ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

Using Dynamic Noise Propagation to Infer Causal Regulatory Relationships in Biochemical Networks

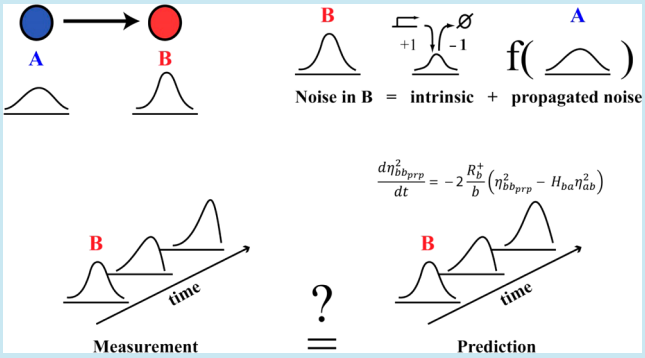
Joanna Lipinski-Kruszka,[†] Jacob Stewart-Ornstein,[‡,#] Michael W. Chevalier,[§,∥] and Hana El-Samad*[,§,∥]

[†]Integrative Program in Quantitative Biology, [‡]TETRAD Graduate Program, and [§]Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California 94143, United States

[∥]The California Institute for Quantitative Biosciences, San Francisco, California 94158, United States

Supporting Information

ABSTRACT: Cellular decision making is accomplished by complex networks, the structure of which has traditionally been inferred from mean gene expression data. In addition to mean data, quantitative measures of distributions across a population can be obtained using techniques such as flow cytometry that measure expression in single cells. The resulting distributions, which reflect a population's variability or noise, constitute a potentially rich source of information for network reconstruction. A significant portion of molecular noise in a biological process is propagated from the upstream regulators. This propagated component provides additional information about causal network connections. Here, we devise a procedure in which we exploit equations for dynamic noise propagation in a network under nonsteady state conditions to distinguish between alternate gene regulatory relationships. We test our approach *in silico* using data obtained from stochastic simulations as well as *in vivo* using experimental data collected from synthetic circuits constructed in yeast.



KEYWORDS: *biochemical network, computational method, dynamic noise propagation, molecular noise, network inference*

Obtaining a predictive understanding of information propagation and decision-making in cellular pathways is one of the paramount goals of systems biology.[1] A first step in generating this understanding is to be able to map the structures of the underlying gene-regulatory networks and the causal relationships between their molecular components.

Learning of networks structures is typically accomplished using bulk data describing the average response of a population. Although informative, these data often fail to establish causal relationships, resulting in nonunique solutions where multiple different topologies can represent the same data pool equally well (Figure 1A).[2]

It has recently been suggested that utilization of cell to cell variability in a population might improve the discriminatory power of network identification methods. Variability in protein expression, or "molecular noise", is a ubiquitous feature of biological systems that results from the probabilistic production, degradation, and collision of biological molecules (Figure 1B).[3−5] Such variability can be accurately quantified by measuring levels of specific proteins in single live cells using genetically encoded fluorescent reporters and high throughput assays such as flow cytometry.

We now have substantial understanding of the nature, sources,[6−8] propagation,[5,9] and information content of gene expression noise. Furthermore, noise at steady-state has been shown to provide information on regulatory pathway membership[10] and used to elucidate regulatory mechanisms.[11,12] These studies used static snapshots of noise in a pathway to gain such understanding, and revealed that despite the new information that could be gleaned from these measurements, they are often

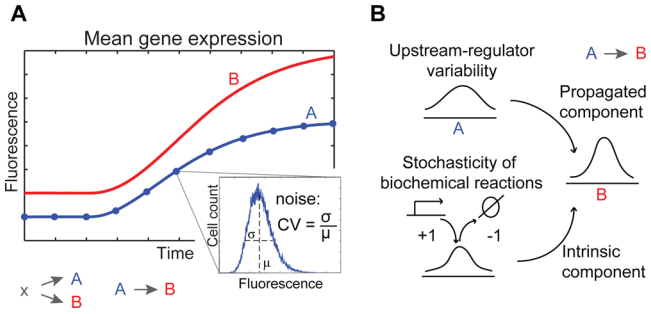


Figure 1. Mean might be insufficient to distinguish between alternate network topologies. (A) Example of mean expression of two genes, A and B, and two alternate network topologies that fit these data equally well. Population variability (inset) provides additional information. (B) Sources of noise can be divided into two components: intrinsic noise due to the stochastic nature of biochemical reactions, and propagated noise from upstream regulators.

Received: January 8, 2014
Published: June 26, 2014

insufficient to reveal relationships between components.[13,14] Namely, while these data could reduce the number of underlying possible topologies, they cannot pinpoint unique causal connectivities. Therefore, dynamic evolution of variability in a pathway, as measured by change of the population's distribution as a function of time, might be necessary to discriminate among alternate regulatory relationships.

This idea was recently explored in a study that used an approach called the Finite State Projection[15] to compute stochastic distributions that can result from different models of the hyperosmolarity pathway in yeast.[16] These distributions were then used to identify a model that was most predictive of the experimentally measured noise dynamics of mRNA expression. This method was capable of identifying a predictive model of the transcriptional dynamics in this system. However, this approach heavily relies on extensive parameter identification, a factor that might limit its scalability.

