

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Using Dynamic Molecular Noise to Infer Gene-Regulatory Networks

Permalink

<https://escholarship.org/uc/item/1nq1x2mp>

Author

Lipinski-Kruszka, Joanna

Publication Date

2014

Peer reviewed|Thesis/dissertation

Using Dynamic Molecular Noise to Infer Gene-Regulatory Networks

by

Joanna Lipinski-Kruszka

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biological and Medical Informatics

in the

GRADUATE DIVISION

Copyright 2014

by

Joanna Lipinski-Kruszka

To Konrad and Sabina:

there is no achievement or award that would make me as proud as you do.

Acknowledgements

None of this work would have been possible without the support and guidance of my advisor Hana El-Samad. I am grateful for her skepticism, diligence and for never letting me stop an investigation until I understand the problem down to its fundamental principles. Without her fervor for scientific rigor this project would have taken on a very different form.

I want to thank my committee, Joe DeRisi and Leor Wineberger, for their insights and support.

Special thanks to Patsy Babbitt for her devotion to students and for being a mentor who remembers that there is more to it than just giving scientific advice.

I appreciate the many science chats, guidance and help I received from my lab-mates, especially Jacob Stewart-Ornstein and Michael Chevalier.

I want to thank those who helped me get through the toughest part of grad school: becoming a parent. I will forever be indebted to Imke Listerman and Danica Fujimori for their support, advice, encouragement, and, most importantly, their friendship.

My time at UCSF would not have been much fun if it wasn't for my classmates. Thank you all, especially Genevieve Erwin, Dan Lu and Chris Fuller, for making my time here so memorable.

Finally and most of all, I want to thank the person without whose encouragement and support I would not even attempt to go to grad school: my husband Ken. Thank you for always being there for me and for always believing in me.

Abstract

Cellular decision making is accomplished by complex networks, the structure of which have traditionally been inferred from mean gene expression data. In addition to mean data, detailed quantitative measures of distributions across a population can be obtained using techniques, such as flow cytometry, which measure protein expression in single cells. These 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 is often referred to as extrinsic noise. When exploited systematically, extrinsic noise provides us with additional information about causal network connections. Here, we devise a procedure in which equations for dynamic noise propagation in a network under non-steady state conditions are exploited to distinguish between alternate regulatory relationships in a network. We validate our method using data obtained from stochastic simulations as well as from data derived from synthetic circuits implemented in yeast.

Table of contents

Chapter 1: Introduction	1
Chapter 2: Mathematical derivations	5
2.1 Using the Chemical Master Equation to derive moment equations	5
2.2. Dynamic noise equations	8
Chapter 3: Strategy for using dynamic noise equations to predict causal relationships in a circuit	13
3.1 Strategy overview.....	13
3.2 Estimation of necessary parameters	13
3.3. Measuring extrinsic noise: <i>in-silico</i> data.....	15
3.4 Estimation of intrinsic noise: <i>in-vivo</i> data.....	17
Chapter 4: Network prediction utilizing noise.....	19
4.1 Test using in-silico data	20
4.2 In vivo test using synthetic circuits	23
Chapter 5: Complex regulatory systems	28
5.1 Multi-node networks	29
5.2 Multi-input systems.....	29
Chapter 6: Discussion and closing remarks	32
Appendix A.....	35
Plasmids and Strain Construction	35
Growth and fluorescence measurements by flow cytometry	35
Flow cytometry data analysis	36

Network inference pseudocode	36
References	40

List of tables

Table 1. Ranges from which specific parameter values were randomly selected for all simulated data.....	6
---	---

List of figures

Figure 1: Mean alone might be insufficient to distinguish between alternate network topologies.....	2
Figure 2: Inclusion of higher moments in the dynamic noise equation does not substantially improve accuracy of predictions.....	7
Figure 3: Noise computed using dynamic equations matches SSA and converges near steady state to established stationary equations.	11
Figure 4: The dynamic noise equations accurately predict trajectories of noise.	12
Figure 5: <i>In silico</i> intrinsic noise for our models is well represented by the dynamic noise equation, as well as, the inverse of the mean ($1/b$) even at non-steady state.....	16
Figure 6: Reconstruction of an in silico network in which gene A activates gene B.	19
Figure 7: Evaluation of the method in silico using data obtained from SSA.	21
Figure 8: Dissection of cases that could not be reconstructed with noise information.....	22
Figure 9: The dynamic propagation of noise depends on parameter values.....	23
Figure 10: Using in vivo data collected for synthetic circuits, mean alone was insufficient for distinguishing between the possible regulatory relationships.....	25
Figure 11: Reconstruction of three distinct in-vivo synthetic networks using noise information.....	27
Figure 12: Reconstruction of a multi-node in silico network.	28
Figure 13: Reconstruction of a two-input in silico network.	31

Figure 14: A linear correlation between the noise profiles of two nodes in a network is not a reliable predictor of their connectivity.....34

Chapter 1: Introduction

Obtaining a predictive understanding of information propagation and decision-making in cellular pathways is one of the paramount goals of systems biology. In order to understand how these tasks are executed, one must carefully map and study structures of the underlying gene-regulatory networks.

Development of high-throughput techniques, such as microarray, propelled ‘reverse-engineering’ of gene regulatory pathways from expression data. Introduction of and improvement to these techniques made new types of data and information available, which, in turn, necessitated the development of new methods and tools for their interpretation and utilization. For example, some of the earliest microarray experiments provided a single time-point snapshot of genome-wide expression. These data were often interpreted using clustering techniques, which allowed for identification of co-expression networks but did not provide any direct evidence of gene-gene interactions (Eisen, 1998). As the microarray technique matured time-series and multi-condition data sets became available. Statistical methods, such as Bayesian networks, and information theory approaches were applied to these larger data sets to identify statistical relationships within genes in a network (Bansal, 2007; Xiang 2007; Imoto, 2003). These data and approaches, however, were unable to identify causal relationships and often resulted in non-unique solutions where many different topologies could represent the same data pool (figure 1a) (Bansal, 2007). Development of techniques, such as flow cytometry and single-molecule fluorescent in situ hybridization (smFISH), allows us to measure gene expression in singles cells and provide us with expression distributions across populations (figure 1b). Recent

improvements to these techniques made time-series experiments possible and allow us to measure how these distributions change over time and under changing conditions. However, methods of interpretation of these data are still in their advent.

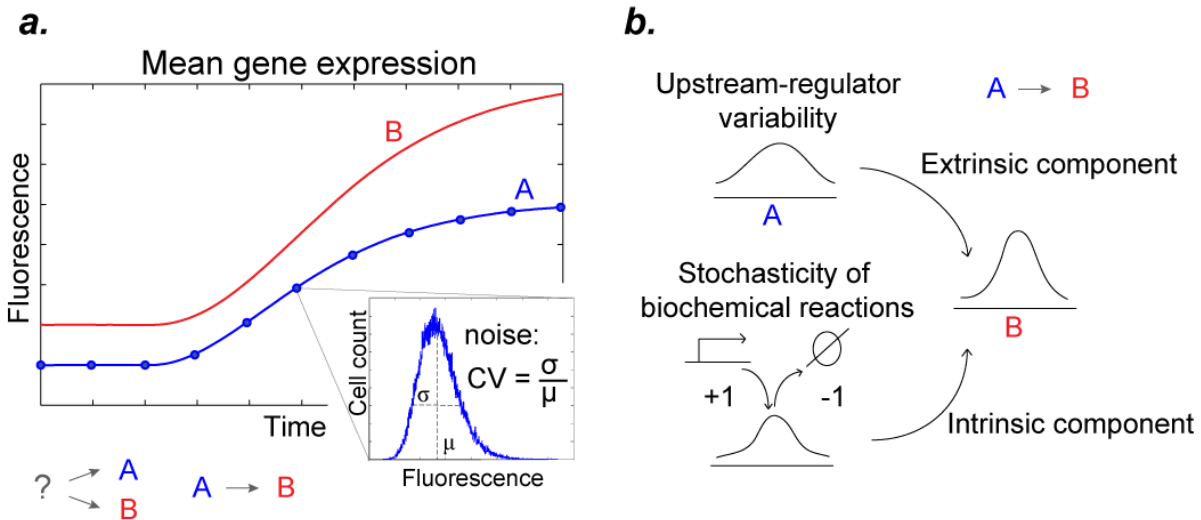


Figure 1: Mean alone might be insufficient to distinguish between alternate network topologies. (a) Example of mean expression of two genes, A and B, and two alternate network topologies that might have produced these data. Population variability (inset), often termed “noise” and quantified by the coefficient of variation (CV), is information rich and can provide additional information. (b) Sources of noise can broadly be divided into two components: intrinsic, which is due to the stochastic nature of biochemical reactions, and extrinsic, which is propagated from upstream regulators.

Variability, or “noise” in protein expression, is a ubiquitous feature of biological systems. This variability stems from the stochastic nature of biochemical processes, such as gene expression and signaling (Elowitz, 2002; Paulsson, 2004; Thattai, 2001). As a result, even genetically identical cells vary in their levels of mRNAs and proteins. While such noise can be disruptive, it also can play an important role in processes such as cell differentiation, state switching, and evolution (Losick, 2008; Eldar, 2010) to name a few. Recently, Cagatay et al. showed that changes in stochastic fluctuation levels alter a circuit's functional control in *Bacillus subtilis* and that network topologies generating similar mean dynamics can have different noise "signatures" (Eldar, 2009). Pedraza, et al. show that, at least for some proteins, network interaction are the key determinants of noise features. Because expression variability is highly dependent on the upstream fluctuation (Paulsson, 2004; Simpson, 2003) noise propagation should be indicative of connectivity between nodes of a network (Pedraza, 2005; Austin, 2006; Cox, 2008; Dunlop, 2008) (figure 1b).

