.1021/sb500263b | *ACS Synth. Biol.* 2015, 4, 673−681

three feed-forward circuits are the first demonstration that highly accurate prediction of circuit behavior is possible and provide a quantitative benchmark for future efforts to be compared against. One of the most remarkable features of our results is that they are accomplished entirely through readily accessible observations of fluorescence, without any detailed biochemical analysis or modeling. This does not argue against the value of detailed biochemical modeling: indeed, we were able to use mechanistic biochemical models for the temporal components of our production and loss functions. Rather, our results demonstrate the power of using precise measurements to inform a composable model that describes precisely those phenomena that can be experimentally observed, no more and no less.

Our results show a number of opportunities where improved modeling or observation could further increase the quality of predictions. For example, Figure 3B and C and Supporting Information Figures 11 and 12 demonstrate that EQuIP generally provides its best predictions when neither circuit copy number nor induction are particularly high. One improvement would be to obtain additional input/output data at higher input levels or to genetically engineer regulatory devices with stronger responses at lower input levels, thus decreasing the amount of extrapolation required (Supporting Information Section 5). Another issue is that extreme fluorescence intensity values can be affected at the low end by cell autofluorescence and at the high end by PMT saturation (though this affects only the <5% of our predictions that are in these ranges); this can be addressed by assays that vary plasmid dosages and instrument settings for different levels of induction. Other possible areas for improvement are understanding the effect of large circuit copy numbers (which may influence cell behavior or create metabolic load), and accounting for retroactivity effects (*e.g.*, from one TALER regulating multiple downstream promoters) and cellular resource sharing (*e.g.*, VP16Gal4 as a driver for multiple hybrid promoters). With appropriate characterization experiments, such effects should be able to be incorporated as new empirical rate terms, similar to the production and loss models already in the EQuIP framework.

More generally, the specific EQuIP implementation presented here can be applied only to circuits that are combinational (meaning there is no feedback) and in which cells do not exhibit strongly divergent behaviors under the same conditions (*e.g.*, noise-induced bistability). Our results, however, provide a basis for extending EQuIP to larger and more complex circuits, and to circuits that include feedback (via initial state assumptions) and divergent populations (using distributions rather than means), as discussed in Supporting Information Section 5.1. The ability to predict circuit behavior is highly valuable for engineering biological systems, as it allows efficient selection of circuit elements and offers guidelines for optimization of devices to obtain a desired function. Accordingly, EQuIP supports the synthetic biology goal of creating libraries of modular, standard, and well-characterized components for rapid development of complex systems. Our framework may also be used for studying natural systems, although accurate predictions may initially be more difficult due to the complexity of many natural regulatory interactions. EQuIP thus forms a basis both for advances in design tools and for new investigations in systems biology. Combining these advances with emerging libraries of biological devices will usher in a new era of exponential growth in our ability to engineer biological systems.

METHODS

Culture Conditions. HEK 293 FT cells (Invitrogen) were cultured in DMEM medium (CellGro), supplemented with 10% FBS (PAA Laboratories), 2 mM L-glutamine (CellGro), 1% Strep/pen (CellGro), 1% non-essential amino acids (NEAA) (HyClone), and 10 000× Fungin (Invivogen) at 37 °C and 5% CO_2. Cells were passaged in a 100 mm dish by removing culture media, adding 2 mL 0.05% trypsin, waiting at room temperature for 2 min and then resuspending the cells in 5 mL of cell culture media and diluting to desired concentration with additional cell culture media.

Transfection. Transfections were carried out with Metafectene Pro (Biontex Laboratories). Cells were seeded 1 day prior at 2×10^5 cells per well in a 24 well plate. 500 ng of DNA was mixed into 60 μL of DMEM (without supplements). 1.5 μL of Metafectene was then added and the tube was gently mixed and kept at room temperature for 15 min to form the DNA−liposome complex. Fresh media was added to the cells directly prior to transfection (500 μL of DMEM with supplements). The DNA−Metafectene solution was then added dropwise to the well. Induction of the circuit was performed at this time as well by the addition of a small molecule (*i.e.*, doxycycline). The media was subsequently changed daily with the appropriate amount of inducer. Each circuit was realized with each transcriptional unit encoded on a separate plasmid, for a total of 6 plasmids (5 plasmid circuits add a blank plasmid), and cotransfected: see Supporting Information Section 2 for details of promoter design and Supporting Information Section 8 for plasmid sequences. The DNA for the circuits transfected were in the ratio 1:3:3:1:1:1; where the transcriptional units that contained the "TRE" promoter were the ones that were transfected at 3× the amount of the others for signal matching purposes. For the time series experiment, we measured EBFP2 with 2000 nM Dox for rtTA activation. EYFP was measured for Gal4 activation with 0 nM Dox, and we used TAL21 for R1. We also assumed that the time dynamics are not significantly affected by choice of repressor. Dose−response data was taken with a logarithmic series of Dox dosages. Feed-forward circuits data was taken in a separate experiment with a slight variation on the protocol. See Supporting Information Section 9 for more details. The cell-to-cell variation due to intracellular variation in copy number was typically small in our experiments (Supporting Information Section 10).

