Using Effective Subnetworks to Predict Selected Properties of Gene Networks

Gemunu H. Gunaratne¹*, Preethi H. Gunaratne²,³, Lars Seemann¹, Andrei Török⁴

¹ Department of Physics, University of Houston, Houston, Texas, United States of America, ² Department of Biology and Biochemistry, University of Houston, Houston, Texas, United States of America, ³ Human Genome Sequencing Center and Department of Pathology, Baylor College of Medicine, Houston, Texas, United States of America, ⁴ Department of Mathematics, University of Houston, Houston, Texas, United States of America

Abstract

Background: Difficulties associated with implementing gene therapy are caused by the complexity of the underlying regulatory networks. The forms of interactions between the hundreds of genes, proteins, and metabolites in these networks are not known very accurately. An alternative approach is to limit consideration to genes on the network. Steady state measurements of these influence networks can be obtained from DNA microarray experiments. However, since they contain a large number of nodes, the computation of influence networks requires a prohibitively large set of microarray experiments. Furthermore, error estimates of the network make verifiable predictions impossible.

Methodology/Principal Findings: Here, we propose an alternative approach. Rather than attempting to derive an accurate model of the network, we ask what questions can be addressed using lower dimensional, highly simplified models. More importantly, is it possible to use such robust features in applications? We first identify a small group of genes that can be used to affect changes in other nodes of the network. The reduced effective empirical subnetwork (EES) can be computed using steady state measurements on a small number of genetically perturbed systems. We show that the EES can be used to make predictions on expression profiles of other mutants, and to compute how to implement pre-specified changes in the steady state of the underlying biological process. These assertions are verified in a synthetic influence network. We also use previously published experimental data to compute the EES associated with an oxygen deprivation network of E.coli, and use it to predict gene expression levels on a double mutant. The predictions are significantly different from the experimental results for less than 30% of genes.

Conclusions/Significance: The constraints imposed by gene expression levels of mutants can be used to address a selected set of questions about a gene network.

Citation: Gunaratne GH, Gunaratne PH, Seemann L, Török A (2010) Using Effective Subnetworks to Predict Selected Properties of Gene Networks. PLoS ONE 5(10): e13080. doi:10.1371/journal.pone.0013080

Introduction

Living systems are typically able to maintain their physiological state under environmental changes and isolated genetic mutations [1]. This robustness, referred to as homeostasis [2] or canalization [3,4], is achieved through feedback within highly connected regulatory networks of genes, proteins, and metabolites [5–10]. For example, an action that reduces the expression of one gene may cause coordinate changes in other nodes to leave the physiological state unaffected. If a genetic mutation blocks one pathway, other avenues on the associated network may take its place. Unfortunately, this systemic stability often makes it difficult to eliminate defects in a biological network, as evidenced by the surprising lack of efficacy of many drugs that were designed to act on single molecular targets [11,12]. The coupling can also lead to side effects from medications. For example, anti-inflammatory COX-2 inhibitors (e.g., Vioxx) cause adverse cardiovascular effects due to a concomitant imbalance of the lipids prostacyclin and thromboxane A2, which lie on the same network [13]. Clearly, the most effective and least detrimental changes in a biological process are implemented by altering the system in its entirety. This task requires predictive mathematical models which can be constructed from experimental data. In this paper, we propose an approach for such a construction.

There are hundreds of genes, proteins, and other molecular participants associated with most biological processes. Gene regulatory networks model all interactions between these nodes. However, the forms of these dependencies, as well as kinetic parameters such as reaction rates and diffusion constants are, at best, only known approximately [14]. It is unlikely that gene regulatory networks which are sufficiently accurate to make quantitative predictions on the underlying biological processes will be available in the near future [15,16].

Many approaches to reduce the complexity of regulatory networks have been proposed [5]. Small modules or network motifs [17,18] associated with specific tasks have been identified. Boolean variables [19] can reduce the complexity, although the coarse-graining will limit predictability to qualitative characteris-
tics such as bifurcations. In gene influence networks [14,20], a gene, its transcript, and protein are represented by a single node, which is quantified by the expression level of the mRNA. Regulatory interactions between nodes of an influence network include actions mediated by other components in the network.

