Discriminating direct and indirect connectivities in biological networks

Taek Kang[a,b], Richard Moore[a,b], Yi Li[a,b], Eduardo Sontag[c,1], and Leonidas Bleris[a,b,d,1]

[a]Bioengineering Department, University of Texas at Dallas, Richardson, TX 75080; [b]Center for Systems Biology, University of Texas at Dallas, Richardson, TX 75080; [c]Department of Mathematics and Center for Quantitative Biology, Rutgers-The State University of New Jersey, Piscataway, NJ 08854; and [d]Electrical Engineering Department, University of Texas at Dallas, Richardson, TX 75080

Edited by Wing Hung Wong, Stanford University, Stanford, CA, and approved September 10, 2015 (received for review April 12, 2015)

Reverse engineering of biological pathways involves an iterative process between experiments, data processing, and theoretical analysis. Despite concurrent advances in quality and quantity of data as well as computing resources and algorithms, difficulties in deciphering direct and indirect network connections are prevalent. Here, we adopt the notions of abstraction, emulation, benchmarking, and validation in the context of discovering features specific to this family of connectivities. After subjecting benchmark synthetic circuits to perturbations, we inferred the network connections using a combination of nonparametric single-cell data resampling and modular response analysis. Intriguingly, we discovered that recovered weights of specific network edges undergo divergent shifts under differential perturbations, and that the particular behavior is markedly different between topologies. Our results point to a conceptual advance for reverse engineering beyond weight inference. Investigating topological changes under differential perturbations may address the longstanding problem of discriminating direct and indirect connectivities in biological networks.

reverse engineering | synthetic biology | direct and indirect connectivities | human cells | nonparametric resampling

A focal point of systems biology is the reverse engineering of gene regulatory networks (1–5). The methods have shifted from intuitive inference of local connectivities to comprehensive analysis of large networks, involving heterogeneous data sets from high-throughput experiments and complex theoretical tools (6–10). Despite significant advances, a fundamental reverse engineering bottleneck is the ability to discriminate between direct and indirect connections. In a simple case, assuming three nodes in a cascade formulation, where an input node is activating an intermediary node which in turn is activating an output node, a reverse engineering algorithm may infer an activating edge from the input node to the output, even though there is no direct biological interaction.

Unfortunately, the limitation in correctly distinguishing the effects stemming from indirect connectivities is pervasive (11–13) and justifies the urgent need for new and reliable methods to eliminate spurious edges. Importantly, remedies to address this problem should not further muddle the interpretation by removing true network edges (14). A number of theoretical approaches have been proposed to overcome this hurdle (4, 15–18), but the ability to experimentally verify the conclusions drawn by reverse engineering tools remains paramount.

The majority of efforts to address the verification issue adopt in silico benchmark suites that are based on biological pathway approximations (19). Although these models do include a number of commonly observed topologies and have provided significant insights, they do not fully capture the complexity of the biological realm and the associated heterogeneity and intrinsic variability. On the other hand, engineered synthetic gene circuits are orthogonal to the endogenous pathways yet operate within the natural cellular context using the available resources. Thus, synthetic networks are a versatile platform for investigating specific connectivities and topological properties and can ultimately guide us to deriving fundamental insights about biological systems and pathways (20–23).

We previously proposed a strategy based upon using a synthetic gene network in human cells as a benchmark for reverse engineering validation and refinement (24). Here, we built three-node synthetic gene regulatory networks that incorporate direct and indirect connectivities and used them as benchmarks in human kidney cells. The first network is the type I coherent feed-forward loop (25, 26), where the origin node (X) activates the target node (Z) directly but also through an intermediate node (Y), with OR logic at the output (Fig. 1*A*). The second network is a cascade motif, where the origin node (X) regulates the target node (Z) indirectly via an intermediary node (Y) (Fig. 1*B*). More specifically, the node-to-node interactions are achieved through inducible transcriptional regulation. The origin (X) and intermediary nodes (Y) contain bidirectional promoter elements that drive the production of a fluorescent reporter protein and a transactivator unit, and the target node (Z) contains a unidirectional promoter for the production of a fluorescent reporter protein only. Each node produces a fluorescent reporter, which allows monitoring its state.

We commenced the experiments confirming the baseline behavior of the synthetic networks under boundary and control conditions. Subsequently, we systematically perturbed each network node using short interfering RNAs (siRNAs) (27); then, we collected and processed the flow cytometry measurements. Using these data, we performed network reconstruction via nonparametric single-cell data resampling followed by modular response analysis (4, 28). The reconstruction results reproduced the benchmark network topologies. Importantly, we identified divergent shifts in predicted interaction strengths under differential perturbations, a feature that can be critical toward discriminating between direct and indirect connectivities.

Results

Design and Assembly of the Benchmark Synthetic Regulatory Networks. The first of the two networks is the type I coherent feed-forward loop (Fig. 1*C*). The plasmid for node X consists of a constitutively

Significance

We used a combination of computational and theoretical approaches coupled to synthetic biology experimentation in mammalian cells to study direct and indirect connectivities in biological networks. After subjecting benchmark circuits to a range of perturbations, we recovered the edge weights using nonparametric single-cell data resampling coupled with modular response analysis. We discovered that inferred weights of specific network edges undergo divergent shifts under differential perturbations, and that the particular behavior is topology dependent. Incorporating this insight in the analysis of high-throughput experiments may provide a sought-after solution to a longstanding reverse engineering problem.

Author contributions: L.B. designed research; T.K., R.M., Y.L., E.S., and L.B. performed research; T.K., Y.L., and L.B. analyzed data; and T.K., R.M., Y.L., E.S., and L.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: bleris@utdallas.edu or eduardo.sontag@rutgers.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1507168112/-/DCSupplemental.

SYSTEMS BIOLOGY

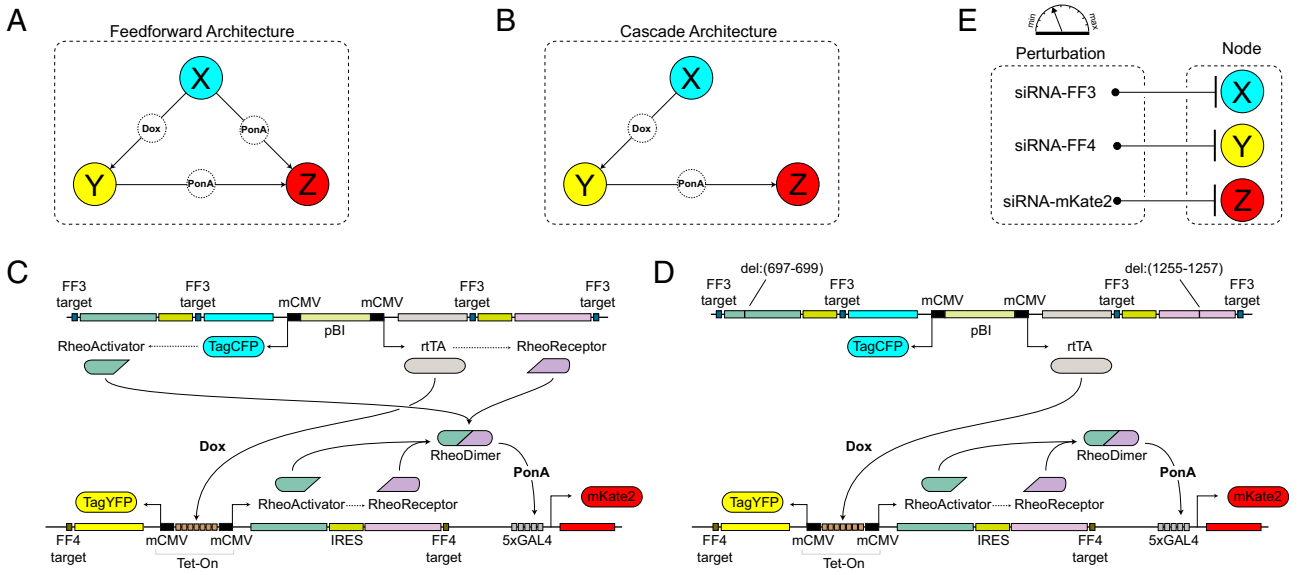

Fig. 1. The benchmark synthetic regulatory networks. (*A*) The first motif is a coherent feed-forward loop where node X regulates node Z in both a direct and an indirect manner. (*B*) The second motif is a cascade, where node X activates node Z only by activating node Y. (*C* and *D*) Detailed information about the synthetic gene networks. The activity of the three nodes X, Y, and Z can be quantified by the output fluorescent proteins TagCFP, TagYFP and mKate2, respectively. The constitutive bidirectional promoter of node X also transcribes rtTA for node Y induction and the RheoSwitch dimers for node Z induction. For the cascade motif, the translation of RheoSwitch dimer protein is prevented by nonsense mutation. In the presence of doxycycline, the constitutively transcribed rtTA induces transcription of RheoSwitch in node Y. When ponasterone A binds to the RheoSwitch dimer, the entire complex serves as a transactivator for the yeast Gal4 domain. Transcription at Gal4 domain results in production of mKate2 to indicate node Z activity. (*E*) Perturbation of each node in the system is performed siRNA. Nodes X and Y are perturbed by synthetic siRNAs (FF3 and FF4, respectively) with the targets located in the 3′ UTR of their corresponding targets. Node Z is perturbed by a custom siRNA that directly targets mKate2. IRES, internal ribosome entry site.

active bidirectional promoter flanked by the reporter fluorescent protein TagCFP and reverse tetracycline-controlled transactivator (rtTA) on either side. The regulatory unit rtTA serves as the transactivator of the tetracycline-inducible expression system (Tet-On) upon forming a homodimer and binding with the ligand doxycycline (Dox). Activation of the downstream target node Y, which consists of tetracycline response element (TRE) enhancer flanked by cytomegalovirus (CMV) promoters on either side, requires binding of active rtTA–Dox complex to the TRE enhancer. Thus, the activation of node Y by node X depends on doxycycline. The activation of node Y results in production of its fluorescent reporter TagYFP and a heterodimeric transactivator composed of the RheoActivator and RheoReceptor domains. The RheoActivator domain consists of a ligand binding domain fused with the viral transactivator VP16, whereas the RheoReceptor domain is a hybrid of insect hormone ecdysone receptor (EcR) fused to yeast GAL4 DNA binding domain for target binding specificity to GAL4 response element. After dimerization of RheoActivator and RheoReceptor, an EcR agonist such as ponasterone A induces conformational change to the RheoReceptor such that the heterodimer bound to GAL4 response element initiates transcription. In our synthetic network, RheoSwitch dimer activates node Z by initiating transcription of its reporter fluorescent protein mKate2. To achieve direct activation of node Z by node X, the node X produces the RheoSwitch proteins in addition to TagCFP and rtTA.

The second of the two networks is a cascade motif (Fig. 1*D*), where node X controls node Z exclusively through the activation of intermediate node Y. To implement this architecture, we modified the coherent feed-forward architecture by inserting a single base pair in both of the node X RheoSwitch genes to induce a nonsense frame-shift mutation. As the RheoSwitch heterodimer genes together constitute ~30% of total plasmid size we selected introducing

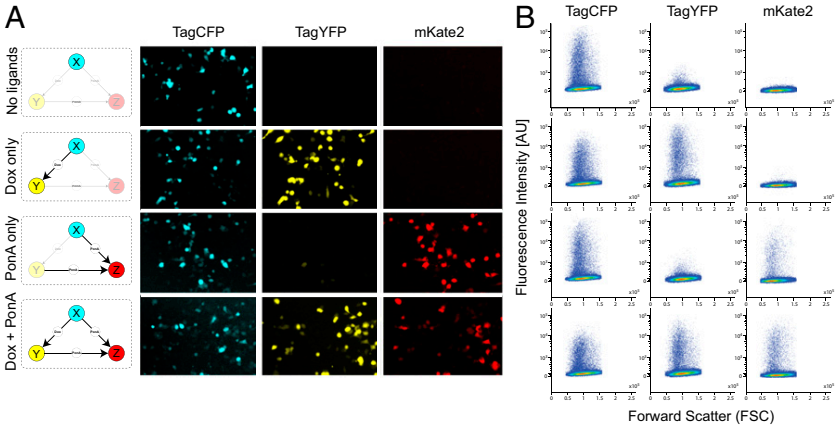

Fig. 2. Validation of the coherent feed-forward architecture. To validate the circuit behavior we tested all combinations of the two small molecules. The result analyzed by fluorescence microscopy (*A*) and flow cytometry (*B*).

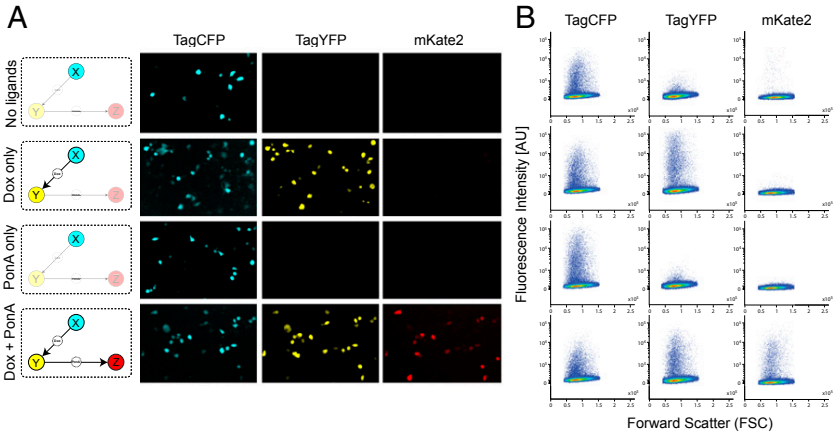

Fig. 3. Validation of the cascade architecture. To validate the circuit behavior we tested all combinations of the two small molecules. The result analyzed by fluorescence microscopy (*A*) and flow cytometry (*B*).

a mutation instead of complete excision to avoid possible discrepancies in transfection and transcription efficiencies due to differences in plasmid and cassette size.

During the design stage we opted for a simple yet effective means of perturbing the individual nodes via RNA interference (Fig. 1*E*). We use a set of siRNA with previously confirmed function (29) for X and Y and a custom siRNA for Z. More specifically, the siRNA-based suppression of node X is achieved through addition of an FF3 target into 3′ untranslated region (UTR) of each transcript produced by the constitutive bidirectional promoter. Similar to X, ubiquitous siRNA-based suppression of Y is made possible by inserting an FF4 target into the 3′ UTR of each transcript produced by the bidirectional TRE enhancer/CMV promoter. Node Z contains a single transcript (reporter protein mKate2), and its activity is modulated by a custom siRNA that directly targets the mKate2 transcript.

