ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

Reverse Engineering Validation using a Benchmark Synthetic Gene Circuit in Human Cells

Taek Kang,[†,§] Jacob T. White,[†,§] Zhen Xie,[∥] Yaakov Benenson,[⊥] Eduardo Sontag,*[,¶] and Leonidas Bleris*[,†,‡,§]

[†]Bioengineering Department, [‡]Electrical Engineering Department, and [§]Center for Systems Biology, University of Texas at Dallas, 800 West Campbell Road, Richardson, Texas 75080, United States

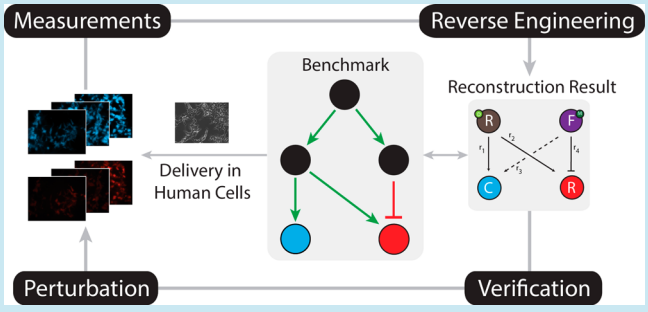
[∥]Center for Synthetic and Systems Biology, Bioinformatics Division, TNLIST, 100084, Tsinghua University, Beijing, China

[⊥]Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule (ETH) Zürich, Mattenstrasse 26, 4058 Basel, Switzerland

[¶]Department of Mathematics, Rutgers-The State University of New Jersey, Hill Center, 110 Frelinghuysen Road, Piscataway, New Jersey 08854, United States

Supporting Information

ABSTRACT: Multicomponent biological networks are often understood incompletely, in large part due to the lack of reliable and robust methodologies for network reverse engineering and characterization. As a consequence, developing automated and rigorously validated methodologies for unraveling the complexity of biomolecular networks in human cells remains a central challenge to life scientists and engineers. Today, when it comes to experimental and analytical requirements, there exists a great deal of diversity in reverse engineering methods, which renders the independent validation and comparison of their predictive capabilities difficult. In this work we introduce an experimental platform customized for the development and verification of reverse engineering and pathway characterization algorithms in mammalian cells. Specifically, we stably integrate a synthetic gene network in human kidney cells and use it as a benchmark for validating reverse engineering methodologies. The network, which is orthogonal to endogenous cellular signaling, contains a small set of regulatory interactions that can be used to quantify the reconstruction performance. By performing successive perturbations to each modular component of the network and comparing protein and RNA measurements, we study the conditions under which we can reliably reconstruct the causal relationships of the integrated synthetic network.



KEYWORDS: *reverse engineering, benchmark synthetic circuits, human cells, modular response analysis*

The reverse engineering question has been pursued increasingly since the advent of molecular biology, and the methods have gradually shifted from manual, intuitive pathway reconstructions to high-throughput computational techniques. The latter methods usually consist of collecting experimental data, performing computer-aided data analysis, and drawing conclusions, which guide further experiments. A successful implementation of this cycle requires high-quality data, adequate models and algorithms, and confidence that the interpretation is correct. There are many possible issues with the experimental techniques used to generate data and with the algorithmic tools designed to interpret these data, but most importantly, uncertainty stems from our inability to independently verify the conclusions suggested by reverse engineering tools.[1,2]

Current reverse engineering (RE) methods differ in every possible dimension of experimental techniques and computational analyses.[3−5] While each method has demonstrated successful network reconstructions on its own, there are no accepted standards to compare their relative strengths and weaknesses due to major differences in the types of data sets and analyses used. In fact, the lack of unifying standards and procedures for validation, along with the high degree of expertise required for computational algorithms, can be regarded as one of the major obstacles that prevents the widespread use of network inference methods. To address this issue, a community-wide effort, DREAM (Dialogue for Reverse Engineering Assessments and Methods), has been initiated to facilitate discussion and refine existing methodologies, resulting in valuable insights about relationships between algorithm performance and experimental parameters.[6−8]

Special Issue: IWBDA 2012

Received: September 21, 2012
Published: March 18, 2013

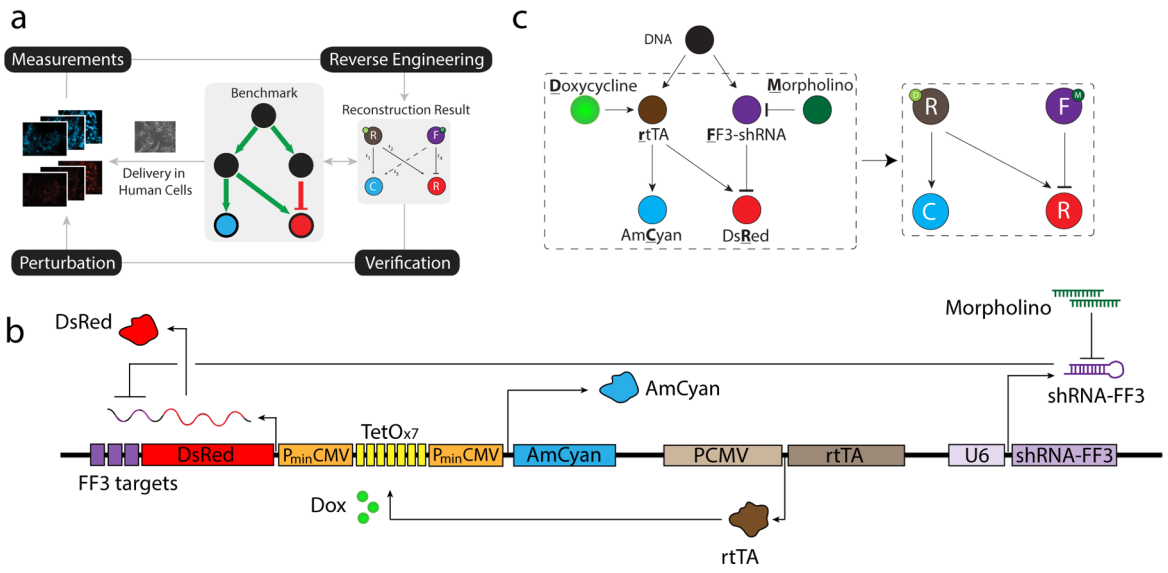


Figure 1. A platform for assessment of reverse engineering using a synthetic circuit. (a) A workflow diagram of the proposed platform. The benchmark architecture is stably integrated in a human cell line. A reverse engineering algorithm undergoes validation using perturbations and by verifying the results using the benchmark. (b) The synthetic circuit delivered to human embryonic kidney cells using FLP recombinase-mediated stable integration (Flp-In). In the presence of doxycycline, a constitutively transcribed reverse tetracycline-induced transactivator (rtTA) induces transcription of both fluorescent proteins by binding to the tetO enhancer region of the bidirectional promoter. A short-hairpin RNA, which is also constitutively transcribed, actively represses translation of DsRed by binding to the target sequence present in 3′ UTR of the DsRed mRNA transcript. Addition of a morpholino oligo reduces the shRNA activity by protecting the 3′ UTR target site of the shRNA. (c) A diagram depicting the benchmark synthetic gene circuit. The circuit consists of two distinct regulatory elements with different mechanisms to control the expression levels of AmCyan and DsRed fluorescent proteins.

The assessment and verification of RE algorithms[9,10] for pathway reconstruction is a critical issue. Results in yeast[11] highlight the usefulness of synthetic circuits for this purpose, but the problem remains largely unsolved in human cells, which are likely the most significant (in terms of the scientific and broader impacts potential) and complex (in terms of the theoretical and experimental issues) platform.

In this paper, we describe an experimental and theoretical procedure to refine and validate biological network inference in human kidney cells. We construct and stably integrate a synthetic gene circuit that resembles a natural network topology, establishing an independent, versatile benchmark system that can assess the performance of a reverse engineering algorithm (Figure 1a). The proposed reverse engineering procedure consists of the following steps. First, a small scale network is stably integrated in mammalian cells. Then, the individual nodes of this network are weakly perturbed from their steady state. The pre- and post-perturbation steady states are measured and fed into a reverse engineering algorithm to predict the network structure. The results of the algorithm are compared against the known network structure and are used to adjust the parameters of the algorithm and set guidelines for future experiments. These parameters include the perturbation magnitudes, the data collection and processing techniques, and the details of computational processing.

As our baseline reverse engineering method we use an approach based on Modular Response Analysis (MRA), where we take experimentally measured steady-state responses following near-linear perturbation of each modular component of the benchmark system.[12−14] The criteria for the selection of MRA as "iteration zero" RE method is based on our experience[12,13,15] with the method and published experimental work that successfully used the method,[16,17] because it reveals network structure with a relatively simple experiment setup, and importantly on the basis of our conviction that the method is indeed best suited for transitioning from benchmark circuits to endogenous pathways.

We believe that the use of synthetic circuits as benchmarks for reverse engineering has important advantages. First, the synthetic circuits can be engineered to be compatible with commonly available perturbation (e.g., siRNA or small molecule) and data acquisition (e.g., qRT-PCR or fluorescent microscopy) methods. Furthermore, the versatility and diversity of building blocks for synthetic circuits allows for engineering a wide range of topologies and functions, while being orthogonal to endogenous signaling networks. Once the network is stably integrated in cells and its topology is experimentally confirmed, it can serve as a basis for validating specific aspects of reverse engineering.

RESULTS AND DISCUSSION

Benchmark Synthetic Network. We engineered a synthetic regulatory network that consists of two fluorescent reporters (AmCyan and DsRed) subject to control by two distinct regulatory elements (Figure 1b). Each of the regulatory elements can be controlled by chemical ligands (at a concentration nontoxic to the cell), and their transcriptional products do not interact with endogenous cellular signaling. This simple circuit allows the study of scenarios where the output fluorescent proteins are subject to a range of heterogeneous inputs: no activation, single source of activation, and combination of activation and repression.