Consider an influence network containing $N$ genes $G = \{ G_1, G_2, \ldots, G_N \}$; denote the expression level of the $K$th gene by $X_K$ and the state of the network by $X = \{ X_1, X_2, \ldots, X_N \}$. The influence network can be modeled by a set of ordinary differential equations $F: \mathbb{R}^N \rightarrow \mathbb{R}^N$

$$\begin{align*}
\dot{X}_1 &= F_1(X) \\
\dot{X}_2 &= F_2(X) \\
\vdots & \quad \vdots \\
\dot{X}_N &= F_N(X).
\end{align*}$$

(1)

Steady states of influence networks can be obtained from DNA microarray experiments [5]. However, most influence networks contain hundreds of genes; thus, even if $F_k(X)$ are assumed to have a simple (e.g., linear) form [21], a prohibitively large number of microarray experiments will need to be conducted in order to compute $F$. Moreover, gene expression levels in microarray experiments have large (~10%) error bars; when $N$ is large, the inversions needed to compute $F$ will exacerbate the uncertainty to a level which will make predictions difficult. One possibility is to only extract partial information on these networks through inference algorithms such as Network Identification by Multiple Regression [14], and Mode-of-action by Network Identification [22].

We propose an alternative approach. Rather than attempting to construct an accurate model of a gene network, we ask what questions on the network can be addressed (perhaps approximately) using low-dimensional and highly simplified effective models constructed from empirical data. What data would be needed for the construction? Will issues addressed through the approach be useful in applications?

We first note that genes in an influence network can be partitioned into strongly coupled subgroups or clusters. This partition can be made either using co-expression under genetic perturbations [23–25], or through the use of the Gene Ontology (GO) database (http://geneontology.org). Our main assumption is that the behavior of all nodes within a cluster can be controlled by imposing suitable changes in a small, specially chosen, subset of its members. The set could include genes that translate to transcription factors, and would hence influence many other genes [26,27]. It may also include microRNAs within the cluster, each of which affect many genes through post-transcriptional regulation [28,29], even though their fold induction on each gene is small.

Suppose we have partitioned the $N$ genes of the influence network into clusters, and identified a small number of genes/microRNAs from each cluster that can be used to control the expression levels of the remaining genes. Denote the set of these nodes by $S$. The number $n$ of nodes in $S$ is much smaller than $N$. We will represent their expression levels by $X_S = \{ X_1, X_2, \ldots, X_n \}$, and re-index the variables so that the remaining expression levels are $\{ X_{n+1}, \ldots, X_N \}$. With the new ordering, we write the state of the network as $X = \{ x_{int}, x_{ext} \}$ where we will refer to $x_{int} = \{ x_1, x_2, \ldots, x_n \}$ and $x_{ext} = \{ X_{n+1}, \ldots, X_N \}$, as “internal” and “external” variables respectively.

In this paper, we limit consideration to networks with steady state solutions. When external perturbations are made on genes within $S$, expression levels of the remaining genes at equilibrium are determined by Eqn. (1). These steady states lie on an $n$-dimensional surface in $\mathbb{R}^n$, which we denote by $\mathcal{S}_S$. Figure 1(a) shows a schematic (2-dimensional) solution surface for the synthetic network introduced in the Results Section.

We make the following observations on solutions of the system $X = F(X)$. First, if the unconstrained system has a unique stable steady state which we denote by $P^{(0)} = \{ P_{m}^{(0)}, P_{ext}^{(0)} \} = \{ P_1^{(0)}, \ldots, P_n^{(0)}, \ldots, P_{N-1}^{(0)} \}$. It satisfies the $N$ equations $F(P^{(0)}) = 0$. The point $P_0$ representing it lies on $\mathcal{S}_S$. Next, consider the single knockout mutant $\Delta G_m$ (assumed to be viable) obtained by knocking out the $m$th gene. Since $x_m$ is set externally, the $m$th equation of (1) is no longer valid. The solution for the expression levels is obtained by solving the remaining $(N-1)$ equations. We denote this equilibrium by $P^{(m)} = \{ P_{m}^{(m)}, P_{m+1}^{(m)}, \ldots, P_{N}^{(m)} \}$ with $P^{(m)} = 0$, and represent it by $P_m$. Since the equilibrium is associated with changes made within the set $S$, $P_m$ lies on $\mathcal{S}_S$.