Validation of the Synthetic Gene Network Behavior. With the exception of node X which relies on a constitutive promoter, the activity of the synthetic networks depends on the presence of the appropriate ligand. In the cascade motif, Y requires doxycycline (Dox), and Z requires an EcR agonist such as Genostat or ponasterone A (PonA). In the type I coherent feed-forward loop, the requirement for activation of Y remains the same, whereas Z can be activated by the combination of Dox and PonA or PonA alone. To confirm these baseline conditions, the circuits were transfected in human embryonic kidney cell line (HEK293), the ligands were introduced at saturating concentrations, and measurements were performed using microscopy and flow cytometry ∼48 h after transfection.

The microscopy measurements of the fluorescent outputs of both benchmark networks show that the inducible transactivators for both architectures function as desired with minimal leakage, thus confirming the designed circuit topologies. In the feed-forward loop,

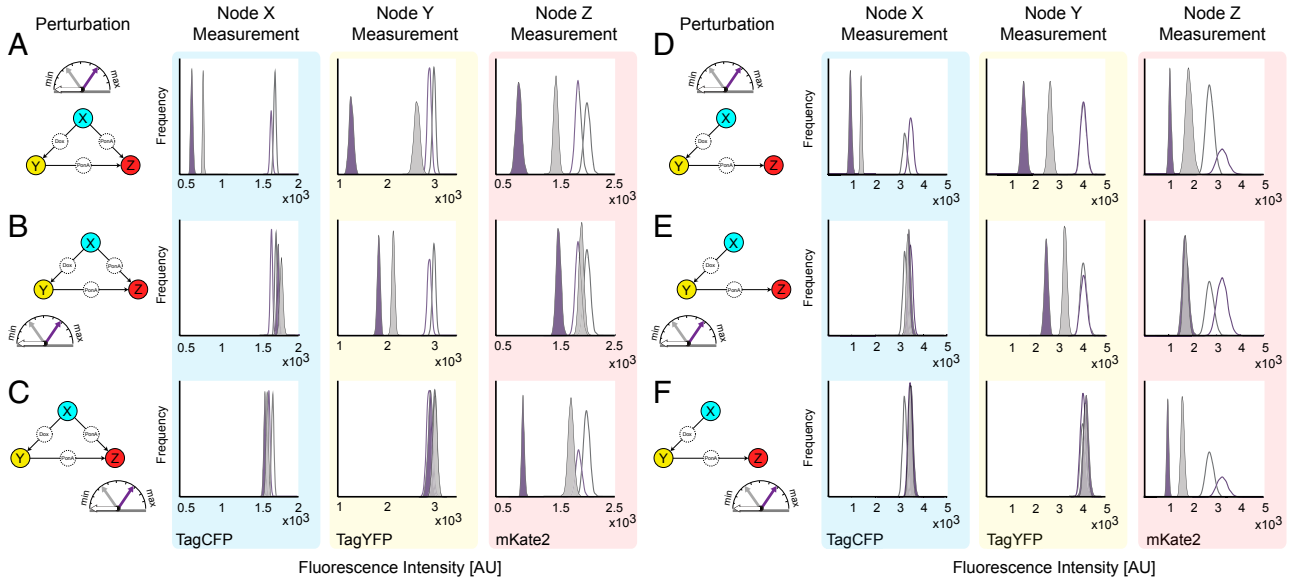

Fig. 4. Resampling the single-cell flow cytometry data after node-wise perturbations. Forty-eight h after siRNA perturbation, the expression level of the fluorescence reporters that represent nodes X, Y, and Z (TagCFP, TagYFP and mKate2, respectively) are measured using flow cytometry. To calculate the mean fluorescence of each population and the associated uncertainty, bootstrap resampling was performed. The resulting probability distributions of the resampled mean before perturbation (empty) and after perturbation (color-filled) are shown for the feed-forward loop (*A*–*C*), and the cascade (*D*–*F*). The colors of the peaks indicate the relative strength of the suppression applied (gray is used to indicate the low and purple the high perturbation). (*A*) The graphical representation of the X-node perturbation and the corresponding nodal responses using the feed-forward architecture. Probability distributions are composed of bootstrapped mean of the fluorescence reporters TagCFP, TagYFP and mKate2 (left to right) following perturbations to node X at two different siRNA concentrations. Color of the peak indicates the relative degree of suppression. (*B* and *C*) The graphical representation of the Y- and Z-node perturbations and the corresponding nodal responses using the feed-forward architecture. (*D*–*F*) Results from the same process using the cascade architecture.

SYSTEMS BIOLOGY

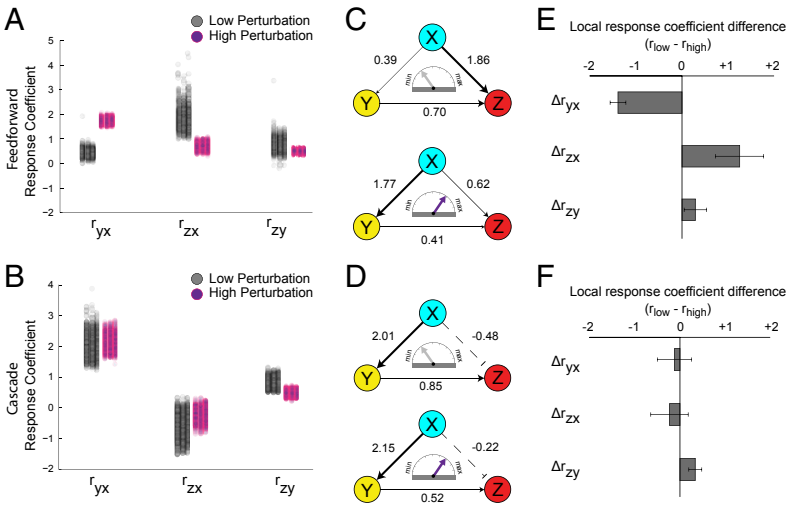

Fig. 5. Reverse engineering of the benchmark topologies using resampled single-cell data. (*A* and *B*) The complete reconstruction of the network with modular response analysis performed after two perturbations. For every set of subsampled means that make up the probability distributions, the MRA results along with the 95% confidence interval of the distribution are plotted as a 1D scatter plot. (*C* and *D*) The graphical representation of the reconstructed synthetic networks. (*E* and *F*) To probe the effect of response coefficient change due to perturbation magnitude shift we calculated the difference between coefficients of equivalent edges (*C* and *D*). The error bars were obtained using a propagation of error among the pair of local response coefficient distributions used to calculate this difference.

node X activity is represented by the constitutively produced fluorescent protein TagCFP and is observed regardless of the ligand conditions (Fig. 2*A*). The addition of doxycycline, which enables X-to-Y interaction by activating the synthetic transactivator rtTA, results in production of the TagYFP fluorescent protein (Fig. 2*A*). The activation of node Z is mediated by the active form of RheoSwitch dimer, which is produced by both nodes X and Y. Due to the constitutive activity of node X, PonA is sufficient to activate node Z in the feed-forward loop (Fig. 2*A*). These observations are confirmed by flow cytometry-based population measurements, which show TagCFP population in all scenarios and a distinct TagYFP population when Dox is present, and mKate2 when PonA is present (Fig. 2*B*). The same control experiments using the cascade network plasmid show an identical response to the ligand combinations except for the node Z activity (Fig. 3*A*). In the cascade motif, sequential activation of node X and node Y are necessary for node Z activation. Thus, mKate2 is only observed when Dox and PonA are present (Fig. 3*A*). Again, we confirm these observations with flow cytometry (Fig. 3*B*).

Subsequently, to probe the parameter space and general behavior of the circuits under perturbations we created mathematical models of our benchmark circuits (*SI Appendix*, *SimBiology Model*, Figs. S1 and S2). The kinetic parameters were selected from literature (21, 30). We performed sensitivity analysis of the output node Z protein concentration against the mRNA species of nodes X and Y (thereby emulating RNAi perturbation). We observe that, in the feed-forward loop, where node X activates node Z in a direct manner as well as an indirect manner, the cumulative sensitivity of the Z node protein to mRNA species of node X was always higher than that of node Y (*SI Appendix*, Fig. S3). Conversely, in the cascade, where node X only activates node Z indirectly through node Y, the production of node Z protein was more sensitive to the node Y mRNA. Based on the simulation results, we hypothesized that the topological differences of the examined architectures may yield divergent responses to different degrees of perturbation. This hypothesis points to an intriguing scenario where the properties and outcome of signal propagation after custom perturbation experiments can be exploited toward distinguishing direct from indirect connectivities.

Modular Response Analysis. An intrinsic difficulty in capturing direct interactions in a biological network is that any perturbation to a particular node may rapidly propagate throughout the network, thus causing global changes which cannot be easily distinguished from direct effects. Rooted in metabolic control analysis, Modular Response Analysis (MRA) uses steady state data obtained from node-wise perturbation to express the network in terms of pair-wise interaction sensitivities. To perform MRA (*SI Appendix*, *Modular Response Analysis Method*), we first calculate the global response coefficients (GRC) from experimentally measured responses to perturbations using $\Delta ln(x_i)$, where x_i represents in our case the quasi steady-state measurement of fluorescent reporter obtained via flow cytometry. Once the functional modules (i.e., perturbation targets) of the target network have been selected, the experimental procedure consists of the following steps: (*i*) measure the steady-state x_i corresponding to the unperturbed set of inputs p_i, (*ii*) perform a perturbation to each p_i individually and measure the new steady-state x_i', (*iii*) calculate the global response coefficients using the steady-state data, and (*iv*) convert global response coefficients to local response coefficients by inversion of the global response matrix.

In higher eukaryotes, perturbation can be achieved through the down-regulation of mRNA, and hence protein levels, using RNA interference (RNAi). This approach has shown to be successful in mapping the positive and negative feedback effects in the Raf/Mek/Erk MAPK network of rat adrenal pheochromocytoma (PC-12) cells (31). Using a variant of the MRA algorithm (32), the authors uncovered connectivity differences depending on whether the cells were stimulated by epidermal growth factor (EGF) or, alternatively, by neuronal growth factor (NGF).

We commenced the experimental reverse engineering process by performing a systematic perturbation of each benchmark architecture node. We first tested the efficacy of siRNA and calibrated the perturbation dosage against the feed-forward loop architecture plasmid (*SI Appendix*, Figs. S4–S6; quantitative RT-PCR results in *SI Appendix*, Fig. S7). As our goal was to find a range of siRNA concentrations that result in moderate yet distinct levels of suppression, we set our maximum siRNA concentration at the manufacturer-recommended dose of 5 pmol and tested five additional concentrations in decreasing magnitudes. To ensure consistency, we cotransfected each siRNA with the network plasmid and measured the circuit activity after 48 h via flow cytometry. Across all three nodes, each of the siRNA was most effective at suppressing the node for which it was designed to disrupt; at least 60% suppression was achieved with the highest siRNA concentration. We then selected the pair of siRNA concentrations that yield the largest difference in the activity of their respective target node, as measured by mean fluorescence level. Specifically, as illustrated in *SI Appendix*, Fig. S8*A*, we selected 1 pmol as "high" perturbation and 0.1 pmol as "low."

After selecting the perturbation magnitude, we performed a node-wise perturbation of the feed-forward circuit using the siRNAs that target each node supplemented with scrambled siRNA to control for the total mass. As before, we used saturating concentrations of the small molecule inducers and applied the predefined set of perturbations based on our calibration results. The three fluorescent reporter protein profiles indicate a response consistent with the benchmark network topologies, confirming the siRNA operation (*SI Appendix*, Fig. S8*B*). Down-regulation in the

fluorescence reporter expression is observed in two cases. Directly by perturbing the actual node or indirectly by perturbing the upstream node responsible for its activation. For node X, a decrease in TagCFP is observed only after direct perturbation; for node Y, a decrease in TagYFP is observed after perturbation of nodes X and Y; and for node Z, a decrease in mKate2 is observed after perturbation of nodes X, Y, and Z (*SI Appendix*, Fig. S8*B*).

We then proceeded with the recovery of the network topology using population data (*SI Appendix*, *Reverse Engineering of the Benchmark Topologies Using Population Data*). For each set of perturbation responses, the global response coefficients were calculated based on the weighted mean fluorescence of gated populations. The pairwise sensitivity coefficients were then obtained via calculating the local response coefficient (LRC) (*SI Appendix*, Figs. S8 *C* and *F* and S9). To determine the significance of the recovered LRC, we performed error propagation using Monte Carlo simulations (31) rendering most of the predicted regulatory connections insignificant (*SI Appendix*, *Monte Carlo Simulation* and Fig. S10). Notably, the reverse engineering recovered a direct inhibitory connection between nodes X and Z for both perturbation magnitudes (*SI Appendix*, Fig. S8*F*), which may be attributed to mild retroactivity effects (33).

Reverse Engineering of the Benchmark Topologies Using Resampled Single-Cell Data. To increase our confidence in predictions we developed a technique based upon bootstrapping, an alternative to the sample statistics obtained from an aggregate population (i.e., mean and SD). Bootstrapping is a nonparametric resampling method (34) designed to estimate the confidence interval of a given statistic, and is particularly useful when the observed population distribution cannot be characterized by typical distributional assumptions such as normality (e.g., typical flow cytometry data). To obtain a bootstrapped mean we: (*i*) resample with replacement the dataset to the same number of times as the original population, (*ii*) calculate the desired statistic from each sample, (*iii*) repeat the process several times to form the probability distribution of the subsampled mean (*SI Appendix*, Fig. S11). For our analysis, we repeated the entire process 2,000 times to form the representative probability distribution of a fluorescent reporter expression.

Subsequently, we produced a unique panel for each of the three fluorescent reporters after perturbations to three different nodes, for a total of nine distinct panels for the feed-forward circuit (Fig. 4 *A*–*C*) and nine for the cascade (Fig. 4 *D*–*F*). Each frequency plot consists of four different probability distributions, from resampled means of fluorescent reporter before (empty) and after the two perturbations (filled with gray for low and purple for high). For every set of subsampled means that make up the probability distributions, we treated them as a unique instance of the perturbation response and fed these values to MRA to calculate the local response coefficients. The results, along with the 95% confidence interval of the distribution, are plotted as a 1D scatter plot and shown in Fig. 5 *A* and *B*. In this case, we were able to successfully recover all relevant regulatory connections of feed-forward loop (Fig. 5*C*) and cascade (Fig. 5*D*) networks with increased confidence. All of the reconstructed edges are included in *SI Appendix*, Fig. S12.