The first of the two regulatory units, rtTA, is based on the tetracycline-inducible expression system (Tet-On)[18] and is responsible for initiating transcription of both fluorescent reporters, AmCyan and DsRed, by binding to a bidirectional promoter. The activation of the TRE enhancer by rtTA can be controlled by varying the amount of doxycycline. The second

dx.doi.org/10.1021/sb300093y | ACS Synth. Biol. 2013, 2, 255−262

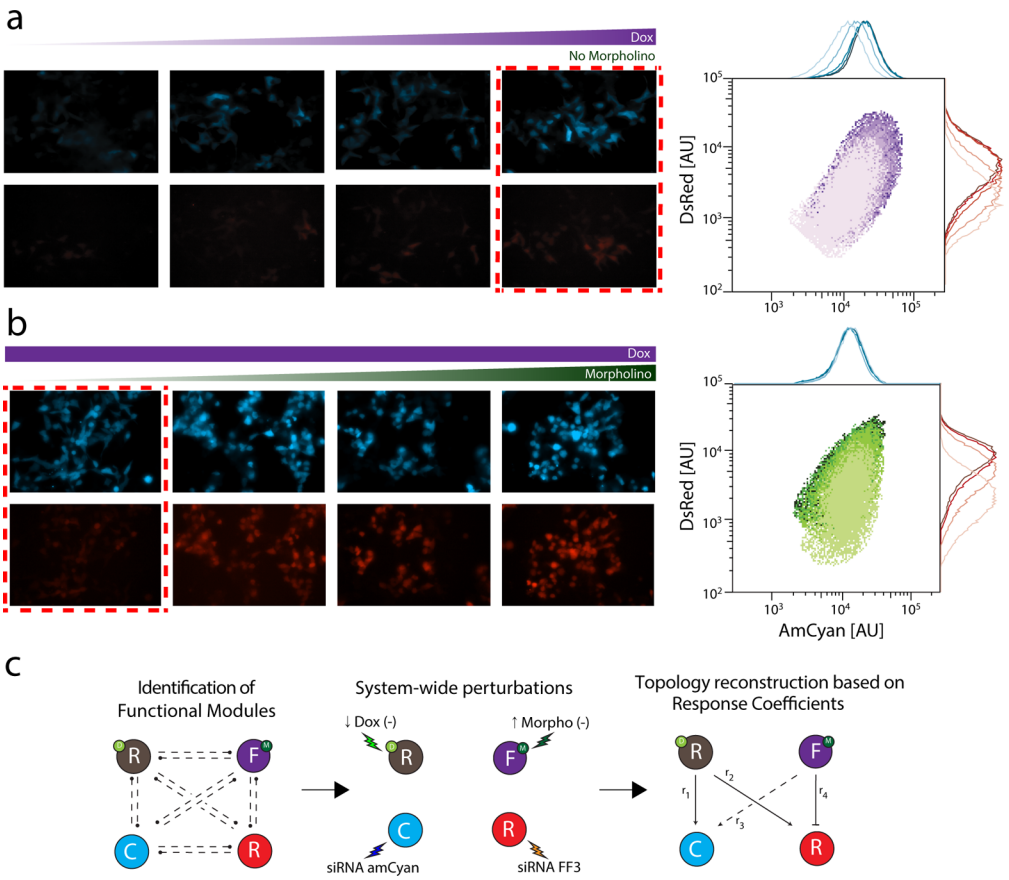


Figure 2. Characterization experiments. (a) Fluorescence microscopy and gated flow cytometry results of doxycycline titration in the absence of morpholino. (b) Fluorescence microscopy and gated flow cytometry results of morpholino titration for fixed saturated doxycycline concentration. Doxycycline concentration ranged from 1 ng/mL to 10 μg/mL, while morpholino concentration ranged from 0 to 5 nmol/mL. Images surrounded by red box shows identical condition: 10 μg/mL doxycycline and 0 nmol/mL morpholino. (c) Schematic representation of modular response analysis. Semiquantitative sensitivity analysis is performed by applying systematic perturbations to each modular component. The resulting change in activity of each module is determined as global response coefficient.

source of regulation is RNA interference in the form of short-hairpin RNA (shRNA). The shRNA was constructed by inserting the FF3 stem-loop in pSiren (Clontech). A gene can be made susceptible to down-regulation by an arbitrary miRNA by incorporating targets into its 3′-untranslated region (UTR).[19,20] By inserting three repeats of FF3 target sequence in the 3′ UTR of the DsRed transcript, we down-regulate the expression of the fluorescent protein. The activity of shRNA, which is constitutively transcribed from a U6 promoter, can be modulated by introducing morpholino oligos (GeneTools). The antisense morpholino protects the transcript from degradation by irreversibly binding to the 3′ UTR target sequence of DsRed.[21,22]

The expected behavior of the network is best illustrated as a four-node system with three edges. A pair of activation edges from DOX-rtTA node to each fluorescent output node are used in order to indicate doxycycline dependence of the bidirectional promoter, and a third inhibition edge connects the shRNA node to DsRed (Figure 1c). We note here that the input variable morpholino represses the inhibitory effects of shRNA, resulting to an overall positive action on the output, but in reporting the reconstruction results we examine the negative connection between the shRNA and the dsRed protein.

Characterization of the Synthetic Circuit. The individual parts of the synthetic circuit (Figure 1b) were first cloned into a single vector (Supporting Information, Cloning), and the cassette was then integrated stably in a FLP-In HEK 293 cell line (Invitrogen). We first examined the behavior of the network in response to titrations of the chemical ligands doxycycline and morpholino (Figure 2). We performed doxycycline titration in the absence of morpholino and morpholino titration for fixed saturated doxycycline concentration (10 μg/mL) (Figure 2a and b). The doxycycline concentration ranged from 1 ng/mL to 10 μg/mL, while morpholino concentration ranged from 0 to 5 nmol/mL. All fluorescence microscopy and flow cytometry measurements were performed between 48 to 56 h after addition of chemical ligands, when the concentration of fluorescent proteins is at quasi-steady state. During the analysis of fluorescence using flow cytometry we discovered that, although stringent selection of cells positive for the circuit integration was possible, a portion of these cells exhibited leaky transgene expression (Supplementary Figure 1). To eliminate this population, we include in the analysis only the cells responsive to both input variables by appropriate gating (Supplementary Figures 1 and 2). The titrations result in output fluorescent protein measurements consistent with the expected behavior of our network topology. Expression levels of both fluorescent reporters were up-regulated in response to increasing doxycycline. The addition of morpholino results in a significant increase in DsRed intensity but not AmCyan, which indicated that the morpholino successfully interferes with the shRNA

dx.doi.org/10.1021/sb300093y | ACS Synth. Biol. 2013, 2, 255−262

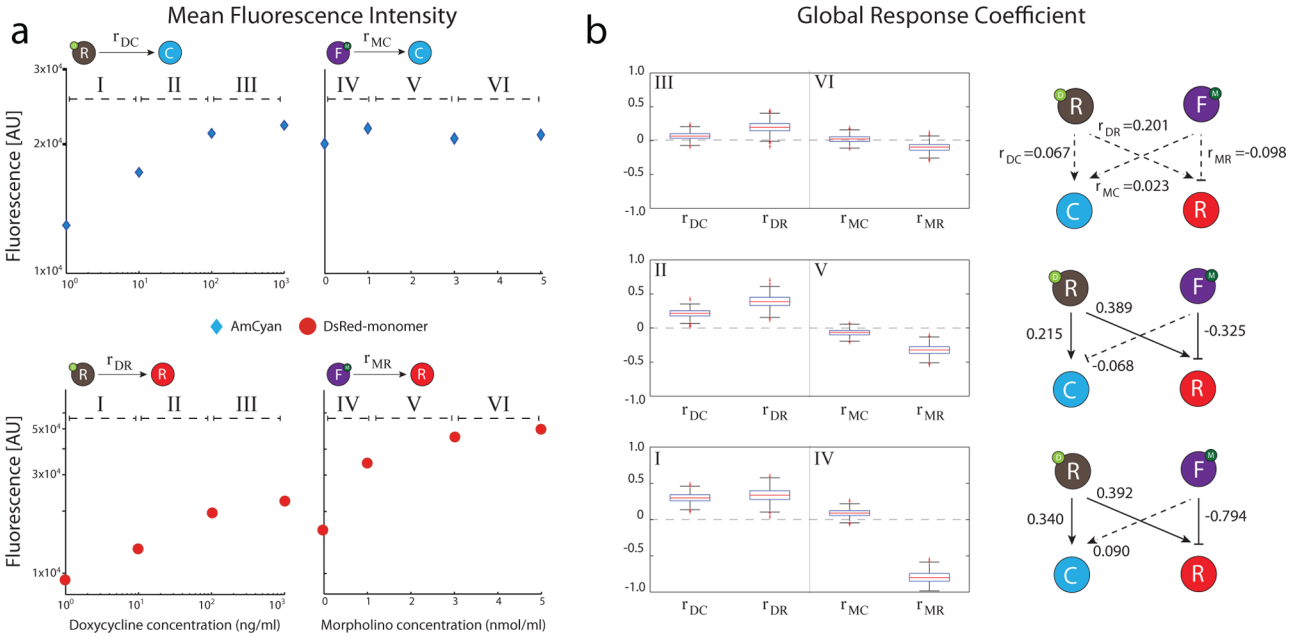


Figure 3. Reverse engineering using flow cytometry protein measurements. (a) Cells harboring the synthetic circuit construct were perturbed with various concentrations of doxycycline and morpholino. After 48 h of induction, cells were harvested, and their expression of AmCyan and DsRed was measured via fluorescence using flow cytometry ($n \geq 100{,}000$). After gating of flow cytometry results to include only the cells responsive to chemical perturbations ($n' \geq 10{,}000$), the mean of AmCyan and DsRed fluorescence is plotted as a function of doxycycline and morpholino concentration. The doxycycline titration was performed while keeping morpholino at 0 nmol/mL, and morpholino titration was performed with doxycycline concentration of 10 μg/mL. To simulate different perturbation conditions, each stepwise change in chemical concentration is defined as intervals I–VI. (b) Global response coefficient of a given interval. Bootstrap resampling method is used to obtain the confidence interval of the response coefficients. For the resampling method, 120 random cells are drawn from pre- and post-perturbation, and their average is used to calculate the response coefficient. Such calculation is performed 1,000 times, and the result is shown as a boxplot showing 95% confidence interval (whiskers). Red + indicates outliers. Networks on the right show complete network reconstruction results. Values denote the median response coefficient confidence interval. Dotted lines represent cases where the 95% confidence interval of the corresponding obtained from simulation includes zero, whereas solid lines indicate cases where it does not.

function and that the shRNA is indeed responsible for DsRed down-regulation.