![Figure 1. Example of an $n$-dimensional solution surface $\mathcal{S}_S$ of (1). The example is chosen from the synthetic network introduced in the Results Section. (a) The surface shown represents the expression levels of the external variable $X$ as the internal variables $X_1$ and $X_2$ are modified. (b) The point $P_0$ representing expression levels of the wildtype and points $P_m, m = 1,2$ representing expression levels of single knockout mutants $\Delta G_m$ lie on this surface. The EES is defined so that its solutions lie on the unique 2-dimensional plane (blue) $\mathcal{H}_S$ passing through $P_0$, and $P_m, m = 1,2$. As can be seen, due to restrictions imposed on the EES, the surfaces $\mathcal{S}_S$ and $\mathcal{H}_S$ are close.

doi:10.1371/journal.pone.0013080.g001
Consequently, \( P_0 \) as well as \( P_m \) for \( m=1,2,\ldots,n \) lie on \( S_S \), see Figure 1(b).

Our goal is to construct a system, referred to as the “effective empirical subnetwork” (EES), that can be computed using the gene expression levels of mutants discussed above, and whose equilibria are close to \( S_S \). Observe that \( P_0 \) and the \( n \) points \( P_m \) define a unique \( n \)-dimensional plane in \( \mathbb{R}^n \), which we denote by \( \mathcal{H}_S \). Figure 1(b) shows the surface for the example above. The EES describes the set \( \mathcal{H}_S \) as parametrized by the internal variables. Since both \( S_S \) and \( \mathcal{H}_S \) contain the points \( P_0 \) and \( P_m \), \( m=1,2,\ldots,n \), we expect them to be close in the region of interest. Observe that the EES: \( \mathbb{R}^n \rightarrow \mathbb{R}^n \) is a function determined by \( P_0 \) and \( P_m \), but is otherwise independent of \( \mathbf{F} \). In particular, each \( X_K \) is a linear function of \( x_1,\ldots,x_n \). Since \( P_0 \) lies on \( \mathcal{H}_S \)

\[
X_k - P_k^{(0)} = \sum_{i=1}^{n} a_{ki} (x_i - P_i^{(0)}), \\
\text{for each } K=(n+1),\ldots,N. \tag{2}
\]

for \( k=1,2,\ldots,n \).

We have thus implemented two significant simplifications. The original system \( \mathbf{F}: \mathbb{R}^n \rightarrow \mathbb{R}^n \) contained a large number (\( N \approx \) several hundred) of nonlinearly coupled variables. In contrast, the EES: \( \mathbb{R}^n \rightarrow \mathbb{R}^n \) has a small number (\( n \approx 10 \)) of internal variables, is linear, and can be constructed using the steady state solutions of the original system (wild-type) and \( n \) single knockout mutants. Clearly, the EES is not an accurate representation of the original network. The issue is whether there are questions about \( \mathbf{F} \) that can be addressed using the EES. As we show below, this is indeed the case due to geometrical constraints imposed on the solution surface. Specifically, the EES can be used to predict, approximately, the expression levels of all nodes in \( \mathbf{F} \), when external changes are made within \( \mathbf{S} \); e.g., double knockout mutants. The validity of the EES construction can be tested by comparing its predictions with microarray data from such mutants. More significantly, the EES can be used to compute how the equilibrium of the system can be moved from its initial state \( P_0 \) to a prespecified set of expression levels defined by a point \( P_{aim} \), see Figure 2. Since \( P_{aim} \) will, in general, not lie on the solution surface, it cannot be reached through changes within \( \mathbf{S} \). Instead, we can use the EES to compute \( P_{aim} \), which is the closest point to \( P_{aim} \) on the plane \( \mathcal{H}_S \), see Figure 2. Since the surfaces \( S_S \) and \( \mathcal{H}_S \) are close, changes imposed on the system by the external actions are expected to be close to those computed from the EES. Below, we verify this proximity in a synthetic influence network.