Using our prior knowledge of our network, we confirmed that the inferred connectivities are consistent with the network topology. Moreover, the inferred interaction strengths and distributions produced by the reverse engineering algorithm reveal features of the network that are not readily apparent and are in agreement with our in silico sensitivity analysis. Specifically, to probe the effect of response coefficient change due to perturbation magnitude shift we calculated the difference between coefficients of equivalent edges (Fig. 5 *E* and *F*).

For the feed-forward architecture we identified "perturbation-sensitive edges." In other words, we discovered edges that undergo distinctively large shift in interaction strengths under differential perturbations. In particular, we observe that for the feed-forward architecture, increasing the perturbation magnitude dramatically alters the inferred interactions arising from node X. Importantly, there is a noticeable reversal in the strengths of activation between node X to nodes Y and Z, whereas the interaction from node Y to node Z remain largely unchanged (Fig. 5*C*). Specifically, after a low perturbation, the recovered topology shows a prominent direct activation of node Z by node X, whereas the topology recovered after high perturbation shows a prominent activation of node Y by node X. We postulate that the low perturbation is buffered as it propagates through the intermediate node Y, therefore the direct connection between X and Z appears to be more important. Aligned with this observation, node Z is less sensitive to disruption of node Y in the context of the feed-forward loop (Fig. 5*B*).

Compared with the feed-forward circuit, we found that the reverse engineering of the cascade is robust to perturbation strengths. In fact, the recovered topology from two perturbation magnitudes are almost indistinguishable except for small decrease in the Y-to-Z interaction strength (Fig. 5*F*), despite the fact that the fluorescence reporter profiles clearly reflect the differences in perturbation magnitude (Fig. 4 *D*–*F*). In contrast with the feed-forward circuit, there is only one possible path of activation in the cascade motif thereby the presumed buffering effect is not critical.

To further explore the effect of differential perturbation on the reverse engineering results, we performed an additional experiment using three perturbation magnitudes (*SI Appendix*, Fig. S13). In this instance, we refer to the perturbation magnitudes as "low," "medium," and "high." We again observe the diverging trend of response coefficient values between the two architectures. In the feed-forward loop, each step-wise increase in perturbation magnitude affects the two edges originating from node X in contrasting manner, highlighting the activation from node X to node Y while reducing the weight of activation from node X to node Z (*SI Appendix*, Fig. S13 *A* and *C*). The recovered topologies of the cascade motif undergo little to no change over same perturbation magnitude intervals (*SI Appendix*, Fig. S13 *B* and *D*). Finally, to qualitatively probe our observations we developed a phenomenological model of the architectures (*SI Appendix*, *Phenomenological Model*). Using this model we analytically calculated the local response coefficients under low and high perturbations and we indeed confirmed the divergent shifts in interaction strengths.

Discussion

Direct and indirect interactions are pervasive in all networks. The inability to disentangle these interactions hampers reverse engineering progress. Recent advancements in high-throughput approaches, combined with algorithm and methodological advances through a host of community-wide efforts (12, 14, 19, 35) have examined these aspects. In fact, attempts to fundamentally address the issue by recognizing and filtering out the effects of indirect interactions at a global scale have begun to surface (11). Meanwhile, parallel developments in synthetic biology (23) have endowed researchers with new tools that allow precise emulation of naturally occurring topologies (21, 22). Networks orthogonal to the cellular milieu can serve as a biomolecular topological "ground truth" (20, 24). Data gathered from benchmark synthetic circuits can complement and inform algorithms, and offer a unique opportunity to correlate topological properties to system identification.

The number of possible networks for a given set of nodes is large and it grows exponentially with the number of nodes, making impractical their exhaustive construction. Fortunately, recent research has uncovered that certain topologies appear more frequently than others. Those topologies were dubbed "network motifs (25, 36). The network topology does not specify the nature of the nodes, and indeed the expectation is that the network behavior will be invariant to the changes in the molecular nature of the nodes and the exact mechanism of the interactions between the nodes.

Here we constructed two synthetic networks that incorporate direct and indirect connectivities. We successfully engineered the benchmark architectures to be inducible with negligible leakage and amenable to simple perturbations to facilitate the reverse engineering analysis. After applying systematic perturbations and a combination of nonparametric single-cell data resampling and modular response analysis, we discovered response patterns that are markedly different between the two topologies.

Using the proposed methodology, individual nodes of a network can be perturbed from their steady-state using transcriptional or posttranscriptional inhibitors [e.g., TALEs/CRISPR (37, 38) or siRNAs]. The pre- and postperturbation steady states can be measured at the mRNA or protein levels, and fed into MRA to predict divergent LRC and accordingly the network structure. Beyond small-scale networks, although motifs are composed of relatively few elements, they are often embedded as "modules" (39–41) in large networks that exhibit complex behavior. The term "modular" in MRA indicates that the same theoretical tools, in principle, scale up to cover large networks that are connected through a small number of "communicating intermediaries" (4, 28).

To conclude, unraveling the complexity of biological networks is central to understanding biology. Our results point to a transformative opportunity in reverse engineering of biological networks. Taking into account inferred topological changes under differential perturbations may provide a solution to the longstanding problem of discriminating between direct and indirect connections.

Methods

Mammalian Cell Culture and Transfections. HEK293 cell line was maintained at 37°C, 100% humidity and 5% (vol/vol) CO_2. Circuit plasmid transfection was performed with jetPRIME (Polyplus) in 12-well plates at a plating density of 200,000 cells. Transfection was performed 24 h after seeding, and each well received 10 ng of plasmid containing node X and 25 ng of plasmid containing nodes Y and Z, with 500 ng of cotransfection junk DNA and varying amounts of siRNA. Detailed information is provided in *SI Appendix, SI Methods*.

Fluorescence Microscopy. Approximately 48 h after transfection of network plasmid, fluorescence images of live cells were captured using an Olympus IX81 microscope. For ambient temperature control, the entire apparatus was housed in a Precision Control environmental chamber. The images were captured using a Hamamatsu ORCA 03 digital camera. Detailed information is provided in *SI Appendix, SI Methods*.

Flow Cytometry. All FACS experiments were performed 48 h after transfection with BD LSRFortessa. Data acquisition was performed using FACS Diva software and subsequent analysis with FlowJo (Treestar). The threshold fluorescence unit for selecting fluorescence-positive population was determined based on untransfected HEK293 cells (*SI Appendix*, Fig. S14). There was no compensation performed (*SI Appendix*, Fig. S15). Detailed information is provided in *SI Appendix, SI Methods*.

Modular Response Analysis. To obtain the pair-wise sensitivities between each node, we performed modular response analysis. In MRA, the local intermodular interactions, described by the local response matrix r_{ij}, are calculated from the global response matrix R_{ij}, which contains the observed change in the steady-state measurement of each node (x_i) due to the experimental perturbation (p_j). Because precise measurement of parameter perturbation size (p_j) is not possible in an experimental setting, the global response matrix is approximated by the fractional change of the steady states $R_{ij} \sim \Delta ln(x_j)$. After obtaining R_{ij} from experimental data, we calculate the local response matrix r_{ij} by solving $r_{ij} = -[dg(R_{ipj}^{-1})]^{-1} \cdot R_{ipj}^{-1}$. Detailed information is provided in *SI Appendix, Modular Response Analysis Method*.

Resampling. To estimate the 95% confidence interval of the obtained local response matrix, bootstrap resampling of the original flow cytometry population is performed. The steps for bootstrap resampling are as follows: From the original flow cytometry population, resample with replacement the same number of cells as the original population. Using the resampled population, compute the desired population statistic (mean), and then calculate the local response matrix using MRA. The bootstrapping and MRA process is repeated 2,000 times to create a distribution of local responses. The 95% confidence interval, which corresponds to values from 2.5th to 97.5th percentile of the calculated values, is used to estimate the error. The process is shown in *SI Appendix*, Fig. S11.

ACKNOWLEDGMENTS. This work was funded by the US National Institutes of Health Grants GM098984, GM096271, CA17001801, National Science Foundation Grant CBNET-1105524, and the University of Texas at Dallas. E.S. partially supported by Air Force Office of Scientific Research Grant FA9550-14-1-0060.

1. Csete ME, Doyle JC (2002) Reverse engineering of biological complexity. *Science* 295(5560):1664–1669.
2. Gardner TS, di Bernardo D, Lorenz D, Collins JJ (2003) Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 301(5629):102–105.
3. Khammash M (2008) Reverse engineering: The architecture of biological networks. *Biotechniques* 44(3):323–329.
4. Kholodenko BN, et al. (2002) Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. *Proc Natl Acad Sci USA* 99(20):12841–12846.
5. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 308(5721):523–529.
6. Kholodenko B, Yaffe MB, Kolch W (2012) Computational approaches for analyzing information flow in biological networks. *Sci Signal* 5(220):re1.
7. Basso K, et al. (2005) Reverse engineering of regulatory networks in human B cells. *Nat Genet* 37(4):382–390.
8. Chen JC, et al. (2014) Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. *Cell* 159(2):402–414.
9. Yeung MK, Tegnér J, Collins JJ (2002) Reverse engineering gene networks using singular value decomposition and robust regression. *Proc Natl Acad Sci USA* 99(9):6163–6168.
10. Tegner J, Yeung MKS, Hasty J, Collins JJ (2003) Reverse engineering gene networks: Integrating genetic perturbations with dynamical modeling. *Proc Natl Acad Sci USA* 100(10):5944–5949.
11. Feizi S, Marbach D, Médard M, Kellis M (2013) Network deconvolution as a general method to distinguish direct dependencies in networks. *Nat Biotechnol* 31(8):726–733.
12. Marbach D, et al. (2010) Revealing strengths and weaknesses of methods for gene network inference. *Proc Natl Acad Sci USA* 107(14):6286–6291.
13. Margolin AA, et al. (2006) ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7(Suppl 1):S7.
14. Marbach D, et al.; DREAM5 Consortium (2012) Wisdom of crowds for robust gene network inference. *Nat Methods* 9(8):796–804.
15. de la Fuente A, Brazhnik P, Mendes P (2002) Linking the genes: Inferring quantitative gene networks from microarray data. *Trends Genet* 18(8):395–398.
16. Friedman N (2004) Inferring cellular networks using probabilistic graphical models. *Science* 303(5659):799–805.
17. Friedman N, Linial M, Nachman I, Pe'er D (2000) Using Bayesian networks to analyze expression data. *J Comput Biol* 7(3-4):601–620.
18. Pe'er D (2005) Bayesian network analysis of signaling networks: A primer. *Sci Signaling* 2005(281):pl4.
19. Prill RJ, et al. (2010) Towards a rigorous assessment of systems biology models: The DREAM3 challenges. *PLoS One* 5(2):e9202.
20. Cantone I, et al. (2009) A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. *Cell* 137(1):172–181.
21. Bleris L, et al. (2011) Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Mol Syst Biol* 7:519.
22. Shimoga V, White JT, Li Y, Sontag E, Bleris L (2013) Synthetic mammalian transgene negative autoregulation. *Mol Syst Biol* 9:670.
23. Lienert F, Lohmueller JJ, Garg A, Silver PA (2014) Synthetic biology in mammalian cells: Next generation research tools and therapeutics. *Nat Rev Mol Cell Biol* 15(2):95–107.
24. Kang T, et al. (2013) Reverse engineering validation using a benchmark synthetic gene circuit in human cells. *ACS Synth Biol* 2(5):255–262.
25. Milo R, et al. (2002) Network motifs: Simple building blocks of complex networks. *Science* 298(5594):824–827.
26. Ma'ayan A, et al. (2005) Formation of regulatory patterns during signal propagation in a Mammalian cellular network. *Science* 309(5737):1078–1083.
27. Fire A, et al. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* 391(6669):806–811.
28. Sontag E, Kiyatkin A, Kholodenko BN (2004) Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data. *Bioinformatics* 20(12):1877–1886.
29. Rinaudo K, et al. (2007) A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat Biotechnol* 25(7):795–801.
30. Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M (2009) A tunable synthetic mammalian oscillator. *Nature* 457(7227):309–312.
31. Santos SDM, Verveer PJ, Bastiaens PIH (2007) Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nat Cell Biol* 9(3):324–330.
32. Andrec M, Kholodenko BN, Levy RM, Sontag E (2005) Inference of signaling and gene regulatory networks by steady-state perturbation experiments: Structure and accuracy. *J Theor Biol* 232(3):427–441.
33. Del Vecchio D, Ninfa AJ, Sontag ED (2008) Modular cell biology: Retroactivity and insulation. *Mol Syst Biol* 4:161.
34. Efron B (1979) Bootstrap methods: Another look at the jackknife. *Ann Stat* 7(1):1–26.
35. Prill RJ, Saez-Rodriguez J, Alexopoulos LG, Sorger PK, Stolovitzky G (2011) Crowdsourcing network inference: The DREAM predictive signaling network challenge. *Sci Signal* 4(189):mr7.
36. Alon U (2007) *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman & Hall/CRC, Boca Raton, FL), p 301.
37. Moore R, et al. (2015) CRISPR-based self-cleaving mechanism for controllable gene delivery in human cells. *Nucleic Acids Res* 43(2):1297–1303.
38. Li Y, Moore R, Guinn M, Bleris L (2012) Transcription activator-like effector hybrids for conditional control and rewiring of chromosomal transgene expression. *Sci Rep* 2:897.
39. Kreimer A, Borenstein E, Gophna U, Ruppin E (2008) The evolution of modularity in bacterial metabolic networks. *Proc Natl Acad Sci USA* 105(19):6976–6981.
40. Bassett DS, et al. (2011) Dynamic reconfiguration of human brain networks during learning. *Proc Natl Acad Sci USA* 108(18):7641–7646.
41. Bullmore E, Sporns O (2009) Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10(3):186–198.