Reverse Engineering Approach. After completing the dose–response profile of the circuit, we attempted to reconstruct the network topology without using our prior knowledge of the circuit. We used a top-down reverse engineering approach to extract interactions from steady-state perturbation experiments. The method is based on modular response analysis[13] and assumes that the biological network is a collection of monotone modules represented by simultaneous output measurements x_i, such as steady-state concentrations of protein or mRNA levels. These quantities are thought of as state variables in a dynamical system, which is the set of differential equations, $dx_i/dt = f(x_i,p_i)$ where p_i is a set of input parameters. One introduces coefficients r_{ij} obtained from partial derivatives of f as a measure of pairwise interaction strengths between nodes. The main objective of modular response analysis is to obtain the signs of the pairwise interactions, which represent the nature of the influence exerted by one node onto another. In cases where no interaction exists, r_{ij} should be identified as zero. For our benchmark architecture and in order to obtain the interactions between nodes, we have to obtain only the global response coefficients (GRC), obtained experimentally by calculating $\Delta\ln(x_i)$, where x_i represents the steady-state concentration of a state variable, such as protein or mRNA. Once the functional modules (i.e., perturbation targets) of the target network have been selected, the experimental procedure consists of the following steps (Figure 2c): (a) measure the steady-state x_i corresponding to the unperturbed set of inputs p_i, (b) perform a perturbation to each p_i individually and measure the new steady state, (c) calculate the global response coefficients using the steady-state data. We use the synthetic circuit as a benchmark to validate the reconstruction results, and we probe specific perturbation and measurement parameters. First, we examine the impact of the experimental perturbation range on the quality of network reconstruction. Second, we probe the consistency of the topology reconstruction between protein (flow cytometry) or mRNA (qRT-PCR) measurements.

Network Reconstruction Using Flow Cytometry Data. For the first question, we defined a set of perturbation ranges to test the general reconstruction performance. By taking advantage of the dose-dependent dynamics of the circuit, we treated each stepwise decrease in concentration of doxycycline and morpholino during the titration experiment as a perturbation and obtained the response coefficients for these intervals by calculating the log fractional change in each fluorescence reading (Figure 3). We observed that as doxycycline concentration increases (Figure 3a) the mean fluorescence of AmCyan and DsRed increase. Moreover, morpholino up-regulates DsRed expression. As a single set of complete topology reconstruction consists of the system's response to each input, we took a pair of single-step intervals (annotated in roman numerals) from each titration data to denote the perturbations and proceeded with the network reconstruction based on the chosen intervals. In total, we defined six intervals to represent equal numbers of separate perturbation scenarios. The three intervals I–III represent 10-

dx.doi.org/10.1021/sb300093y | ACS Synth. Biol. 2013, 2, 255–262

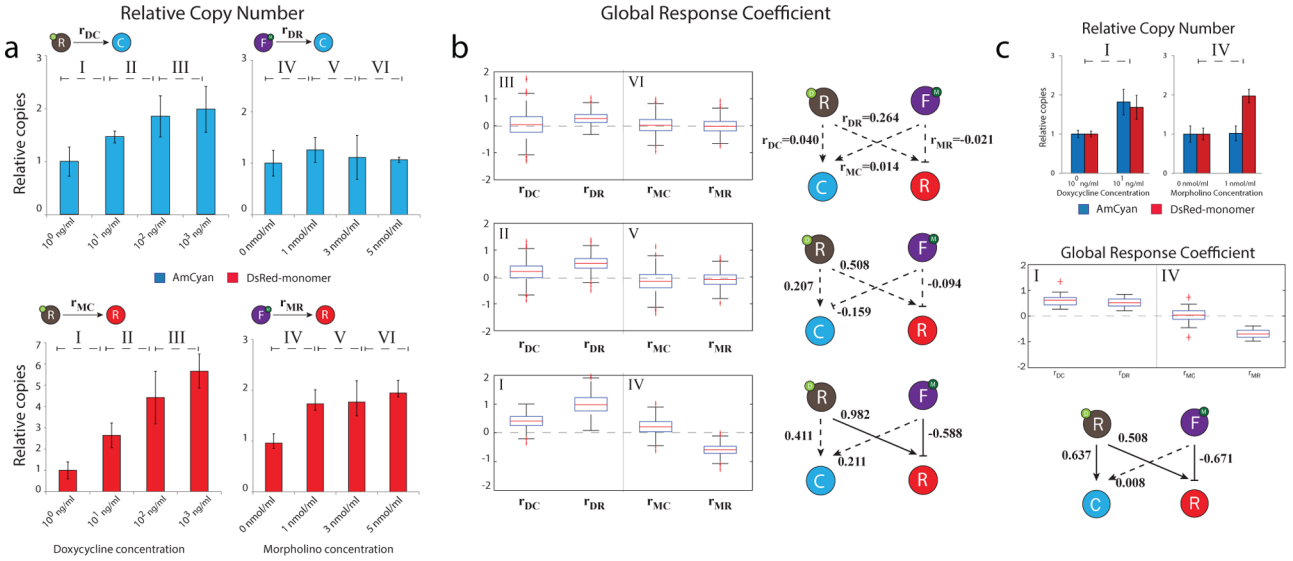


Figure 4. Reverse engineering using qRT-PCR mRNA measurements. (a) Cells stably integrated with the synthetic circuit were perturbed with various concentrations of doxycycline and morpholino. Total mRNA was extracted 48 h post-induction, and relative abundance of fluorescent output mRNA was measured using qRT-PCR with $\Delta\Delta C_t$ method. The data shown consist of three separate experiments with each with triplicate PCR results, and error bars indicate standard error of the mean. Each stepwise decrease in perturbation agent is labeled as a single interval (denoted by roman numerals) to simulate perturbation conditions for acquisition of response coefficients. (b) Global response coefficients of each simulated perturbation interval. To quantify the accuracy of the result, Monte Carlo simulation was used to build a confidence interval of the response coefficients. The results are displayed as boxplots showing 95% confidence intervals (whiskers). Networks on the right shows complete network reconstruction results. Values denote the median response coefficient confidence interval. Dotted lines represent cases where the 95% confidence interval of the corresponding obtained from simulation includes zero, whereas solid lines indicate cases where it does not. (c) Cells grown in pre- and post-perturbation intervals I and IV were sorted for high fluorescence using FACS prior to extracting mRNA. Harvested mRNA of the selected population was then used for qRT-PCR (top) and subsequent network reconstruction (bottom) as in panel b.

fold increase in doxycycline concentration (starting from 10^0 ng/mL to 10^3 ng/mL) and were used to calculate response coefficients relating Dox-rtTA activity to each of the fluorescent output nodes (r_{DC} and r_{DR}). Intervals IV–VI represent increase in morpholino concentration (starting from 0 nmol/mL to 1, 3, 5 nmol/mL) and were used to calculate the response coefficients relating the activity of shRNA-FF3-morpholino to the fluorescence output nodes (r_{MC} and r_{MR}). For each set of reconstruction result, intervals were paired based on their relative pre-perturbation concentration for each drug (i.e., I and VI, II and V, III and VI). As a control, we perturbed the fluorescent proteins using siRNA to verify that there is no crosstalk between them (Supplementary Figure 2).

In order to estimate the significance of the reconstruction results, we calculated the confidence interval of the global response coefficients and plotted the results using a boxplot showing 2.5 and 97.5 percentiles. More specifically, if the 95% confidence interval does not intersect 0, then the response coefficient is statistically significant, indicating a connection between two nodes. In other words, we accept calculated coefficients as positive only if $P(r_{ij} > 0) > 0.975$ and negative if $P(r_{ij} < 0) > 0.975$.

To generate the confidence interval for the inferred response coefficient using the flow cytometry measurements, we selected the method of bootstrap resampling (Methods, Bootstrap).[23] Briefly, each round of bootstrap estimation consists of (1) randomly drawing a number N of resamples (with replacement) from flow cytometry data of pre- and post-perturbation states, (2) calculating the mean of fluorescent readings, and (3) performing response coefficient calculation by calculating fractional change $\Delta\ln(x)$ for each protein.

The bootstrap resampling size is a critical parameter as it has an inverse-squared relationship with the size of the confidence interval, which in turn directly affects our conclusions regarding the reconstructed network topology. We selected a resampling size such that the variations in the calculated global response coefficients reflect typical experimental errors and used our knowledge of the benchmark circuit to validate the selection. In particular, for proteins AmCyan and DsRed, we observed the standard errors of 6.5% and 4.4%, respectively (Supplementary Table 1). If we select 4.4% as the standard error, we obtain a resampling rate of 120 (Methods, Resampling Rate Selection). When we apply this rate to analyze the results of perturbations outside of protein saturation, the response coefficient representing an edge that is absent (r_{MC}) includes zero in its confidence interval (and thus considered statistically negligible), while all other response coefficient (r_{DC}, r_{DR}, r_{MR}) do not include zero in their 95% confidence interval (Supplementary Figure 3). Even when we widen the 95% confidence interval by accepting 6.5% standard error and repeating the procedure (N = 60), the same topology is retained and identical conclusions can be made.

We proceed to the network reconstruction for the various perturbation scenarios using the resample rate N = 120. The confidence interval of the response coefficient was constructed with 1,000 repetitions of bootstrap estimates. For the interval pair III and VI, which represent perturbations applied at or near saturation of both fluorescent proteins, the 95% confidence interval for every response coefficient includes zero, yielding a circuit topology that suggests negligible functional relationship between regulator nodes and output nodes (Figure 3b, top). When we perform the same calculation for the other intervals, we observe that the response coefficients increase in magnitude.

dx.doi.org/10.1021/sb300093y | ACS Synth. Biol. 2013, 2, 255–262

Only in a case where no functional relationship exists between two nodes (r_{MC}), the response coefficient shows negligible variation in magnitude across all three intervals ($\bar{r}_{MC} = 0.060 \pm 0.034$). Furthermore, we observe a noticeable difference between the response coefficients r_{DC} and r_{DR} despite the fact that rtTA co-regulates AmCyan and DsRed expression via a bidirectional promoter. The comparison of median response coefficients for r_{DR} and r_{DC} indicates that r_{DR} is consistently larger, particularly for higher doxycycline concentration ($r_{DR}/r_{DC} = 3.00, 1.81, 1.16$ for intervals I, II and III, respectively). Such differences can potentially be used for identifying secondary connections to coregulated genes, in the particular case the constant but incomplete suppression of DsRed transcript by the shRNA.

