

# Multiple Functions of a Feed-Forward-Loop Gene Circuit

Michael E. Wall<sup>1,2\*</sup>, Mary J. Dunlop<sup>3</sup>, William S. Hlavacek<sup>4</sup>

<sup>1</sup>*Computer and Computational Sciences Division and* <sup>2</sup>*Bioscience Division, Mail Stop B256, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

<sup>3</sup>*Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA*

<sup>4</sup>*Theoretical Division, Mail Stop K710, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

Keywords: *Escherichia coli*; genetic regulatory circuit; network motif; biological design principles; controlled mathematical comparison

Abbreviations used: FFL, feed-forward loop; cAMP, cyclic adenosine monophosphate; I/O, input/output; IPTG, isopropyl- $\beta$ -D-thiogalactopyranoside; TF, transcription factor

\* *Corresponding author*

Tel: +1 505 665 4209

Fax: +1 505 667 1126

E-mail address of the corresponding author: mewall@lanl.gov

## Abstract

The feed-forward loop (FFL), a network motif in genetic regulatory networks, involves two transcription factors (TFs): one regulates the expression of the second, and both TFs regulate the expression of an effector gene. Analysis of FFL design principles has been initiated, but the functional significance of the FFL is still unclear. In theoretical studies so far, the TFs are assumed to interact with different signals, which is common. However, we have found examples of FFLs in *Escherichia coli* in which both TFs interact with the same signal. These examples belong to the type 2 incoherent class of FFLs, in which each TF acts exclusively as a repressor of transcription. Here, we analyze mathematical models of this class of circuits, examining a comprehensive array of subclasses that differ in the way a signal modulates the activities of the TFs. Through parameter variation, we statistically characterize how I/O behavior and temporal responsiveness are predicted to depend on the wiring of signal interactions in a circuit. We find that circuits can exhibit any of 13 qualitatively distinct steady-state input/output (I/O) patterns, including inducible and repressible patterns. Some subclasses exhibit as many as six patterns. Transient pulses are also possible, and the response of a circuit to a signal may be either faster or slower than that of a gene circuit in which there is only one TF. Our results provide a catalog of functions for a class of FFL circuits, whose subclasses have different breadths of possible behaviors and different typical behaviors.

## Introduction

A genetic regulatory circuit, or gene circuit, is comprised of the genes and gene products involved in the cellular response to a signal, which is often a metabolite. Theoretical studies have provided insight into some of the ecological factors and performance criteria that might be important for understanding the evolution of gene circuits through natural selection, or for engineering gene circuits in the interest of practical applications. In particular, the design principles of inducible and repressible elementary gene circuits in bacteria, which involve just one transcription factor (TF), have been studied extensively.<sup>1-3</sup>

Although elementary gene circuits are important, other, more complicated, types of circuits are also found in genetic regulatory networks, and we know little about the design principles of these circuits. Alon and co-workers found that there are several types of circuits, involving more than one TF, that are more common in the genetic regulatory network of *Escherichia coli*<sup>4</sup> than might be expected. These circuits are distinguished by network motifs, which are recurring patterns of protein-DNA interactions (or rather TF activities at promoters of genes) involved in regulating gene expression. The motifs are presumed to have evolved for functional reasons, and it has been speculated that they might represent basic units of regulatory control, with defined information-processing roles.<sup>4,5</sup> One of the more prevalent motifs is the feed-forward loop (FFL), which involves two TFs.

Interestingly, among gene circuits that share the FFL motif, there is still diversity in design. For example, TFs may act with either an activator or repressor mode of control in FFL gene circuits. In recent work,<sup>5</sup> the functions of several types of FFLs were delimited theoretically, but at present, much remains to be known about the design principles of FFL gene circuits. For example, how do interactions between signals and

TFs affect the functions of FFL gene circuits? Here, our goal is to better understand how the features of a FFL gene circuit constrain its function and affect its performance, particularly temporal responsiveness to an environmental signal.

In the FFL gene circuit, two TFs (X and Y) regulate the expression of an effector gene product (Z), such as an enzyme, and one of the two TFs (X) also regulates the expression of the other TF (Y). Eight classes of FFLs have been defined by Mangan & Alon<sup>5</sup> based on the possible influences of TFs at the promoters of genes  $y$  and  $z$ , encoding proteins Y and Z. The classification scheme is based on considering a TF to have the ability to act as either a repressor or activator of transcription at the promoter of each of the genes under its control. Because of the two possible modes of TF activity (repressor or activator), the two activities of X at  $y$  and  $z$ , and the activity of Y at  $z$ , there are  $2^3=8$  classes. These classes of FFLs have been called type 1-4 coherent FFLs and type 1-4 incoherent FFLs. For reasons explained shortly, we will focus on circuits of the type 2 incoherent FFL class, in which only the repressor mode of control is used by the two TFs.

In the theoretical study of Mangan & Alon,<sup>5</sup> mathematical models for the eight classes of FFL circuits were analyzed. For each class, the qualitative steady-state input/output (I/O) behavior was studied and characterized. The inputs considered were step increases and decreases in the level of a signal  $S_X$  in the presence of either a high or low fixed level of a signal  $S_Y$ . The signal  $S_X$  interacts with X, influencing the activities of X at  $y$  and  $z$ , and the signal  $S_Y$  interacts with Y, influencing the activity of Y at  $z$ . The output considered was the level of effector protein Z. Temporal responsiveness to a signal was also evaluated by comparing a circuit of each class with a reduced form of the circuit in which Y is absent. The results of this study provide design principles that perhaps can be used to guide the construction of synthetic gene circuits. For example, the coherent FFLs were predicted to function as sign-sensitive delays (or persistence

detectors): these circuits have a delayed response to a step change in the concentration of  $S_X$  in one direction (*e.g.*, an increase) but not in the opposite direction (a decrease). The incoherent FFLs, including the type 2 incoherent FFLs that we analyze below, were predicted to function as sign-sensitive accelerators.

The mathematical models of FFLs analyzed so far have encompassed all possible modes of control of X and Y at y and z. However, not all possible effects of the signals  $S_X$  and  $S_Y$  on these activities have been considered. A signal that interacts with a TF can negatively or positively modulate the activity of the TF. The signal may act as an inducer of gene expression by blocking repression or enabling activation. Conversely, the signal may act as a co-repressor of gene expression by enabling repression or by blocking activation. In addition, the possibility of a single signal that affects the activities of both X and Y has not yet been considered, which is significant because we have found examples of FFLs in *E. coli* that involve this type of signaling.