We now have substantial understanding of the nature and sources (Elowitz, 2002; Paulsson, 2004; McAdams, 1997), propagation (Pedraza, 2005), and information content of gene expression noise at steady-state. Furthermore, noise at steady-state has been shown to provide information on regulatory pathway membership (Stewart-Ornstein, 2012; Suel, 2006) and elucidate regulatory mechanisms (Zenklusen, 2008). However, a static snapshot is often insufficient to reveal causal relationships between components of a pathway (Dunlop, 2008; Munsky, 2012), suggesting that dynamic evolution of the population's distribution might be necessary to discriminate among alternate regulatory relationships.

This idea was recently explored in a study by Neuert et al who used an approach called the Finite State Projection (Munsky, 2006) to compute stochastic distributions for models of

various complexity of the hyperosmolarity pathway in yeast (Neuert, 2013). They then used these models to identify the one that was most predictive of the experimentally measured noise dynamics of mRNA expression. This method was capable of identifying causalities and predictive models of transcriptional dynamics in this system. However, reliance of this approach on extensive parameter identification limits its scalability.

In this work, we present a new approach for the identification of regulatory connections in a network using dynamic noise. 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 (Dunlop, 2008; Pedraza, 2005; Rosenfeld, 2005). This propagation results in a time-dependent and link-specific relationship between noise profiles of the two nodes. To exploit this feature, we present a mathematical formalism describing noise propagation under non-steady 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 this approach using *in vivo* data collected from synthetic networks expressed in budding yeast.

Chapter 2: Mathematical derivations

2.1 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 (Gillespie, 1977; McQuarri, 1967). 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 $\tau \rightarrow 0$, is $w_r(X(t)) \tau + o(\tau)$. 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 specie. The probability of the system being in state x at time t can be represented by the joined probability function $P(x, t)$. The chemical master equation (CME) gives us how this probability evolves over time:

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

We can express propensities $w_r(x)$ in the form of a first-order Taylor series around the expected value $x = m$

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

If we assume that the higher order terms in the Taylor Series expansion are negligible, the time dependent mean equation for the i th species is given by (Engblom, 2006):

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

Finally, for any two species, i and j , we obtain a first-order approximation of the derivative of their covariance C_{ij} (Engblom, 2006):

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

Table 1. Ranges from which specific parameter values were randomly selected for all simulated data.

Parameter	Range
α_a	(0, 250000]
α_b	(0, 10^{13}]
γ_a, γ_b	(0, 15]
K	(0, 10^6]
n	(0, 5]

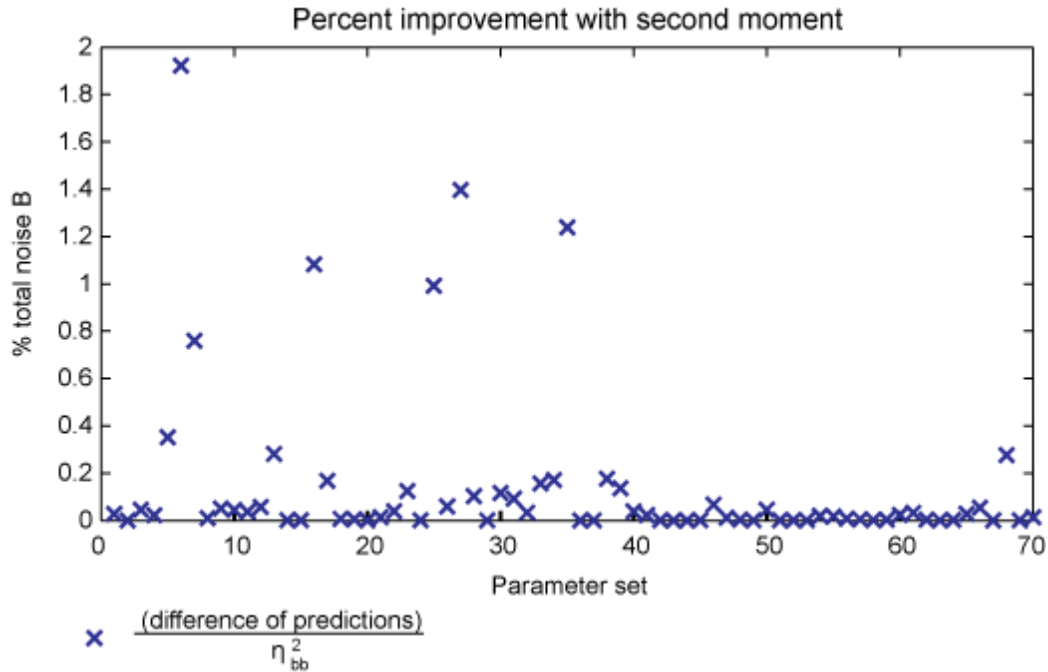


Figure 2: Inclusion of higher moments in the dynamic noise equation does not substantially improve accuracy of predictions. Error incurred by not including higher moments is quantified as percent of total noise and plotted for each topology and parameter set tested. The values reported are averaged over the entire trajectory of the downstream protein.

These equations provide a predictive model that links the topology of a network to the dynamic evolution of its mean behavior across a population, and the time-dependent evolution of the second moment of its output distribution. Our strategy below is to check the solution generated by the second moment equation, expected for a given 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.

2.2. Dynamic noise equations

We consider two simple transcriptional systems in which protein A, constitutively expressed at rate: $R_a^+ = \alpha_a$, either activates: $R_b^+ = \frac{\alpha_b a^n}{a^{n+K}}$, or inhibits: $R_b^+ = \frac{\alpha_b}{a^{n+K}}$, 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 $R_b^- = \gamma_b b$. In our model, each reaction produces or degrades a single molecule at a time.

Based on previously published work by Engblom (Engblom, 2006 and equation 4) we derived the following covariance equations:

$$\frac{dC_{aa}}{dt} = -2 \frac{\partial R_a^-}{\partial a} C_{aa} + R_a^+ + R_a^- \quad (5)$$

$$\frac{dC_{ab}}{dt} = -\frac{\partial R_a^-}{\partial a} C_{ab} + \frac{\partial R_b^+}{\partial a} C_{aa} - \frac{\partial R_b^-}{\partial b} C_{ab} \quad (6)$$

$$\frac{dC_{bb}}{dt} = 2 \frac{\partial R_b^+}{\partial a} C_{ab} - 2 \frac{\partial R_b^-}{\partial b} C_{bb} + R_b^+ + R_b^- \quad (7)$$

In our derivations of covariance, we assume that linearized model is a sufficient approximation of the system and, hence, that contributions of higher moments are minor and can be ignored. To verify, we investigated how the addition of the second term ($s_2 = -\eta_{aa}^2 H_{ba} \frac{R_b^+}{2b^2} \frac{K-nK+a^n+na^n}{a^{n+K}}$) impacted noise predictions. We compared how well equations with and without the second moment term predicted dynamic noise obtained from SSA simulations. We quantified contributions of higher moments as the difference of the two predictions normalized by the total noise of B. To obtain a single measure for each network we averaged over the entire time trajectory. We found that for the 70 randomly chosen sets of parameters (Table 1) inclusion of the second moment changed the prediction by less 2% and in most cases by less than 0.5% of

total noise of B (Figure 2). We, therefore, conclude that the linearized model is indeed a reasonable approximation of the system.

We next define noise of A or B as the squared coefficient of variation, $\eta_{aa}^2 = \frac{C_{aa}^2}{a^2}$ and $\eta_{bb}^2 = \frac{C_{bb}^2}{b^2}$, respectively and derive dynamic equations:

$$\frac{d\eta_{aa}^2}{dt} = \frac{d}{dt} \left(\frac{C_{aa}}{a^2} \right) = -2\eta_{aa}^2 \frac{R_a^+}{a} + \frac{R_a^+}{a^2} + \frac{R_a^-}{a^2} \quad (8)$$

$$\frac{d\eta_{bb}^2}{dt} = \frac{d}{dt} \left(\frac{C_{bb}}{b^2} \right) = -2 \frac{R_b^+}{b} (\eta_{bb}^2 - H_{ba}^* \eta_{ab}^2) + \frac{R_b^+}{b^2} + \frac{R_b^-}{b^2} \quad (9)$$

Here, H_{ba}^* is the susceptibility of B to A as defined at steady state: $H_{ba}^* = \frac{\partial \ln b}{\partial \ln a} \approx \frac{\partial \ln(R_b^-/R_b^+)}{\partial \ln a}$ (Savageau, 1971; Paulsson, 2004). For the activation system, the susceptibility is: $H_{ba}^* = \frac{nK}{a^{n+K}}$, and for the inhibitory link: $H_{ba}^* = -\frac{na^n}{a^{n+K}}$.

We also derive an equation for the shared noise, $\eta_{ab}^2 = \frac{C_{ab}^2}{ab}$, which is given by:

$$\frac{d\eta_{ab}^2}{dt} = \frac{d}{dt} \left(\frac{C_{ab}}{ab} \right) = -\eta_{ab}^2 \left(\frac{R_a^+}{a} + \frac{R_b^+}{b} \right) + \eta_{aa}^2 H_{ba}^* \frac{R_b^+}{b} \quad (10)$$

We decompose our noise equations of A and B into intrinsic and extrinsic 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 extrinsic components which sum up to the total noise: $\eta_{bb}^2 = \eta_{bb_{int}}^2 + \eta_{bb_{ext}}^2$ (Paulson, 2004; Elowitz, 2002). The dynamic evolutions of $\eta_{bb_{int}}^2$ and $\eta_{bb_{ext}}^2$ can be extracted from Equation (9). Evidently, the intrinsic noise of B does not depend on shared noise, η_{ab}^2 (Elowitz, 2002). Terms containing η_{ab}^2 reflect noise propagated from A to B and, therefore, are ascribed to the extrinsic noise. The resulting dynamic intrinsic and extrinsic noise equations are:

$$\frac{d\eta_{bb_{int}}^2}{dt} = -2\frac{R_b^+}{b}\eta_{bb_{int}}^2 + \frac{R_b^+}{b^2} + \frac{R_b^-}{b^2} \quad (11)$$

$$\frac{d\eta_{bb_{ext}}^2}{dt} = -2\frac{R_b^+}{b}(\eta_{bb_{ext}}^2 - H_{ba}^*\eta_{ab}^2) \quad (12)$$

The derived dynamic noise equations converge at steady-state to known expressions (Paulsson, 2004) (figure 3). Furthermore, we validate these equations in the dynamic regime using data obtained from stochastic simulations (SSA) (Gillespie, 1977) of the regulatory circuits using several different, randomly chose parameter sets (Table 1) for both activation and inhibition motifs (Figure 4).