We also look for similar defining features following a network reconstruction process with $N = 60$ as a resampling rate (Supplementary Figure 4). The changes in confidence interval due to the smaller resampling rate do not affect our reconstruction conclusions. We also discover that selecting this confidence interval has negligible effect on the observed median values. Lastly, it is worth mentioning that the reconstruction results for the three different interval pairs (Figure 3b) point to the possibility of using a "consensus-based reconstruction". Under this strategy the consensus graph would consist of the significant connections that appear most of the times. This strategy yields a perfect reconstruction for our benchmark architecture (Supplementary Figure 5).

Network Reconstruction Using mRNA Measurements. We proceeded with mRNA measurements using qRT-PCR and performed the network reconstruction in a similar fashion. We performed titration of doxycycline and morpholino on cells harboring the benchmark circuit, but this time harvested the total mRNA of each population 48 h after induction instead of directly measuring the fluorescence (Figure 4a). The perturbation intervals are also defined in the same fashion, with each Roman numeral signifying logarithmic increase in doxycycline and linear increase in morpholino from different starting points. We performed qRT-PCR to measure mRNA levels of each fluorescent protein and compared their relative values by using $\Delta\Delta C_t$ method with expression of GAPDH as the normalization factor for each gene. In this case we assume a Gaussian distribution for the biological triplicates of threshold cycle readings and employed Monte Carlo simulations to generate the confidence interval associated with each response coefficient. In these simulations, we draw random instances of $\Delta\Delta C_t$ values from a simulated normal distribution with mean and standard deviation identical to that of corresponding values obtained from three biological replicates. The simulated value is then used to calculate the global response coefficient in the same fashion as we did using flow cytometry data, by calculating the fractional change $\Delta\ln(x_i)$ where x_i is the normalized copy number of mRNA. After performing 1,000 replicates of the simulation, we once again plotted the distribution on a boxplot showing 2.5 and 97.5 percentile of the data; the reconstruction results with median response coefficient are displayed in Figure 4b.

The plot of relative mRNA copy number versus doxycycline and morpholino concentration shows a trend consistent with that of the dose-dependent response of protein expression; both AmCyan and DsRed are up-regulated in response to doxycycline, while only DsRed is sensitive to changing morpholino concentration. As a consequence, we observe that overall reconstruction results of two different types of data share a common trend of increased sensitivity as pre-perturbation concentration decreases, albeit with increased uncertainty due to standard deviation associated with $\Delta\Delta C_t$ calculation. Of the 6 non-zero edges (r_{DC} and r_{DR} from I and II and r_{MR} from IV and V) that were identified as accurate based on confidence interval derived from flow cytometry data (Figure 3b), only 2 of the corresponding edges (r_{DR} and r_{MR} from intervals I and IV, respectively) derived from qPCR data fit this criteria (Figure 4b). It is worth noting that, of the 4 that remain inconclusive, r_{DC} from interval I and r_{DR} from interval II was closest to being conclusive, with $P(r_{ij} > 0) = 0.958$ and 0.969, respectively (0.975 is the minimum requirement).

In order to examine further the noise associated with mRNA measurements, we performed a triplicate measurement of a benchmark cellular component via qRT-PCR. We chose the housekeeping gene GAPDH (also used for our qRT-PCR data normalization). After calculating the standard error of its expression among three separate experiments, we obtained the standard error to be 4.5%, a rate comparable to that of flow cytometry experiments but smaller than that of the AmCyan and DsRed mRNA measurements. Therefore, we anticipate that it is possible (e.g., primer selection and experimental repeats) to reduce the error in qRT-PCR measurements and potentially improve the decisiveness of the reconstruction results.

A difference between the protein and mRNA experiments is that the reverse engineering is applied to a gated population for the former, whereas we use RNA harvested from all cells for the latter. Therefore, we hypothesize that we can improve the qRT-PCR-based RE results by appropriately sorting a population of cells. Accordingly, we chose a pre- and post-perturbation interval most likely to result in decisive reconstruction results (intervals I and IV of Figure 3) and sorted the cells prior to the mRNA harvesting. We used the same criteria as we did in gating flow cytometry measurement (Supplementary Figure 6). We then proceeded with qRT-PCR and subsequent network reconstruction. The newly recovered topology confirmed our hypothesis, successfully predicting the signs of all of the expected interactions with 95% confidence (Figure 4c). This was also reflected in our calculation of relative mRNA copy number analysis using $\Delta\Delta C_t$ calculation.

Conclusion. One of the main challenges in validating a biological network inference algorithm lies in defining suitable "ground truth" experiments. Traditionally, this task involved verification against known data[24−26] or using in silico simulations,[27−29] which restricts possible tests and makes comparison across methods nontrivial. In this study, we show that genetic circuits integrated in human cells can serve as a benchmark for reverse engineering validation. After stably integrating a synthetic circuit with three different connections in human cells, we used a simple reverse engineering approach and were able to assess its performance in two different fronts: evaluation of reconstructed network based on defining features of the network, and comparison of reconstruction results using different species to represent the synthetic network activity. We chose to utilize protein and mRNA as the two representative species and took steps to determine the appropriate statistical procedure for each measurement in order to increase the prediction confidence. Although both flow cytometry and qPCR data reflect defining features of the network, we had to tailor our analysis for each data type in order to obtain a coherent reconstruction result with reasonable confidence.

A critical issue with our reconstruction results was the inability to pick up statistically significant connections due to noise. While stochasticity is an inherent part of synthetic biological systems, other issues associated with transgenes, such as leakiness of expression and unwanted epigenetic silencing, are obstacles that limit their functions and utility for reverse engineering validation. In fact, the difficulties in using transgenic expression data for reverse engineering purpose has been highlighted by the second DREAM challenge, where a majority of the state-of-the-art reverse engineering algorithms failed to reliably reconstruct a yeast synthetic gene network.[30−32] Perhaps, the most promising aspect of this study is its contribution toward unraveling previously unknown endogenous signaling and disease-related networks. Unraveling biological networks is central to understanding biology in general and human biology in particular. Many human diseases are essentially network-level phenomena. While engineering a large representative set of small to medium scale synthetic circuits in mammalian cells is a daunting task,[33,34] we argue that a few well-characterized circuits can be instrumental toward improving the theory and algorithms. We hope that eventually the validated RE methods will lead to the improvement in understanding and treating human diseases.

METHODS

Bootstrap. For a standard calculation of experimental error, several empirical measurements are averaged. If the probability density of the individual measurements has finite moments, then the true average is a *t*-distributed random variable with mean equal to the sample mean and standard deviation equal to the sample standard deviation divided by the square root of the number of samples (called the standard error). The 95% confidence interval means that the probability is less than 5% that the true value of the average lies more than 1.96 standard deviations from the sample mean. Bootstrapping is used to estimate the confidence interval for experiments where the *t*-distribution does not apply or cannot be assumed. For a simple example of bootstrapping, suppose that one is interested in estimating the median of a set of experimental replicates. Then a number of these replicates are drawn at random with replacement (i.e., replacing each number after each draw), the median is calculated, and this is repeated several times to obtain a distribution of medians. Then the 95% confidence interval of the true median is the range from the 2.5 to the 97.5 percentile of the set of medians. In the main text we use bootstrapping to estimate the experimental error for the global response coefficients.

Stable Transfection and Cell Line. The synthetic gene circuit construct was stably integrated into a HEK293 cell line using Flp-In 293 System (Invitrogen) according to the manufacturer's instructions. The cells were maintained at 37 °C, 100% humidity, and 5% CO_2. The complete growth medium consists of Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% Fetal Bovine Serum (Invitrogen), 0.1 mM MEM nonessential amino acids (Invitrogen), 0.045 units/mL of penicillin and 0.045 units/mL of streptomycin (Invitrogen). Hygromycin B (Invitrogen) at 50 μg/mL was added as a selection agent. When the culture reached 75−90% confluency, it was passed first by washing with PBS (Mediatech) and then trypsinized with 0.25% Trypsin-EDTA (Invitrogen). New culture was plated at 40% seeding density.

Fluorescence Microscopy. Fluorescence images of live cells were captured 48−56 h post-perturbation with doxycycline and morpholino. The live cells were grown on 12-well plates (Greiner Bio-One) in the complete medium. Cells were imaged using the Olympus IX81 microscope and a Precision Control environmental chamber. The images were captured using a Hamamatsu ORCA-03 Cooled monochrome digital camera. The filter sets (Chroma) are as follows: ET436/20x (excitation) and ET480/40m (emission) for AmCyan, ET560/40x (excitation) and ET630/75m (emission) for DsRed. Data collection and processing was performed in software package Slidebook 5.0. All images within a given experimental set were collected with the same exposure times and underwent identical processing.

Flow Cytometry. For FACS experiment, cells were prepared as follows: 48 h after perturbation, cells were trypsinized with 0.3 mL of 0.25% trypsin-EDTA for 3 min and pelleted by centrifugation at 4000 rpm for 2 min. The pellet was resuspended with 0.4 mL of PBS (Mediatech). Analysis was performed with BD LSRFortessa. AmCyan protein was detected with a 445 nm laser and a 515/20 band-pass filter, and DsRed with a 561-nm laser, 610 emission filter, and 610/20 band-pass filter. For each culture representing different conditions, 100,000 events were collected. Subsequent gating and analysis of the flow cytometry data was performed in FlowJo (Treestar). For sorting, sample preparation was performed the same, but the analysis and sorting were performed with a BD FACSAria cell sorter. For each condition, 200,000 cells that fell within the defined gate were collected as sorted population.