Flow Cytometry. Flow cytometry data was taken at 72 h post transfection. Cells were again trypsinized as previously described. The cells were then centrifuged at 150*g* for 10 min at 4 °C. The supernatant was removed and the cells were resuspended in 1× PBS that did not contain calcium or magnesium. A BD LSR Fortessa was used to take flow cytometry measurements with the following settings: EBFP2, measured with a 405 nm laser and a 450/50 filter, EYFP, measured with a 488 nm laser and a 530/30 filter, and mKate, measured with a 561 nm laser and a 610/20 filter. Flow cytometry data was analyzed as described in Supporting Information Sections 3, 11, and 12. EQuIP also included internal cross validation and checks to determine the quality of the data collected and identify potential experimental problems (Supporting Information Section 13).

Cloning. Creation of the plasmids used for this project was carried out using the Gateway system from Invitrogen. We used a multisite cloning strategy with two entry vectors. One entry

dx.doi.org/10.1021/sb500263b | *ACS Synth. Biol.* 2015, 4, 673−681

vector contained the promoter and the other contained the transcription factor or gene. The destination vector was modified from its original sequence to contain an insulator 5′ to L4 and a polyadenylation signal 3′ to the R1 site.

■ ASSOCIATED CONTENT

Supporting Information

Details of repressor/promoter design, sequences of all constructed used, details on calibration of flow cytometry data, measurement and estimation of cell division rates, details of modeling and prediction, details of experimental results, and additional experimental and analytical method details. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rweiss@mit.edu.

Author Contributions

[‖]N.D. and J.B. contributed equally. N.D. designed and performed repressor characterization and two-stage cascade experiments, analyzed data, and wrote the manuscript. J.B. designed experiments, developed and applied computational analysis and prediction techniques, and wrote the manuscript. S.K. designed and performed feed-forward circuit experiments, analyzed data, and wrote the manuscript. A.A. and F.Y. developed and applied computational analysis and prediction techniques. Y.L. and Z.X. built initial versions of some DNA constructs. R.W. designed experiments, helped develop the computation framework, analyzed data, and wrote the manuscript.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Lucks, J. B., Qi, L., Whitaker, W. R., and Arkin, A. P. (2008) Toward scalable parts families for predictable design of biological circuits. *Curr. Opin. Microbiol. 11*, 567−573.

(2) Alon, U. (2007) Network motifs: Theory and experimental approaches. *Nat. Rev. Genet. 8*, 450−461.

(3) Purnick, P. E. M., and Weiss, R. (2009) The second wave of synthetic biology: From modules to systems. *Nat. Rev. Mol. Cell Biol. 10*, 410−422.

(4) Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R.-Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z.-Q., Segall-Shapiro, T. H., Calvey, C. H., Parmar, P. P., Hutchison, C. A., 3rd, Smith, H. O., and Venter, J. C. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science 329*, 52−56.

(5) Carr, P. A., and Church, G. M. (2009) Genome engineering. *Nat. Biotechnol. 27*, 1151−1162.

(6) Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods 6*, 343−345.

(7) Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009) Golden gate shuffling: A one-pot DNA shuffling method based on Type IIs restriction enzymes. *PloS One 4*, e5553.

(8) Guye, P., Li, Y., Wroblewska, L., Duportet, X., and Weiss, R. (2013) Rapid, modular, and reliable construction of complex mammalian gene circuits. *Nucleic Acids Res. 41*, e156.

(9) Litcofsky, K. D., Afeyan, R. B., Krom, R. J., Khalil, A. S., and Collins, J. J. (2012) Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. *Nat. Methods 9*, 1077−1080.

(10) Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C., and Voigt, C. A. (2012) Genetic programs constructed from layered logic gates in single cells. *Nature 491*, 249−253.