In this work, we present a new approach for the identification of regulatory connections in a network using dynamic noise data. Our approach is based on the premise that if a regulatory link between two nodes in a network is present and active, then variability in the upstream node should propagate downstream. This propagation results in a time-dependent and link-specific relationship between noise profiles of the two nodes.[5,13,17] To exploit this feature, we present a mathematical formalism describing noise propagation under nonsteady state conditions. By comparing model predictions and experimental measurements of noise, we can provide evidence for or against a putative regulatory interaction. Conveniently, our method requires estimation of only two kinetic parameters, both of which can uniquely be determined from single cell gene expression data. We first illustrate how this methodology can extract regulatory connectivity in a circuit using *in silico* data. Then, we demonstrate the usefulness of our approach using *in vivo* data collected from synthetic networks constructed in the budding yeast *S. cerevisiae*.

RESULTS AND DISCUSSION

Using the Chemical Master Equation to Derive Moment Equations. The formulation that we assume in our model consists of a homogeneous system in which each cell is treated as a well-mixed bag of N molecular species.[18,19] The state of the system is represented by a N-length integer vector $X(t)$ denoting the number of molecules of each species at time t. The M possible reactions that can occur among these species are represented by state transitions in a Markov chain. Transitions occur in discrete steps at random time intervals and depend only on the previous state of the system ("memoryless" process). The probability that a reaction r will happen in the next time interval, $(t,t+\tau)$ as $\tau \to 0$, is $R_r(X(t))\tau + o(\tau)$. Occurrence of reaction r changes state $X(t)$ according to the stoichiometric vector ϑ_r, which defines how the reaction changes number of each reactant species. The probability of the system being in state x at time t can be represented by the joint probability function $P(x,t)$. The chemical master equation (CME) gives us how this probability evolves over time:

$$\frac{dP(x,t)}{dt} = \sum_{r=1}^{M} [R_r(x-\vartheta_r)P(x-\vartheta_r,t) - R_r(x)P(x,t)] \quad (1)$$

Based on previously published work by Engblom, we can express propensities $R_r(x)$ in the form of a first-order Taylor series expended around the expected value $x = m$

$$R_r(x) = R_r \sum_{k}^{N} \frac{\partial R_r(m)}{\partial x_k}(x_k - m_k) + \text{Higher order terms} \quad (2)$$

If we assume that the higher order terms in the Taylor series expansion are negligible, the time dependent mean equation for the ith species is given by[20]

$$\frac{dm_i}{dt} \cong \sum_{r=1}^{M} \vartheta_r^i R_r(m) \quad (3)$$

Finally, for any two species i and j, we can also obtain a first-order approximation of the derivative of their covariance C_{ij}:[20]

$$\frac{dc_{ij}}{dt} \cong \sum_{r=1}^{M}\left(\vartheta_r^i \sum_k \frac{\partial R_r(m)}{\partial x_k} C_{kj} + \vartheta_r^j \sum_l \frac{\partial R_r(m)}{\partial x_l} C_{il}\right) + \sum_{r=1}^{M} \vartheta_r^i \vartheta_r^j R_r(m) \quad (4)$$

These equations provide a predictive model that links the topology of a network to the dynamic evolution of its mean behavior and to the time-dependent evolution of the second moment of its distribution across a population of cells. The strategy we propose below is to check the solution generated by an equation based on the second moment and the mean for a given expected network connectivity against data to test whether this topology is likely. In this way, we augment the information from the mean with that from variability to discriminate between different possible connectivities in a network. Importantly, our approach relies on measurements during dynamic network operation, therefore exploiting regimes where noise propagation is most likely to occur.

Dynamic Noise Propagation Equations. As a proof of principle, we consider two simple transcriptional systems in which protein A, constitutively expressed at rate: $R_a^+ = \alpha_a$, either activates: $R_b^+ = \alpha_b a^n/(a^n + K)$, or inhibits: $R_b^+ = \alpha_b/(a^n + K)$, expression of gene B. Here, a and b represent the mean copy number of proteins A and B, respectively. The proteins A and B are degraded at first order, linear rates, $R_a^- = \gamma_a a$, and $R_b^- = \gamma_b b$.

We define noise of A or B as the squared coefficient of variation, $\eta_{aa}^2 = C_{aa}^2/a^2$ and $\eta_{bb}^2 = C_{bb}^2/b^2$, respectively, and derive dynamic equations for these quantities as described above (eqs 3 and 4; See Supporting Information for specific derivations):

$$\frac{d\eta_{aa}^2}{dt} = -2\eta_{aa}^2\frac{R_a^+}{a} + \frac{R_a^+}{a^2} + \frac{R_a^-}{a^2} \quad (5)$$

$$\frac{d\eta_{bb}^2}{dt} = -2\frac{R_b^+}{b}(\eta_{bb}^2 - H_{ba}^*\eta_{ab}^2) + \frac{R_b^+}{b^2} + \frac{R_b^-}{b^2} \quad (6)$$

Here, H_{ba}^* is the susceptibility of B to A as defined at steady state: $H_{ba}^* = (\partial \ln b)/(\partial \ln a) \approx \partial \ln(R_b^-/R_b^+)/(\partial \ln a)$.[7,21,22] For the activation system, the susceptibility is $H_{ba}^* = nK/(a^n + K)$, and for the inhibitory link: $H_{ba}^* = na^n/(a^n + K)$.