\section*{Results}

\subsection*{A Synthetic Influence Network}

In the model system, \( F_K(\mathbf{X}) \) is a linear combination of sigmoidal Hill functions; specifically,

\[
F_K(\mathbf{X}) = X_K \sum_{i=1}^{N} a_{ki} \left[ H(X_i; \epsilon_{ki}) - H\left(P_i^{(0)}; \epsilon_{ki}\right) \right], \tag{4}
\]

where \( H(X; \epsilon) = X^h / (X^h + \epsilon^h) \) is the Hill function and the Hill index \( h \) is chosen to be 2. The action of the \( i^{th} \) gene on the \( K^{th} \) one is characterized by parameters \( a_{ki} \) and \( \epsilon_{ki} \), which are assumed to be independent of the state \( \mathbf{X} \) of the system. The action is activating if \( a_{ki} > 0 \) and inhibiting if \( a_{ki} < 0 \). The system is constructed so that \( P^{(0)} \) is a steady state of Eqns. (1). Numerically, we find that model systems defined by Eqns. (1) and (4), have at most one stable equilibrium. We suspect that this is due to restrictions imposed by the fact that the sign of each partial derivative \( \partial F_K / \partial X_i \) is independent of the state of the system.

In order to compute the solutions to the knockout mutant \( \Delta G_K \), we set \( x_k = 0 \), and solve the remaining equations of (1) as a nonlinear least squares problem. When the normalized residue fails to fall below \( 10^{-10} \), it is assumed that the corresponding solution does not exist.

We report on a model system containing 21 nodes and shown schematically in Figure 3. We start with the three subnetworks, each of size 7. The vector \( P^{(0)} \) for each of these subgroups consists of random entries between 0.5 and 1.5, and the matrix \( (\epsilon_{ki}) \) contains random values between 0 and 2. The entries of the Jacobian of the system given by Eqns. (1) and (4) at \( P^{(0)} \) are \( J_{ki} = a_{ki} H\left(P^{(0)}; \epsilon_{ki}\right) \); thus \( a_{ki} \) can be computed for a given set.
of JKi's. Since we require P0 to be stable, we insist that all eigenvalues of the Jacobian be negative. This is guaranteed by starting with a diagonal matrix with negative entries and performing a (random) orthonormal transformation. Once three such subnetworks are computed, their nodes are coupled by sparse, weak interactions. Each node in a subnetwork is coupled to only one from each of the other subnetworks, and the mean coupling strength is chosen to be 0.1 of the average coupling within subgroups.

The EES is to be constructed using the expression levels of single knockout mutants. As illustrated in Figure 3, mutants ΔG1, ΔG3, ΔG6, ΔG10, ΔG15, ΔG16, ΔG20, and ΔG21 in our example are not viable; i.e., the corresponding set of equations do not have a solution. Genes shown in black are not viable; i.e., the corresponding set of equations do not have a solution. Genes shown in red are used to construct the effective empirical subnetwork.

doi:10.1371/journal.pone.0013080.g003

Results for the first group (genes in S) are as follows. Of the 56 comparisons, 46 EES predictions are within 1% of the expression levels computed from (1); 3 others are between 1–5% and 3 between 5–10%. Results for the second group (genes outside of S) are as follows. Of the 210 expression levels to be compared, 170 EES predictions are within 1% of the expression levels computed from (1); 30 more are between 1–5%, and 7 others are between 5–10%.