Quantitative RT-PCR. Forty-eight hours after perturbation, total RNA of the population was harvested using the RNeasy Mini kit (Qiagen) according to manufacturer's suggestion. One microgram of total RNA was reverse transcribed to cDNA using QuantiTect Reverse Transcription kit (Qiagen). Quantitative PCR assays were performed with the Mastercycle ep realplex thermal cycler (Eppendorf) using the KAPA SYBR FAST qPCR kit (KAPA Biosystems). The relative mRNA expression levels of each node in the synthetic gene circuit were quantified with $\Delta\Delta C_t$ method, using GAPDH as normalization factor. Forward primer used to amplify GAPDH sequence was 5′-AATCCCATCACCATCTTCCA-3′, and the reverse primer was 5′-TGGACTCCACGACGTACTCA-3′. Amplification started with an enzyme activation step at 95 °C for 3 min, followed by 45 cycles consisting of 3 s of denaturation step at 95 °C and 20 s of annealing/extension at 60 °C.

Resampling Rate Selection. For proteins AmCyan and DsRed, we calculated the standard errors of 6.5% and 4.4%, respectively (Supplementary Table 1). The 95% confidence interval of the mean is defined as $\bar{x} \pm (\mathrm{SE} \times 1.96)$, and by substituting SE with the observed value we can determine the limits of 95% confidence interval. For example, the observed standard error of 4.4% translates to $\bar{x} \pm 8.8\%$. In other words, the "radius" of the 95% confidence interval (distance from mean to upper and lower boundary) in this case should be approximately 0.09 for the calculated global response coefficients. Supplementary Figure 3, the calculated global response coefficients vs resampling rate (N), illustrates that increasing the resampling rate decreases the 95% confidence interval at a rate of $1/\sqrt{N}$. Therefore, $N = 120$ gives us a confidence interval radius of 0.09 ($1/\sqrt{120} = 0.0912$).

dx.doi.org/10.1021/sb300093y | *ACS Synth. Biol.* 2013, 2, 255−262

■ ASSOCIATED CONTENT

Supporting Information

This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: bleris@utdallas.edu; sontag@math.rutgers.edu.

Author Contributions

T.K., L.B., and Z.X. performed the experiments. T.K., J.W., and L.B. analyzed the data. T.K., L.B., J.W., Y.B., and E.S. prepared the manuscript. L.B., E.S., and Y.B. conceived the project. L.B. supervised the project. Requests for materials should be addressed to L.B. (bleris@utdallas.edu).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by the US National Institutes of Health (NIH) grant R21GM098984, TxACE SRC grant P12095, and the University of Texas at Dallas. We would like to thank Yi Li and Richard Moore for discussions.

■ REFERENCES

(1) Kholodenko, B., Yaffe, M. B., and Kolch, W. (2012) Computational approaches for analyzing information flow in biological networks. *Sci. Signaling 5*, re1.

(2) Marbach, D., Prill, R. J., Schaffter, T., Mattiussi, C., Floreano, D., and Stolovitzky, G. (2010) Revealing strengths and weaknesses of methods for gene network inference. *Proc. Natl. Acad. Sci. U.S.A. 107*, 6286−6291.

(3) Sprinzak, D., and Elowitz, M. B. (2005) Reconstruction of genetic circuits. *Nature 438*, 443−448.

(4) Bansal, M., Belcastro, V., Ambesi-Impiombato, A., and di Bernardo, D. (2007) How to infer gene networks from expression profiles. *Mol. Syst. Biol. 3*, 78.

(5) Albert, R., Dasgupta, B., and Sontag, E. (2010) Inference of signal transduction networks from double causal evidence. *Methods Mol. Biol. 673*, 239−251.

(6) Marbach, D., Costello, J. C., Kuffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., Allison, K. R., Kellis, M., Collins, J. J., and Stolovitzky, G. (2012) Wisdom of crowds for robust gene network inference. *Nat. Methods 9*, 796−804.

(7) Prill, R. J., Marbach, D., Saez-Rodriguez, J., Sorger, P. K., Alexopoulos, L. G., Xue, X., Clarke, N. D., Altan-Bonnet, G., and Stolovitzky, G. (2010) Towards a rigorous assessment of systems biology models: The DREAM3 challenges. *PLoS One 5*, e9202.

(8) Prill, R. J., Saez-Rodriguez, J., Alexopoulos, L. G., Sorger, P. K., and Stolovitzky, G. (2011) Crowdsourcing network inference: The DREAM predictive signaling network challenge. *Sci. Signaling 4*, mr7.

(9) Stolovitzky, G., Monroe, D., and Califano, A. (2007) Dialogue on reverse-engineering assessment and methods: The DREAM of high-throughput pathway inference. *Ann. N. Y. Acad. Sci. 1115*, 1−22.

(10) He, F., Balling, R., and Zeng, A. P. (2009) Reverse engineering and verification of gene networks: Principles, assumptions, and limitations of present methods and future perspectives. *J. Biotechnol. 144*, 190−203.

(11) Cantone, I., Marucci, L., Iorio, F., Ricci, M. A., Belcastro, V., Bansal, M., Santini, S., di Bernardo, M., di Bernardo, D., and Cosma, M. P. (2009) A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. *Cell. 137*, 172−181.

(12) Kholodenko, B. N., Kiyatkin, A., Bruggeman, F. J., Sontag, E., Westerhoff, H. V., and Hoek, J. B. (2002) Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. *Proc. Natl. Acad. Sci. U.S.A. 99*, 12841−12846.

(13) Sontag, E. D. (2008) Network reconstruction based on steady-state data. *Essays Biochem. 45*, 161−176.

(14) Kholodenko, B. N. (2007) Untangling the signalling wires. *Nat. Cell Biol. 9*, 247−249.

(15) Sontag, E., Kiyatkin, A., and Kholodenko, B. N. (2004) Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data. *Bioinformatics. 20*, 1877−1886.

(16) Santos, S. D., Verveer, P. J., and Bastiaens, P. I. (2007) Growth factor-induced MAPK network topology shapes erk response determining PC-12 cell fate. *Nat. Cell Biol. 9*, 324−330.

(17) Bruggeman, F. J., Westerhoff, H. V., Hoek, J. B., and Kholodenko, B. N. (2002) Modular response analysis of cellular regulatory networks. *J. Theor. Biol. 218*, 507−520.

(18) Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science 268*, 1766−1769.

(19) 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.

(20) Rinaudo, K., Bleris, L., Maddamsetti, R., Subramanian, S., Weiss, R., and Benenson, Y. (2007) A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat. Biotechnol. 25*, 795−801.

(21) Summerton, J., and Weller, D. (1997) Morpholino antisense oligomers: Design, preparation, and properties. *Antisense Nucleic Acid Drug Dev. 7*, 187−195.

(22) Choi, W. Y., Giraldez, A. J., and Schier, A. F. (2007) Target protectors reveal dampening and balancing of nodal agonist and antagonist by miR-430. *Science 318*, 271−274.

(23) Efron, B., and Tibshirani, R. J. (1993) *An Introduction to the Bootstrap*, 1st ed., Chapman & Hall/CRC, New York.

(24) Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A., and Nolan, G. P. (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science 308*, 523−529.

(25) Xu, T. R., Vyshemirsky, V., Gormand, A., von Kriegsheim, A., Girolami, M., Baillie, G. S., Ketley, D., Dunlop, A. J., Milligan, G., Houslay, M. D., and Kolch, W. (2010) Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species. *Sci. Signaling 3*, ra20.

(26) Pe'er, D., and Hacohen, N. (2011) Principles and strategies for developing network models in cancer. *Cell 144*, 864−873.

(27) Dalle Pezze, P., Sonntag, A. G., Thien, A., Prentzell, M. T., Godel, M., Fischer, S., Neumann-Haefelin, E., Huber, T. B., Baumeister, R., Shanley, D. P., and Thedieck, K. (2012) A dynamic network model of mTOR signaling reveals TSC-independent mTORC2 regulation. *Sci. Signaling 5*, ra25.

(28) Marbach, D., Schaffter, T., Mattiussi, C., and Floreano, D. (2009) Generating realistic in silico gene networks for performance assessment of reverse engineering methods. *J. Comput. Biol. 16*, 229−239.

(29) Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005) Reverse engineering of regulatory networks in human B cells. *Nat. Genet. 37*, 382−390.

(30) Stolovitzky, G., Prill, R. J., and Califano, A. (2009) Lessons from the DREAM2 Challenges. *Ann. N.Y. Acad. Sci. 1158*, 159−195.

(31) Marbach, D., Mattiussi, C., and Floreano, D. (2009) Replaying the evolutionary tape: biomimetic reverse engineering of gene networks. *Ann. N.Y. Acad. Sci. 1158*, 234−245.

(32) Baralla, A., Mentzen, W., and De La Fuente, A. (2009) Inferring gene networks: dream or nightmare? Part 1: Challenges 1 and 3. *Ann. N.Y. Acad. Sci. 1158*, 246−256.

(33) Yokobayashi, Y., Weiss, R., and Arnold, F. H. (2002) Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. U.S.A. 99* (26), 16587−16591.

(34) Ellis, T., Wang, X., and Collins, J. J. (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol. 27* (5), 465−471.