We also derive an equation for the shared noise, $\eta_{ab}^2 = C_{ab}^2/ab$, which is a measure of covariation of A and B. It is given by

dx.doi.org/10.1021/sb5000059 | ACS Synth. Biol. 2015, 4, 258−264

$$\frac{d\eta_{ab}^2}{dt} = -\eta_{ab}^2\left(\frac{R_a^+}{a} + \frac{R_b^+}{b}\right) + \eta_{aa}^2 H_{ba}^* \frac{R_b^+}{b} \quad (7)$$

The noise equations of A and B (eqs 5 and 6) can be decomposed into intrinsic and propagated components. Noise of A, η_{aa}^2, has an intrinsic component only originating from stochastic expression and degradation of the protein. Because the expression of B is regulated by A, its noise (η_{bb}^2) has both intrinsic and propagated components which sum up to the total noise: $\eta_{bb}^2 = \eta_{bb_{int}}^2 + \eta_{bb_{prp}}^2$.[6,7] The dynamic evolutions of $\eta_{bb_{int}}^2$ and $\eta_{bb_{prp}}^2$ can be extracted from eq 6. Evidently, the intrinsic noise of B does not depend on the shared noise η_{ab}^2.[6] Terms containing η_{ab}^2 reflect noise propagated from A to B. The resulting dynamic intrinsic and propagated noise equations for B are

$$\frac{d\eta_{bb_{int}}^2}{dt} = -2\frac{R_b^+}{b}\eta_{bb_{int}}^2 + \frac{R_b^+}{b^2} + \frac{R_b^-}{b^2} \quad (8)$$

$$\frac{d\eta_{bb_{prp}}^2}{dt} = -2\frac{R_b^+}{b}(\eta_{bb_{prp}}^2 - H_{ba}^*\eta_{ab}^2) \quad (9)$$

The derived dynamic noise equations converge at steady-state to known expressions (Paulsson[7]). Furthermore, we validate these equations in the dynamic regime using data obtained from stochastic simulations (SSA)[18] of the regulatory circuits using different parameter sets (Figure 2, and Supporting Information Figure S1 and Table 1).

Strategy for Using Dynamic Noise Equations to Predict Causal Relationships in a Circuit. Noise transmission depends on the regulatory relationship between two genes. Therefore, the propagated noise equation (eq 9) offers an opportunity to test for the existence of a causal connection between two components of a circuit. Specifically, if expression of A and B are measured simultaneously in single cells as a function of time, we can determine how their means (a and b), downstream propagated noise ($\eta_{bb_{prp}}^2$) and shared noise (η_{ab}^2) evolve over time. By using these measured values in the noise equation for either the activation or inhibitory model (eq 9), we can calculate for every time point the rate at which propagated noise should be changing ($d\eta_{bb_{prp}}^2/dt$) and subsequently compute the entire time-course trajectory of the propagated noise $\eta_{bb_{prp}}^2(t)$. If for a given tested model, this predicted trajectory coincides with the experimentally measured trajectory, it is an indication that this model is likely to represent the causal relationship present in the network. Evidently, we can repeat this procedure to test for activating or inhibitory interactions, as well as for all permutations of the circuit (e.g., A activates B or B activates A).

Estimation of Necessary Parameters. To implement the strategy outlined above, we must first estimate parameters that are necessary to compute H_{ba}^* and R_b^+ in eq 9 and that cannot be directly measured from dynamic expression data. There are three such parameters, all defining the production rate of the downstream protein: the Michaelis–Menten parameter K, the hill coefficient n and the maximum rate of synthesis α_b. All other values in eq 9, the means and noise of the proteins, are directly measurable.

At steady-state, mean expression of B is linearly related to the steady-state susceptibility H_{ba}^*: $\omega b = n - H_{ba}^*$ where ω is a constant (for activation $\omega = n\gamma_b/\alpha_b$ and for inhibition $\omega = nK\gamma_b/\alpha_b$; see Supporting Information for detailed derivations).