We finally demonstrate how the equilibrium of the system can be moved (near to) a pre-specified set of expression levels. The original equilibrium of our example is

\[
\mathbf{P}^{(0)} = \left\{ \mathbf{P}_{int}^{(0)}, \mathbf{P}_{ext}^{(0)} \right\},
\]

with

\[
\mathbf{P}_{int}^{(0)} = \{0.89,0.97,1.17,0.63,0.51,0.51\},
\]

\[
\mathbf{P}_{ext}^{(0)} = \{1.47,0.74,0.83,1.24,0.58,1.03,0.85,1.17,
0.96,1.39,1.40,0.53,1.21,0.68,1.15\}.
\]

We want to find out how the expression levels of genes in S need to be changed so that the system moves to, or as close as possible to, a pre-specified set of expression levels for all genes. As an example, we attempt to change the equilibrium of the system to \(P_{aim} \) (see Figure 2) given by

\[
P_{aim}^{(30)} = \{\mathbf{P}_{int}^{(aim)}, \mathbf{P}_{ext}^{(aim)}\},
\]

where

\[
P_{aim}^{(int)} = \{0.8,0.6,1.3,0.7,0.6,0.6\},
\]

\[
P_{aim}^{(ext)} = \{1.6,0.8,0.9,1.3,0.5,1.1,0.9,1.2,0.9,1.2,1.5,0.4,1.3,0.6,1.1\}.
\]

Since we have computed the EES, we can calculate the projection \(P_{lin} \) of \(P_{aim} \) on \(H_S \). It is given by

\[
P_{lin}^{(30)} = \{\mathbf{P}_{lin}^{(int)}, \mathbf{P}_{lin}^{(ext)}\},
\]

where

\[
P_{lin}^{(int)} = \{0.87,0.75,1.30,0.75,0.57,0.45\},
\]

\[
P_{lin}^{(ext)} = \{1.49,0.73,0.84,1.08,0.56,1.03,0.89,1.18,
0.87,1.22,1.48,0.46,1.22,0.67,1.15\}.
\]

Finally, we use the model system Eqns. (1) and (4) to compute the external variables when internal variables are fixed at \(P_{lin}^{(30)} \). It is found to be

\[
P_{sys}^{(30)} = \{\mathbf{P}_{sys}^{(int)}, \mathbf{P}_{sys}^{(ext)}\}, \text{where } P_{sys}^{(int)} = P_{lin}^{(int)}, \text{ and } \]

\[
P_{sys}^{(int)} = \{1.48,0.75,0.85,1.11,0.58,1.00,0.91,1.21,
0.95,1.25,1.39,0.54,1.21,0.67,1.14\}.
\]

The Euclidean distances between the points are

\[
d(P_0,P_{aim}) = 0.55, \quad d(P_0,P_{lin}) = 0.40, \quad d(P_{aim},P_{lin}) = 0.38, \quad \text{ and } \]

\[
d(P_{lin},P_{sys}) = 0.15. \quad \text{Thus, we attempted to move the equilibrium from } P_0 \text{ to a point } P_{aim} \text{ that was a distance 0.55 away, but were only able to move it on } H_S \text{ to a point } P_{lin} \text{ which is a distance 0.40 away from } P_{aim}. \quad 
\]

\[
\text{However, } P_{lin} \text{ is only a distance 0.15 from the point } P_{sys}, \text{ which is the solution of the original system when expression levels of the internal variables are set to } P_{lin}^{(30)}. \quad 
\]

We have found that \(P_{lin} \) and \(P_{sys} \) are close in studies of several model systems and for many points \(P_{aim} \). Thus, the EES can be used to pre-determine, approximately, the equilibrium of the original network when changes made within S.

Transcriptional Regulatory Network in E.coli

The EES can be constructed using microarray data from the wildtype and single knockout mutants of genes in S. It can then be used to predict gene expression levels of other mutants. This observation is of interest due to the availability of previously

Figure 3. A schematic of the synthetic network. The 21 genes in the system consists of 3 groups, each with 7 genes. Genes within a cluster are coupled by interactions whose intensity is chosen randomly. Genes between clusters are weakly coupled. The "mutants" ΔGi shown in black are not viable; i.e., the corresponding set of equations do not have a solution. Genes shown in red are used to construct the effective empirical subnetwork.

Using Effective Subnetworks

PLoS ONE | www.plosone.org 4 October 2010 | Volume 5 | Issue 10 | e13080
published data on an oxygen deprivation network in E.coli [30,31]. Ref. [32] reports gene expression levels in the wildtype and in single knockout mutants of key transcriptional regulators in the oxygen response, namely ΔarcA, ΔappY, Δfnr, ΔoxyR, and ΔsoxS, as well as in the double knockout mutant ΔarcAΔfnr, in aerobic and anaerobic glucose minimal medium conditions. Since the oxygen deprivation network is not fully active under anaerobic conditions, we focus on the behavior of E.coli under anaerobic conditions.