Alon and co-workers compiled a list of 42 FFLs in *E. coli*.<sup>4</sup> They did not document the effects of signals on TFs, but in many of the FFLs, the TFs clearly recognize different signals, as in the mathematical models considered by Mangan & Alon.<sup>5</sup> An example is the type 1 coherent FFL circuit in which the TFs CRP and AraC activate expression of the *araBAD* operon; as is well known, CRP and AraC interact with different metabolites, cyclic adenosine monophosphate (cAMP) and arabinose, respectively.<sup>6,7</sup> To better determine how signals influence the TFs in FFLs, we surveyed information available about TF-signal interactions in *E. coli*, a large amount of which has been cataloged in electronic databases: RegulonDB,<sup>8</sup> EcoCyc,<sup>9</sup> and EcoTFs.<sup>3</sup> We were able to identify known or putative signals (or factors that influence activity) for most of the TFs in the 42 FFLs (<http://EcoTFs.lanl.gov/FFLs.html>): in all cases for which a signal is known for each of the TFs in a FFL, there are two signals, one for each TF.

In the course of our survey of TF-signal interactions, however, we identified new FFLs in which the TFs interact not with different signals but with the same signal. We found two pairs of TFs that recognize a common signal and comprise a FFL circuit: GalR (=X) and GalS (=Y), which both recognize the inducer galactose and both repress expression of *galETKM* (=z);<sup>10</sup> and ExuR (=X) and UxuR (=Y), which both recognize the inducer fructuronate and both repress expression of *uxuAB* (=z).<sup>11-13</sup> Consistent with the FFL motif, GalR regulates expression of *galS*,<sup>14</sup> and ExuR regulates expression of *uxuR*.<sup>15</sup> In both of these cases, regulation is exerted through the repressor mode of control. Thus, these newly identified FFLs (*galR-galS-galETKM* and *exuR-uxuR-uxuAB*) belong to the type 2 incoherent FFL class, for which no members had previously been identified. We also found another member of this class: *gntRKU-idnDOTR-gntKU*.<sup>16-19</sup> In this case, as is typical of FFLs, GntR is regulated by gluconate,<sup>20</sup> whereas IdnR (also called GntH) is regulated by a different signal, idonate and/or 5-ketogluconate.<sup>21,22</sup> However, because gluconate and 5-ketogluconate can be interconverted through the enzymatic activity of the *idnO* gene product,<sup>21</sup> the different signals affecting GntR and IdnR may well be correlated, which is a possibility for a number of other FFLs (see <http://EcoTFs.lanl.gov/FFLs.html>).

Because of the above findings, we were motivated to perform a systematic investigation of signal interactions in type 2 incoherent FFL circuits. Here, we present a study in which we analyze mathematical models of this class of circuits which encompass all combinations of enabling and blocking effects of a signal (Fig. 1).

## Model

The models of type 2 incoherent FFLs that we analyze here are illustrated in Fig. 1. Let  $X$  and  $Y$  be the levels of TFs  $X$  and  $Y$ ,  $Z$  be the level of an effector protein  $Z$ , and  $S$  be the level of a signal  $S$ . As in previous work,<sup>5</sup> we assume that  $X$  is unaffected by  $S$  or  $Y$ ,

and we treat  $S$  as an independent variable. The latter assumption is equivalent to considering the response of a gene circuit to a gratuitous inducer, such as the response of the *lac* circuit in *E. coli* to isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG). The models assume that protein concentrations are homogeneous, stochastic effects are negligible, and the effect of a signal is “all-or-none,” determined by comparison to a threshold, as described below. We also neglect TF autoregulation, as in previous work.<sup>5</sup> The general model for a type 2 incoherent FFL that we consider in our analysis is

$$\begin{aligned} dY/dt &= B_Y + \alpha_Y h_{YX}^r(s_{YX}, X, S) - \beta_Y Y \\ dZ/dt &= B_Z + \alpha_Z h_{ZX}^r(s_{ZX}, X, S) h_{ZY}^r(s_{ZY}, Y, S) - \beta_Z Z \end{aligned} \quad (1)$$

$B_i$ ,  $\alpha_i$ , and  $\beta_i$  are rate constants, and each  $h_{ij}^r$  is a Hill function that models repression of protein  $i \in [Y, Z]$  by TF  $j \in [X, Y]$ . The constants in Eq. 1 represent the steady unregulated rate of protein expression ( $B_i$ ), the maximal change in the rate of expression through regulation ( $\alpha_i$ ), and the rate of protein decay ( $\beta_i$ ), which assumes first-order degradation and/or dilution due to exponential growth. The Hill function  $h_{ij}^r$  has the following form:

$$h_{ij}^r(s_{ij}, u_j, S) = \frac{K_{ij}^{n_{ij}}}{K_{ij}^{n_{ij}} + [s_{ij}(S)u_j]^{n_{ij}}}, \quad (2)$$

where  $K_{ij}$  and  $n_{ij}$  are positive real-valued constants,  $s_{ij}$  is a function that has a value of 0 or 1, and  $u_j$  is either  $X$  ( $j=X$ ) or  $Y$  ( $j=Y$ ). The Hill constant  $K_{ij}$  gives the value of  $u_j$  at which  $h_{ij}^r = 1/2$ ; it can be related to the affinity of TF  $j$  for the DNA binding sites from which it regulates transcription of the gene encoding protein  $i$ . The Hill number  $n_{ij}$  determines the sensitivity of  $h_{ij}^r$  to changes in  $u_j$ ; it can be related to the cooperativity of TF  $j$  binding to DNA (*e.g.*,  $n_{ij} = 2$  if TF  $j$  binds DNA as a dimer). The signal interactions of the FFL determine the function  $s_{ij}$ , which describes in detail how  $S$  modulates the regulation of expression of protein  $i$  by TF  $j$ . The value of  $s_{ij}$  is taken to be either 0 or 1

based on comparison of  $S$  and a constant  $T_j$ , which we interpret as a threshold value of  $S$  related to the affinity of  $S$  for TF  $j$ . Above  $T_j$ ,  $S$  affects regulation of  $i$  by  $j$ ; below this threshold, the signal has no effect. The functional form of  $s_{ij}$  depends on the mode of the signal interaction: if binding of signal enables regulation, then  $s_{ij}(S) = \theta(S - T_j)$ , where  $\theta$  is the unit step function ( $\theta(v) = 1, v \geq 0$ ;  $\theta(v) = 0, v < 0$ ); if binding of signal blocks regulation, then  $s_{ij}(S) = \theta(T_j - S)$ . If regulation of  $i$  by  $j$  does not depend on signal then  $s_{ij} = 1$  regardless of the value of  $S$ .