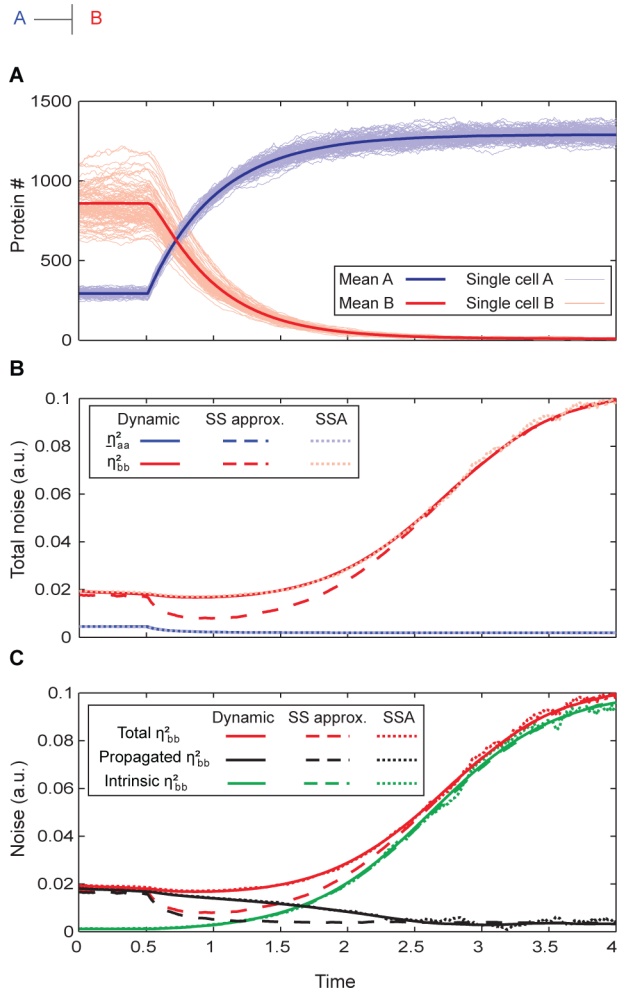


Figure 2. Noise computed using dynamic equations matches SSA and converges near steady state to established stationary equations. (A) Example of protein expression of a two-node system in which A (α_a = 567 and 2491, γ_a = 1.9) inhibits B (α_b = 4.796E10, γ_b = 2.25, K_b = 870, n_b = 3); population mean (dark solid lines), trajectories of individual cells (n = 1000) obtained from SSA (thin, light lines). (B) Total noise in A and B: measured (dotted lines) and noise computed using dynamic equations (solid lines) and steady-state approximation (dashed lines). Solutions converge as the system approaches steady state. (C) Noise in B decomposed into intrinsic and propagated components.

Since both the mean and susceptibility can be accurately obtained from distributions, this relationship allows us to uniquely identify n from two steady-state measurements: one at the steady-state before circuit induction and another at the new steady-state after induction. Furthermore, inspection of eq 9 reveals that at steady-state, propagated noise is given by $\eta_{bb_{prp}}^2 = H_{ba}^*\eta_{ab}^2$, where for activation $H_{ba}^* = nK/(a^n + K)$ and for inhibition $H_{ba}^* = -na^n/(a^n + K)$. With n calculated as above, and $\eta_{bb_{prp}}^2$, η_{ab}^2, and a experimentally measured, we can also uniquely determine K using $K = a^n\eta_{bb_{prp}}^2/(n\eta_{ab}^2 - \eta_{bb_{prp}}^2)$ for activation or $K = -a^n(n\eta_{ab}^2 + \eta_{bb_{prp}}^2)/\eta_{bb_{prp}}^2$ for an inhibitory connection.

Using the obtained values of n and K, the trajectory of $\eta_{bb_{prp}}^2$ for a given assumed topology can be determined from the noise equations (eq 9). Calculated and measured values of $\eta_{bb_{prp}}^2$ can then be compared for all time points, for example by looking at the linear correlation between these two quantities. It is worth

dx.doi.org/10.1021/sb5000059 | ACS Synth. Biol. 2015, 4, 258–264

noting that due to the structure of the equations, measured and estimated $\eta^2_{bb_{prp}}$ will differ by a constant scaling factor corresponding to the synthesis rate of the upstream node (either α_a or $\alpha_{b)}$ whose exact value does not need to be determined since it has no bearing on the quality of the correlation between these two quantities (Figure 3B lower panel).

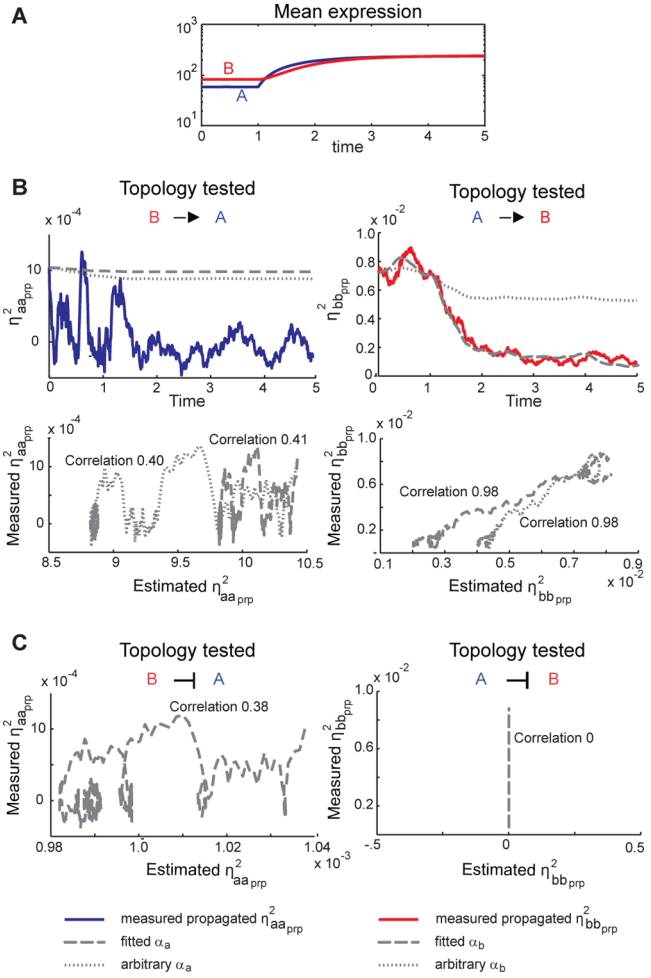


Figure 3. Reconstruction of an *in silico* network in which gene A activates gene B. (A) Mean expression of proteins A (α_a = 80 and 320, γ_a = 1.3) and B (α_b = 1207, γ_b = 1.8, K_b = 428, n_b = 1) as a function of time. (B) Test of an activation link between A and B. Comparison of measured noise of either A or B and dynamic noise predictions for different permutation of an activation model. Quality of prediction is quantified as a correlation between measured and estimated noise for the particular topology (lower panels). (C) Test of an inhibitory link between A and B. In both cases, correlation between estimated and measured noise is poor. The zero correlation indicates that the numerical integrator failed to solve the noise equation.