It should be noted that gene expression levels in E.coli are unlikely to be in a steady state; rather, the expression levels reported in Ref. [32] are averages from a group of cells in various stages in the cell cycle. The analysis in this Section assumes that the computation of the EES and its predictions are valid for these averages. Preliminary results from our current work on systems exhibiting circadian rhythms validate this assumption.

We construct G as follows. In the Gene Ontology classification assigned by Affymetrix, the five genes arcA, appY, fnr, oxyR, and soxS have a common term “GO:0006355, Regulation of soxS” assigned by Affymetrix. The full list of 299 genes is given in Supporting Information S1. The proximity of the predicted and experimental values is not due to a lack of variability in the expression levels of genes in G. We verify this by computing the differential expression of each gene in the double knockout, the corresponding predictions of the EES, and the test statistic t are given in Supporting Information S1. Since the Student’s distribution associated with the comparison has two degrees of freedom, the null hypothesis is rejected when t>4.303. The histogram of the test statistic for the 299 genes is shown in Figure 4(a). In Supporting Information S1, we highlight the genes for which the null hypothesis is rejected. We emphasize that, unlike in many prior studies whose assertions are limited to whether genes in mutants are up/down regulated, our predictions are quantitative.

The proximity of the predicted and experimental values is not due to a lack of variability in the expression levels of genes in G. We verify this by computing the differential expression of genes in the mutants. Figure 4(b) shows the histogram of the largest deviations from the wildtype, normalized by the standard deviation (between replicates) in the wildtype. Expression levels of over half the genes deviate by more than 2 standard deviations.

### Discussion

An accurate model of the gene regulatory network associated with a hereditary disease can be used to compute the most effective and least detrimental treatment to prevent its onset. Unfortunately these networks contain hundreds of genes, proteins, and other molecules whose interactions are only partially known [5,8–10]. It is unlikely that detailed models of such networks will be available in the near future. The question raised in the paper is whether information needed to move the steady state of a network can be deduced from an analysis of highly simplified, empirically determined models. The data used for analysis is obtained from microarray experiments.

Our approach is as follows. We first identify a (relatively) small set S of n nodes (internal variables) in the influence network which
can be used to affect the remaining genes. (Mathematically, for each external variable $X_i$, we require one or more of $\partial F_j / \partial x_k$, where $x_k$ are the internal variables, to be non-vanishing.) Next, we limit consideration only to steady states of the network as internal variables are modified. Finally, this solution surface $S_S$ is approximated using the unique $n$-dimensional plane $\mathcal{H}_S$ defined by the gene expression levels of the wildtype and the $n$ single knockout mutants in $S$; the model system whose solutions lie on the plane is the EES.

Some genes may be critical in the sense that the organism may not be viable when they are knocked out. There were similar nodes (shown in black in Figure 3) in our synthetic model. In our approach they cannot be used as internal nodes. However, if they need to be utilized, the EES can be computed using heterozygous mutants or those where the expression level is up/down regulated to a value other than zero through, for example, transfection [33,34].

We emphasize that, due to the reduced dimensionality and its linearity, we do not expect the EES to be an accurate model of the original system. However, because of the geometrical constraints, it is possible to use the EES to (approximately) compute answers to a very limited set of questions about the system. Specifically, they are questions on gene expression levels when external changes are made within $S$. As an example, the EES can predict gene expression levels in double knockout mutants. We tested the predictions using previously published data on a double knockout mutant in an oxygen deprivation network of *E.coli*. (Here, as in most cases, the underlying network is unknown.) We identified the group of 299 genes to be studied using the Gene Ontology database. The EES was computed using the expression levels of five single knockout mutants, and used to predict their expression levels in the double mutant. The predictions were significantly different from the experimentally obtained expression levels for less than 30% of genes.