The main difference between Eq. 1 and the model considered by Mangan & Alon<sup>5</sup> is the introduction of the functions  $s_{YX}(S)$ ,  $s_{ZX}(S)$ , and  $s_{ZY}(S)$ . These functions involve two thresholds,  $T_X$  and  $T_Y$ . We will consider different relative values for these thresholds:  $T_X > T_Y$ , and  $T_X < T_Y$ . We are motivated to consider cases where  $T_X \neq T_Y$ , in part because of the example of the *galR-galS-galETKM* circuit,<sup>10</sup> in which regulation by GalS is affected by lower concentrations of the inducer, galactose, than is regulation by GalR.<sup>23</sup> These thresholds define three qualitative levels of signal: low [ $S < \min(T_X, T_Y)$ ], intermediate [ $\min(T_X, T_Y) < S < \max(T_X, T_Y)$ ], and high [ $S > \max(T_X, T_Y)$ ]. We will only consider circuits with  $T_X \neq T_Y$ ; the response of a circuit to a change from a low to high level of signal (and *vice versa*) is the same as that of a circuit with  $T_X = T_Y$  that is otherwise the same. We will also consider all possible combinations of modes of signal interactions. Each combination defines a subclass of type 2 incoherent FFL circuits. We use a three-symbol code to refer to each subclass: the first symbol indicates whether signal enables (+), blocks (-) or has no influence (0) on regulation of  $Z$  by  $X$ . The second symbol indicates the mode of the signal in regulation of  $Y$  by  $X$ , and the third regulation of  $Z$  by  $Y$ . Thus, the triplet (+,-,0) should be understood to refer to a type of circuit in which the signal  $S$  enables repression of  $Z$  by  $X$ , blocks repression of  $Y$  by  $X$ , and has no influence on repression of  $Z$  by  $Y$ .

## Results

In the following sections, we analyze the various subclasses of type 2 incoherent FFL gene circuits and characterize their functions. Because its function has been the subject of a previous theoretical study,<sup>5</sup> we begin by analyzing the type-(+,+,0) circuit, characterizing its steady-state behavior and dynamic response to signal. We then characterize the behaviors of each subclass of circuits.

### Analysis of the (+,+,0) subclass

For this subclass, signal activates the  $X \rightarrow Z$  and  $X \rightarrow Y$  regulatory interactions, and does not influence the  $Y \rightarrow Z$  regulatory interaction (Fig. 1(e)). Mangan & Alon<sup>5</sup> concluded that the type-(+,+,0) circuit should have a repressible steady-state behavior, a derepression rise time faster than that of an equivalent circuit without the  $X \rightarrow Y$  regulatory interaction, and a repression decay time equivalent to that of the alternative circuit without Y.

We examined the response of type-(+,+,0) circuits to a signal for a wide range of parameter combinations (Methods). Representative time courses are illustrated in Fig. 2. The steady-state behavior depends on parameter choice: the majority of the circuits (50%) exhibit an unresponsive behavior (Table 1). Of the rest, most are repressible (37%), but a significant minority (13%) are inducible. We compared the temporal responsiveness of the repressible type-(+,+,0) circuit to that of the circuit without the  $X \rightarrow Y$  regulatory interaction, which is equivalent to an elementary circuit with just one TF (Table 2). As measured by the rise time  $t_r$ , derepression of the type-(+,+,0) circuit was either faster than or similar to that of the elementary circuit. However, we found that systems with a short rise time may exhibit significant overshoot, and have a larger settling time  $t_s$ . Comparing  $t_s$  instead of  $t_r$ , the results are not clear-cut. For  $K_{ZY} \geq 1$ ,  $t_s$  of

the type-(+,+,0) circuit is mostly either similar or smaller; for  $K_{ZY} < 1$ , however, the fraction of type-(+,+,0) circuits with a significantly larger value of  $t_s$  (42%) is much greater than that with a significantly smaller value of  $t_s$  (12%). We conclude, consistent with earlier results, that the derepression rise time for a type-(+,+,0) circuit is less than or equal to that for an equivalent elementary circuit, but significant overshoot is possible, which may be detrimental.

Comparing repression decay times, some circuits have an accelerated decay time, as shown in Fig. 2, and all have a decay time that is at least as fast as that of a circuit without the  $X \rightarrow Y$  regulatory interaction (Table 2). Considering repression settling times, some type-(+,+,0) circuits are accelerated at the expense of overshoot; however, most accelerated circuits exhibit a faster response without overshoot. We conclude that the type-(+,+,0) circuit can exhibit faster repression than the circuit without the  $X \rightarrow Y$  regulatory interaction. This conclusion disagrees with an earlier study, in which the two circuits were only found to exhibit the same repression decay times.<sup>5</sup> We found that repression decay times were similar for many, but not all, of the sampled parameter values.

## **I/O behaviors of type 2 incoherent FFLs**

To better understand the relation between the structure and function of FFLs, we characterized the behaviors of subclasses of the type 2 incoherent FFL with all possible signal interactions. In each subclass, the signal may have one of three effects on each of three genetic regulatory interactions (+, -, or 0), leading to 27 different subclasses. Type-(0,0,0) circuits have a Null response, and the responses of type-(0,0,-) and type-(0,0,+) circuits to signal are like those of elementary circuits; we do not analyze these subclasses here. Because of symmetry in the model, of the remaining 24 subclasses, only half need to be considered in detail — we focus on the 12 subclasses illustrated in

Fig. 1(a)-(d). For example, the response of the type-(-,-, 0) circuit (Fig. 1(a)) to increasing signal is the same as the response of the type-(+,+, 0) circuit (Fig. 1(e)) to decreasing signal. Cases in which  $T_X < T_Y$  or  $T_X > T_Y$  were treated separately. We subjected each circuit to the same input signal (Fig. 3, inset) and classified the steady-state I/O behavior for 681,472 parameter combinations (Methods). The resulting behavior distributions are summarized in Table 1 and are discussed in the following sections, grouped according to the signs of the first two signal interactions: ([effect on  $X \rightarrow Z$ ], [effect on  $X \rightarrow Y$ ], [effect on  $Y \rightarrow Z$ ]). We use a wildcard “\*” to indicate all possible effects of a signal on a genetic regulatory interaction.