(11) Tabor, J. J., Salis, H. M., Simpson, Z. B., Chevalier, A. A., Levskaya, A., Marcotte, E. M., Voigt, C. A., and Ellington, A. D. (2009) A synthetic genetic edge detection program. *Cell 137*, 1272−1281.

(12) Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R., and Benenson, Y. (2011) Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science 333*, 1307−1311.

(13) Ausländer, S., Ausländer, D., Müller, M., Wieland, M., and Fussenegger, M. (2012) Programmable single-cell mammalian biocomputers. *Nature 487*, 123−127.

(14) Lu, T. K., Khalil, A. S., and Collins, J. J. (2009) Next-generation synthetic gene networks. *Nat. Biotechnol. 27*, 1139−1150.

(15) Hooshangi, S., Thiberge, S., and Weiss, R. (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl. Acad. Sci. U.S.A. 102*, 3581−3586.

(16) Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., de Mora, K., Glieberman, A. L., Monie, D. D., and Endy, D. (2009) Measuring the activity of BioBrick promoters using an *in vivo* reference standard. *J. Biol. Eng. 3*, 4.

(17) Canton, B., Labno, A., and Endy, D. (2008) Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol. 26*, 787−793.

(18) Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. (2005) Gene regulation at the single-cell level. *Science 307*, 1962−1965.

(19) Beal, J.; Weiss, R.; Yaman, F.; Davidsohn, N.; Adler, A. (2012) *A Method for Fast, High-Precision Characterization of Synthetic Biology Devices*; Computer Science and Artificial Intelligence Laboratory Technical Report; Massachusetts Institute of Technology, Cambridge, MA.

(20) Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods 10*, 354−360.

(21) Liu, C. C., Qi, L., Lucks, J. B., Segall-Shapiro, T. H., Wang, D., Mutalik, V. K., and Arkin, A. P. (2012) An adaptor from translational to transcriptional control enables predictable assembly of complex regulation. *Nat. Methods 9*, 1088−1094.

(22) Kosuri, S., Goodman, D. B., Cambray, G., Mutalik, V. K., Gao, Y., Arkin, A. P., Endy, D., and Church, G. M. (2013) Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A. 110*, 14024−14029.

(23) Salis, H. M. (2011) The ribosome binding site calculator. *Methods Enzymol. 498*, 19−42.

(24) Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. (2007) Accurate prediction of gene feedback circuit behavior from component properties. *Mol. Syst. Biol. 3*, 143.

(25) Becskei, A., and Serrano, L. (2000) Engineering stability in gene networks by autoregulation. *Nature 405*, 590−593.

(26) Hooshangi, S., and Weiss, R. (2006) The effect of negative feedback on noise propagation in transcriptional gene networks. *Chaos 16*, 026108.

(27) Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S., and Hasty, J. (2008) A fast, robust, and tunable synthetic gene oscillator. *Nature 456*, 516−519.

(28) Ellis, T., Wang, X., and Collins, J. J. (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol. 27*, 465−471.

(29) Lou, C., Stanton, B., Chen, Y.-J., Munsky, B., and Voigt, C. A. (2012) Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol. 30*, 1137−1142.

(30) Strang, G. (2003) *Introduction to Linear Algebra*. Society for Industrial and Applied Mathematics (SIAM), Bangkok, Thailand.

(31) *Measuring Molecules of Equivalent Fluorescein (mefl), pe (mepe), and rpe-cy5 (mepcy) Using Sphero Rainbow Calibration Particles*;

Technical Report SpheroTechnical Notes: STN-9; (2001) SpheroTech, Libertyville, IL.
(32) Das, A. T., Zhou, X., Vink, M., Klaver, B., Verhoef, K., Marzio, G., and Berkhout, B. (2004) Viral evolution as a tool to improve the tetracycline-regulated gene expression system. *J. Biol. Chem. 279*, 18776−18782.
(33) Markusic, D., Oude-Elferink, R., Das, A. T., Berkhout, B., and Seppen, J. (2005) Comparison of single regulated lentiviral vectors with rtTA expression driven by an autoregulatory loop or a constitutive promoter. *Nucleic Acids Res. 33*, e63−e63.
(34) Bleris, L., Xie, Z., Glass, D., Adadey, A., Sontag, E., and Benenson, Y. (2011) Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Mol. Syst. Biol. 7*, 519.
(35) Lechardeur, D., Verkman, A. S., and Lukacs, G. L. (2005) Intracellular routing of plasmid DNA during non-viral gene transfer. *Adv. Drug Delivery Rev. 57*, 755−767.