Interestingly, the EES can be used to compute how expression levels of genes within $S$ need to be changed so that the equilibrium of the entire network is moved from its initial state $P_0$ to, or as close as possible to, a pre-specified position $P_{am}$. We showed through an example that the solution computed using the EES is close to that of the full network. However, the efficacy of the move depends on the proximity of $P_{am}$ to the surface $\mathcal{H}_S$. If $P_{am}$ is far from $\mathcal{H}_S$, then the set of internal variables need to be expanded in order to find acceptable solutions.

Before concluding, we briefly address a few issues; the first is the observation that, in the parameter range considered, the model system given by Eqns. (1) and (4) have at most one stable steady state. Even though we required $P^{(0)}$ to be stable (by an appropriate choice of eigenvalues of the linearization), non-linear systems can, in general, be expected to have additional solutions. However, our model has a special feature: the signs of the partial derivatives $\partial F_j / \partial X_f$ are independent of the state of the system. The analogous biological statement is that, if nodes $J$ and $K$ are isolated, the action of node $J$ on node $K$ increases in magnitude as $X_J$ increases. Is this condition, combined with the choice of eigenvalues, sufficient to guarantee a unique stable solution? We are currently studying this question. It should be noted that the uniqueness of solutions has been proven for several other classes of monotonic nonlinear systems [35–37].

The second issue involves the partitioning of genes into clusters and the choice of internal variables. Internal variables in the oxygen deprivation network of *E.coli* were already determined from the experiments reported in Ref. [32]. We used the GO classification to identify nodes belonging to the network. Different approaches can be used to partition genes into clusters when biological classifications are not available. For example, one could use topological (e.g., persistent homology [38,39]) or graph theoretic (e.g., spectral clustering [23], community clustering [24]) methods. Integrated genomic analysis, which successfully identified subtypes of glioblastoma [40], can also be used in clustering genes through the use of heat maps [41,42]. The choice of internal variables requires biological input. Mathematically, the requirement is that each node in the cluster can be affected by suitable changes in internal variables. As we mentioned, genes that translate to transcription factors, or microRNAs [28,29] within the cluster, could act as internal nodes.

Third, can one estimate the proximity of $\mathcal{H}_S$ to the solution surface $S_S$? Differences in gene expression levels of double knockout mutants are one measure of the proximity. Alternatively, we could use the corresponding differences in heterozygous single knockouts (whose expression levels are roughly half of the wildtype) and the predictions of the EES.

We believe that approaches similar to those outlined here can prove useful in treating complex genetic diseases by helping identify optimal combinations of up/down regulation of genes (or optimal combinations of single target drugs) that have minimal

---

**Figure 4. Comparison of EES predictions with experimental gene expression levels of ΔarcAΔfnr.** (a) The histogram of the test statistic of the Student's distribution (two degrees of freedom) for the 299 genes chosen for the study. (b) The good agreement in (a) is not due to lack of variation in the gene expression levels between the wildtype and mutants. The histogram shows the largest differential expression level of mutants, normalized by the standard deviation for the wildtype (computed from the four replicates given in the data set GSE1121 of the GEO database [32]). doi:10.1371/journal.pone.0013080.g004
side effects and are most effective in moving the equilibrium of the network in its entirety to a preferred state. We hope our work motivates studies on this issue.

Methods

Construction of the EES

As illustrated in Figures 1, the EES is constructed so that, as internal variables are modified, the solutions of the system lie on the n-dimensional plane \( \mathcal{P}_n \). Thus the external variables are linear in \( x_k \)'s, and consequently, have the form given by Eqns. (2). We need to compute the coefficients \( a_{ki} \) for \( K = (n+1)(n+2), \ldots , N \) and \( j=1, \ldots , n \). This is done by noting that the expression levels \( p^{(0)} \) of each of the \( n \) mutants \( G_1, G_2, \ldots , G_n \) satisfies Eqn. (2), thus providing the conditions necessary to compute \( a_{ki} \)’s.