**Type-(-,-,\*) circuits.** As a group, these three subclasses exhibit all 13 types of steady-state I/O behaviors (Table 1). One example of this type of circuit in *E. coli* is the *gntRKU-idnDOTR-gntKU* system, in which repression of both *idnDOTR* and *gntKU* by GntR is blocked by gluconate.<sup>18,20</sup> Gluconate does not affect IdnR repression of *gntKU*. Thus, we classify *gntRKU-idnDOTR-gntKU* as a type (-,-,0) circuit. Another example in *E. coli* is the *exuR-uxuR-uxuAB* system, in which repression of *uxuAB* by both ExuR and UxuR is blocked by fructuronate.<sup>11-13</sup> We assume that fructuronate also blocks the weak ExuR-mediated repression of *uxuR*,<sup>15</sup> although the effect of fructuronate has not been studied. Thus we classify *exuR-uxuR-uxuAB* as a type (-,-,-) circuit. Some temporal responses of this subclass are illustrated in Fig. 3. 80% of the type(-,-,-) circuits are inducible; none are repressible (Table 1). 37% of the type(-,-,0) circuits are inducible, but 13% are repressible. The type(-,-,+) circuits are split evenly between inducible (33%) and repressible (33%) circuits.

**Type-(-,0,\*) circuits.** These circuits exhibit every I/O behavior except HML, and only rarely exhibit HLH (Table 1). The response of the type(-,0,0) circuit to signal is equivalent to that of an elementary gene circuit, in which there is only  $X \rightarrow Z$  regulation; the type(-,0,+) and type(-,0,-) circuits are equivalent to coincidence circuits, in

which there is no cascade. An example of this type of circuit in *E. coli* is the *galR-galS-galETKM* system, in which repression of *galETKM* by both GalR and GalS is blocked by galactose, but in which galactose has no effect on GalR repression of *galS*.<sup>14</sup> We classify *galR-galS-galETKM* as a type  $(-,0,-)$  circuit. Representative responses are illustrated in Fig. 4. 79% of type- $(-,0,-)$  circuits are inducible, and 46% of type- $(-,0,0)$  circuits are inducible; neither subclass exhibits a repressible behavior (Table 1). These results contrast with those for type- $(-,0,+)$  circuits, of which 41% are inducible, and 27% are repressible.

**Type- $(-,+,* )$  circuits.** Like the type- $(-,0,* )$  circuits, these circuits also exhibit every I/O behavior except HML, and only rarely exhibit HLH (Table 1). Overall, these circuits have behaviors that are very similar to those of the corresponding type- $(-,0,* )$  circuits (Table 1). Representative responses are illustrated in Fig. 5.

**Type- $(0,+,* )$  circuits.** These circuits exhibit only the Null, HLH, HHL, HLL, HLM, LLH, LHH, and LMH behaviors (Table 1). The response of this type of circuit to signal is equivalent to that of a two-step cascade, in which there is no feed-forward genetic regulatory interaction. Representative responses are illustrated in Fig. 6. 39% of type- $(0,+,-)$  circuits are inducible, and 16% of type- $(0+,0)$  circuits are inducible; neither subclass exhibits a repressible behavior (Table 1). 33% of type- $(0+,+)$  circuits are repressible, and none are inducible. In contrast to prior assumptions about two-step cascades,<sup>5</sup> type- $(0+,+)$  circuits are capable of generating pulses (Fig. 7).

**Clustering of type 2 incoherent FFL behaviors.** A total of 10,080 time courses in response to a two-level signal input were clustered into 15 groups (Methods), 12 of which had 10 or more time courses. These 12 clusters show inducible and repressible I/O behaviors, with examples of overshoot and pulses (Fig. 8). Given the earlier finding that increases in gene expression are accelerated for the type 2 incoherent FFL,<sup>5</sup> we

were surprised to find a cluster that appeared to have a delayed induction without a delayed de-induction (see cluster (D) in Fig. 8). Because this cluster is primarily associated with the type- $(-,+,0)$  circuits (Supplemental Table S1), we analyzed the temporal responsiveness of this subclass in the same manner as for the type- $(+,+,0)$  circuits. This analysis confirmed the initial observation: induction rise times and settling times of the  $(-,+,0)$  subclass were either the same as or slower than those for equivalent circuits without the  $X \rightarrow Y$  regulatory interaction, and de-induction decay times were the same as or slower than those for equivalent circuits (Supplemental Table S2).

A total of 20,160 time courses in response to a three-level signal input were automatically clustered into 30 groups (Methods), 20 of which had 10 or more time courses. These 20 clusters exhibit all of the steady-state I/O behaviors except LHL and HLH (Fig. 9). The distributions of time courses among the 20 clusters are shown in Supplemental Table S3. Similar to the results found for a two-level input signal, type- $(-,+,0)$  circuits can also have a delayed induction in response to the three-level input signal (see cluster (I) in Fig. 9, and Supplemental Table S3,  $T_X > T_Y$ ).

### **Robustness of type 2 incoherent FFL behaviors**

To quantify the robustness of the behaviors of different subclasses of the type 2 incoherent FFL, the Shannon entropies of the steady-state I/O behavior distributions were calculated (Methods). Results are summarized in the last column of Table 1. The lower the number, the more predictable is the behavior type for a subclass assuming the parameter-sampling statistics of this study, and the more robust is the qualitative behavior of the subclass to parameter changes. The most robust subclass is  $(0,+,0)$ , with an entropy of 0.62. The least robust subclass is  $(-,-,+)$ , with an entropy of 2.56 for cases where  $T_X < T_Y$ , and 2.31 for cases where  $T_X > T_Y$ . Results for the clustered time courses are similar: the Pearson correlation between the entropies of the steady-state

behaviors (Table 1) and the entropies of the three-level clustered time courses (Supplemental Table S3) is 0.76 for  $T_X < T_Y$ , and 0.67 for  $T_X > T_Y$ , which indicates that both classification schemes yield similar categorizations.