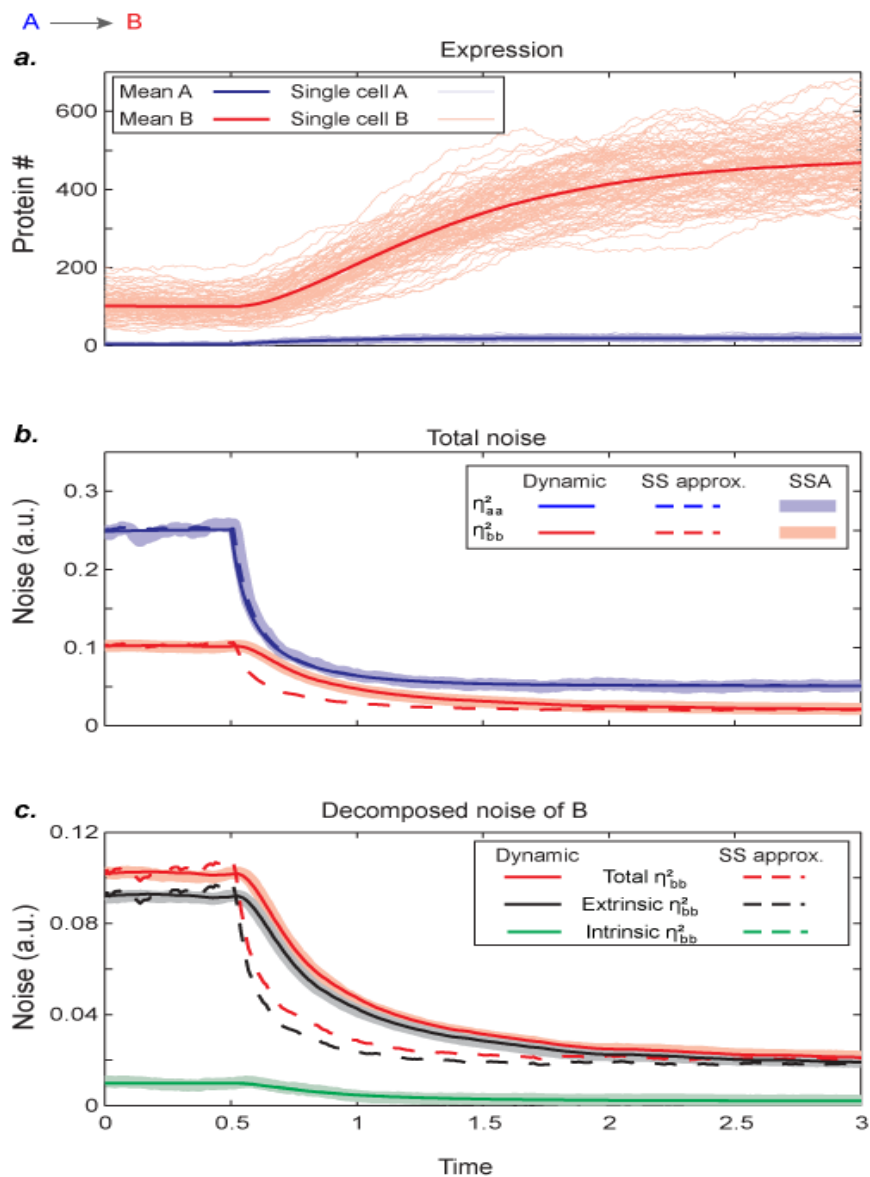


Figure 3: Noise computed using dynamic equations matches SSA and converges near steady state to established stationary equations. **(a)** Example of protein expression of simple two-node systems in which A activates B; population mean (dark lines), trajectories of individual cells ($n = 1000$) obtained from SSA (thin, light lines). **(b)** **(c)** Measured (light, wide lines) and noise predicted using dynamic equations (solid lines) and approximations using steady-state expressions (dashed lines). Solutions converge as the system approaches steady state; (b) total noise, (c) noise decomposed into intrinsic and extrinsic components.

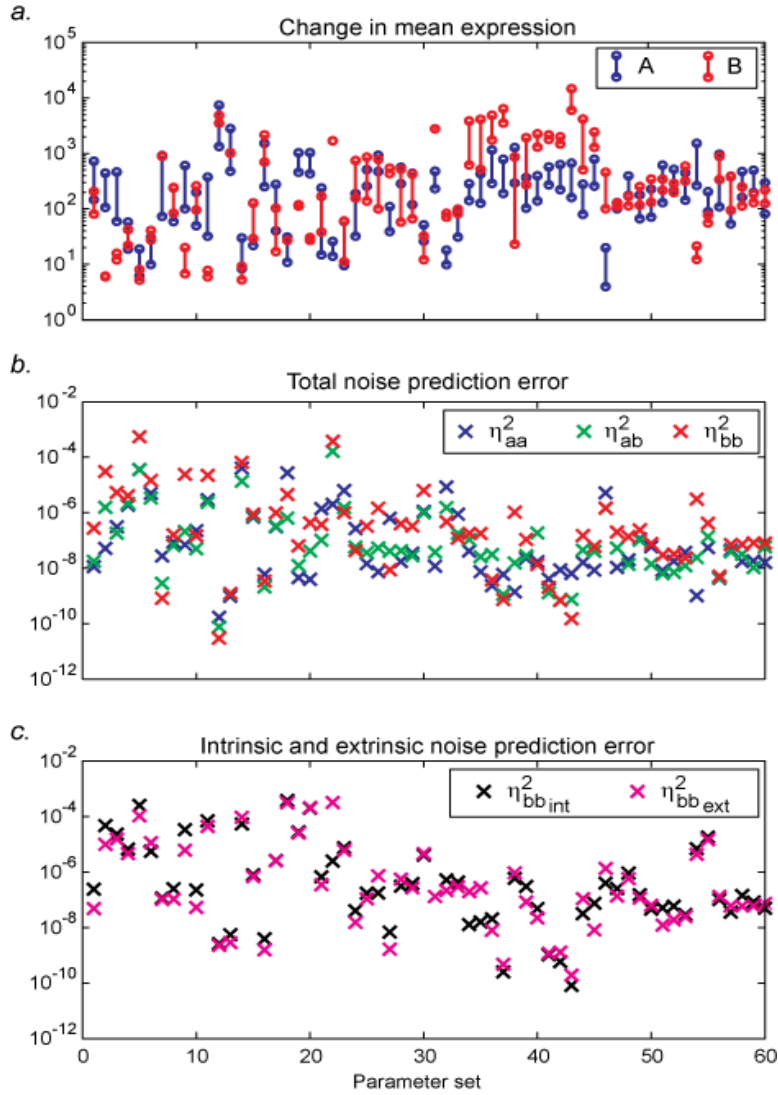


Figure 4: The dynamic noise equations accurately predict trajectories of noise. Prediction accuracy was computed using 60 SSA simulations of activation and inhibition networks using different, randomly chosen parameter sets. (a) Change in mean expression of A and B for each of the 60 tested networks. (b) (c) Prediction error was quantified as mean square error and averaged over all time-point (251 or 501 time-points). (b) total noise of A and B and of their shared noise; (c) error for the decomposed noise of B.

Chapter 3: Strategy for using dynamic noise equations to predict causal relationships in a circuit

3.1 Strategy overview

Because noise propagation depends on the regulatory relationship between two genes, the extrinsic noise equation (Equation 12) offers an opportunity to test for the existence of causal connections in a circuit.

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 extrinsic noise ($\eta_{bb_{ext}}^2$) and shared noise (η_{ab}^2) evolve over time. By using these measured values in the extrinsic noise equation for either the activation or inhibitory model (Equation 12), we can calculate for every time point the rate at which extrinsic noise should be changing, $\frac{d\eta_{bb_{ext}}^2}{dt}$, and subsequently compute the entire time-course trajectory of the extrinsic noise $\eta_{bb_{ext}}^2$. If for a given tested model, this predicted trajectory coincides with experimentally measured trajectory, then this is an indication that this model represents the causal relationship present in the network. Clearly, we can repeat this procedure to test for interactions for all permutations of the circuit.

3.2 Estimation of necessary parameters

To implement the strategy outlined above, we must first approximate the Michaelis-Menten parameter K and the hill coefficient n . For our model system, at steady-state the mean of

B can be computed as: $b = \frac{\alpha_b a^n}{\gamma_b(a^n+K)}$ and $b = \frac{\alpha_b}{\gamma_b(a^n+K)}$ for activation and inhibition,

respectively. We represent these expressions as a function of H_{ba}^* . Specifically for activation:

$$\begin{aligned}
b &= \frac{\alpha_b a^n}{\gamma_b(a^n + K)} \\
\frac{\gamma_b}{\alpha_b} b &= \frac{a^n}{a^n + K} \\
\frac{n\gamma_b}{\alpha_b} b &= \frac{na^n}{(a^n + K)} \\
\frac{n\gamma_b}{\alpha_b} b &= \frac{na^n + nK - nK}{(a^n + K)} \\
\frac{n\gamma_b}{\alpha_b} b &= \frac{n(a^n + K)}{(a^n + K)} - \frac{nK}{(a^n + K)} = n - \frac{nK}{(a^n + K)}
\end{aligned} \tag{13}$$

We represent all constant on the left-hand side as $\omega = \frac{n\gamma_b}{\alpha_b}$, and obtain the final linear relationship between susceptibility and the mean:

$$\omega b = n - H_{ba}^*$$

Inhibition:

$$\begin{aligned}
b &= \frac{\alpha_b}{\gamma_b(a^n + K)} \\
\frac{n\gamma_b}{\alpha_b} b &= \frac{n}{a^n + K} \\
\frac{n\gamma_b K}{\alpha_b} b &= \frac{nK}{a^n + K} \\
\frac{n\gamma_b K}{\alpha_b} b &= \frac{nK + na^n - na^n}{(a^n + K)} \\
\frac{n\gamma_b K}{\alpha_b} b &= \frac{n(a^n + K)}{(a^n + K)} - \frac{na^n}{(a^n + K)} = n - \frac{na^n}{(a^n + K)}
\end{aligned} \tag{14}$$

Again, we represent all constants on the left hand side as $\omega = \frac{nK\gamma_b}{\alpha_b}$:

$$\omega b = n + H_{ba}^*$$

These equations relating the mean and susceptibility have two unknown constants, n and ω , which can uniquely be identified from distribution information collected at two different steady-states.