Test Using *In Silico* Data. We first tested our method *in silico* using data obtained from stochastic simulations of activation and inhibition motifs. We randomly sampled the parameters of these motifs (see Supporting Information Table 1) and generated time-dependent distributions. We used these distributions to extract propagated and shared noise values as a function of time, to which we then applied the procedure detailed above. For the correct regulatory relationship and directionality, we were mostly able (~75%) to accurately predict how propagated noise fluctuates over time in the downstream gene. Noise trajectories predicted for the incorrect regulatory relationship (for example, activation instead of inhibition) or reversed topology (B upstream of A instead of A upstream of B) failed to match the *in silico* data (Figure 3, and Supporting Information Figure S2). As expected, the networks for which we were unable to deduce the correct regulatory relationships corresponded to regimes where noise either was insignificant or did not propagate between the two nodes (Supporting Information Figures S3 and S4).

***In Vivo* Test Using Synthetic Circuits.** We next subjected our method to an *in vivo* test. For this purpose, we designed and built synthetic networks implementing transcriptional activation and inhibition motifs in the yeast *S. cerevisiae*. To build the activation circuit, we placed the transcription factor MSN2 tagged with YFP under the galactose responsive promoter, pGAL1 in a Δ*msn2/4* strain, allowing the fusion protein to provide the sole Msn2 activity in the cell. In the same strain, we integrated an RFP protein under the control of the Msn2-responsive HSP12 promoter. In the inhibitory circuit, the pGAL1 promoter was used to drive expression of the TetR protein tagged with RFP. To monitor the activity of TetR we integrated GFP under the control of a TetR-repressible Adh1 promoter ($Adh1^{tet}$). As a control, we implemented a third network in which reporter proteins, GFP and RFP driven by pGAL1 promoter were integrated at separate loci. This final stain has no direct interactions between the two reporters, but they are coregulated by the transcription factor Gal4 (Figure 4A) .

All three strains were grown in noninducing raffinose containing media and then induced by addition of galactose. We subsequently measured single cell abundance of the fluorescent proteins in ~5000 cells every 20 min for 12 h by flow cytometry. These data were processed and the mean and standard deviation of the per-cell fluorescence signal and the correlation between the RFP and GFP signals computed for each time point. Using these data along with the analytical propagated noise equation (eq 9), we then tested for regulatory relationships.

First, we tested whether the information contained in the mean alone could uniquely identify the underlying networks. To do so, we used ODE models of different regulatory mechanisms (causal, i.e., activation or inhibition, or noncausal, i.e., having no relationship between A and B) to mimic the behavior of the data. We found that the data could be fit equally well by all models (Supporting Information Figure S5), indicating that mean information alone cannot discriminate between the possible alternate topologies.

We next moved to testing whether the measured distributions could be exploited to provide discrimination using our noise propagation methodology. Using eq 9, we indeed determined that the propagated noise trajectory predicted using the topology that correctly reflects the true relationship (Msn2 activates Hsp12) matches the experimental results (correlation of 0.98 between predicted and measured propagated noise along the trajectory of the system). At the same time, noise trajectories predicted by assuming the incorrect, reverse topology (Hsp12 activates Msn2) cannot recapitulate the data (correlation of −0.15). Furthermore, predictions made assuming the wrong regulatory mechanism (inhibition instead of activation) do not match experimental results regardless of circuit permutation (Figure 4C, D top).