Supplementary Information

Reverse Engineering Validation using a Benchmark Synthetic Gene Circuit in Human Cells

Taek Kang[1,3], Jacob T. White[1,3], Zhen Xie[4], Yaakov Benenson[5], Eduardo Sontag[*6,] and Leonidas Bleris[*1,2,3]

[1]Bioengineering Department, University of Texas at Dallas, 800 West Campbell Road, Richardson TX 75080 USA

[2]Electrical Engineering Department, University of Texas at Dallas, 800 West Campbell Road, Richardson TX 75080 USA

[3]Center for Systems Biology, University of Texas at Dallas, NSERL 4.708, 800 West Campbell Road, Richardson TX 75080 USA

[4]Center for Synthetic and Systems Biology, Bioinformatics Division, TNLIST, 100084, Tsinghua University, Beijing, China

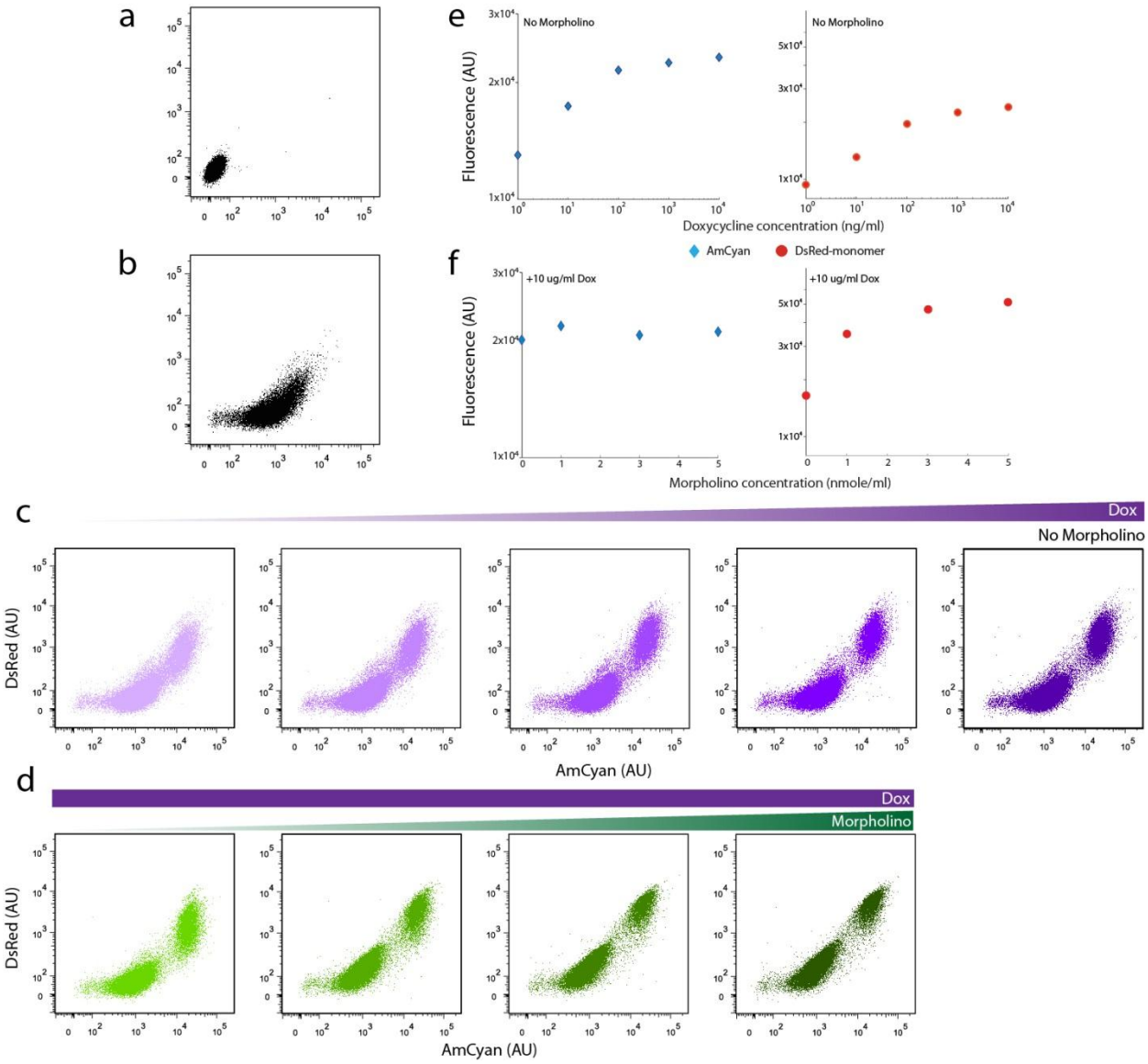
[5]Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule (ETH) Zürich, Mattenstrasse 26, 4058 Basel, Switzerland

[6]Department of Mathematics, Rutgers-The State University of New Jersey, Hill Center, 110 Frelinghuysen Rd., Piscataway NJ 08854 USA

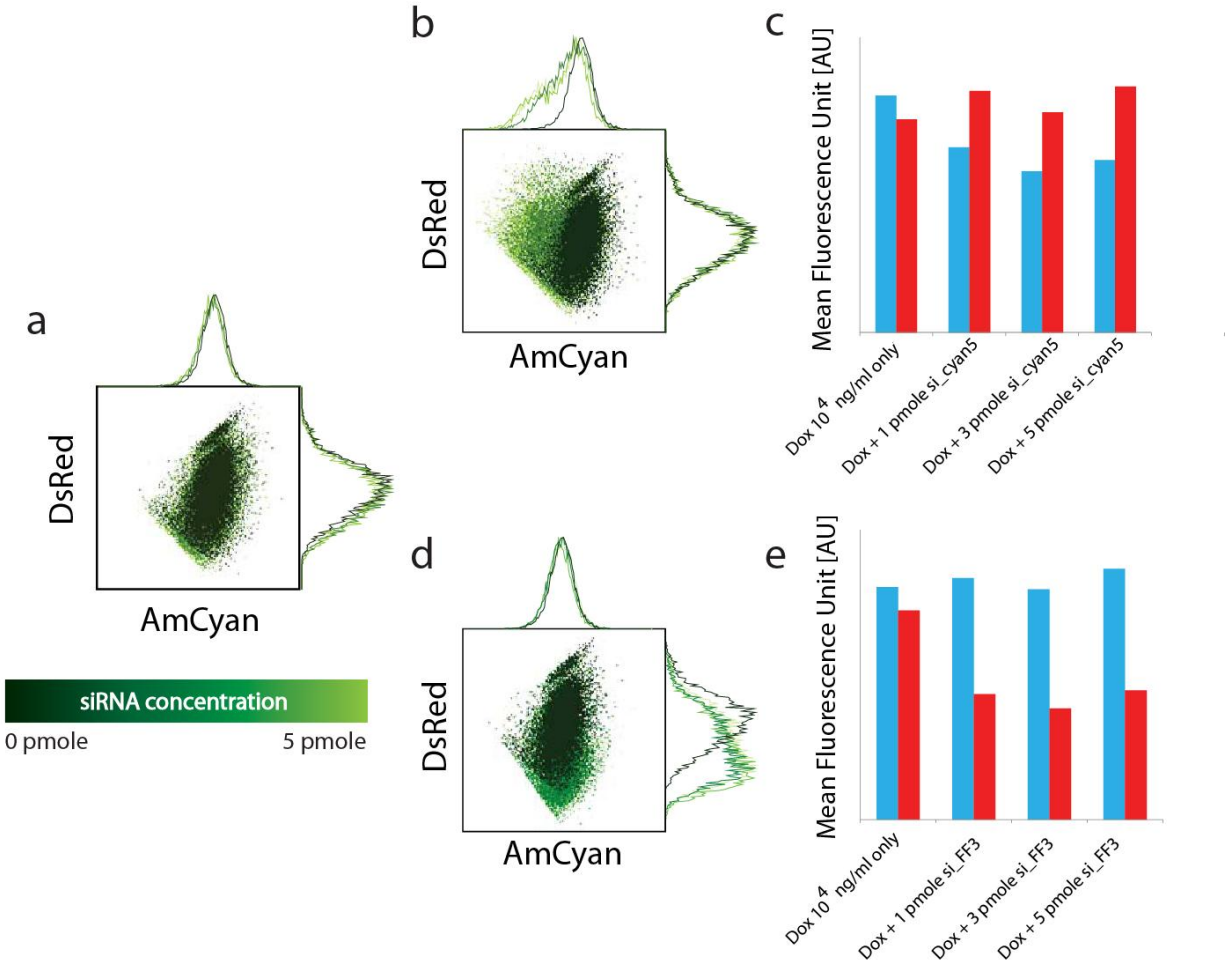
[*]Corresponding authors: L.B. (bleris@utdallas.edu) and E. S. (sontag@math.rutgers.edu)

Keywords: Reverse Engineering, Benchmark Synthetic Circuits, Human Cells, Modular Response Analysis.