We use this linear relationship between the mean of B and susceptibility: $\omega b = n - H_{ba}^*$ to find the necessary parameters. Since both, the mean and susceptibility, can be accurately obtained from distributions, we can uniquely identify n from measurements taken at two, or more, discrete steady-states. Inspection of Equation 12 reveals that at steady-state extrinsic noise is given by: $\eta_{bb_{ext}}^2 = H_{ba}^* \eta_{ab}^2$ (activation $H_{ba}^* = \frac{nK}{a^{n+K}}$; inhibition: $H_{ba}^* = \frac{-na^n}{a^{n+K}}$). Since n , $\eta_{bb_{ext}}^2$, η_{ab}^2 , and a are experimentally measured, we can also uniquely determine K ($K = \frac{a^n \eta_{bb_{ext}}^2}{n \eta_{ab}^2 - \eta_{bb_{ext}}^2}$ or $K = -\frac{a^n (n \eta_{ab}^2 + \eta_{bb_{ext}}^2)}{\eta_{bb_{ext}}^2}$).

Using the obtained values of n and K , trajectory of $\eta_{bb_{ext}}^2$ for a given assumed topology can be determined from the noise equations (Equation 12). Calculated and measured values of $\eta_{bb_{ext}}^2$ can then be compared, for example by looking at the linear correlation between these two quantities. It is worth noting that due to the structure of the equations, measured and estimated $\eta_{bb_{ext}}^2$ will differ by a constant scaling factor corresponding to the synthesis rate of the upstream node (either α_a or α_b) whose exact value has no bearing on the quality of the correlation between these two quantities (Figure 6b).

3.3. Measuring extrinsic noise: *in-silico* data

Because intrinsic and extrinsic noise sum up to the total noise: $\eta_{bb}^2 = \eta_{bb_{int}}^2 + \eta_{bb_{ext}}^2$ (Paulson, 2004; Elowitz, 2002), we can estimate the extrinsic component, $\eta_{bb_{ext}}^2$, from measurements of

the total and intrinsic noise. We measure the total expression noise of a protein, B , as the squared coefficient of variation, $\eta_{bb}^2 = \frac{C_{bb}^2}{b^2}$, where b denotes the mean and C_{bb}^2 variance of the population. For our model system in which intrinsic noise stems just from the random birth and death events of individual proteins, we defined intrinsic noise as: $\eta_{bb_{int}}^2 = \frac{1}{b}$ (Paulson, 2004 & 2005). Using this definition $\left(\frac{1}{b}\right)$ we derive its time derivative and show that it results in the same expression as the dynamic equation for intrinsic noise:

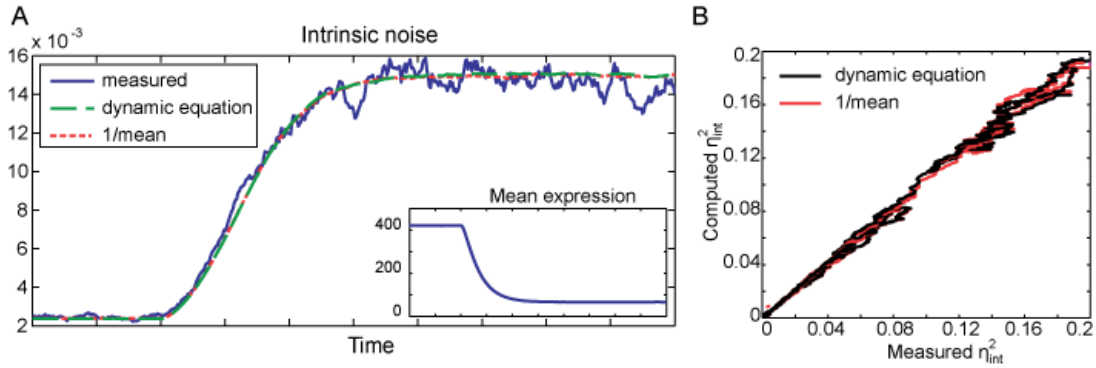


Figure 5: *In silico* intrinsic noise for our models is well represented by the dynamic noise equation, as well as, the inverse of the mean ($1/b$) even at non-steady state. **(A)** Example of an intrinsic noise trajectory of protein B calculated by an *in silico* noise decomposition of SSA results (blue), the dynamic noise equation (green) or the relationship $1/b(t)$ (red) for a circuit in which $A \rightarrow B$. Inset: mean behavior B . **(B)** Comparison of intrinsic noise to its measured value computed using all three methods for 40 different parameter sets.

$$\begin{aligned}
\frac{d}{dt}\left(\frac{1}{b}\right) &= -\frac{db/dt}{b^2} = -\frac{R_b^+ - R_b^-}{b^2} = \\
&= -\frac{R_b^+ - R_b^-}{b^2} - \frac{R_b^+}{b^2} + \frac{R_b^+}{b^2} = -2\frac{R_b^+}{b^2} + \frac{R_b^+}{b^2} + \frac{R_b^-}{b^2} = \\
&= -2\frac{R_b^+}{b}\eta_{bb_{int}}^2 + \frac{R_b^+}{b^2} + \frac{R_b^-}{b^2} = \frac{d\eta_{bb_{int}}^2}{dt}
\end{aligned} \tag{15}$$

This shows that intrinsic noise can be measured as the inverse of the mean, $\eta_{bb_{int}}^2 = \frac{1}{b}$, also at non-steady state. To verify, we used dynamic stochastic simulations (Gillespie, 1977) of three node circuits in which one node regulated two identical downstream nodes, A and B. We obtained time-varying measurements of A and B's distributions and covariance which allowed us to compute their total and shared noise. For either node, measured intrinsic noise was defined as the uncorrelated (not shared) portion of its total noise (Elowitz, 2002). We compared the measured intrinsic noise to predictions obtained using the steady-state definition, $\eta_{bb_{int}}^2 = \frac{1}{b}$, as well as our dynamic equation (as described in main paper). For the tested parameter sets, these predictions were equivalent and matched well the trajectories of the experimentally measured intrinsic noise, verifying that in our model intrinsic noise scales with protein copy number and is well represented by $\eta_{bb_{int}}^2 = \frac{1}{b}$ even under non-steady state conditions (Figure 5).

3.4 Estimation of intrinsic noise: *in-vivo* data

In the case of the *in vivo* data, intrinsic noise encompasses two key components: noise that, as in our model, comes from the stochastic creation and degradation of individual proteins and noise that stems from mRNA copy-number fluctuations. The contribution of mRNA fluctuations to the

protein noise have been shown to scale as the inverse of the protein copy number: $\eta_{bb_{mRNA}}^2 = \frac{C_m}{b}$ where C_m is the average number of proteins made per transcript (Bar-Evan, 2006). We assume, C_m is similar for all of genes in our networks (Bar-Evan, 2006; Stewart-Ornstein, 2012). We represent intrinsic noise as the sum of protein and mRNA contribution: $\eta_{bb_{int}}^2 = \frac{1}{b} + \frac{C_m}{b}$, and simplify it: $\eta_{bb_{int}}^2 = \frac{C}{b}$. However, we did not have a direct measure of protein copy number but rather fluorophore intensity, f , which is thought to be proportional to the number of transcripts in a cell (Elowitz, 2002). Therefore, to compute intrinsic noise we used $\eta_{bb_{int}}^2 = \frac{C_f}{f}$, where C_f is a fluorophore-specific scalability factor. We estimated C_f for each fluorophore, GFP and RFP, from experimental steady state data collected for a circuit in which the two reporters were driven by two copies of the pGAL1 promoter. Using these data we were able to estimate the scaling constants from: $\frac{C_{f_i}}{f_i} = \eta_{f_i}^2 - \eta_{f_i f_j}^2$, where $\eta_{f_i}^2$ is the total noise of either GFP or RFP computed from measurements of their respective intensities, f_i and f_j , and $\eta_{f_i f_j}^2$ is the fluorophores' shared noise. Estimation of the fluorophore scaling constants was done at the same time as collection of all the other data to ensure identical calibration of the flow cytometer.