Supplementary Information

Discriminating direct and indirect connectivities in biological networks

Taek Kang[1,2], Richard Moore[1,2], Yi Li[1,2], Eduardo Sontag[4], and Leonidas Bleris[*1,2,3]

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

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

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

[4]Department of Mathematics and Center for Quantitative Biology, Rutgers-The State University of New Jersey, Hill Center, 110 Frelinghuysen Rd., Piscataway NJ 08854 USA

*Corresponding author: L.B. (bleris@utdallas.edu)

Keywords: Reverse Engineering, Benchmark Synthetic Circuits, Human Cells, Modular Response Analysis, Direct and Indirect Connectivities, Nonparametric Resampling, Bootstrap.

Supplementary Figures

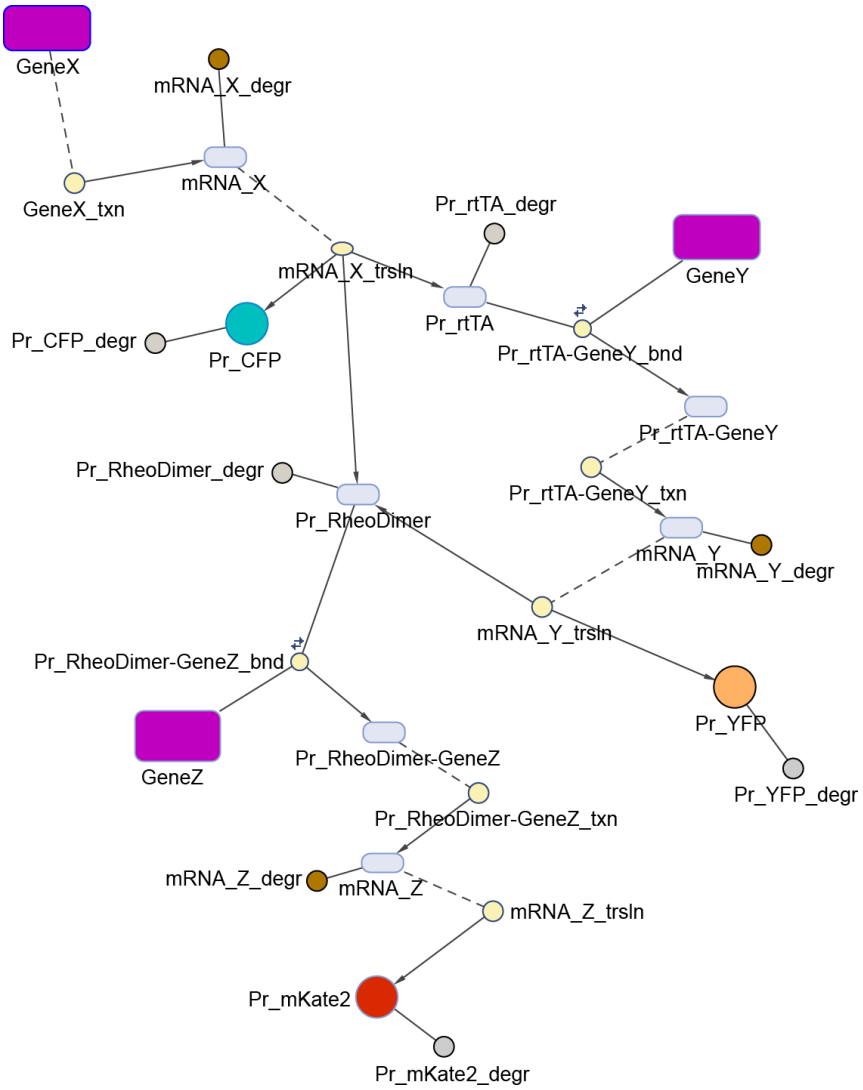

Figure S1. SimBiology model diagram of the synthetic feedforward architecture. To simulate the network behavior and their response to perturbations, the synthetic networks were recreated as mathematical models using MATLAB SimBiology. Each model is composed of distinct nodes that represent DNA, RNA and protein species and reactions that connect these nodes. The published kinetic parameters are listed in Table S4.

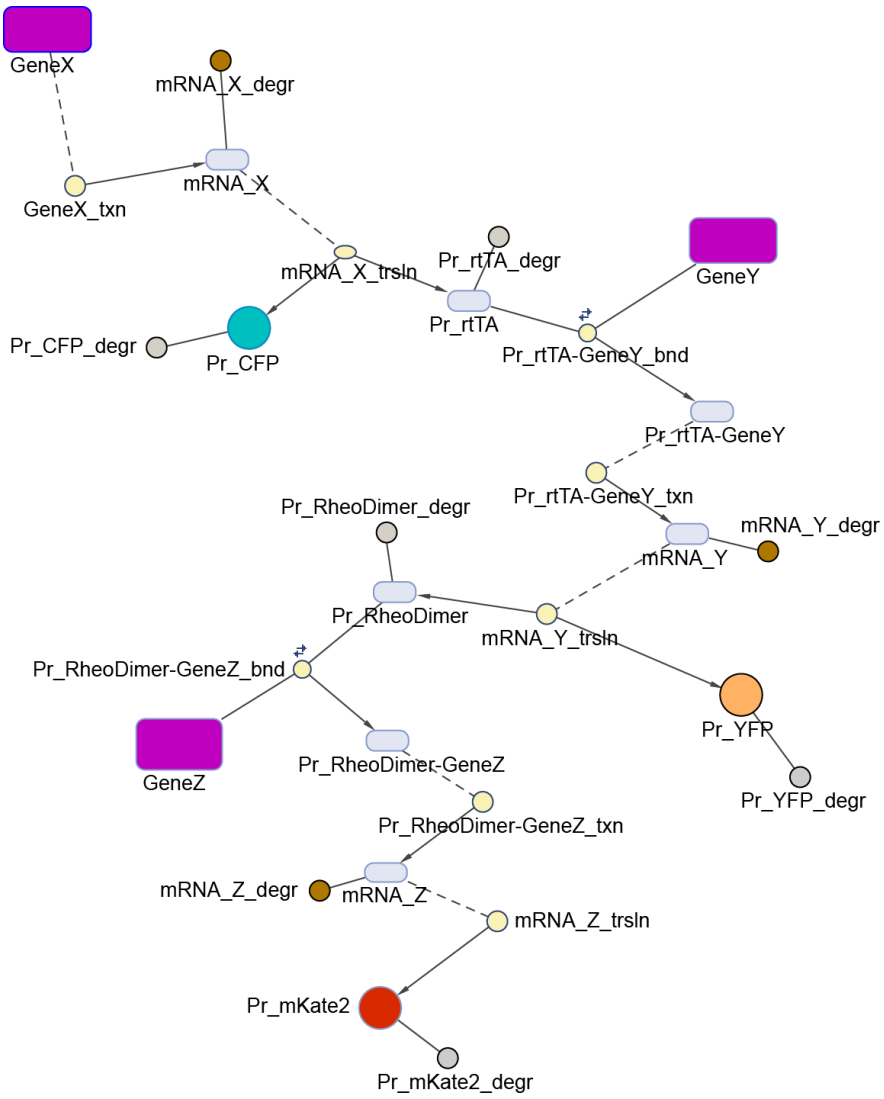

Figure S2. SimBiology model diagram of the synthetic cascade network. Mathematical model of synthetic cascade network was created using MATLAB SimBiology. See Supplementary Figure 1 for details.

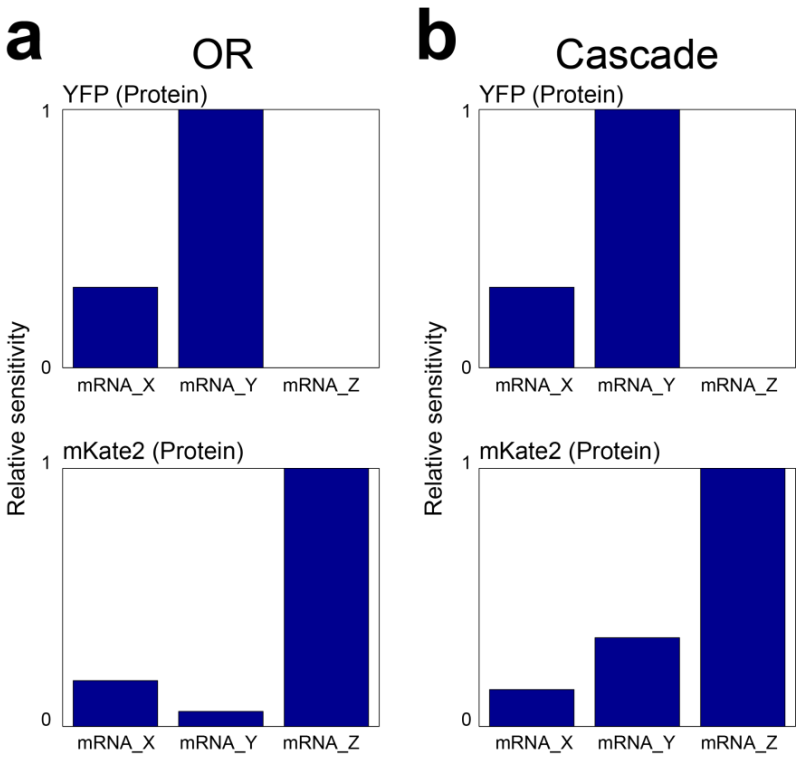

Figure S3. In silico sensitivity analysis of the model benchmark networks. Using the mathematical model of the synthetic networks as presented in Supplementary Figure 1, sensitivity analysis of YFP and mKate2 protein against mRNA species of each node was performed.

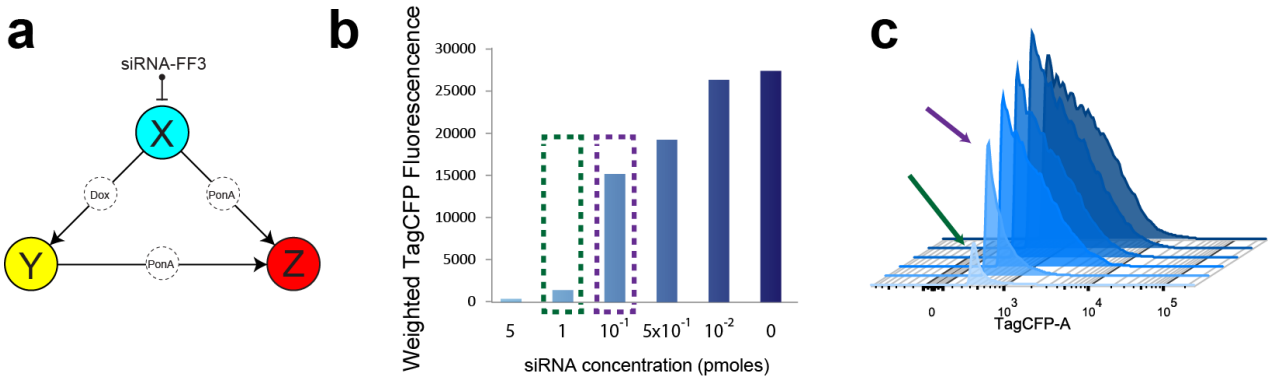

Figure S4. Node X response after perturbation with siRNA that target node X. Each node of the synthetic feedforward loop is perturbed via RNA interference. Ubiquitous suppression of nodes X and Y are achieved by incorporating synthetic siRNA target sites in the 3' UTR of the transcripts. For node Z, a synthetic siRNA that targets the mKate2 mRNA was used. To measure the efficacy of each siRNA, titrations were performed. (a) The activity of node X consists of fluorescent reporter TagCFP, the transactivator proteins rtTA and components of the RheoDimer (RheoActivator and RheoReceptor), and its activity is suppressed via siRNA FF3. (b) Weighted fluorescence of TagCFP obtained from fluorescence-positive cells after siRNA titration. Measurement surrounded by the purple box refers to the siRNA concentration that was included in a "weak" perturbation set, and the green box indicates siRNA concentration belonging to a "strong" perturbation set. (c) Histogram of the fluorescence-positive population, with the arrows indicating the corresponding population from (b).

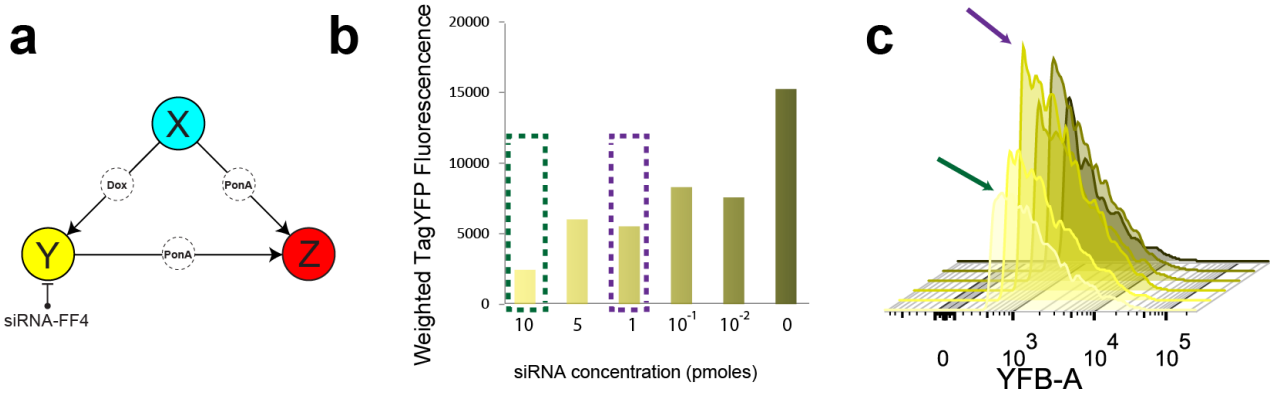

Figure S5. Synthetic network response following perturbation with siRNA. Node Y consists of the fluorescent reporter TagYFP and components of RheoDimer transactivator, and its activity is suppressed by FF4 siRNA. See Figure S4 for description of the experiment.