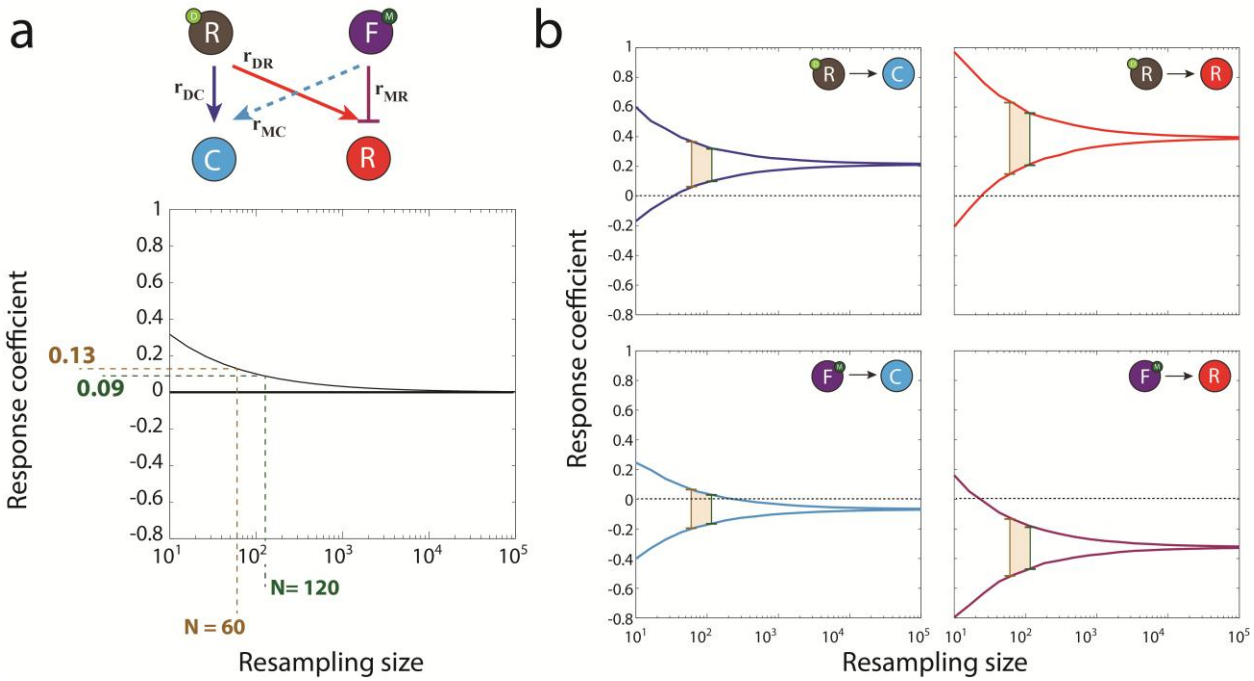
Supplementary Figures



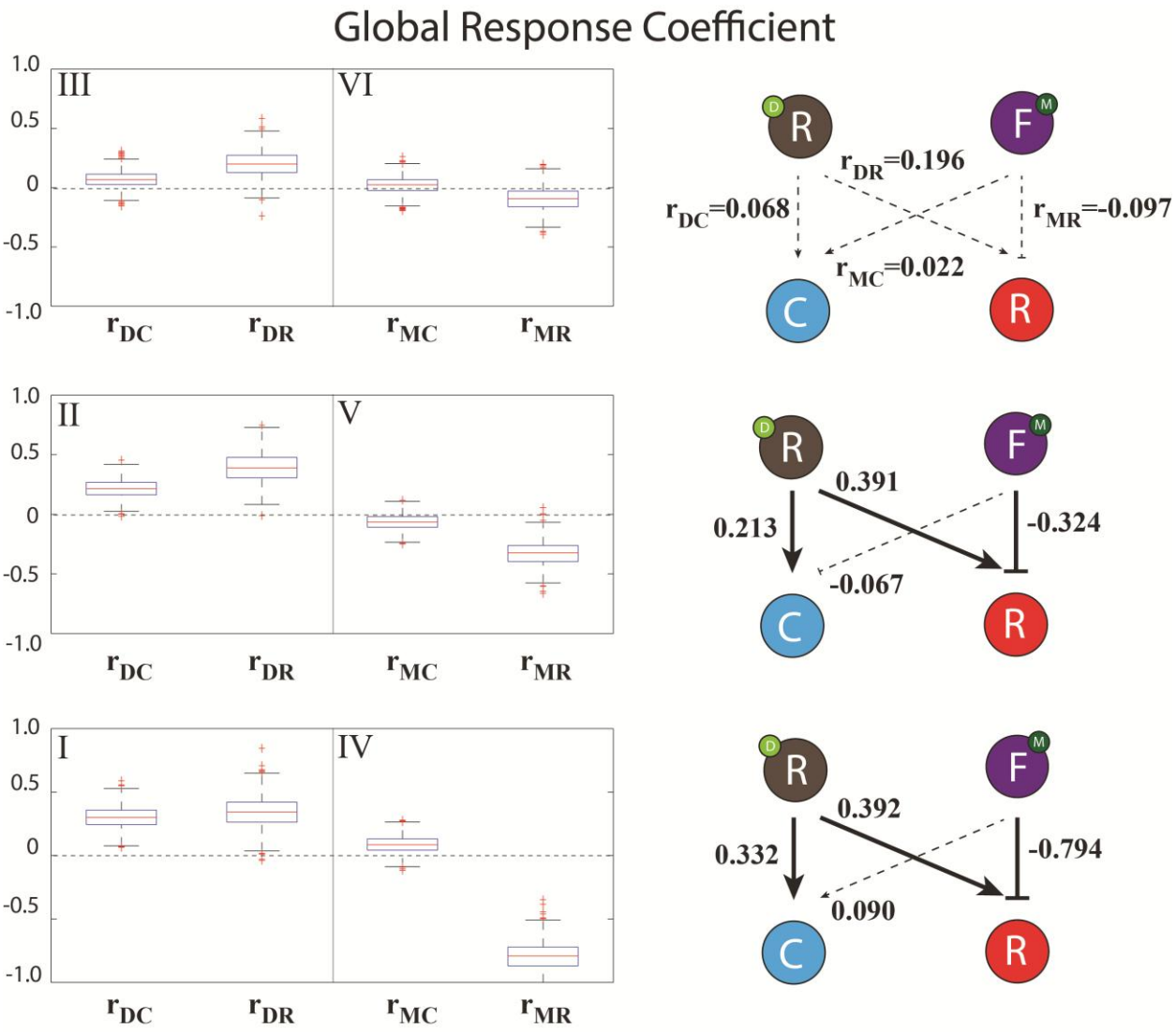
Supplement figure 1. Flow cytometry data. Expression of fluorescent proteins AmCyan and DsRed measured via flow cytometry (n=100,000). All measurements were performed 48 hours post-induction. **(a)** Wild-type HEK 293 cells with no induction. **(b)** FLP cells stably integrated with the synthetic gene circuit construct, with no induction. **(c)** Flow cytometry results of doxycycline titration with 10^0, 10^1, 10^2, 10^3, 10^4 ng/ml (left to right) on cells stably integrated with the circuit construct. Morpholino concentration was kept constant at 0 nmol/ml. **(d)** Flow cytometry results of morpholino titration with 0, 1, 3, 5 nmol/ml (left to right) on cells stably integrated with the circuit construct. Doxycycline concentration was kept constant at 10^4 ng/ml. **(e,f)** Mean fluorescence of AmCyan and DsRed plotted against concentration of doxycycline and morpholino (respectively). Each point represents mean fluorescence obtained from gated population of corresponding position in the titration experiment.



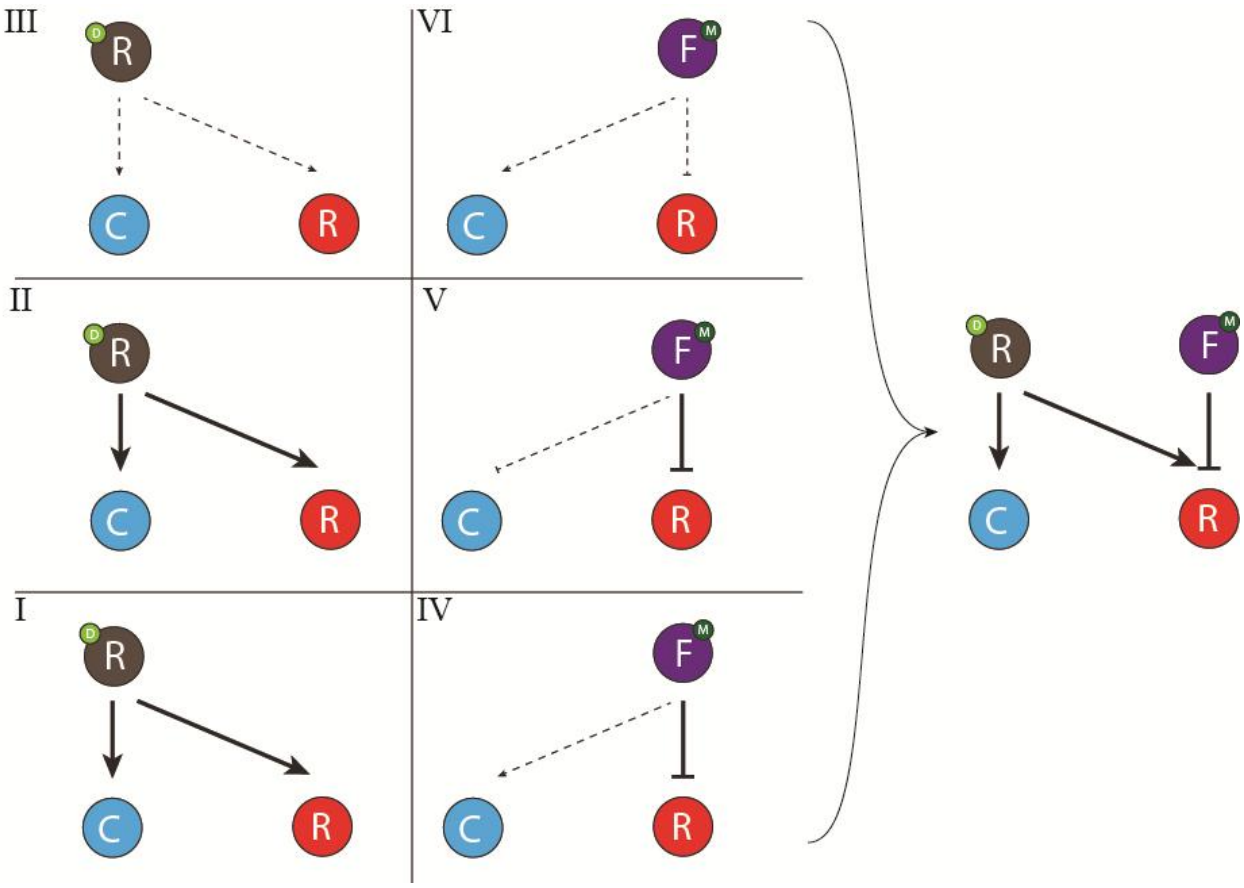
Supplement figure 2. Perturbations of fluorescent output nodes with RNAi. In order to measure the dependence of one fluorescent protein expression on another, direct perturbations to each protein (AmCyan and DsRed) were performed in separate experiments with siRNA targeting the corresponding mRNA transcripts. To assess the effect of each siRNA, cells harboring the synthetic circuit were induced with 10^4 ng/ml of doxycycline and transfected with variable concentrations (1, 3, 5 nmol/ml) of the siRNA. 48 hours after perturbation, fluorescence was measured via flow cytometry (n=100,000). For the scatter plots shown, darkest color represents absence of siRNA, while lightest color represents maximum siRNA. **(a)** Flow cytometry results after transfection of scrambled siRNA. **(b)** Flow cytometry results after transfection of siRNA targeting AmCyan transcript. **(c)** Mean fluorescence of both DsRed and AmCyan for each population**. (d)** Fluorescence readings obtained after transfection of a siRNA targeting FF3-shRNA binding site. **(e)** Mean fluorescence of both DsRed and AmCyan.