Chapter 4: Network prediction utilizing noise

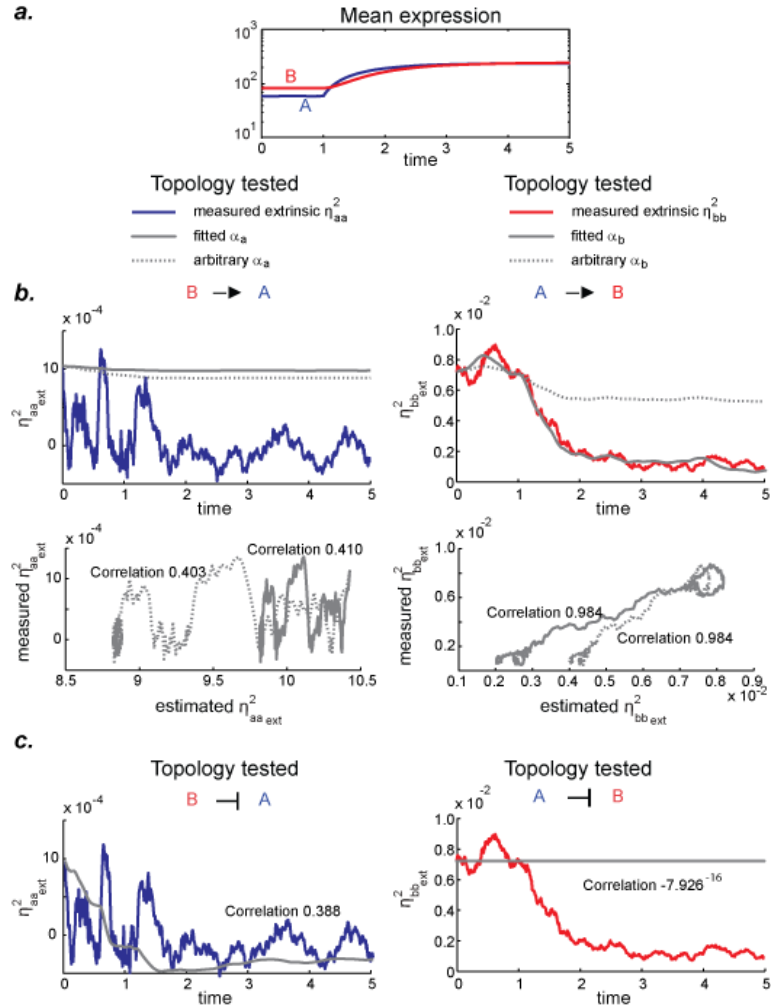


Figure 6: Reconstruction of an in silico network in which gene A activates gene B. (SSA data, population size 1000). (a) Dynamic mean expression of proteins of a two node network in which protein A activates protein B. (b)(c) Predicted noise (gray lines) fits the measured extrinsic noise (solid color lines) only when the correct regulatory relationship (activation) and directionality (A upstream of B) is assumed. Fit is evaluated as a linear correlation which allows for an arbitrary choice of the scaling factor α_b . In panel (b) dynamic noise was predicted using fitted (gray solid line) and arbitrary (dashed line) values for α_b . Regardless of the choice of the parameter, correlation is preserved.

4.1 Test using *in-silico* data

We first tested our method *in silico* using data obtained from stochastic simulations of activation and inhibition motifs. Specifically, we randomly sampled the parameters of the activation or inhibition motifs (Table 1) and generated time-dependent distributions. We used these distributions to extract extrinsic 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 extrinsic noise fluctuates over time in the downstream gene. Importantly, 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 7). Figure (6) shows an example of a reconstruction of one of the many networks that we tested. The networks for which we were unable to deduce the correct regulatory relationships (Figure 8) corresponded to regimes where extrinsic noise either was insignificant or did not propagate between the two nodes (Figure 9).

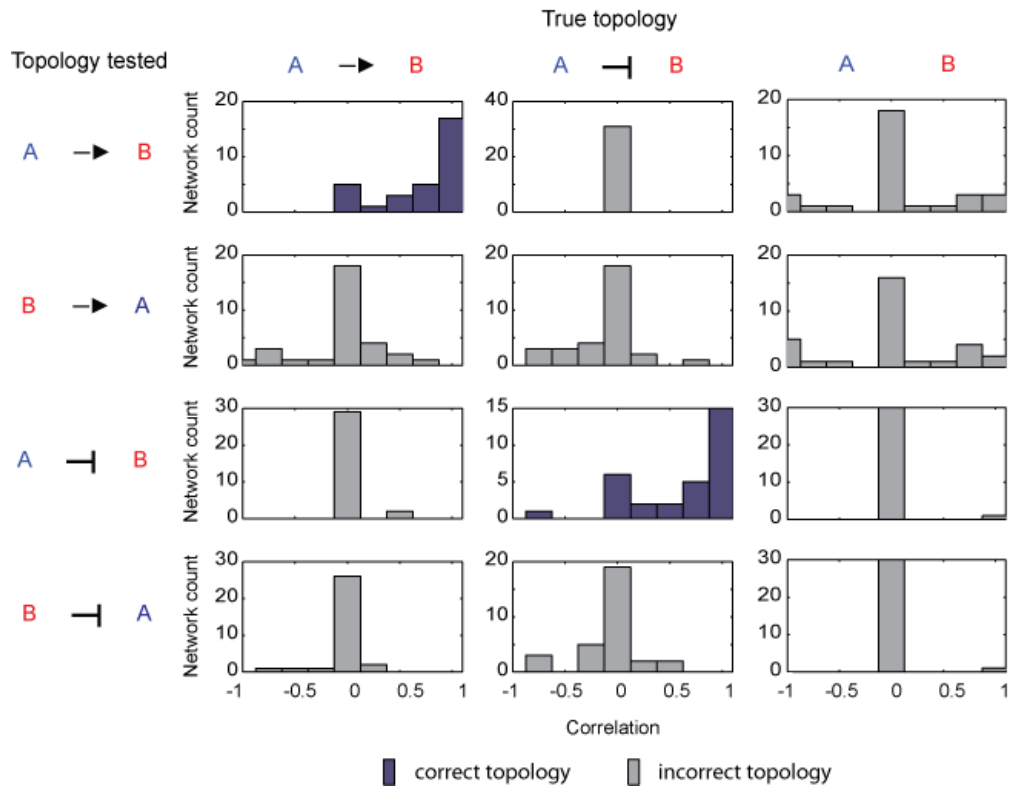


Figure 7: Evaluation of the method in silico using data obtained from SSA. Reconstruction of regulatory motifs, A activates B, A inhibits B, and co-regulation of A and B, obtained for 93 different, randomly chosen parameter sets. Each motif was tested for all regulatory permutations between A and B (rows). The results are reported as histograms of the correlation between measured and estimated noise for that particular topology. A low or negative correlation value for the correct topology indicates error in reconstruction.

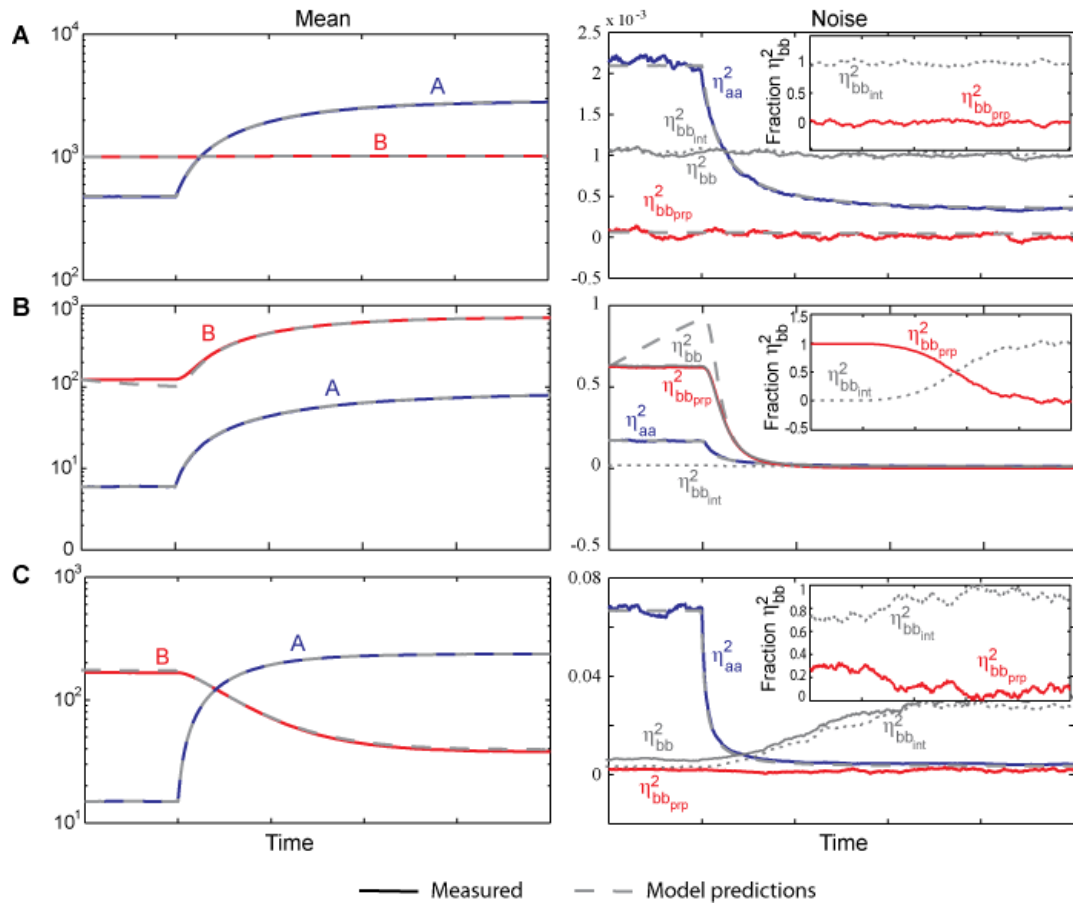


Figure 8: Dissection of cases that could not be reconstructed with noise information. (A) In this case, B and its noise were insensitive to changes in concentration of the upstream protein A (left panel) and hence noise did not propagate (right panel). (B) In this case, the approximations inherent in the model were inadequate for both mean and noise. (C) In this case shutdown synthesis of B by A made its noise dominated by its degradation alone, and no noise from A was propagated. Insets in (A), (B) and (C): propagated noise and intrinsic noise in B expressed as a fraction of total noise.