## Discussion

Our results provide a catalog of the repertoire of functional capabilities of a simple type of gene circuit. These results may be useful for the construction of synthetic gene circuits with desired steady-state I/O behavior and dynamic properties. More importantly, we have shown how diversity in signal interactions expands the range of possible behaviors. Clearly, a documentation of TF-signal interactions is just as critical to understanding a genetic regulatory network as TF-DNA interactions.

An examination of the type 2 incoherent FFL reveals a more complex picture of function than was found in an earlier mathematical modeling study of FFLs.<sup>5</sup> Instead of being purely repressible, type-(+,+,0) circuits may be either inducible or repressible, depending on parameter values. Instead of only exhibiting an accelerated increase in expression of Z, different subclasses of type 2 incoherent FFL circuits may exhibit many different behaviors, including a delayed increase in expression of Z. Because of the effect of signal interactions on FFL behavior, knowledge of TF activities at the promoters of genes is insufficient to characterize the functions of FFL gene circuits.

In the absence of specific model parameters, some subclasses of circuits examined have a function that is more predictable than other types. For example, subclasses in which signal inhibits regulation of Z by Y may be inducible but not repressible. Similarly, most subclasses in which signal does not influence regulation of Z by Y — all but type-(-,-,0) circuits, which may be inducible or repressible — are non-repressible. Subclasses in which signal enables regulation of Z by Y tend to have less

predictable functions: type-(-,-,+) circuits were equally likely to be inducible or repressible, and all others except the type-(0,+,+) circuits were found to have a majority of inducible systems, but a significant number of repressible systems. Type-(0,+,+) circuits are predictably repressible. Thus, the qualitative steady-state I/O behavior of the type 2 incoherent FFL circuit is robust to parameter changes when signal either blocks or does not influence regulation of Z by Y, but is sensitive to parameter changes when signal enables regulation of Z by Y.

The temporal responsiveness of FFL circuits with faster rise times has been described as accelerated,<sup>5</sup> but we found that such circuits can exhibit significant overshoot for a wide range of parameters, leading to a longer settling time. Comparisons of the temporal responsiveness thus depend on whether one is considering the rise time or settling time: systems with accelerated rise times may have longer settling times than equivalent reference systems, depending on the degree of overshoot in the response. Because de-induction may exhibit undershoot (Fig. 2), similar arguments hold for interpretation of accelerated decay times.

In a large number of subclasses that represent half a comprehensive set of possible signal interactions for the type 2 incoherent FFL, most of the subclasses were found to exhibit induction upon a change from low to high signal levels. For the other half of the set, in which all signal effects are inverted, most of the subclasses exhibit repression by signal. Again, because of symmetry, among all subclasses considered in our analysis, half of them are inducible, and half of them are repressible.

Allowing for a different affinity of signal for X and Y enables steady-state I/O behaviors with intermediate responses to signal. Depending upon the signal interactions and the relative affinities, the intermediate response to signal may change. Circuits in

which signal does not influence regulation of Z by Y do not exhibit an intermediate response to signal, because signal does not affect Y.

Although the *galR-galS-galETKM* circuit has a type 2 incoherent FFL pattern of TF-DNA interactions, the interactions with galactose are those of the  $(-,0,-)$  subclass, and the response to galactose is that of a coincidence circuit and not a FFL. (Recall that galactose affects the activities of both GalR and GalS at *galETKM*, but does not affect the activity of GalR at *galS*.<sup>10,14</sup>) Thus, a circuit with the requisite TF-DNA interactions of the FFL might behave as a simpler type of circuit. Signal interactions can reduce the complexity of FFL circuits: type- $(+/-,0,+/-)$  circuits respond as coincidence circuits; type- $(0,+/-,*)$  circuits respond as two-step cascades; type- $(+/-,0,0)$  and type- $(0,0,+/-)$  circuits respond as elementary circuits; and type- $(0,0,0)$  circuits do not respond to signal. In this list,  $+/-$  is a wildcard for either + or -, and \* is a wildcard for either +, -, or 0. Only in type- $(+/-,+/-,*)$  circuits do all TF-DNA interactions of the FFL contribute to the response to a signal.

Our results enable predictions of the behaviors of type 2 incoherent FFL gene circuits. As a member of the  $(-,0,-)$  subclass, the *galR-galS-galETKM* circuit is predicted to be capable of LLH, LHH, and LMH steady-state I/O behaviors (which behaviors are all inducible) in response to galactose (Table 1). The temporal response is predicted to be that of a coincidence circuit, which, because the TFs act independently, is similar to that of an elementary circuit. Because the activities of GalR and GalS have different sensitivities to galactose, they might act together to increase the range of galactose levels over which the circuit may respond to galactose. Further studies of the *galR-galS-galETKM* circuit might provide clues about the functional significance of GalR repression of *galS*, which contributes to the classification of this circuit as a FFL. The *exuR-uxuR-uxuAB* circuit, which we classify as a member of the  $(-,-,-)$  subclass, is predicted to be capable of the LLH, LLH, LMH, HLH, and MLH behaviors. The

*gntR**KU-idnDOTR-gntK**U* circuit, which we classify as a member of the  $(-, -, 0)$  subclass, is predicted to be capable of the HLL or LHH behaviors. Both the *exuR-uxuR-uxuAB* and *gntR**KU-idnODTR-gntK**U* circuits are predicted to be capable of accelerated induction (Fig. 3).

Our study of the functions of type 2 incoherent FFL circuits complement earlier studies of elementary gene circuits, in which gene-circuit performance was found to be influenced by how the signal controls TF self-regulation,<sup>1,2</sup> or by the position of the signal among substrates, intermediates, or products in the metabolic pathway catalyzed by the regulated effector gene products.<sup>24</sup> Although here our identification of two systems in *E. coli* in which the same signal interacts with two TFs has led us to study the class of FFLs with all negative genetic regulatory interactions, examples of gene circuits from the other seven FFL classes are known (<http://EcoTFs.lanl.gov/FFLs.html>). It might be important to consider signal interactions of the types examined here for the other FFL classes (for which two signals are generally involved), because even though the signals that interact with each TF are different, their levels might be correlated through metabolism, making them similar to gene circuits in which a signal interacts with both TFs.