dx.doi.org/10.1021/sb5000059 | *ACS Synth. Biol.* 2015, 4, 258−264

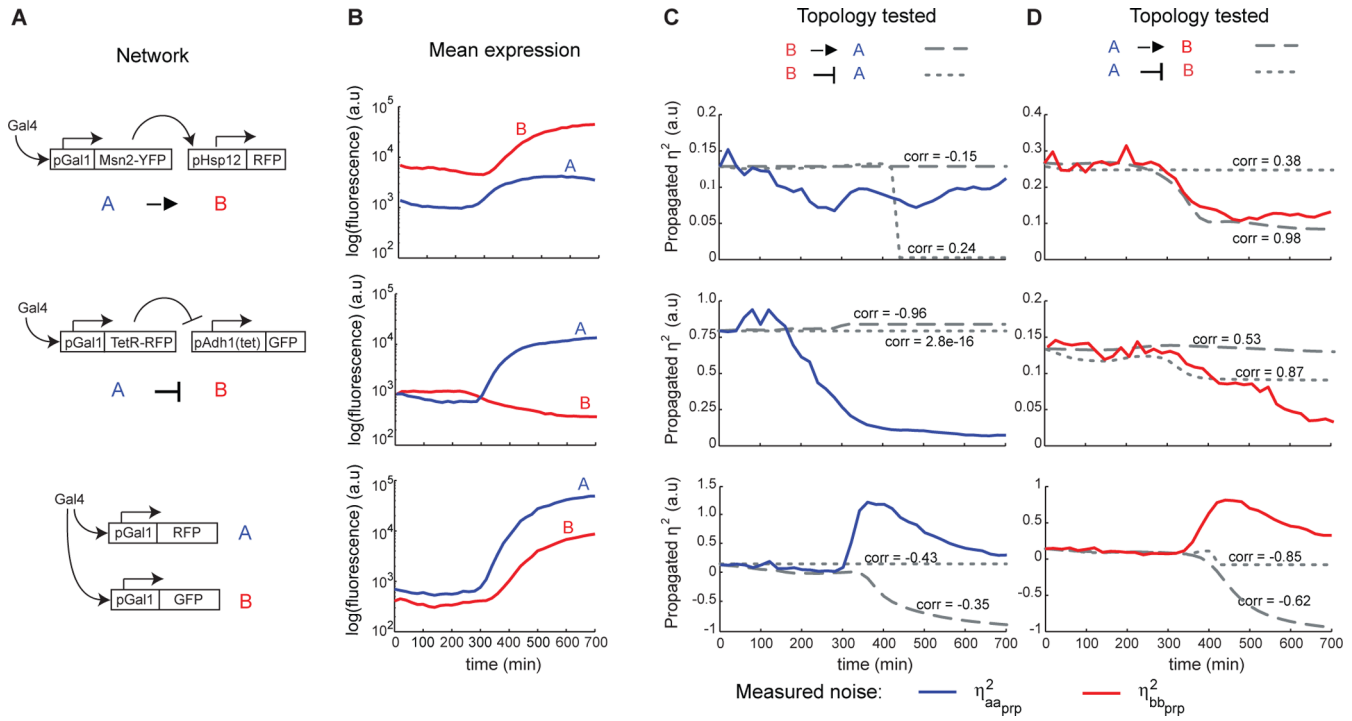


Figure 4. Reconstruction of three distinct *in vivo* synthetic networks using noise information. (A) Schematics of the three networks in which A activates B (top), A inhibits B (center), and A and B are coregulated by the same transcription factor (bottom). (B) Mean expression profiles of proteins in each of the three networks measured over a course of 12 h. (C) Noise trajectories predicted using dynamic equations for topologies in which B is assumed to either activate or inhibit A. (D) Noise trajectories predicted using dynamic equations for topologies in which A is assumed to either activate or inhibit B. In the circuit in which A and B are coregulated, we were not able to predict noise correctly for either circuit permutation, suggesting that A and B have no direct regulatory relationship (bottom).

Our methodology was equally efficient at pinpointing the right regulatory relationship for the inhibitory synthetic circuit. There again, we could discriminate between the correct topology (correlation of 0.87), TetR-RFP inhibits GFP, and other possible network permutations. Notably, the predicted trajectory for the reversed inhibitory relationship, GFP inhibits TetR-RFP, shows clear mismatch with the data (correlation 2.8×10^{-16}) (Figure 4C, D center). Similarly, predictions using the activation model fail to match experimental results regardless of network permutation.

For the control network in which GFP and RFP were coregulated, predicted propagated noise does not match the experimental data regardless of the assumed regulatory mechanism or network permutation, correctly indicating that these genes have no causal interactions (Figure 4C, D bottom). However, in such cases, we cannot in general rule out the existence of a regulatory relationship between the two genes since the relationship might not manifest itself in the data due to poor noise propagation or inactivity of the regulatory link under the tested conditions. Despite the ability of our method to provide hints about the lack of causal relationship between the circuit components, in this case, further tests are needed using different input schemes to uniquely determine the correct topology.

Conclusion. Technologies that provide expression measurements in single cells are ubiquitous, but the measured population variability data are seldom meaningfully exploited. This variability, its magnitude and frequency of fluctuations, can be information-rich, and when analyzed rigorously can be particularly useful for informing the structure of gene regulatory networks.[13,14,24]

Some early studies tested for regulatory relationships by attempting to directly score the linear correlation between the noise trajectories of a pair of genes. Such correlations can potentially pinpoint active connections particularly when taking time dynamics into consideration.[13] However, the relationship between noise in different components of a circuit is governed by potentially complex relationships as depicted by eqs 6 and, therefore, might be poorly quantified by linear pairwise noise correlation (Figure 5). This is because the fidelity with which noise propagates depends on factors such as the susceptibility of a gene to the upstream fluctuations, the amount of the upstream noise, and rates at which the protein is able to respond to upstream change. All of these factors change over

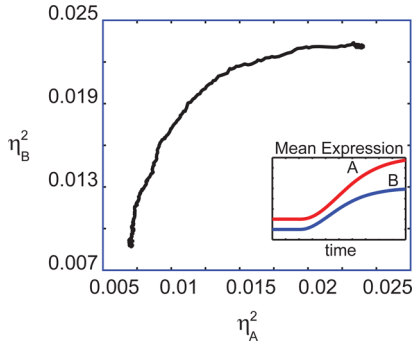


Figure 5. Linear correlation between the noise profiles of two nodes in a network is not a reliable predictor of their connectivity. Noise of A and B in a simple model: $(db/dt) = (\alpha_b a/a + K) - \gamma_b b$ shows a nonlinear relationship. Inset: mean expression of A (α_a = 65 and 192, γ_a = 1) and B (α_b = 2732, γ_b = 2, K_b = 696) as a function of time.

dx.doi.org/10.1021/sb5000059 | ACS Synth. Biol. 2015, 4, 258−264

time, most rapidly in the dynamic range where proteins concentrations change the most, conditions under which most experiments are usually conducted.