We note, however, that the internal variables themselves are interrelated. For example, in the single knockout mutant \( G_1 \), all expression levels (other than \( x_1 \)) are determined by solving the last (\( N-1 \)) equations of (1). Thus, we need to derive relationships between the internal variables. Consider for example, the dependence of \( x_0 \) on the remaining internal variables. In order to find its form, let us reduce the set of internal variables to \( \{x_1, x_2, \ldots , x_{n-1} \} \); \( x_n \) is now an external variable. Hence, with the approximations used in the paper, \( x_n \) is a linear combination of the remaining internal variables. Since \( p^{(0)} \) is one solution of the system

\[
x_k - p_k^{(0)} = \sum_{i=1}^{n-1} a_{ki} \left(x_i - p_i^{(0)}\right). \tag{6}
\]

Similar relationships are obtained for the other internal variables.

References

1. Wagner A (2005) Robustness and Evolvability in Living Systems. Princeton University Press, Princeton, New Jersey.
2. Bernard C (1927) An Introduction to the Study of Experimental Medicine. Macmillan, New York: Macmillan, New York.
3. Waddington C (1959) Canalization of Development and Genetic Assimilation of Acquired Characters. Nature 183: 1654–1655.
4. Waddington C (1960) Experiments on Canalizing Selection. Genetical Research 1: 140–150.
5. Hecker M, Lambeck S, Toepfer S, van Someren E, Guthke R (2009) Gene networks inference: Data integration in dynamic models-A review. Biosystems 96: 86–103.
6. Barabasi A, Oltvai Z (2004) Network biology: Understanding the Cell’s Functional Organization. Nature Reviews Genetics 5: 101–U15.
7. Ravasz E, Somera A, Mongru D, Oltvai Z, Barabasi A (2002) Hierarchical Organization of Modularity in Metabolic Networks. Science 297: 1551–1555.
8. Oltvai Z, Barabasi A (2002) Life’s Complexity Pyramid. Science 298: 763–764.
9. Kitano H (2007) Towards a Theory of Biological Robustness. Molecular Systems Biology 3: 1–7.
10. Hauser K, Abdollahi A, Huber PE (2009) Inverse system perturbations as a new methodology for identifying transcriptional signaling participants in balanced biological processes. Cell Cycle 8: 2718–2722.
11. Zimmermann GR, Lehar J, Keith CT (2007) Multi-target therapeutics: when the whole is greater than the sum of the parts. Drug Discovery Today 12: 34–42.
12. Bairoil E (2004) Flaying Dirty. Nature 437: 942–943.
13. von Luxburg U (2007) A tutorial on spectral clustering. Statistics and Computing 17: 395–416.
14. Barabasi A, Oltvai Z (2002) Hierarchical Organization of Modularity in Metabolic Networks. Science 297: 1551–1555.
15. Oltvai Z, Barabasi A (2002) Life’s Complexity Pyramid. Science 298: 763–764.
16. Fodor PA, Ananiev E, Strumpf J, Bucigrossi S, Leuckart J (1991) DNA arrays for high density hybridization to study gene expression. Science 252: 439–444.
17. Bouvier ME, Balch WE (1983) Dissection of the transcriptions network regulation of Escherichia coli. Nature Genetics 31: 64–68.

Supporting Information

Supporting Information S1 Tables showing (1) The set G of 299 genes chosen to study the oxygen deprivation network of E.coli. These genes have the common biological function 0006355 “regulation of transcription, DNA-dependent” (2) Mean values of the 299 genes in the wildtype and the mutants. (3) Standard deviation of 299 genes in the wildtype and mutants. (4) The coefficients of the Effective Empirical Network. (5) Comparison between the predicted and experimental gene expression levels for the double knockout of fur and arcA. The experimental data are normalized by the corresponding mean value of the wildtype replicates (item (2)).

Found at: doi:10.1371/journal.pone.0013080.s001 (0.20 MB XLS)

Acknowledgments

The authors would like to thank Tim Cooper, Chad Creighton, John Miller, and Gregg Roman for discussions.

Author Contributions

Conceived and designed the experiments: GHG PHG. Performed the experiments: GHG LS. Analyzed the data: LS AT. Contributed reagents/materials/analysis tools: PHG. Wrote the paper: GHG AT.