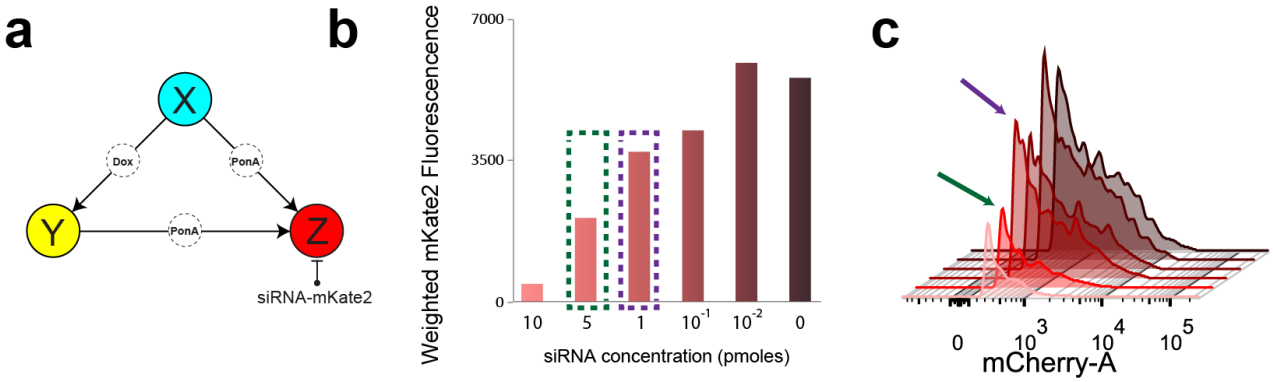

Figure S6. Synthetic network response following perturbation with siRNA. Node Z consists of fluorescent reporter mKate2, and its activity is suppressed by a custom siRNA that targets the mKate2 mRNA. See Figure S4 for description of the experiment.

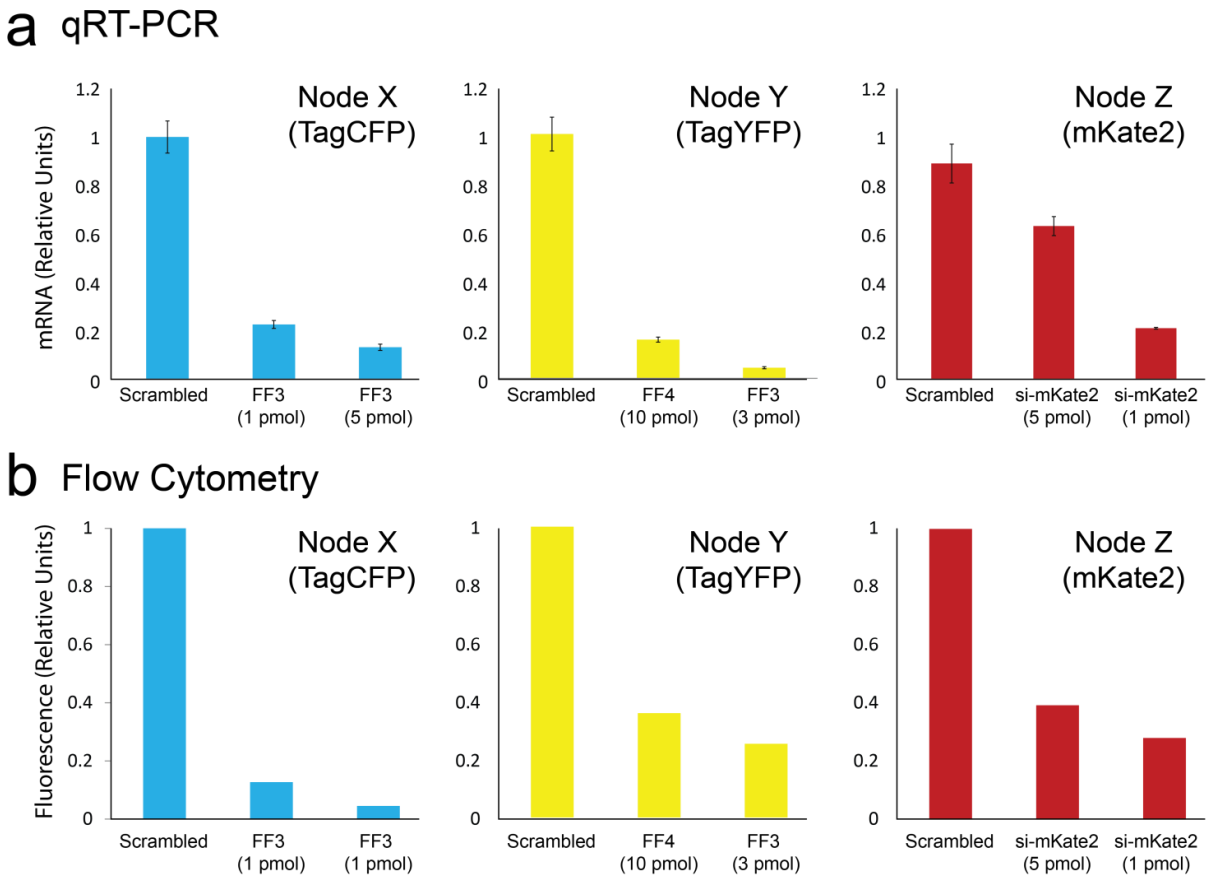

Figure S7. qRT-PCR of the synthetic network following systemic perturbation. (a) To directly test for the effectiveness of the siRNA on its target, qRT-PCR was performed 48 hours after co-transfection of the circuit cassette and the siRNA that target each node. (b) Flow cytometry was performed to confirm the relative protein expression after each perturbation condition. The protocol for qRT-PCR is described in SI Appendix, Supplementary Methods.

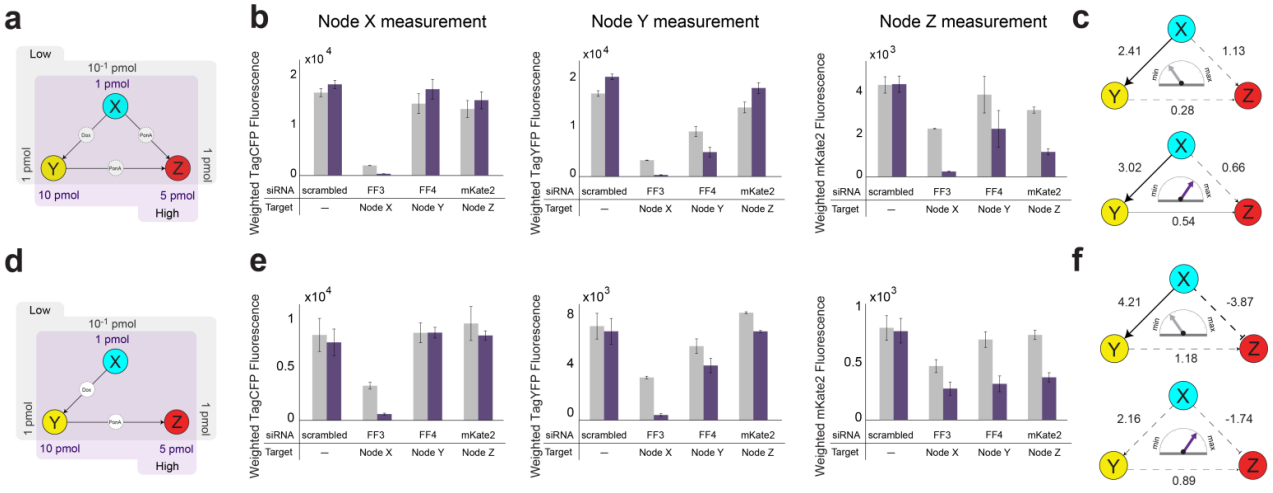

Figure. S8 Network reverse engineering using population data. The nodes X, Y, and Z of the benchmark synthetic networks are quantified by fluorescent reporters TagCFP, TagYFP and mKate2, respectively. (a) Graphical representation of the perturbation properties. Systematic perturbation to the network is performed twice, each with a different set of siRNA concentrations. Each siRNA concentrations are categorized as either "weak" or "strong" to indicate the relative strength of suppression. The set with lower overall concentration of siRNAs ("low") is in grey and the set with higher concentration ("high") is in purple. The same color scheme is used to distinguish these two experiments in subsequent analysis. (b) Perturbation responses of the gated population to perturbations to nodes X, Y and Z. To evaluate the fluorescence profile of the architecture post-perturbation, flow cytometry was performed 48 hours after transfection of the circuit. Weighted fluorescence corresponds to the percentage of the fluorescence-positive gated cells multiplied the mean fluorescence of the population. Error bars indicate standard deviation among triplicate experiments. (c) Graphical representation of the recovered circuit topology using population data. Network inference using MRA was performed with the weighted fluorescence profile shown in (b). Uncertainty of each local response coefficient was calculated by Monte Carlo simulations. Dotted lines are used to indicate that the 95% confidence interval of the corresponding distribution includes zero, which renders the connection as statistically insignificant. The corresponding results from the same experiment using the cascade architecture are shown in panels (d – f).

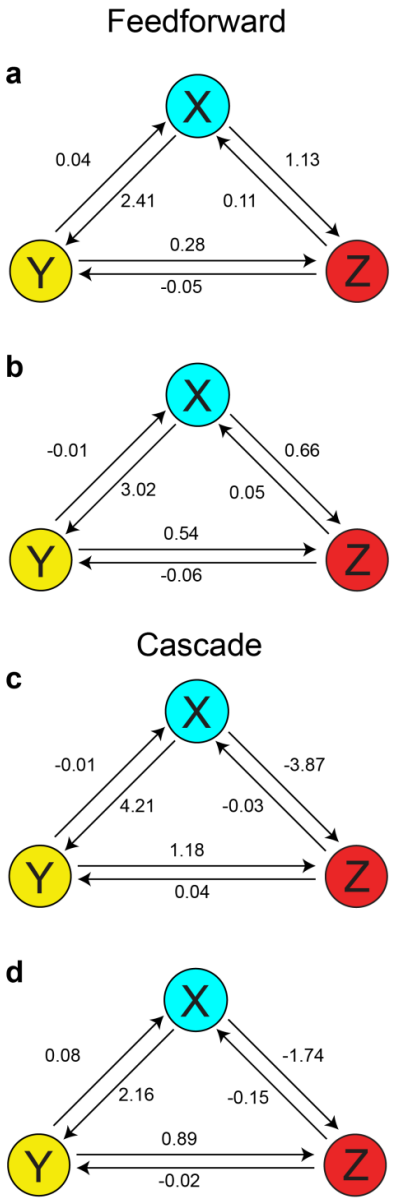

Figure S9. Graphical representation of the complete circuit topology derived from population-level statistics. Interaction coefficients for all regulatory links were calculated using modular response analysis, using mean fluorescent values of each population. Circuit topology of the feedforward architecture following weak and strong perturbation are shown in (a,b), respectively. Corresponding results from the cascade architecture are shown in (c,d).

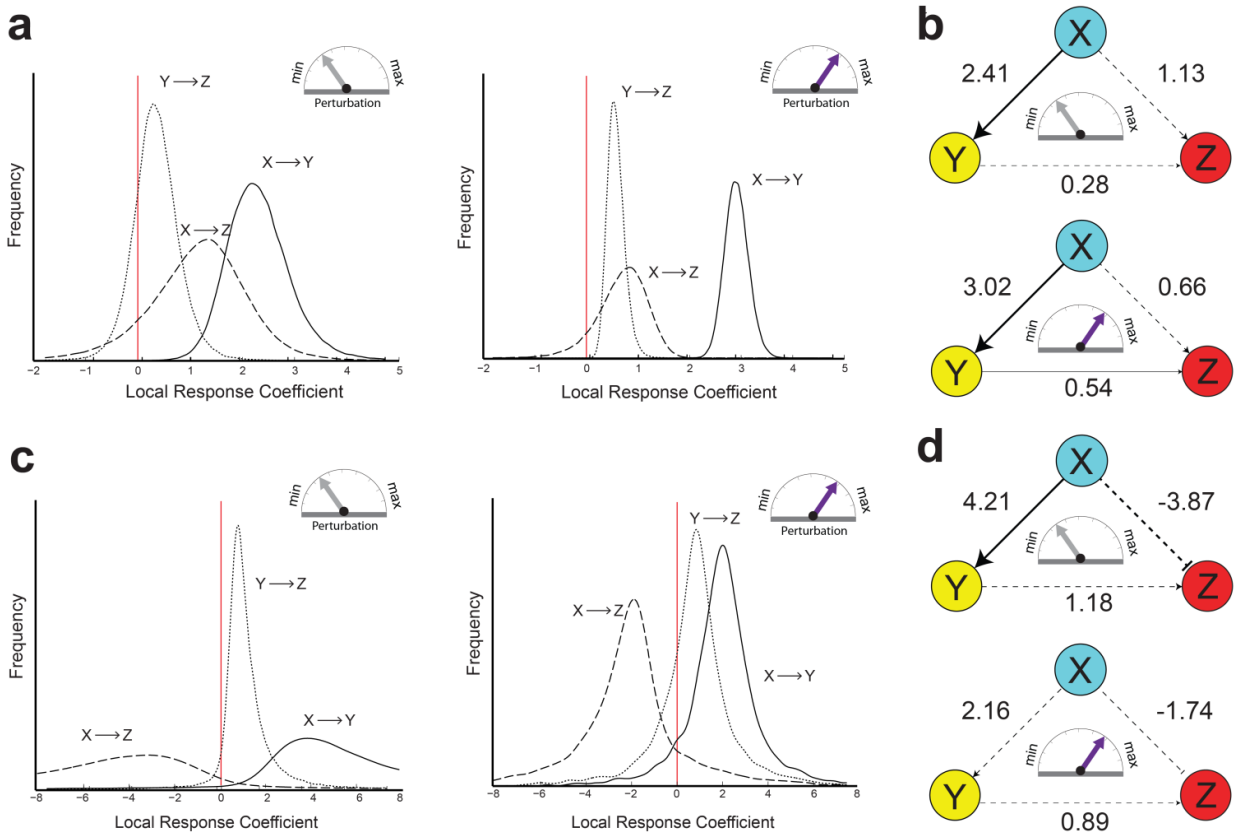

Figure S10. Monte Carlo error propagation analysis of modular response analysis. (a) To confirm the significance of the calculated local response coefficients, error propagation analysis using Monte Carlo simulation was performed. Each peak represents the probability distribution of calculated local response coefficients (LRCs) of the connections that make up the feedforward loop topology after weak perturbation (left) and strong perturbation (right). (b) Graphical representation of the recovered circuit topology, with mean LRCs from (a). (c) Monte Carlo simulation results after perturbation of the cascade networks and the corresponding recovered circuit topology (d).