By contrast, our approach takes into account the noise dynamics, allowing us to integrate how fluctuations in gene expression are amplified, dissipate and propagate for a postulated network topology. Our computational investigations and experimental data both support the notion that propagated noise, if sampled at intervals that capture the dynamics of the system (Supporting Information Figure S9), can be sufficient to discriminate between alternate topologies when causal relationships exist between different network components. Our approach is similar to that used by Cox et al. in that both approaches rely on comparing measured noise to that predicted based on an assumed model of the underlying system. This comparison is then used to provide evidence for or against a tested system topology. Nonetheless, two main differences exist between the two approaches. While our study only uses the magnitude of the noise, the Cox *et al.* study scores both magnitude and frequency content of the noise. On the other hand, by exploiting the structure of the noise equations, breaking the identification process into pairwise comparisons, and leveraging different quantities that can be directly measured, we are able to proceed without estimation of a model's kinetic parameters. This advantage is not afforded by the Cox et al. approach. We envision that a method marrying the two that proceeds by matching both the measured magnitude and frequency of the noise to models of pairwise components combinations will capitalize on the strengths of both approaches.

In this work, we only presented results pertaining to genes that are regulated by a single input. However, genes are often regulated by more than one upstream component. In the case where multi-input regulation is competitive, only one regulatory link is active at a time, our method can be directly applied to test which of these regulatory links is active under changing conditions (condition dependent rewiring). For cases where multiple inputs simultaneously regulate expression, the necessary equations can be derived. This method is easily scalable when the two inputs are additive because, in this case, noise is also additive and we can sum up the contributions of both inputs in eq 9 to predict the dynamic trajectory of the propagated noise.

As a proof of concept, we demonstrated reconstruction of networks with two nodes. Isolating and considering only two genes at a time allows us to test for regulatory relationships without a priori knowledge or assumptions about the network's topology. Because our method relies on solving differential equations, it has low computational cost and we expect it to scale for larger, multinode networks. We envision that it can be extended by carrying out combinatorial, pairwise connectivity tests for many genes simultaneously. Furthermore, because our approach provides a rigorous, mathematically supported method to exploit noise information, it can be incorporated into existing mean-based network inference methods to facilitate reconstruction of complex, multigene regulatory structures.

In summary, as the development of increasingly sophisticated single cell measurement techniques[25,14] accelerates, there is increasing need for approaches that utilize population distribution information. Our approach provides a solid first step in that direction.

METHODS

Plasmids and Strain Construction. Galactose responsive constructs were constructed by amplification of the Gal1 promoter (defined as −1 to −625 nucleotides relative to the Gal1 ORF) from the yeast genome by PCR followed by restriction enzyme cloning into a single integration Trp1 or His3 marked vector upstream of Venus (YFP), yeGFP (GFP), or mKate2 (RFP). Msn2 was amplified from the genome and cloned in front of a Gal1 promoter with a Venus C-terminal tag in a Trp1 marked vector. TetR was amplified and cloned in front of a Gal1 promoter with a mCherry C-terminal tag in a His3 marked vector. The Hsp12 promoter consisting of 700bp directly upstream of the HSP12 start codon was amplified from the genome and inserted upstream of mKate2 (RFP) in a Ura3 marked vector. The Adh1(tet) promoter was cloned as described previously using amplification of the 700 bp upstream of the ATG and cloning in front of GFP in a His3 marked vector.[23]

W303A yeast were transformed serially with combinations of the above constructs using standard LioAc protocols. Transformants were selected on appropriate drop-out media and single colonies were picked for downstream use.

Growth and Fluorescence Measurements by Flow Cytometry. Yeast strains were grown to saturation overnight at 30 °C in 3 mL of synthetic complete media with 2% rafinose (SCraf) as a carbon source. Cells were diluted 1:100 into deep 96 well plates (Corning) and grown for 6−8 h at 30 °C on orbital shakers (Elim) to an OD of ∼0.1. All data were collected using a high-throuput automated flow cytometry system.[26] To induce expression of constructs, galactose (Sigma, 20% stock) was added to the media to a final concentration of 1%. Samples were taken from the primary culture every 20 min and an equal volume of fresh media added.

Cytometry measurements were made on a Becton Dickinson LSRII flow cytometer, along with an autosampler device (HTS) to collect data over a sampling time of 4−10 s, typically corresponding to 2000−10000 cells. GFP and YFP were excited at 488 nm, and fluorescence was collected through HQ530/30 bandpass filters (Chroma), mCherry and mKate2, were excited at 561 nm and fluorescence collected through 610/20 bandpass filter (Chroma) .

Flow Cytometry Data Analysis. Data analysis was done using custom MATLAB software. In order to minimize error due to uneven sample flow through the cytometer, we removed the first second and last 0.25 s of data. To control for cell aggregates, as well as cell size and shape, we excluded the bottom and top 5% of the forward (FSC) and side (SSC) scatter.[27]

ASSOCIATED CONTENT

Supporting Information

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

AUTHOR INFORMATION

Corresponding Author

*Email: hana.el-samad@ucsf.edu.