Our analysis of the functions of FFLs with a variety of signal interactions adds support the perspective that, to understand the genetic regulatory network of an organism, it is necessary to place equal weight on documentation of both genome-wide genetic regulatory interactions and genome-wide signal interactions.<sup>25,26</sup> Based on this perspective, the EcoTFs database of *E. coli* TFs and signals was recently developed,<sup>3</sup> and, independently, signal interactions were incorporated into the well-known RegulonDB database.<sup>8,25</sup> Important features of the EcoTFs database, available at (<http://EcoTFs.lanl.gov>), include annotation of the signal(s) influencing the activities of each TF, and classification of each TF based on (1) the response to a

change in a signal (induction or repression of regulated effector genes), (2) the mode of regulation at the promoters of regulated effector genes (repressor or activator control), and (3) the co-regulation of TF and effector gene products in response to signals. The most recent version of RegulonDB (version 4.0) provides similar information, but without documenting the response to a change in signal (upregulation and downregulation of genes are documented for many growth conditions, but this information is not linked to specific TFs and signals).<sup>8</sup>

The need to accurately document genetic regulatory interactions and signal interactions has become more urgent because computational and experimental methods have accelerated the rate of discovery of such interactions. Computational methods for genome-wide prediction of TF binding sites are being developed, and similar methods must be developed for systematic prediction of interactions between signals and TFs.<sup>26</sup> In addition to sequence-based approaches, predictions of signal-TF interactions might include methods of structural biology, such as homology modeling, searching for ligand-binding structural motifs, and molecular docking simulations.

## Methods

### Parameter values

For systematic characterization of the behaviors of the models, a reference set of parameter values was defined as follows:  $B_Y = B_Z = 0.1$ ;  $\alpha_Y = \alpha_Z = 0.9$ ;  $\beta_Y = \beta_Z = 1$ ;  $K_{YX} = 1$ ,  $K_{ZX} = 1$ ,  $K_{ZY} = 1$ ; and  $n_{YX} = n_{ZX} = n_{ZY} = 2$ . The value of the independent variable  $X$  was taken to be 1. These values are the same as those used by Mangan & Alon,<sup>5</sup> except  $B_Y = B_Z = 0$  and  $\alpha_Y = \alpha_Z = 1$  in their study. We modified  $B_Y$  and  $B_Z$  to consider realistic cases in which there is a nonzero unregulated rate of protein expression, and modified  $\alpha_Y$  and  $\alpha_Z$  to ensure that the maximal protein expression rates  $B_Y + \alpha_Y$  and  $B_Z + \alpha_Z$  were

the same as those in Mangan & Alon – the qualitative results of the present study are the same either with or without these modifications. As in their study, these reference parameter values were selected in the interest of exploring model behaviors rather than in the interest of understanding the dynamics of specific natural systems. We characterized the behaviors of circuits for a large number of parameter value combinations by varying  $n_{YX}$ ,  $n_{ZX}$ , and  $n_{ZY}$  systematically, and randomly sampling values of  $K_{YX}$ ,  $K_{ZX}$ , and  $K_{ZY}$ . To sample parameters for the temporal responsiveness studies and for the classification of steady-state I/O behaviors, we consider each possible combination of integer values for  $n_{YX}$ ,  $n_{ZX}$ , and  $n_{ZY}$ , where  $n_{ij} \in [1,2,3,4,5,6,7,8]$  ( $8^3 = 512$  combinations). For each combination  $(n_{YX}, n_{ZX}, n_{ZY})$ , 1331 combinations of values for  $K_{YX}$ ,  $K_{ZX}$ , and  $K_{ZY}$  were generated by random log-uniform sampling of each  $K_{ij}$  in the range  $[0.01,100]$ . A comparison of temporal responsiveness was performed for each combination of parameter values,  $512 \times 1331 = 681,472$  comparisons in total. To sample parameters for the clustering of time courses, we systematically sampled values of  $n_{ij}$  from the set  $n_{ij} \in [2, 4, 6, 8]$ . Odd numbered values of  $n_{ij}$  yielded results that are qualitatively similar (unpublished material). For each combination  $(n_{YX}, n_{ZX}, n_{ZY})$ , 210 combinations of  $K_{ij}$  values were generated by random log-uniform sampling in the range  $[0.01, 100]$ . A total of 840 time courses with different parameter combinations were sampled for each circuit.

### **Comparisons of temporal responsiveness**

We compared the temporal responsiveness of type-(+,+,0) and type-(-,+,0) FFLs with that of circuits lacking the second TF, Y. The input signal for type-(+,+,0) circuits was a two-level high-low-high pattern with the time dependence illustrated in the inset in Fig. 2; the input signal for type-(-,+,0) circuits had the opposite pattern — low-high-low — with the same time dependence. The comparisons were performed using the method of controlled mathematical comparison, which involves constraints of internal and external

equivalence.<sup>27,28</sup> Internal equivalence was enforced by requiring all parameters that appear in both of any two models being compared, except  $B_Z$  and  $\alpha_Z$ , to be the same. These conditions guarantee that all processes involving production of  $Y$  are the same for alternative circuits. They also guarantee that the combined effects of degradation and dilution for  $Z$  are the same for alternative circuits. External equivalence was enforced by requiring the minimal and maximal steady-state values of  $Z$  in an alternative system to be the same as those in the reference system. The steady-state values of  $Y$  and  $Z$ , which we refer to as  $Y_\infty$  and  $Z_\infty$ , are determined analytically from Eq. 1 by setting  $dY/dt = dZ/dt = 0$ .

To satisfy external equivalence for alternative systems that are capable of the same I/O behavior, we determine the values of  $\alpha_Z$  and  $B_Z$  by using the following equations:

$$\begin{aligned} \alpha_Z &= \alpha'_Z \left( \frac{R'_{Z,\max} - R'_{Z,\min}}{R_{Z,\max} - R_{Z,\min}} \right), \\ B_Z &= B'_Z + \alpha'_Z R'_{Z,\min} - \alpha_Z R_{Z,\min} \end{aligned} \quad (3)$$

where  $R_Z = h_{ZX}^r(s_{ZX}(S)X/K_{ZX})h_{ZY}^r(s_{ZY}(S)Y_\infty/K_{ZY})$ . The primed quantities refer to the reference system, and the labels “min” and “max” indicate evaluation at levels of signal  $S$  at which  $Z_\infty$  is minimal and maximal, respectively.