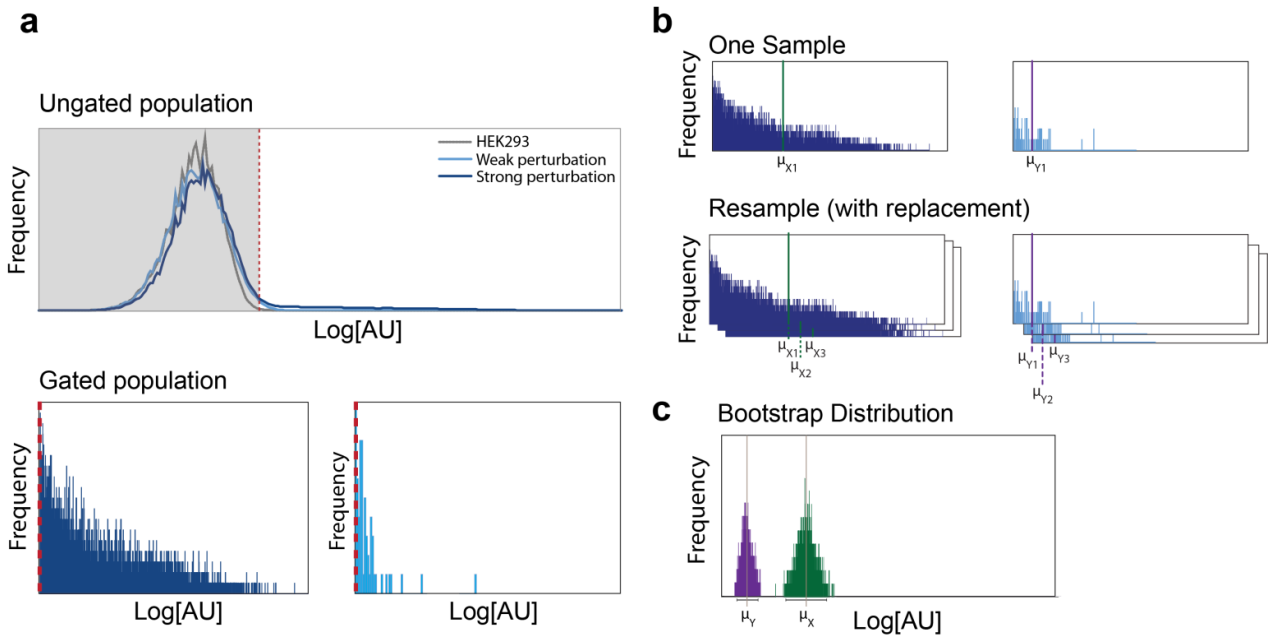

Figure S11. Bootstrap resampling workflow. To calculate the error of the obtained population statistic of the flow cytometry data, bootstrap distribution is used. (a) After gating for fluorescence-positive cells, we randomly sample (with replacement) to create a resampled dataset. (b) The size of a bootstrap resampled dataset is identical to the original population from which we sample from. We then calculate the desired population statistic (mean). (c) The repetition of this process produces a distribution and the confidence interval of means, which we can use as an estimator for the error of the calculated mean.

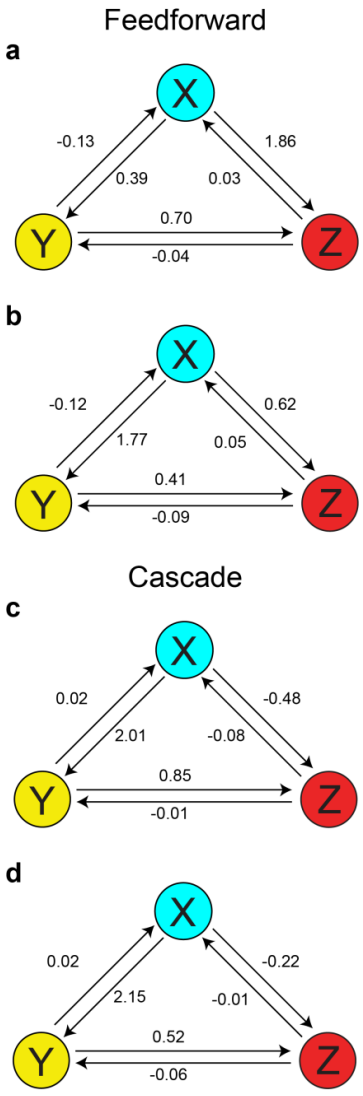

Figure S12. Graphical representation of the complete circuit topology derived from single cell data. Interaction coefficients for all regulatory links were calculated using modular response analysis using resampled fluorescence data. Circuit topology of the feedforward architecture following weak and strong perturbation are shown in (a, b), respectively. Corresponding results from the cascade architecture are shown in (c, d).

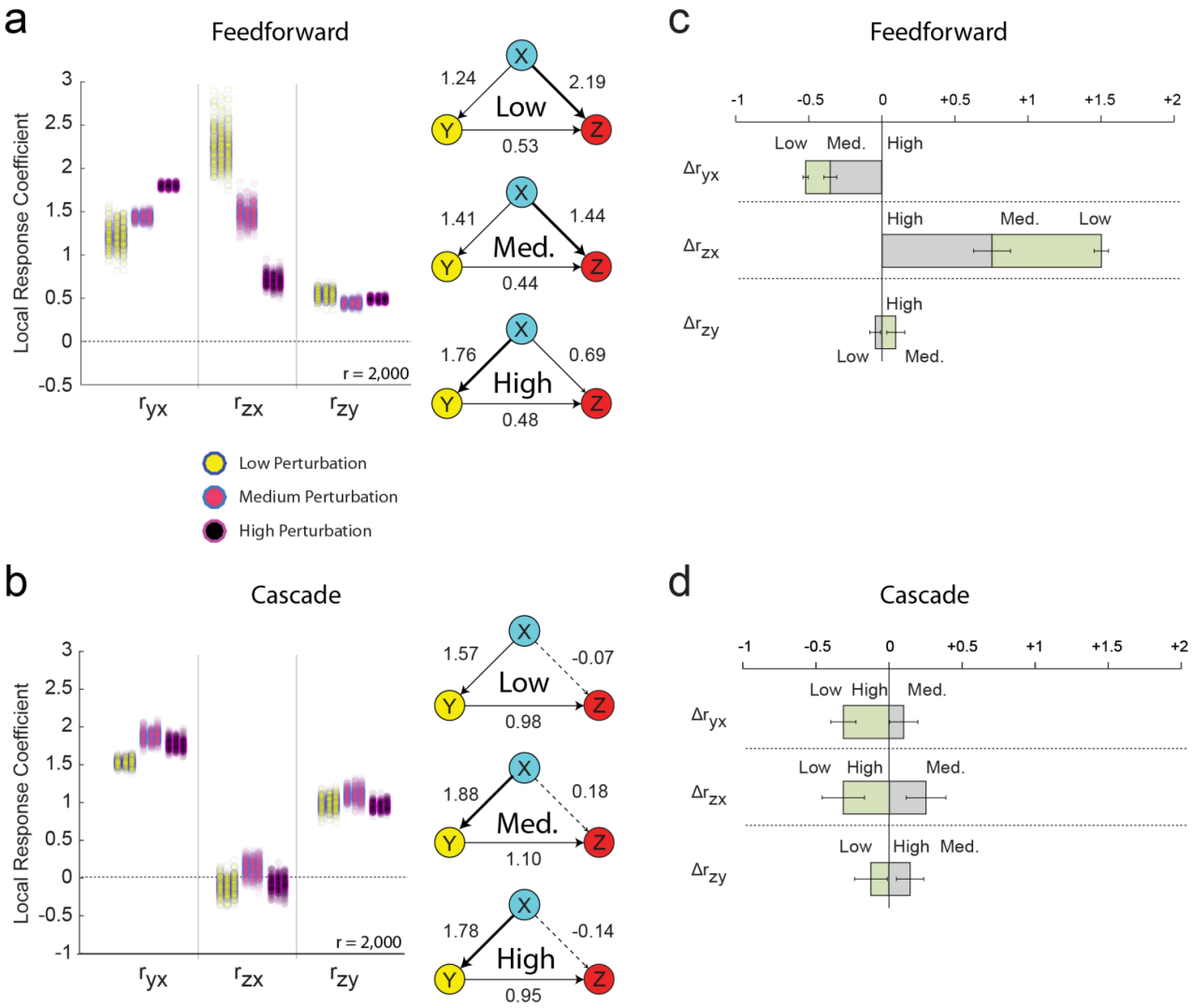

Figure S13. Reverse engineering of the benchmark topologies using resampled single-cell data. The complete reconstruction of the feedforward (a) and cascade (b) with modular response analysis performed after three different magnitudes of perturbations. For every set of subsampled means that make up the probability distributions, local response coefficients are calculated. This process cycle is performed 2,000 times, and the resulting local response coefficient distribution is plotted as a 1 dimensional scatter plot, and the corresponding graphical representation of the reconstructed synthetic networks with the mean values of these distributions are shown on the right (From top to bottom: low, medium and high perturbations, respectively). After reconstruction of the synthetic networks using three distinct sets of systemic perturbation, the change in response coefficients of equivalent edges are calculated after subsequent decrease in perturbation magnitude for feedforward (d) and cascade (d). The response coefficients recovered after the strongest perturbation sets ("High") are used as a reference point. The uncertainty associated with these values was obtained by error propagation based on the standard deviations of the original distributions.

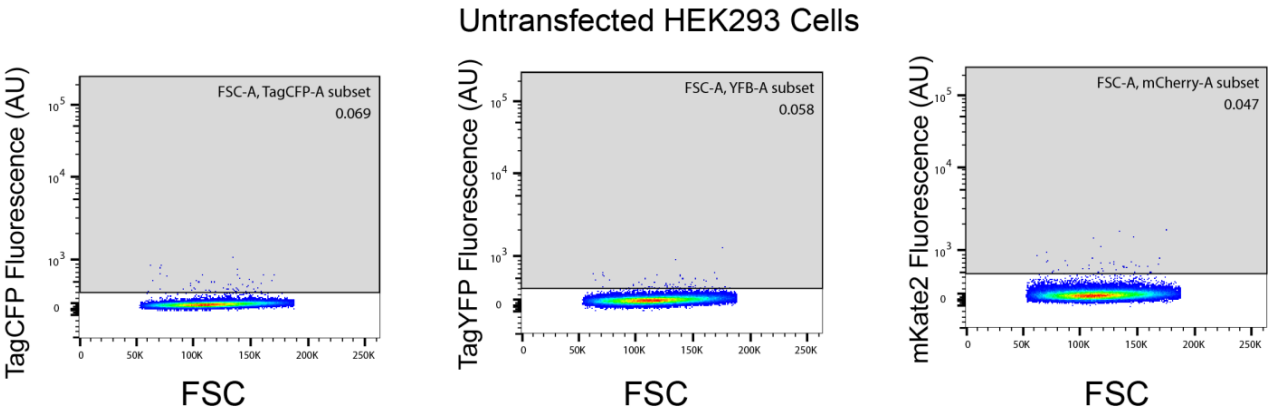

Figure S14. Fluorescence threshold used for gating. The threshold units used to select for fluorescence-positive cells were selected based on fluorescence profile of the untransfected HEK 293 cells.

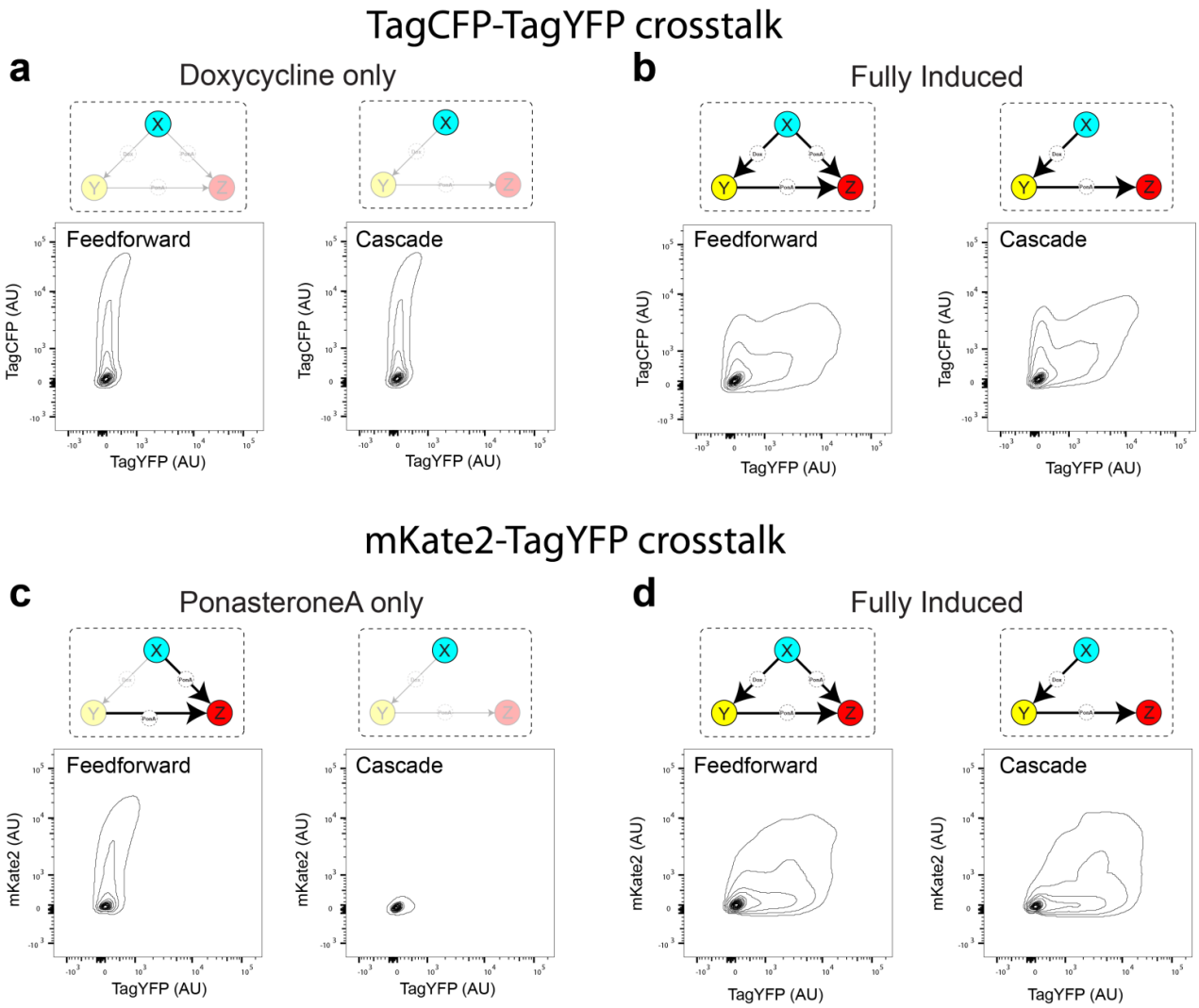

Figure S15. Crosstalk observed at working plasmid concentration. To check for potential cross-talk between the three fluorescent reporters, TagCFP and mKate2 proteins were expressed in the absence of TagYFP (a,c) and with active expression of TagYFP (b,d).