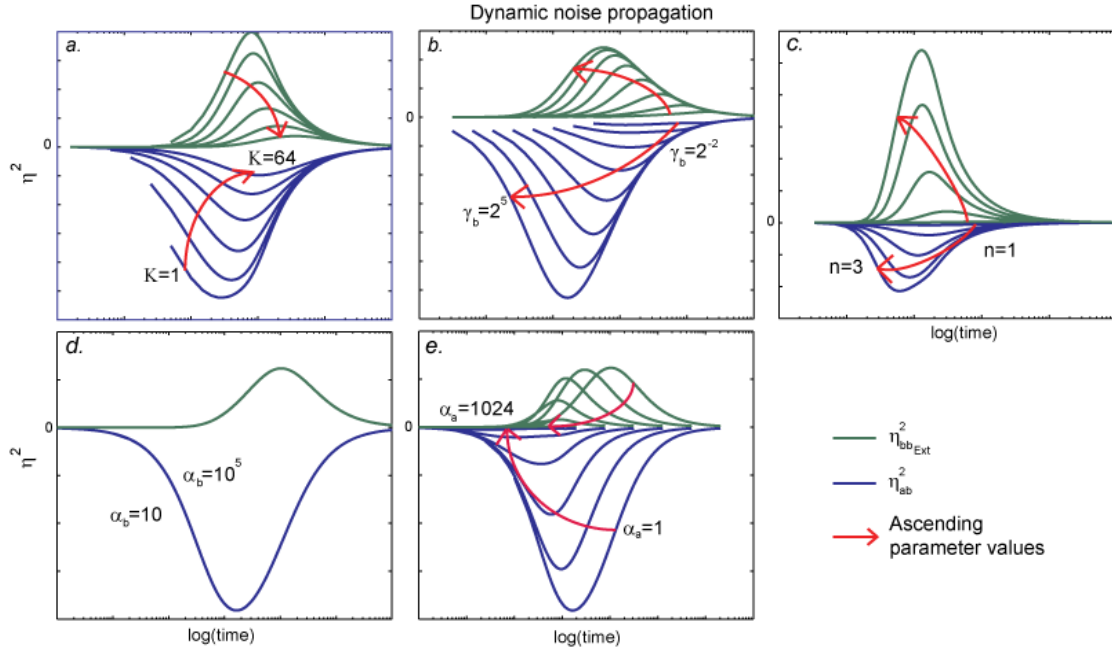


Figure 9: The dynamic propagation of noise depends on parameter values. Each plot is the propagated (green) and shared (blue) noise of a downstream node of an inhibitory network versus time. Each trace shows noise trajectory for different parameter value for **(a)** Michaelis-Menten parameter K , which is a key determinant of susceptibility, **(b)** degradation rate of the downstream protein, γ_b , which is a key determinant of time averaging **(c)** hill coefficient n , **(d)** synthesis rate of the downstream protein, α_b , and **(e)** upstream rate of activation, α_a . Red arrow indicates ascending parameter values.

4.2 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 $\Delta msn2/4$ 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 strain has no direct interactions between the two reporters but they are indirectly linked through co-regulation by the transcription factor Gal4 (Figure 10) (Appendix A).

All three strains were grown in non-inducing 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 minutes for 12-hours 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 extrinsic noise equation (Equation 12) 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 simple ODE models of different regulatory mechanisms (causal, i.e., activation or inhibition, or non-causal, i.e., having no relationship between A and B) to mimic the behavior of the data. We found, however, that the data could be fit equally well by both causal and non-causal models (Figure 10), indicating that mean information alone cannot discriminate between the possible alternate topologies.

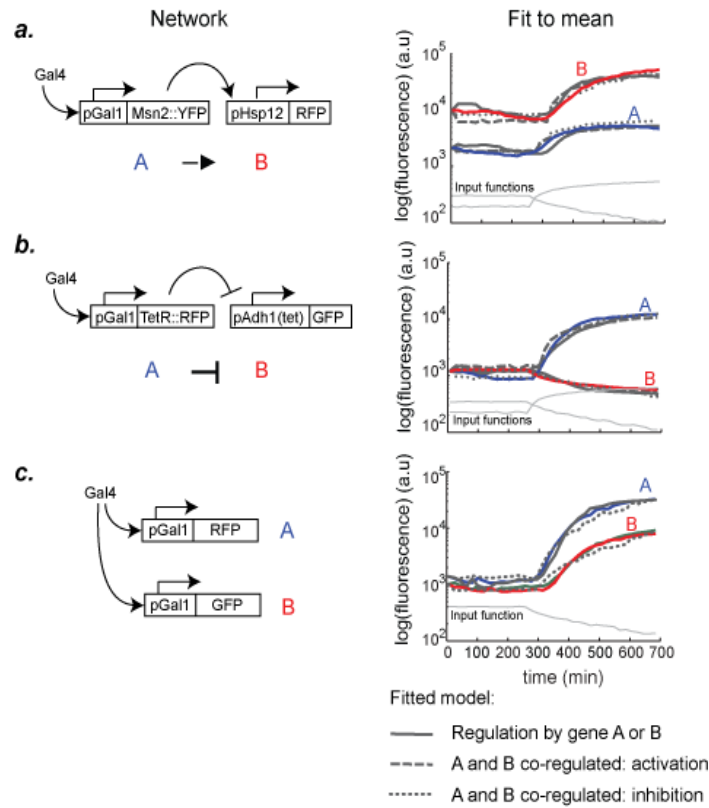


Figure 10: Using in vivo data collected for synthetic circuits, mean alone was insufficient for distinguishing between the possible regulatory relationships. (A) Activation circuit. Plot of means for the topology $A \rightarrow B$, $B \rightarrow A$ and A and B are co-regulated versus time. (B) Inhibition circuit. Plot of means for the topology $A \dashv B$, $B \dashv A$ and A and B are co-regulated versus time. (C) Co-regulation circuit. Plot of means for the topology $A \rightarrow B$, $B \rightarrow A$ and A and B are co-regulated versus time.

We next moved to testing whether the measured distributions could be exploited to provide such discrimination using our noise propagation methodology. Using Equation 12, we indeed determined that the extrinsic noise trajectory predicted using the topology that correctly

reflects the true causal relationship (Msn2 activates Hsp12) matches the experimental results (correlation of .97693 between predicted and measured extrinsic 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 -.14733). Furthermore, predictions made assuming the wrong regulatory mechanism (inhibition instead of activation) do not match experimental results regardless of circuit permutation (Figure 10c, d top).

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, TetR-RFP inhibits GFP, and other possible network permutations (correlation of 0.86930). Notably, the predicted trajectory for the reversed inhibitory relationship, GFP inhibits TetR-RFP, shows clear mismatch with the data (correlation $2.8103e-16$) (Figure 11 c,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 co-regulated, predicted extrinsic noise does not match the experimental data regardless of the assumed regulatory mechanism or network permutation, correctly suggesting that these genes have no causal interactions (Figure 11 c,d bottom). However, we cannot 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.

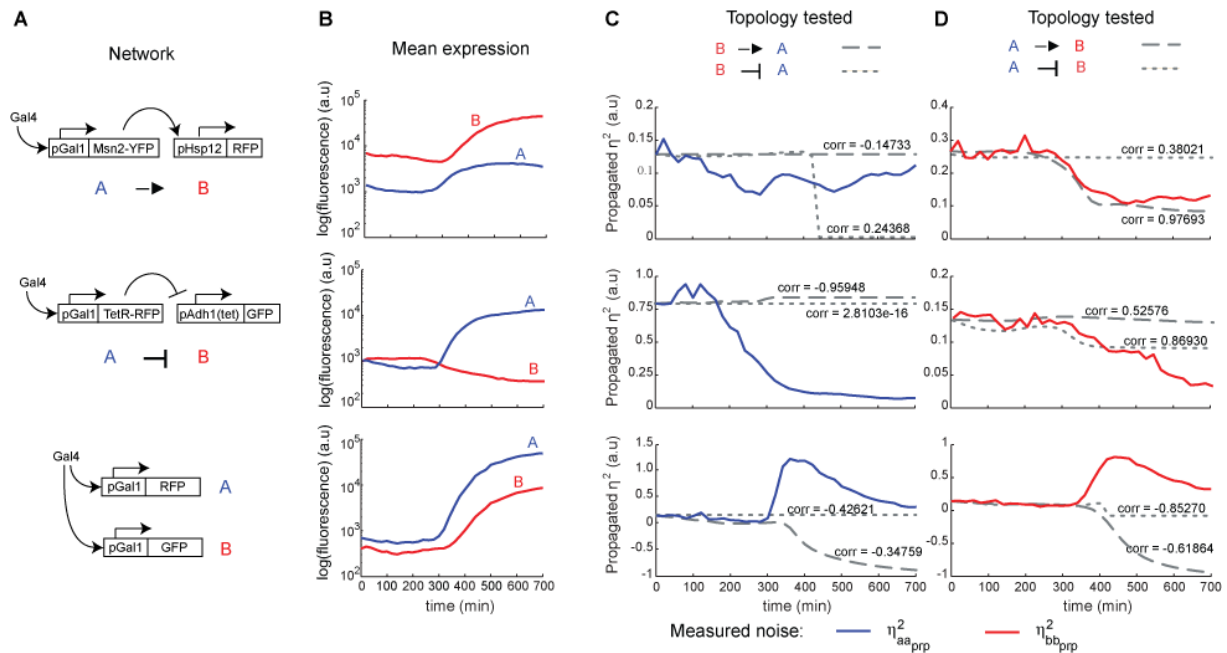


Figure 11: Reconstruction of three distinct in-vivo synthetic networks using noise information. (A) Schematics of the three networks in which A activates B (top), A inhibits B (center), and A and B are co-regulated by the same transcription factor (bottom). (B) Mean expression profiles of proteins in each of the three networks measured over a course of 12 hours. (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 co-regulated, we were not able to predict noise correctly for either circuit permutation, suggesting that A and B have no direct regulatory relationship (bottom).

Chapter 5: Complex regulatory systems

Thus far we have focus on simple, two node circuits. In the next two section we focus on more complex systems.