Equations were integrated numerically to obtain system dynamics and summary measures of temporal responsiveness: decay times ( $t_d$ ), rise times ( $t_r$ ), and settling times ( $t_s$ ). The values of  $t_d$  and  $t_r$  are obtained by calculating the time required for the value of  $Z$  to first come within  $\Delta_Z$  of its final steady-state value after a change in signal, where  $\Delta_Z = 5\%$  of the difference between the steady-state levels before and after the change in signal. The value of  $t_s$  is obtained by calculating the time required for the value of  $Z$  to settle within  $\Delta_Z$  of its final steady-state value after a change in signal. The settling time  $t_s$  is greater than or equal to  $t_r$  (or  $t_d$ ), being equal only when overshoot (or undershoot)

is less than  $\Delta_z$ . For each model considered, comparisons were performed for 681,472 parameter combinations (Parameter values).

## Behavior distributions

To determine whether the steady-state behavior of a model is inducible or repressible, we calculate the value of  $Z_\infty$  at both low [ $S < \min(T_X, T_Y)$ ] and high [ $S > \max(T_X, T_Y)$ ] signal levels. The system is termed inducible (repressible) if  $Z_\infty$  increases (decreases) when the signal changes from low to high; alternatively,  $Z_\infty$  may be unresponsive to such a change in signal. We additionally calculate  $Z_\infty$  when the signal level is intermediate [ $\min(T_X, T_Y) < S < \max(T_X, T_Y)$ ]. A three-letter code was used to classify the behavior according to the relative values of  $Z_\infty$  at low, intermediate, and high signal levels: the first letter is L, M, or H to indicate whether the value of  $Z_\infty$  at low signal is low, medium, or high, respectively; the second and third letters similarly indicate the values of  $Z_\infty$  at intermediate and high signal. For example, the code HML is a repressible system for which the value of  $Z_\infty$  is high at low signal, medium at intermediate signal, and low at high signal. There are  $3^3 = 27$  possible 3-letter codes, but there are just 13 distinguishable types: HLH, LHL, HHL, HLL, HML, HLM, MHL, LLH, MLH, LHH, LHM, LMH, and Null. Some circuits, which we term pulsers, are capable of exhibiting a dynamic response without a change in steady state. If the value of  $Z_\infty$  does not change significantly for different levels of signal, either because the circuit exhibits only transient behavior or because it is simply insensitive to changes in the level of a signal, the response is assigned the code Null. We consider a change to be insignificant if it is less than 5% of the minimum value of  $Z_\infty$ . For each type of circuit,

we calculated behavior distributions that indicate how I/O classification depends on parameter values. A distribution was obtained by calculating the fraction  $f_k$  of 681,472 distinct parameter combinations sampled (Parameter values) that yield each type of behavior  $k$  (e.g.,  $f_{\text{HML}}$  is the fraction of parameter combinations that yield the HML steady-state I/O behavior). Automatic clustering was used to obtain similar behavior distributions, based on the dynamic response of a circuit to signal (below).

### **Automatic clustering**

To obtain function classes that are determined by dynamics, we used a clustering algorithm, with time courses like those illustrated in Figs. 2-6 as input data.<sup>29</sup> To obtain time courses, two types of signal input were applied to the 12 circuits in Figs. 1(a)-1(d). Parameter values were sampled as described above. For each circuit type, two cases were considered:  $T_X < T_Y$  and  $T_X > T_Y$ . For the first type of signal input, the signal level  $S$  switches sharply between two levels: a low [ $S < \min(T_X, T_Y)$ ] and high [ $S > \max(T_X, T_Y)$ ] value. This type of signal input had a low-high-low time dependence, which is the opposite of that illustrated in the inset of Fig. 2. For each of the 12 circuits, 840 time courses were generated (Parameter values), for a total of 10,080 time courses, which were organized into 15 clusters. Because the distinction between  $T_X < T_Y$  and  $T_X > T_Y$  does not influence the response to such a two-step input, it was not necessary to consider these cases separately. In another type of signal input, illustrated as insets in Figs. 3-6, the signal level  $S$  assumes one of three levels: a low [ $S < \min(T_X, T_Y)$ ], intermediate [ $\min(T_X, T_Y) < S < \max(T_X, T_Y)$ ], or high [ $S > \max(T_X, T_Y)$ ] value. For the three-level-input cases, two sets of 10,080 time courses were obtained, one each for  $T_X < T_Y$  and  $T_X > T_Y$ , for a total of 20,160 time courses, which were organized into 30

clusters. For both the two-level input and three-level input cases, all time courses within a cluster were similar, and different clusters exhibited different temporal response patterns. In each analysis, the number of clusters chosen captured a wide range of dynamic behaviors.

**Clustering algorithm.** A greedy approximation algorithm<sup>30</sup> is used to cluster time courses of the value of  $Z$  in response to changes in the level of signal, which we call a functional response. The distance used for clustering is defined as follows. Let each response be a vector  $\underline{z}_i$  of length  $N$ , which contains the values of  $Z$  running from  $t = 0$  to  $t = N - 1$ . The distance function between two vectors,  $\underline{z}_1$  and  $\underline{z}_2$ , is defined as half the Pearson correlation distance:

$$d(\underline{z}_1, \underline{z}_2) = \frac{1}{2} - \frac{\frac{1}{N} \sum_{i=1}^N (z_{1i} - \bar{z}_1)(z_{2i} - \bar{z}_2)}{2 \sqrt{\frac{1}{N} \sum_{i=1}^N (z_{1i} - \bar{z}_1)^2} \sqrt{\frac{1}{N} \sum_{i=1}^N (z_{2i} - \bar{z}_2)^2}}. \quad (4)$$

This distance function is designed so that  $d(\underline{z}_1, \underline{z}_2) = 0$  if  $\underline{z}_1$  and  $\underline{z}_2$  are perfectly correlated,  $d(\underline{z}_1, \underline{z}_2) = 1/2$  if they are uncorrelated, and  $d(\underline{z}_1, \underline{z}_2) = 1$  if they are perfectly anti-correlated.