SimBiology Model

The SimBiology reaction array models for Supplementary Figures 1 and 2:

GeneX -> mRNA_X + GeneX

mRNA_X -> null

mRNA_X -> Pr_CFP + Pr_RheoDimer + Pr_rtTA + mRNA_X

Pr_CFP -> null

Pr_rtTA -> null

Pr_RheoDimer -> null

Pr_rtTA + GeneY <-> [Pr_rtTA-GeneY]

[Pr_rtTA-GeneY] -> mRNA_Y + [Pr_rtTA-GeneY]

mRNA_Y -> mRNA_Y + Pr_YFP + Pr_RheoDimer

mRNA_Y -> null

Pr_YFP -> null

Pr_RheoDimer + GeneZ <-> [Pr_RheoDimer-GeneZ]

[Pr_RheoDimer-GeneZ] -> [Pr_RheoDimer-GeneZ] + mRNA_Z

mRNA_Z -> Pr_mKate2 + mRNA_Z

mRNA_Z -> null

Pr_mKate2 -> null

Modular response analysis (MRA) method

We define the differential equations describing the system's dynamics as the following general form:

$$\dot{x}_i = F_i(x_1, \ldots x_N, p_i), \quad i = 1, \ldots, N$$

At steady state, the same equation can be written as:

$$\dot{x}_i = F_i(x_1, \ldots x_N, p_i) = 0$$

The MRA method amounts to the following procedure. We are interested in finding the direct dependencies of each variable x_i on each other variable x_j. First, we describe the system based on the implicit function theorem:

$$x_i = x_i(x_1(p_k), x_2(p_k), \ldots, x_{i-1}(p_k), x_{i+1}(p_k), \ldots x_N(p_k), p_i), \; k = 1, \ldots, N$$

When we take the derivative of the above with respect to p_j when $j \neq i$ and follow the multivariable chain rule, we obtain:

$$\frac{dx_i}{dp_j} = \sum_{k \neq i} \frac{\partial x_i}{\partial x_k} \cdot \frac{dx_k}{dp_j} + \frac{\partial x_i}{\partial p_i} \cdot \frac{dp_i}{dp_j} \tag{1}$$

Since p_i and p_j are mutually independent perturbations placed on different species, $\frac{dp_i}{dp_j} = 0$. Now (1) can be rewritten as:

$$\frac{dx_i}{dp_j} = \sum_{k \neq i} \frac{\partial x_i}{\partial x_k} \cdot \frac{dx_k}{dp_j} \qquad \forall\, j \neq i \tag{2}$$

Invoking the chain rule once again, we arrive at:

$$\frac{dx_i}{dx_j} \cdot \frac{dx_j}{dp_j} = \sum_{k \neq i} \frac{\partial x_i}{\partial x_k} \cdot \frac{dx_k}{dx_j} \cdot \frac{dx_j}{dp_j} \qquad \forall\, j \neq i \tag{3}$$

Since $\frac{dxj}{dpj}$ is usually non-zero, we divide this term from both sides of equation (3), which becomes:

$$\frac{dx_i}{dx_j} = \sum_{k \neq i} \frac{\partial x_i}{\partial x_k} \cdot \frac{dx_k}{dx_j} \qquad \forall\, j \neq i \tag{4}$$

In (4), the partial differential term $\frac{\partial x_i}{\partial x_k}$ indicates the "localized" response of species x_i when an infinitesimal perturbation is introduced on species x_k around its steady state, while the activity of other species are "clamped" and kept as constants. Thus, we define the local response coefficients as:

$$r_{ik} = \frac{\partial x_i}{\partial x_k}, \text{ and } r_{ii} = -1 \tag{5}$$

Moreover, the total differential term $\frac{dx_k}{dx_j}$ indicates the "global" response of species x_k when a perturbation is introduced on the species x_j around its steady state. Thus we define the global response coefficients as:

$$R_{kj} = \frac{dx_k}{dx_j}, \text{ and } R_{jj} = 1 \tag{6}$$

Now we plug the definitions (5) and (6) into equation (4), it becomes:

$$R_{ij} = \sum_{k \neq i} r_{ik} \cdot R_{kj} \qquad \forall\, j \neq i \tag{7}$$

Equation (7) relates the global response coefficients with local response coefficients, through which local response coefficients can be solved when the global response coefficients are available.

Reverse engineering of the benchmark topologies using population data

For each set of perturbation responses, the global response coefficients were calculated based on the weighted mean fluorescence of gated populations. The pairwise sensitivity coefficients were then obtained via calculating the local response coefficient (LRC) (Fig. S8C and S8F).

For both perturbation magnitudes, we were able to validate the positive regulatory link from node X to nodes Y and Z, as well as a positive link from node Y to node Z. All other edges had negligible interaction coefficients (Fig. S9). To determine the significance of the recovered LRC we performed error propagation analysis using Monte Carlo simulations, in order to probe whether the 95% confidence interval includes zero (Fig. S10). In this case, we observe that several regulatory links are not statistically significant (Fig. S10, dashed line edges).

In contrast to the feedforward loop where node Z is under the regulation of both nodes X and Y, the node Z in a cascade architecture has a single activation source (node Y). Therefore, for the cascade network, we hypothesize that node Z will be more sensitive to node Y disruption. Indeed, after performing the node-wise perturbation of the cascade using the siRNAs, we were able to confirm this pattern. We observe that the perturbation of node Y in the cascade architecture (Fig. S8E) has strong impact on mKate2 as compared to the same experiment on the feedforward loop (Fig. S8B).

The inference results point to higher importance of the connection between nodes Y and Z, as compared to the feedforward loop. Interestingly, MRA also recovered a direct inhibitory connection between nodes X and Z for both perturbation magnitudes which may be attributed to mild retroactivity effects. Finally, in this case, the error propagation analysis rendered most of the predicted regulatory connections as insignificant.

Monte Carlo simulation

For MRA performed using statistics obtained at a population level, the errors of calculated local response coefficient r_{ij} were projected using Monte Carlo simulations. In each cycle of the simulation, the global response coefficient was calculated using a hypothetical value of the steady-state concentrations x_i. This hypothetical value was obtained by randomly drawing from a normally distributed population with experimentally obtained mean and standard deviation values of the corresponding triplicate steady state measurement. The local response coefficient was calculated from the simulated global response matrix as described above. The simulation was repeated for 2,000 times, and the resulting distribution of the local response coefficients are plotted as a probability density function. Lastly, 95% confidence interval of each simulated local response coefficients, an interval bound by the 2.5^{th} and 97.5^{th} percentile values of the distribution, is calculated to determine whether a simulated local response coefficient is significantly different than zero.

Phenomenological Model

Consider the following set of equations to describe our synthetic three-node feedforward system:

$$\dot{x} = 1 - p \cdot x \tag{8}$$

$$\dot{y} = f(x) - q \cdot y \tag{9}$$

$$\dot{z} = g(x) + h(y) - z \tag{10}$$

The functions f, g and h represent rates of production for nodes x, y and z, respectively. In our specific experimental design, function f signifies activity by Tet-on transactivator and functions g and h indicate activities of RheoSystem dimer. We assume f(0) = g(0) = h(0), indicating that no production occurs if there is no substrate. Furthermore, p and q denote parameters for modulation of nodes x and y, which were experimentally achieved by the transient transfection of siRNAs. In other words, for an unperturbed system, p = 1 and q = 1, respectively. Therefore, a "weak" perturbation corresponds to $p \approx 1$ and $q \approx 1$, while a "strong" perturbation means that p and q will approach ∞. Notice that cascade motif is described by the special variation of this model where g = 0.

First, we compute the local response coefficient under weak perturbation. At the nominal steady state, $\bar{x} = 1$, $\bar{y} = f(1)$, and $\bar{z} = g(1) + h(f(1))$. Based on the definitions of local response coefficients provided in (5), we can obtain the following expression for local response coefficients:

$$r_{yx} = f'(1) \tag{11}$$

$$r_{zy} = g'(1) \tag{12}$$

$$r_{zx} = h'(f(1)) \tag{13}$$

To calculate the local response coefficient following a strong perturbation, we calculate the differences between the steady state coordinates when p = 1, q = 1 minus steady state coordinates when p = p' (p' approaches ∞), q = 1. Same rationale is applied to the perturbation of parameter q. The general expressions for the differences in each node are follows:

$$D_{xp} = \frac{1}{p'} - 1; D_{xq} = 0 \tag{14}$$

$$D_{yp} = f\left(\frac{1}{p'}\right) - f(1); D_{yq} = \frac{f(1)}{q'} - f(1) \tag{15}$$

$$D_{zp} = g(\frac{1}{p'}) + h(f\left(\frac{1}{p'}\right)) - g(1) - h\big(f(1)\big); D_{zq} = h(\frac{f(1)}{q'}) - h(f(1)) \tag{16}$$

As p' and q' both approach ∞, (14), (15) and (16) can be simplified:

$$D_{xp} = -1; D_{xq} = 0$$

$$D_{yp} = -f(1);\ D_{yq} = -f(1)$$

$$D_{zp} = -g(1) - h\big(f(1)\big);\ D_{zq} = -h(f(1))$$

Thus, the corresponding global response coefficients are:

$$R_{xx} = \frac{D_{xp}}{D_{xp}};\ R_{yx} = \frac{D_{yp}}{D_{xp}}; R_{zx} = \frac{D_{zp}}{D_{xp}};$$

$$R_{xy} = \frac{D_{xq}}{D_{yq}};\ R_{yy} = \frac{D_{yq}}{D_{yq}}; R_{zy} = \frac{D_{zq}}{D_{yq}};$$

From these we can obtain local response coefficients of the connections that make up our network:

$$r_{yx} = \frac{R_{yx}}{R_{xx}} = f(1) \tag{17}$$

$$r_{zx} = \frac{R_{zx} - r_{zy} \cdot R_{yx}}{R_{xx}} = g(1) \tag{18}$$

$$r_{zy} = \frac{R_{zy}}{R_{yy}} = \frac{h(f(1))}{f(1)} \tag{19}$$

Assume functions f, g and h can be described by Hill function:

$$\alpha(x) = \frac{x^n}{(K + x^n)}$$

For a faithful approximation of the system, we choose biologically relevant parameters. For example, for function f we choose $n_f = 2$, since the rtTA transactivator forms dimers. For functions g and h, which describe the activity of RheoDimer transactivator, we choose $n_g = n_h = 3$. Since rtTA is constitutively transcribed and upstream of RheoDimer proteins, we assume that $K_f > K_g = K_h$. We chose $K_f = 0.1$ and $K_g = K_h = 3$. With these parameters, we can estimate the local response coefficients. The results, alongside the values obtained from our experiments, are summarized in Table 1.

We emphasize that while the qualitative behavior in is the same, no quantitative agreement is sought nor is even possible, since we are using a phenomenological model. To highlight the qualitative changes we use color indicators in Tables S1: green indicates increase in calculated local response coefficient after decrease in perturbation magnitude, and red indicates decrease in calculated local response following the same change.

	Estimated			Experimental		
	Low Perturbation	High Perturbation	Difference (r_{low} - r_{high})	Low Perturbation	High Perturbation	Difference (r_{low} - r_{high})
r_{yx}	0.17	0.91	-0.74	0.39	1.77	-1.38 ± 0.17
r_{zx}	0.56	0.25	0.31	1.86	0.62	1.24 ± 0.52
r_{zy}	0.53	0.22	0.31	0.70	0.41	0.29 ± 0.24

Table S1. Local response coefficient following varying perturbation magnitudes for feedforward loop

We repeated the same procedure for estimation of local response coefficients in cascade motif. As mentioned previously, the cascade model is a special variation of the feedforward loop where (8) and (9) are the same, but g = 0 in (10). Thus, the only change in our local response coefficient after weak perturbation is in r_{zx}, as described in (11), which is now zero. Similarly, the differences in our local response coefficient following a strong perturbation are in (18), which also becomes zero.

	Estimated			Experimental		
	Low Perturbation	High Perturbation	Difference (r_{low} - r_{high})	Low Perturbation	High Perturbation	Difference (r_{low} - r_{high})
r_{yx}	0.17	0.91	-0.74	2.01	2.15	-0.14 ± 0.38
r_{zx}	0	0	0	-0.48	-0.22	0.26 ± 0.41
r_{zy}	0.53	0.22	0.31	0.85	0.52	0.33 ± 0.14

Table S2. Local response coefficient following varying perturbation magnitudes for cascade

Supplementary Methods

Mammalian Cell Culture and Transfections: HEK293 cell line was maintained at 37C, 100% humidity and 5% CO_2. The complete growth medium for the cells consists of Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% Fetal Bovine serum (Atlanta Biologicals), 0.1mM nonessential amino acids (Invitrogen) and 0.045units/mL of penicillin and streptomycin (Invitrogen). The cells were grown in a T-75 flask (Corning) and was passed when it reached 80% confluency. To pass the cells, the culture was washed with PBS (Life Technologies) and then trypsinized with 0.25% Trypsin-EDTA (Invitrogen). The cells were then reseeded at 40% density in a new flask. Circuit plasmid transfection was performed with jetPRIME (Polyplus) in 12-well plates at a plating density of 200,000 cells. Transfection was performed 24 hours after seeding, and each well received 10ng of plasmid containing node X and 25ng of plasmid containing nodes Y and Z, with 500ng of co-transfection junk DNA and varying amounts of siRNA. Inducer ligands doxycycline and ponasterone A, were added immediately following transfection. Perturbations to the synthetic network were applied by co-transfecting the network plasmid with three different siRNAs. For unperturbed control population, cells were co-transfected with a scrambled siRNA. Fluorescent reporter activity after various perturbation scenarios were acquired by flow cytometry.