Present Address

#Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, 02115, United States.

Author Contributions

J.L.K. derived the mathematical models with input from M.W.C., developed the computational methods, and performed all data analysis. J.S.O. designed and constructed the synthetic circuits. J.L.K. and J.S.O. collected the flow cytometry data. H.E.S. oversaw the research. J.L.K. and H.E.S. wrote the manuscript. All authors have read and approved the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge Raj Bhatnagar and Leor Weinberger for critical reading of the manuscript. This work was funded by the National Institute of General Medical Sciences (NIGMS) system biology center (P50 GM081879) and the Paul G. Allen Foundation to H.E.S.

■ REFERENCES

(1) Chuang, H. Y., Hofree, M., and Ideker, T. (2010) A decade of systems biology. *Annu. Rev. Cell Dev. Biol. 26*, 721−744.
(2) Bansal, M., Belcastro, V., Ambesi-Impiombato, A., and di Bernardo, D. (2007) How to infer gene networks from expression profiles. *Mol. Sys Biol. 3*, 78.
(3) Thattai, M., and van Oudenaarden, A. (2001) Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A. 98* (15), 8614−9.
(4) Raser, J. M., and O'Shea, E. K. (2005) Noise in gene expression: Origins, consequences, and control. *Science 309* (5743), 2010−3.
(5) Pedraza, J. M., and van Oudenaarden, A. (2005) Noise propagation in gene networks. *Science 307*, 1965−1969.
(6) Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002) Stochastic gene expression in a single cell. *Science 297*, 1183−1186.
(7) Paulsson, J. (2004) Summing up the noise in gene networks. *Nature 427*, 415−418.
(8) Raj, A., and van Oudenaarden, A. (2008) Stochastic gene expression and its consequences. *Cell 135* (2), 216−226.
(9) Austin, D. W., Allen, M. S., McCollum, J. M., Dar, R. D., Wilgus, J. R., Sayler, G. S., Samatova, N. F., Cox, C. D., and Simpson, M. L. (2006) Gene network shaping of inherent noise spectra. *Nature 439* (7076), 608−611.
(10) Stewart-Ornstein, J., Weissman, J. S., and El-Samad, H. (2012) Cellular noise regulations underlie fluctuations in *Saccharomyces cerevisiae*. *Mol. Cell 45* (4), 483−493.
(11) Golding, I., Paulsson, J., Zawilski, S. M., and Cox, E. C. (2005) Real-time kinetics of gene activity in individual bacteria. *Cell 123* (6), 1025−1036.
(12) Zenklusen, D., Larson, D. R., and Singer, R. H. (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol. 15* (12), 1263−1271.
(13) Dunlop, M. J., Cox, R. S., III, Levine, J. H., Murray, R. M., and Elowitz, M. B. (2008) Regulatory activity revealed by dynamic correlations in gene expression noise. *Nat. Genet. 40*, 1493−1498.
(14) Munsky, B., Neuert, G., and van Oudenaarden, A. (2012) Using gene expression noise to understand gene regulation. *Science 336*, 183−187.
(15) Munsky, B., and Khammash, M. (2006) The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys. 124* (4), 044104.
(16) Neuert, G., Munsky, B., Tan, R. Z., Teytelman, L., Khammash, M., and van Oudenaarden, A. (2013) Systematic identification of signal-activated stochastic gene regulation. *Science 339*, 584−587.
(17) Simpson, M. L., Cox, C. D., and Sayler, G. S. (2003) Frequency domain analysis of noise in autoregulated gene circuits. *Proc. Natl. Acad. Sci. U.S.A. 100* (8), 4551−4556.
(18) Gillespie, D. T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem. 81* (25), 2340−2361.
(19) McQuarrie, D. A. (1967) Stochastic approaches to chemical kinetics. *J. Appl. Prob. 4*, 413−478.
(20) Engblom, S. (2006) Computing the moments of high dimensional solutions of the master equation. *Appl. Math. Comput. 180*, 498−515.
(21) Savageau, M. A. (1971) Parameter sensitivity as a criterion for evaluating and computing the performance of biochemical systems. *Nature 229*, 542−544.
(22) Paulsson, J. (2005) Models of stochastic gene expression. *Phys. Life Rev. 2*, 157−175.
(23) Murphy, K. F., Balázsi, G., and Collins, J. J. (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc. Natl. Acad. Sci. U.S.A. 104* (31), 12726−12731.
(24) Cox, C. D., McCollum, J. M., Allen, M. S., Dar, R. D., and Simpson, M. L. (2008) Using noise to probe and characterize gene circuits. *Proc. Nat. Acad. Sci. 105* (31), 10809−10814.
(25) Soon, W. W., Hariharan, M., and Snyder, M. P. (2013) High-throughput sequencing for biology and medicine. *Mol. Syst. Biol.*, 9−640.
(26) Zuleta, I. A., Aranda-Díaz, A., Li, H., and El-Samad, H. (2014) Dynamic characterization of growth and gene expression using high-throughput automated flow cytometry. *Nat. Methods 4 11* (4), 443−448.
(27) Newman, J. R., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L., and Weissman, J. S. (2006) Single-cell proteomic analysis of *S. cerevisiae* reveal the architecture of biological noise. *Nature 441*, 840−846.