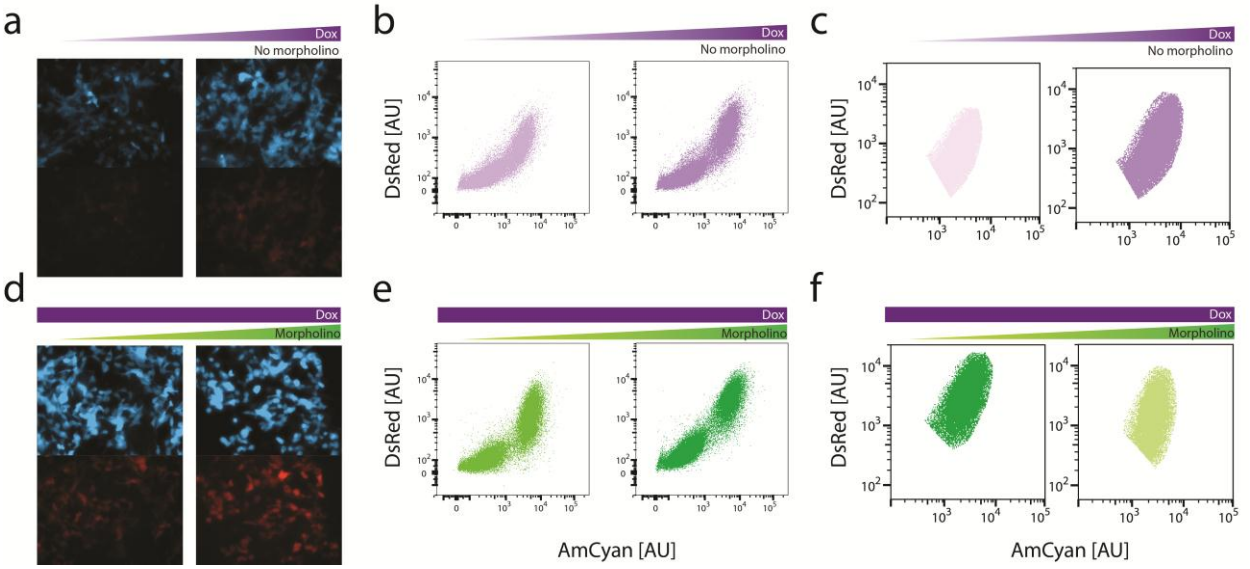
Supplement figure 3. Resampling rate and network topology reconstruction. (a) When using a resampling method such as bootstrapping the width of the sample confidence interval converge to zero as resampling size increases. This decrease occurs at a rate of $1/N^{1/2}$, where N is the resampling rate. Using the standard error observed from three separate measurements of proteins AmCyan and DsRed (Supplement Table 1), we determined that the resampling rate should be between 60 and 120. **(b)** We evaluated the effect of altering resampling size on the network reconstruction results by plotting 97.5^{th} and 2.5^{th} percentiles of calculated global response coefficient versus the resampling rate. The plots shown are results after perturbations of 10^1ng/ml to 10^2ng/ml doxycycline and 1 nmol/ml to 3 nmol/ml of morpholino (intervals II and V in the main text). Green and tan lines indicate the 95% confidence interval using the 60 and 120 resampling rates.



Supplement figure 4. Network topology reconstruction results using N=60 for the resampling rate. Complete reconstruction results of bootstrapped resampling method with N = 60 as a resampling size. Boxplots indicate 95% confidence interval of the calculated result. Values shown in right indicate the median GRC.



Supplement figure 5. Consensus-based reverse engineering. The response coefficients between regulator nodes and fluorescent output nodes are calculated after six different perturbations using doxycycline (I~III) and morpholino (IV~VI) are performed. Dotted lines in the reconstruction results indicate that the calculated global response coefficient includes zero in its 95% confidence interval. Based on these results, we construct a final consensus network, consisting of edges that are observed most frequently.



Supplement figure 6. Population Separation with FACS. (a) Fluorescence microscopy results of doxycycline titration with cells harboring stable integration of the benchmark synthetic gene circuit. Cells were induced with 1 ng/ml (left) and 10 ng/ml of doxycycline (right). This corresponds to interval 'I' in Figure 3 and 4 of manuscript. **(b,c)** Flow cytometry results of the same populations, before (b) and after (c) separation by FACS. Color of the scatter plot corresponds to the level of doxycycline as shown in gradient above. **(d)** Fluorescence microscopy results of morpholino titration in the presence of saturating concentration of doxycycline. Cells were induced with 10 µg/ml of doxycycline alone (left) and 10 µg/ml of doxycycline with 1nmol/ml of Morpholino (right). **(e,f)** Flow cytometry results of the same populations, before (e) and after (f) separation by FACS. Color of the scatter plot corresponds to the level of doxycycline as shown in gradient above. The mRNA was harvested from each of the eight populations represented with scatter plot (b,c,e,f) for subsequent qRT-PCR.

Cloning

PtreTightBI-amCyan-dsRed-F3x3: AmCyan gene was amplified from the AmCyan-C1 plasmid (Clontech) using primers P1 and P2 with NheI and MluI restriction sites. The product was digested with NheI and MluI and repurified. In parallel, pTRE-Tight-BI (Clontech) was digested with NheI and MluI and gel-purified. The digested insert was ligated into the digested vector at 1:2 ratio at 12 ^{o}C overnight, transformed and explanded. The construct was sequenced using primers P1 and P2. The DsRed-monomer gene was amplified from the pCAGOP-DsRed-Monomer-N1 plasmid (Rinaudo et al, 2007) using primers P3 and P4 with XbaI and AgeI restriction sites. The PCR product was digested with XbaI and AgeI enzymes. In parallel, the above pTRE-Tight-BI-AmCyan plasmid was digested with XbaI and AgeI and gel-purified. The digested DsRed-monomer insert was ligated into the digested vector at 1:1 ratio at 14 ^{o}C overnight, transformed and expanded. Correct constructs were identified by double EagII and BglII digestion and by observing fluorescence in transfected cells. The FF3x3 sense and antisense sequences (Supplement Table 3) were diluted annealed. The double-stranded FF3x3 insert with xbaI-compatible sticky ends was phosphorylated and ligated into xbaI-digested and purified pTRE-Tight-BI-AmCyan-DsRed backbone at 1:3 ratio at room temperature for 10 min. The ligation product was transformed and expanded.

CMVrtTA-U6shRNA: The FF3 shRNA stem-loop was first inserted into pSiren (Clontech) using BamHI and EcoRI. The U6-shRNA-FF3 sequence was then amplified using with HindIII and PciI flanking primers and ligated to digested with HindIII and PciI enzymes pTET-ON-advanced (Clontech) at 1:3 ratio at 14^{o}C overnight, transformed and expanded. Correct constructs were identified by sequencing.

PtreTightBI-amCyan-dsRed-F3x3-CMVrtTA-U6shRNA: The CMVrtTA-U6shRNA sequence was amplified from CMVrtTA-U6shRNA with PciI flanking primer P7 and P8, subsequently digested with PciI and phosphorylated. The PtreTightBI-amCyan-dsRed-F3x3 was digested with PciI and ligated with CMVrtTA-U6shRNA at 1:2 ratio at 14 ^{o}C overnight, transformed and expanded. Correct constructs were identified by sequencing.

PtreTightBI-amCyan-dsRed-F3x3-CMVrtTA-U6shRNA-FRT-Hygro: The FRT-Hygro sequence was amplified from the pcDNA5/FRT (Invitrogen) with AvrII flanking primer s P9 and P10, subsequently digested with AvrII and phosphorylated. The PtreTightBI-amCyan-dsRed-F3x3-CMVrtTA-U6shRNA was digested with AvrII, dephosphorylated and ligated to FRT-Hygro at 1:1 ratio at 14^{o}C overnight, transformed and expanded. Correct constructs were identified by sequencing.

Cloning kits: PCR product purifications were performed using the Qiagen PCR purification kit. Gel-purifications were performed using Qiagen Gel Purification kit from 1% agarose gel. DNA isolations were performed using Qiagen MiniPrep kit. Sequencing was performed by Genewiz.

Morpholino oligonucleotides

Morpholino oligo (Gene Tools) was re-suspended in distilled H_2O to final concentration of 1mM, as instructed by manufacturer. The following antisense Morpholino (complimentary to the 3' UTR shRNA target site of DsRed) was used: 5'-TTTGTATTCAGCCCATATCGTT-3'. To facilitate the delivery of Morpholino oligos into the cytosol of adherent cell line, 1μM of Endo-Porter solution was simultaneously added with the injection of Morpholino oligo.

Supplement Tables

Supplement Table 1

Triplicate flow cytometry measurement used for calculating the standard error

Protein	Fluorescence (AU)			Mean	Standard Deviation	Standard Error
AmCyan	2189	1744	1999	1977	223	128.75(6.51%)
DsRed	167	144	161	157	11.9	6.87 (4.38%)

Supplement Table 2

ShRNA FF3 insert

Type	Sequence
FF3	TTTGTATTCAGCCCATATCGTT

Supplement Table 3

FF3 target sequence

Construct	Strand	Sequence (5' -3')
FF3	Sense	AACGATATGGGCTGAATACAAAAACGATATGGGCTGAATACA AAAACGATATGGGCTGAATACAAA
	Antisense	TTTGTATTCAGCCCATATCGTTTTTGTATTCAGCCCATATCGTTT TTGTATTCAGCCCATATCGTT

Supplement Table 4

Primers used for cloning purposes

Primer P1	ACATATTTGAATGTATTTAGAAAAAT
Primer P2	AGTGAGCGAGGAAGCTCGGGGCAG
Primer P3	GTTGTCCATGGTGGCGAGACCGGTTGG
Primer P4	CCATCTAGATTACTACTGGGAGCCGGAGTG
Primer P5	CCAACATGTTTACCAGTAATGACCTCAGAACTCCA
Primer P6	CCAAAGCTTGCCTTACGCGTGTAAAACGAC
Primer P7	CCAACATGTCATGTCCAACATTACCGCCATG
Primer P8	GGAAAAACGCCAGCAACGC
Primer P9	CCACCTAGGGAAGTTCCTATTCCGAAGTTCCTATTC
Primer P10	CCACCTAGGCAGACATGATAAGATACATTGATGAGTTT