Fluorescence Microscopy: The cells were grown on 12-well plates (Greiner Bio-One) in complete media for transfection. Approximately 48 hours after transfection of network plasmid, fluorescence images of live cells were captured using the Olympus IX81 microscope. For ambient temperature control, the entire apparatus was housed in 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 TagCFP, ET500/20x (excitation) and ET535/30m (emission) for TagYFP, and ET560/40x (excitation) and ET630/75m (emission) for mKate2. Image acquisition and post-acquisition analysis was performed using Slidebook 5.0.

Flow Cytometry: All FACS experiments were performed 48 hours after transfection with BD LSRFortessa. To prepare live cells for flow cytometry, they were trypsinized with 0.25ml of 0.25% trypsin-EDTA for 3 minutes and pelleted by centrifugation at 4000 rpm for 2 minutes. The pellet was resuspended in 0.4 ml of PBS (Life Technologies). TagCFP protein was detected with a 445nm laser and a 470/20 band-pass filter, TagYFP with a 488-nm laser, 525 long-pass filter and 545/35 band-pass filter, mKate2 with a 561-nm laser 600 long-pass filter with 610/20 band-pass filter. At least 200,000 events were collected from each well. Data acquisition was performed using FACS Diva software. Subsequent analysis, including population gating, of the flow cytometry was performed with FlowJo (Treestar). The threshold fluorescence unit for selecting fluorescence-positive population was determined based on untransfected HEK293 cells (Figure S14). There was no compensation performed (Figure S15).

Quantitative reverse transcription PCR (qRT-PCR): 48 hours after co-transfection of circuit plasmid and node-specific siRNA, total RNA of the population was harvested using the RNeasy Mini kit (Qiagen) according to manufacturer's suggestion. 1 μg 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 ΔΔCt method, using GAPDH as normalization factor. Primers used to amplify GAPDH and transcripts from each node are listed in Table S3. Amplification started with an enzyme activation step at 95°C for 3 minutes, followed by 45 cycles consisting of 3 seconds of denaturation step at 95°C and 20 seconds of annealing/extension at 60°C.

Recombinant DNA plasmids

All restriction enzymes are from New England Biolabs unless otherwise stated. The sequences for the synthetic parts and primers are provided in tables S3 and S4.

FF3-TagCFP-pBI-CMV4-rtTA-IRES-RheoReceptor-FF3 The backbone plasmid containing the constitutive bidirectional promoter pBI-CMV4 (Clontech) was digested with EcoRI and XbaI and gel purified. TagCFP open reading frame was amplified from pTagCFP-N (Evrogen) using primers P1 and P2 that contain EcoRI and XbaI restriction sites. This PCR product was digested with EcoRI and XbaI and re-purified. The digested insert TagCFP was ligated into the digested backbone to create pBI-CMV4-TagCFP. This plasmid serves as the backbone for the rtTA-IRES-RheoReceptor insert. To generate the insert, rtTA-IRES was amplified from pTRE-Tight-BI-LacI-IRES-Green-LacO-rtTA-IRES-dsRED (Bleris, unpublished) using P3 and P4. RheoReceptor open reading frame was amplified using P5 and P6 from pNEBR-R1 (New England Biolabs) and reamplified using P5 and P7 to miR-FF3 target at 3' end. To create rtTA-IRES-RheoReceptor insert, rtTA-IRES was combined with RheoReceptor via overlap PCR with P3 and P7. The overlap PCR product was digested with NheI and EcoRV and ligated with pBI-CMV-TagCFP digested with the same enzymes.

FF3-RheoActivator-IRES-FF3-TagCFP-pBI-CMV4-rtTA-IRES-RheoReceptor-FF3 The aforementioned plasmid FF3-TagCFP-pBI-CMV4-rtTA-IRES-RheoReceptor-FF3 was amplified using P8 and P9 and digested with AgeI and NotI to serve as the backbone. The same plasmid was also to purify IRES sequence using P10 and P11. In parallel, miR-FF3 target was added to the 3' of RheoAcceptor open reading frame from pNEB-R1 (New England Biolabs) using P12 and P13. Next, we performed overlap PCR with P10 and P13 to create the amplicon RheoAcceptor-IRES insert. This amplicon was digested with AgeI and NotI and ligated with the backbone.

FF3-RheoActivator-IRES-FF3-TagCFP-pBI-CMV4-rtTA-FF3-IRES-RheoReceptor-FF3 The miR-FF3 target was added to the open rtTA reading frame with P14 and P15, using FF3-TagCFP-pBI-CMV4-rtTA-IRES-RheoReceptor-FF3 as the template. As the common enzyme site for ligation only existed at 5' end (MluI), this amplicon was extended with another fragment purified with P16 and P17 using same template to add a common enzyme site at 3' end (BlpI). The amplicons were combined via overlap PCR with P14 and P17, and was digested with MluI and BlpI. FF3-TagCFP-pBI-CMV4-rtTA-IRES-RheoReceptor-FF3 digested with MluI and BlpI was used as a backbone and ligated with the combined fragment.

FF3-RheoActivator-IRES-FF3-TagCFP-pBI-CMV4-rtTA-FF3-IRES-Δ1255_1257del_RheoReceptor-FF3
Mutation to the RheoReceptor open reading frame was introduced in the same procedure as the miR-FF3 target site addition to rtTA open reading frame (see above). The miR-FF3 target was added to the open rtTA reading frame with P14 and P15, using FF3-TagCFP-pBI-CMV4-rtTA-IRES-RheoReceptor-FF3 as the template. As the common enzyme site for ligation only existed at 5' end (MluI), this amplicon was extended with another fragment purified with P16 and P18 using same template to add a common enzyme site at 3' end (BlpI). In addition to

extending the amplicon for ligation, this purification step introduced base pair additions in RheoReceptor open reading frame designed to induce nonsense frameshift mutation. The amplicons were combined via overlap PCR with P14 and P18, and was digested with MluI and BlpI. FF3-TagCFP-pBI-CMV4-rtTA-IRES-RheoReceptor-FF3 digested with MluI and BlpI was used as a backbone and ligated with the combined fragment.

FF3-Δ697_699del_RheoActivator-IRES-FF3-TagCFP-pBI-CMV4-rtTA-FF3-IRES-Δ1255_1257del_RheoReceptor-FF3 Mutation to the RheoActivator open reading frame was initiated by purification of RheoActivator with an extraneous base pair introduced by the 3' end primer. The PCR reaction was performed with P19 and P20, using FF3-RheoActivator-IRES-FF3-TagCFP-pBI-CMV4-rtTA-FF3-IRES-RheoReceptor-FF3 as the template. The amplicon was digested with AgeI and Bsu36I and ligated with the backbone plasmid FF3-RheoActivator-IRES-FF3-TagCFP-pBI-CMV4-rtTA-FF3-IRES-Δ1255_1257del_RheoReceptor-FF3, which was also digested with AgeI and Bsu36I.

5xGal4-mKate2 pNEB-X1 Hygro (New England Biolabs) digested with HindIII-HF and NheI-HF and was used as a backbone. Fluorescent protein mKate2 open reading frame was purified from pmKate2 (Evrogen) using P21 and P22. The amplicon was digested with HindIII and NheI and ligated with the backbone.

FF4-TagYFP-pTRE-TightBI Fluorescent protein TagYFP open reading frame was amplified from pTagYFP (Evrogen) using P23 and P24 and digested with EcoRI-HF and XbaI. pTRE-TightBI (Clonetech) was digested with XbaI and EcoRI and ligated with the digested TagYFP amplicon.

FF4-TagYFP-PTREBI-RheoAcceptor-FF4 We amplified RheoAcceptor from pNEB-R1 (New England Biolabs) using oligos PT and PU and digested with KpnI-HF and NotI-HF. FF4-TagYFP-PTREBI was cut with the same enzymes and ligated with the digested amplicon to.

FF4-TagYFP-PTREBI-RheoAcceptor-IRES-RheoReceptor-FF4—5xGal4-mKate2 FF3-TagCFP-BICMV4-rtTA-IRES-RheoReceptor-FF3 as template IRES-RheoReceptor was amplified with oligos P25 and P26 and digested with NotI-HF and EcoRV-HF. We digested FF4-TagYFP-PTREBI-RheoAcceptor-FF4 with NotI-HF and EcoRV-HF and ligated with the digested IRES-RheoReceptor amplicon to generate FF4-TagYFP-PTREBI-RheoAcceptor-IRES-RheoReceptor-FF4. We amplified FF4-TagYFP-PTREBI-RheoAcceptor-IRES-RheoReceptor-FF4 using oligos P27 and P28 (including Ampicillin resistance and origin of replication), digested with AgeI-HF and CIP treated. Next, we amplified 5xGal4-mKate2 with P29 and P30 and digested with AgeI-HF. Both digested PCR amplicons were ligated to yield FF4-TagYFP-PTREBI-RheoAcceptor-IRES-RheoReceptor-FF4—5xGal4-mKate2.

Table S3. Primers

P1	CAGTACACGCGTGCTAGCGCCACCATGTCTAG	TagCFP Forward
P2	CCAGTAGAATTCGCCACCATGAGCGGGGGCGAGGAGCTGTTCG	TagCFP Reverse
P3	CCAGTAGCTAGCATGTCTAGACTGGACAAGAGCAAAG	rtTA Forward
P4	ATAGAAGACAGTAGCTTCATGGTTGTGGCCATATTATCAT	IRES Rheo Receptor Overlap
P5	ATGATAATATGGCCACAACCATGAAGCTACTGTCTTCTAT	IRES Rheo Receptor Overlap (Reverse)
P6	TTTGTATTCAGCCCATATCGTTCTAGAGATTCGTGGGGGACTCGA	Rheo Receptor Reverse with FF3
P7	CCAGTAGATATCTTTGTATTCAGCCCATATCGTT	FF3 (Rheo Receptor Reverse)
P8	CCAGTAGCGGCCGCCTAGATTTGTATTCAGCCCATATCGTTTTAGCGG	TagCFP Reverse
P9	CCAGTAACCGGTAGGATCATAATCAGCCATACCACATTT	SV40 Poly A Reverse
P10	CCAGTAGCGGCCGCTAGCGCTACCGGACTCAGATCTCGAG	IRES Reverse
P11	CGCTTCTTTTTAGGGCCCATGGTTGTGGCCATATTATCAT	RheoActivator IRES Overlap
P12	ATGATAATATGGCCACAACCATGGGCCCTAAAAAGAAGCG	RheoActivator IRES Overlap (Reverse)
P13	CCAGTAACCGGTTTTGTATTCAGCCCATATCGTTTTATGAATCAGAAGGTGATT	RheoActivator Reverse with FF3
P14	CAGTACACGCGTGCTAGCGCCACCATGTCTAG	rtTA Forward
P15	TTTGTATTCAGCCCATATCGTTTTACCCGGGGAGCATGTCAA	rtTA reverse + FF3
P16	CCGGGTAAAACGATATGGGCTGAATACAAACCTGCATTAGCGCTACCGGA	rtTA_IRES_forward
P17	CAGTACGCTCAGCTGGTTCAGGATATAGATGCG	RheoR_reverse
P18	CAGTACGCTCAGCATGGTTCAGGATATAGATGCG	RheoR_forward
P19	ACCGGTTTTGTATTCAGCCC	RheoA_FF3_forward
P20	CCTGAGGAGCAATCATTCTGTTTAATCCAGAG	RheoA_reverse
P21	CCAGTAGCTAGCATGGTGAGCGAGCTGATTAAGGAGAACA	mKate2 FWD
P22	CCAGTAAAGCTTTTATCTGTGCCCCAGTTTGCTAGGGAGG	mKate2 REV
P23	CCAGTATCTAGATTTAATTAAAGACTTCAAGCGGTTAGCGGTACAGCTCGTCCATGC	YFP Fwd with FF4 (Rev comp)
P24	CCAGTAGAATTCATGGTTAGCAAAGGCGAGGAGCTGTTCG	YFP Rev (Rev comp)
P25	CCAGTAGCGGCCGCTAGCGCTACCGGACTCAGATCTCGAG	IRES Reverse
P26	CCAGTAGATATCTTTAATTAAAGACTTCAAGCGGCTAGAGATTCGTGGGGGACTCGA	Rheo Receptor FF4

		Reverse
P27	CCAGTAACCGGTGCGAGCGGTATCAGCTCACTCAAAGG	Yp Origin of Replication FWD
P28	CCAGTAACCGGTCGCAGCCGAACGACCGAGCGCAGCGAG	SV40 Poly A Reverse (Node Yp)
P29	CCAGTAACCGGTCACGACGTTGTAAAACGACGGACATG	5xGal4 FWD
P30	CCAGTAACCGGTACGCGTTAAGATACATTGATGAGTTTGGAC	mKate2 Poly A Reverse
P31	AACAACGCCAAGTCATTCCG	qRT-PCR: Node X Forward
P32	TCAGCGACAACGTGTACATC	qRT-PCR: Node X Reverse
P33	ACGTTCCTGATGGAGATGCTTG	qRT-PCR: Node Y Forward
P34	TCGACTGCAGAATTCGAAGC	qRT-PCR: Node Y Reverse
P35	AACCACCACTTCAAGTGCAC	qRT-PCR: Node Z Forward
P36	TTTGCTGCCGTACATGAAGC	qRT-PCR: Node Z Reverse
P37	AATCCCATCACCATCTTCCA	qRT-PCR: GAPDH Forward
P38	TGGACTCCACGACGTACTCA	qRT-PCR: GAPDH Reverse

Table S4.

Specific parameters used for the SimBiology simulations[1]:

Parameter	Value	Units
k_GeneX_txn	0.01	1/second
k_mRNA_X_trsln	0.0001	1/second
k_mRNA_X_degr	0.00002	1/second
kON_PrX-GeneY	0.00000001	1/(molecule*second)
kOFF_PrX-GeneY	0.0001	1/second
kON_PrX-GeneZ	0.00000001	1/(molecule*second)
kOFF_PrX-GeneZ	0.0001	1/second
k_mRNA_Y_trsln	0.0001	1/second
k_mRNA_Y_degr	0.00002	1/second
k_mRNA_Z_degr	0.00002	1/second
k_PrX-GeneY_txn	0.01	1/second
k_PrX-GeneZ_txn	0.01	1/second
k_mRNA_Z_trsln	0.0001	1/second
k_Pr_degr	0.00002	1/second

References

1. 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.