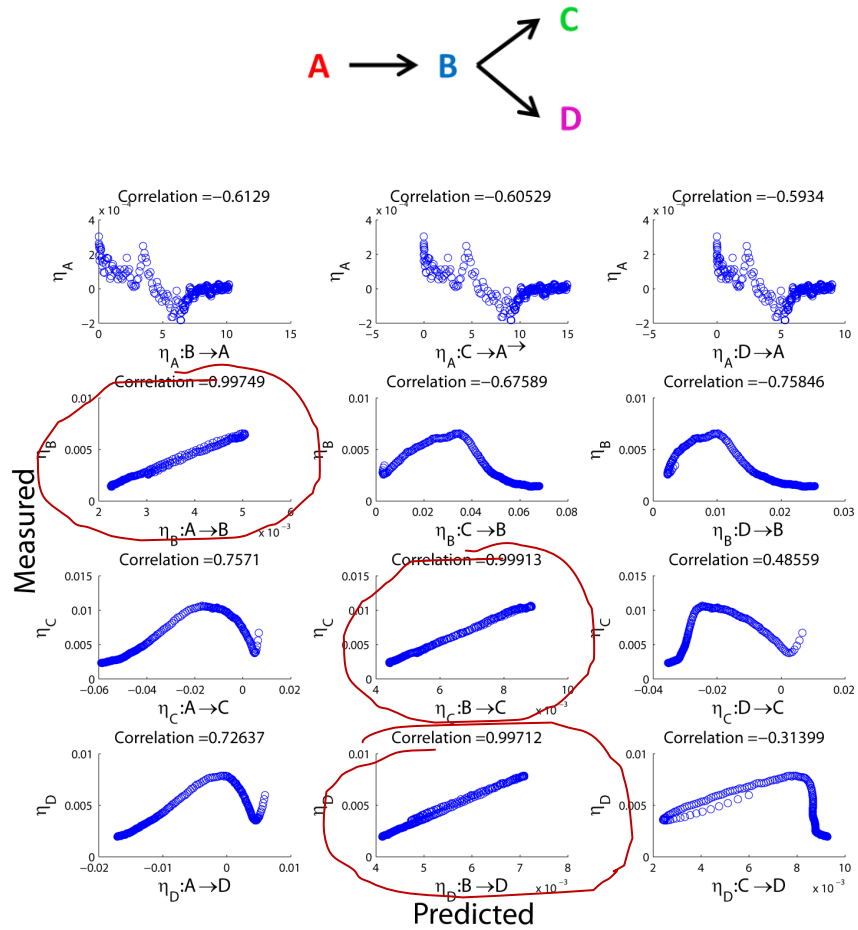


Figure 12: Reconstruction of a multi-node in silico network. The methods was able to identify all three regulatory connections present in the network (circled).

5.1 Multi-node networks

In order to verify that our methods can be applied to larger regulatory networks we used SSA data of four node networks. Nodes were connected by either activation or inhibition link such that node A was regulating B and B was regulating two nodes, C and D. Figure 12, shows an example of such network and its reconstruction. All possible regulatory permutations between the nodes of the network were tested but only those that matched the true relationships were able to match the measured noise.

5.2 Multi-input systems

Because gene expression is often governed by multiple transcription factor we investigate noise propagation in a multi-input systems. In our model expression of gene B was regulated by two proteins A and C. In order to simplify the analysis we used a slightly simplified model of activation which we describe as a first order, linear process: $R_b^+ = \alpha_b a$. The mean expression for this system is:

$$\frac{da}{dt} = \alpha_a - \gamma_a a \quad (16)$$

$$\frac{dc}{dt} = \alpha_c - \gamma_c c \quad (17)$$

$$\frac{db}{dt} = \alpha_{b_a} a + \alpha_{b_c} c - \gamma_b b \quad (18)$$

where the term $\alpha_{b_a} a$ describes activation of B by A and $\alpha_{b_c} c$ activation by C. Using this simplified model we derive equation of dynamic noise propagation:

$$\frac{d\eta_{aa}^2}{dt} = -2\eta_{aa}^2 \frac{R_a^+}{a} + \frac{R_a^+}{a^2} + \frac{R_a^-}{a^2} \quad (19)$$

$$\frac{d\eta_{cc}^2}{dt} = -2\eta_{cc}^2 \frac{R_c^+}{c} + \frac{R_c^+}{c^2} + \frac{R_c^-}{c^2} \quad (20)$$

$$\frac{d\eta_{bb}^2}{dt} = -2\eta_{bb}^2 \left(\frac{R_{ba}^+}{b} + \frac{R_{bc}^+}{b} \right) + 2H_{ba}^* \eta_{ab}^2 \frac{R_{ba}^+}{b} + 2H_{bc}^* \eta_{bc}^2 \frac{R_{bc}^+}{b} + \frac{R_{ba}^+}{b^2} + \frac{R_{bc}^+}{b^2} + \frac{R_b^-}{b^2} \quad (21)$$

Finally we separate terms of Equation (21) to obtain expressions for intrinsic and propagated components in nodes B:

$$\frac{d\eta_{bb_{int}}^2}{dt} = -2\eta_{bb_{int}}^2 \left(\frac{R_{ba}^+}{b} + \frac{R_{bc}^+}{b} \right) + \frac{R_{ba}^+}{b^2} + \frac{R_{bc}^+}{b^2} + \frac{R_b^-}{b^2} \quad (22)$$

$$\frac{d\eta_{bb_{ext}}^2}{dt} = -2\eta_{bb_{ext}}^2 \left(\frac{R_{ba}^+}{b} + \frac{R_{bc}^+}{b} \right) + 2H_{ba}^* \eta_{ab}^2 \frac{R_{ba}^+}{b} + 2H_{bc}^* \eta_{bc}^2 \frac{R_{bc}^+}{b} \quad (23)$$

Inspection of Equation (23) reveals that the propagated noise of B is a sum of contributions of the upstream nodes, A and C. This can be verified by setting either one of the contributions to zero, for example, if we let $\alpha_c = 0$ then:

$$\frac{d\eta_{bb_{ext}}^2}{dt} = -2\eta_{bb_{ext}}^2 \left(\frac{R_{ba}^+}{b} + 0 \right) + 2H_{ba}^* \eta_{ab}^2 \frac{R_{ba}^+}{b} + 0 = -2 \frac{R_{ba}^+}{b} (\eta_{bb_{ext}}^2 - H_{ba}^* \eta_{ab}^2) \quad (24)$$

which gives us the same expression as the dynamic single-input extrinsic noise expression, Equation (12).

We then tested whether we can reconstruct a system in which a node has two regulatory inputs. We found that if we take only one regulatory link of B into consideration we are unable to infer the correct topology. However, if the noise predictions of the two upstream influences are added up the resulting prediction matches the measured noise of B very closely (correlation 0.98139) (Figure 13).

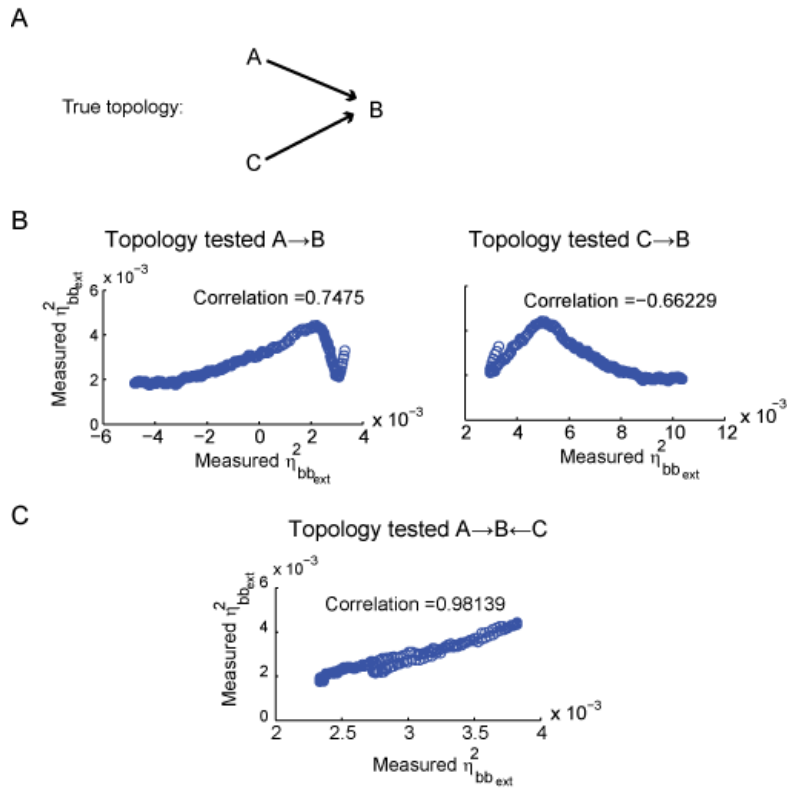


Figure 13: Reconstruction of a two-input in silico network. **(A)** Schematic of the network in which expression of B is regulated by A and C. **(B)** Test for regulatory relationship of B considering only a single regulatory link with either (left) A or (right) C. **(C)** Prediction using a two-input model. The resulting noise prediction is a sum of individual contributions of A and C.

Chapter 6: Discussion and closing remarks

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

Some early studies explored this idea 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 (Dunlop, 2008). However, the relationship between noise in different components of a circuit is governed by potentially complex relationships as depicted by Equations 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 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. When compared to experimentally measured variability, this information allows us to provide evidence for or against this topology. Our computational investigations and experimental data both support the notion that noise propagation alone can be sufficient to

discriminate between alternate topologies when causal relationships exist between different network components.

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 particularly 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 equation 9 to predict the dynamic trajectory of the propagated noise.

As a proof of concept, we demonstrated reconstruction of networks with two nodes. However, because our method relies on solving differential equations, it has low computational cost. Therefore, we envision that it can be extended to multi-node systems, for example by carrying out combinatorial, pair-wise 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, multi-gene regulatory structures.

In summary, as the development of increasingly sophisticated single cell measurement techniques (Soon, 2013; Munsky, 2012) accelerates, there is increasing need for approaches that utilize population distribution information. Our approach provides a solid first step in that direction.

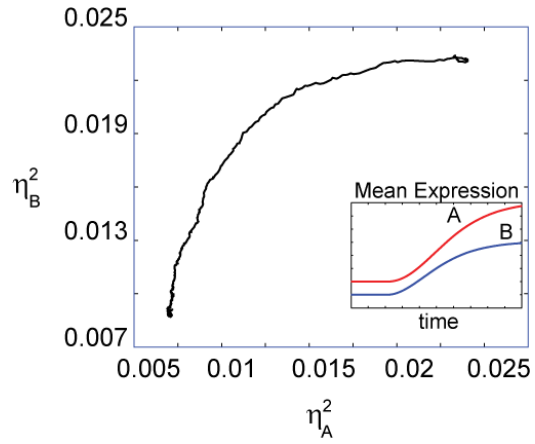


Figure 14: A linear correlation between the noise profiles of two nodes in a network is not a reliable predictor of their connectivity. Noise of A and B in a simple model: $\frac{db}{dt} = \frac{\alpha_b a}{a+K} - \gamma_b b$ shows a nonlinear relationship. Inset: mean expression of A and B as a function of time.

Appendix A

Plasmids and Strain Construction

Galactose responsive constructs were constructed by amplification of the Gal1 promoter from the yeast genome by PCR followed by restriction enzyme cloning into a single integration Trp1 or His3 marked vector upstream of Venus (YFP) 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 an Ura3 marked vector. The Adh1(tet) promoter has been described previously (Murphy et al., 2007), the 700bp upstream of the ATG was amplified and cloned in front of GFP in a His3 marked vector.

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 for downstream use.

Growth and fluorescence measurements by flow cytometry

Yeast strains were grown to saturation overnight at 30C in 3ml of synthetic complete media with 2% raffinose (SCraf) as a carbon source. Cells were diluted 1:100 SCraf into deep 96 well plates (Corning) and grown for 6-8hrs at 30C on orbital shakers (Elim) to an OD of ~0.1. To induce expression of Gal1 regulated 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 minutes 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 seconds, typically corresponding to 2000-10000 cells. GFP and YFP were excited at 488nm, and fluorescence was collected through a 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 seconds 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 (Newman, 2006).

Network inference pseudocode

Inputs:

data - time-series population measurements for i nodes of the network

Procedure:

```
measuredNoise = getMeasuredNoise(data);
```

```
for each node in network
```

```

childData = data(i);

for each possible parent(j)

    parentData = data(j);

    // compute expected noise for the hypothetical parent-child relationship

    predictedNoise = predictNoise(childData, parentData);

    // evaluate prediction using correlation

    score(predictedNoise, measuredNoise(i));

end

end

//-----

getMeasuredNoise(data)

for each node(i)

    noiseTotal = var/mean^2;

    noiseInt =  $\varepsilon$ /mean; //  $\varepsilon$  estimated as described in sections 3.3 and 3.4

    noiseExt = noiseTotal - noiseInt;

end

```



```

    return noiseExt;

end

//-----

predictNoise(childData, parentData)

    sharedNoise = cov(parentData,childData)/(meanChild * meanParent);

    childInfo.meanFunc = fit(mean(childData));

    childInfo.noiseFunc = fit(noise(childData));

    parentInfo.meanFunc = fit( mean(parentData ) );

    parentInfo.sharedNoiseFunc = fit (sharedNoise)

    // find the necessary parameters  $K$  and  $n$  as described in section 3.2

    parentInfo.params = findParams(childNoise, sharedNoise)

    noise = noiseODE(t, x0, childInfo, parentInfo);

    return noise;

end

```

```

%-----

% model:

%    dB/dt = alphaB/(A + K) - gammaB * B;

%-----

function dxdt = noiseODE(t, x, child, parents)

    a = feval(parents(1).meanFunc, t-1); // measured mean value of A

    b = feval(child(1).meanFunc, t-1); // measured mean value of B

    sharedNoise = feval(parents(1).sharedNoiseFunc, t-1); // measured shared noise

    extNoiseB = feval(child(1).noiseFunc, t-1); // measured extrinsic noise

    [alphaB, K, n]= parents(1).params;

    dxdt(1) = -2 * Rb+/b * (extNoiseB + Hba * sharedNoise );

    return dxdt

end

```

References

D. W. Austin, M. S. Allen, J. M. McCollum, R. D. Dar, J. R. Wilgus, G. S. Saylor, N. F. Samatova, C. D. Cox, M. L. Simpson. Gene network shaping of inherent noise spectra. *Nature* 439 (2), 608:611 (2006)

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

A. Bar-Evan, J. Paulsson, N. Maheshari, M. Carmi, E. O'Shea, Y. Pilpel, N. Barkai. Noise in protein expression scales with natural protein abundance. *Nat. Genet.* 38, 636 (2006)

Cagatay T, Turcotte M, Elowitz M, Garcia-Ojalvo J, Suel GM. Architecture-Dependent Noise Discriminates Functionally Analogous Differentiation Circuits. *Cell* 139, 512-522, 2009.

C. D. Cox, J. M. McCollum, M. S. Allen, R. D. Dar, M. L. Simpson. Using noise to probe and characterize gene circuits. *PNAS* 31 (105): 10809–10814 (2008)

M. J. Dunlop, R. S. Cox III, J. H. Levine, R. M. Murray, M. B. Elowitz, Regulatory activity revealed by dynamic correlations in gene expression noise. *Nat. Genet.* 40, 1493 (2008).

M Eisen, P Spellman, P Brown, D Botstein. Cluster analysis and display of genome-wide expression patterns. *PNAS USA* 95: 14863–14868 (1998).

Eldar A, Elowitz MB. Functional Roles of Noise in genetic Circuits. *Nature* 467, 167-173, 2010.

M. B. Elowitz, A. J. Levine, E. D. Siggis, P. S. Swain. Stochastic Gene Expression in a single cell. *Science* 297, 1183 (2002).

S. Engblom. Computing the moments of high dimensional solutions of the master equation. *App. Math. and Comp.* 180, 498 (2006).

D. T. Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Physical Chemistry* 81 (25): 2340–2361 (1977).

S. Imoto, T Higuchi, T Goto, S Kuhara, S Miyano. Combining Microarrays and Biological Knowledge for Estimating Gene Networks via Bayesian Networks. *IEEE CSB'03*, 104-113 (2003).

R. Losick, C. Desplan. Stochasticity and Cell Fate. *Science* 320, 65-68 (2008)

H.H. McAdams, A. Arkin. Stochastic mechanisms in gene expression. *PNAS* 94, 814–819 (1997)

B. Munsky, M. Khammash. The Finite State Projection Algorithm for the Solution of the Chemical Master Equation . *J. Chem. Phys.* 124, 044104 (2006)

B. Munsky, G. Neuert, A. van Oudenaarden. Using gene expression noise to understand gene regulation. *Science* 336, 183 (2013).

K.F. Murphy, G. Balázsi, J.J. Collins. Combinatorial promoter design for engineering noisy gene expression. *Proc Natl Acad Sci U S A.* 104(31):12726-31 (2007).

G. Neuert, B. Munsky, R.Z. Tan, L. Teytelman, M. Khammash, A. van Oudenaarden. Systematic identification of signal-activated stochastic gene regulation. *Science* 339, 584 (2013)

Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS. Single-cell proteomic analysis of *S. cerevisiae* reveal the architecture of biological noise. *Nature* 441, 840-846 (2005)

J.R. Newman, S. Ghaemmaghami, J. Ihmels, D.K. Breslow, M. Noble, J.L. DeRisi, J.S. Weissman. Single-cell proteomic analysis of *S. cerevisiae* reveal the architecture of biological noise. *Nature* 441, 840-846 (2006).

J. Paulsson. Summing up the noise in gene networks. *Nature* 427, 415 (2004).

J. Paulsson. Models of stochastic gene expression. *Physics of Life Reviews* 2, 157-175 (2005).

J. M. Pedraza, A. van Oudenaarden, Noise propagation in gene networks. *Science* 307, 1965 (2005).

C. A. Penfold, D. L. Wild. How to infer gene networks from expression profiles, revisited. *Interface Focus* 1, 857-870 (2011).

N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, M. B. Elowitz. Gene Regulation at the Single-Cell Level. *Science* 307: 1962-1965 (2005)

M. A. Savageau. Parameter Sensitivity as a Criterion for evaluating and computing the performance of biochemical systems. *Nature* 229, 542-4 (1971).

M. L. Simpson, C. D. Cox, S. S. Gary S. Sayler. Frequency domain analysis of noise in autoregulated gene circuits. *PNAS* 8 (10): 4551-4556 (2003)

W. W. Soon, M. Hariharan, M. P. Snyder. High-throughput sequencing for biology and medicine. *Mol Syst Biol* 9, 640. (2013)

J. Stewart-Ornstein, J. S. Weissman, H. El-Samad, Cellular noise regulons underlie fluctuations in *Saccharomyces cerevisiae*. *Mol. Cell* 45, 483 (2012).

G. M. Suel, J. Garcia-Ojalvo, M. L. Liberman, M. B. Elowitz. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 440: 545-550 (2006)

M . Thattai, A. van Oudenaarden. Intrinsic Noise in Gene Regulatory Networks. *PNAS* 98: 8615-861 (2001)

L. S. Weinberger, R. D. Dar, M. L Simpson. Transient-mediated fate determination in a transcriptional circuit of HIV. *Nature Genet* 40 (4): 466:470 (2008)

Xiang Z, Minter RM, Bi X, Woolf PJ, He Y. MiniTUBA: Medical Inference by Network Integration of Temporal Data Using Bayesian Analysis. *Bioinformatics*, 23, 2423-2432 (2007).

D. Zenklusen, D. R. Larson, R. H. Singer, Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* 15, 1263 (2008).

Publishing Agreement

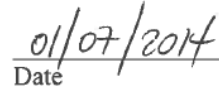
It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature



Date