**Implementation.** The clustering algorithm is used to compare a large set of responses in an automatic fashion. The input to the algorithm was a set of either 10,080 responses, calculated for a low-high-low two-level signal input (a pattern opposite to Fig. 2, inset), or 20,160 responses, calculated for a three-level signal input with an intermediate level (Figs. 3-6, insets). The set of responses is calculated from models that span the 12 subclasses of circuits illustrated in Figs. 1(a)-1(d). For three-level input responses, two

cases were considered for each subclass:  $T_X < T_Y$  and *vice versa*. Parameter sampling was performed as in the “Parameter values” section.

**Calculation of representative responses.** Singular value decomposition (SVD) was used to generate a representative response for each cluster. SVD has been used in other applications to summarize and interpret multivariate biological data.<sup>31</sup> We form an  $M \times N$  matrix  $A$  from the response in a cluster (the mean of each response is subtracted). The  $M$  rows correspond to the different responses, and the  $N$  columns are the time-point samples for each response. The SVD is defined as  $A = USV^T$ , where  $S$  is the diagonal matrix of singular values,  $U$  is the matrix of left-singular vectors, and  $V$  is the matrix of right-singular vectors. The left-most column of  $V$  is the right-singular vector associated with the largest singular value, and thus accounts for most of the variance of the responses. It is this vector that was used to plot the representative responses in Figs. 8 and 9.

## Entropies

The Shannon entropy of each behavior distribution was calculated as  $H = -\sum_k f_k \log_2 f_k$  (see text for definition of  $f_k$ ) and was used as a measure of robustness of the type of behavior of a model to parameter changes; the lower the entropy, the more robust is the type of behavior. The smallest possible value of  $H$  is 0, which corresponds to the case in which a circuit exhibits the same behavior for all parameter values. For the distribution of steady-state I/O behaviors, the largest possible value of  $H$  is  $\log_2 13 = 3.7$ , corresponding to the case in which all 13 types of behavior are each equally represented in parameter space. The two-level-input time courses were

grouped into 15 clusters, and the largest possible value of  $H$  for this case is  $\log_2 15 = 3.91$ , corresponding to the case in which all 15 clusters have the same number of time-courses. The three-level-input time courses were grouped into 30 clusters, leading to a largest possible value of  $\log_2 30 = 4.91$ .

## **Acknowledgment**

This work was supported by the Department of Energy and the National Institutes of Health.

## Literature cited

1. Hlavacek, W. S. & Savageau, M. A. (1996). Rules for coupled expression of regulator and effector genes in inducible circuits. *J Mol Biol* **255**, 121-39.
2. Wall, M. E., Hlavacek, W. S. & Savageau, M. A. (2003). Design principles for regulator gene expression in a repressible gene circuit. *J Mol Biol* **332**, 861-876.
3. Wall, M. E., Hlavacek, W. S. & Savageau, M. A. (2004). Design of gene circuits: lessons from bacteria. *Nat Rev Genet* **5**, 34-42.
4. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* **31**, 64-8.
5. Mangan, S. & Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A* **100**, 11980-5.
6. Schleif, R. (2000). Regulation of the L-arabinose operon of *Escherichia coli*. *Trends Genet* **16**, 559-65.
7. Harman, J. G. (2001). Allosteric regulation of the cAMP receptor protein. *Biochim Biophys Acta* **1547**, 1-17.
8. Salgado, H., Gama-Castro, S., Martinez-Antonio, A., Diaz-Peredo, E., Sanchez-Solano, F., Peralta-Gil, M., Garcia-Alonso, D., Jimenez-Jacinto, V., Santos-Zavaleta, A., Bonavides-Martinez, C. & Collado-Vides, J. (2004). RegulonDB (version 4.0): transcriptional regulation, operon

- organization and growth conditions in *Escherichia coli* K-12. *Nucleic Acids Res* **32 Database issue**, D303-6.
9. Karp, P. D., Arnaud, M., Collado-Vides, J., Ingraham, J., Paulsen, I. T. & Saier, M. H. (2004). *ASM News* **70**, 25-30.
  10. Weickert, M. J. & Adhya, S. (1993). The galactose regulon of *Escherichia coli*. *Mol Microbiol* **10**, 245-51.
  11. Robert-Baudouy, J. M., Portalier, R. C. & Stoeber, F. R. (1974). [Regulation of hexuronate metabolism in *Escherichia coli* K12. Kinetics of enzyme induction of the hexuronate system]. *Eur J Biochem* **43**, 1-15.
  12. Robert-Baudouy, J., Portalier, R. & Stoeber, F. (1981). Regulation of hexuronate system genes in *Escherichia coli* K-12: multiple regulation of the *uxu* operon by *exuR* and *uxuR* gene products. *J Bacteriol* **145**, 211-20.
  13. Ritzenthaler, P., Blanco, C. & Mata-Gilsinger, M. (1983). Interchangeability of repressors for the control of the *uxu* and *uid* operons in *E. coli* K12. *Mol Gen Genet* **191**, 263-70.
  14. Weickert, M. J. & Adhya, S. (1993). Control of transcription of *gal* repressor and isorepressor genes in *Escherichia coli*. *J Bacteriol* **175**, 251-8.
  15. Ritzenthaler, P. & Mata-Gilsinger, M. (1983). Multiple regulation involved in the expression of the *uxuR* regulatory gene in *Escherichia coli* K-12. *Mol Gen Genet* **189**, 351-4.
  16. Tong, S., Porco, A., Isturiz, T. & Conway, T. (1996). Cloning and molecular genetic characterization of the *Escherichia coli* *gntR*, *gntK*, and

- gntU* genes of GntI, the main system for gluconate metabolism. *J Bacteriol* **178**, 3260-9.
17. Izu, H., Adachi, O. & Yamada, M. (1997). Gene organization and transcriptional regulation of the *gntRKU* operon involved in gluconate uptake and catabolism of *Escherichia coli*. *J Mol Biol* **267**, 778-93.
  18. Tsunedomi, R., Izu, H., Kawai, T. & Yamada, M. (2003). Dual control by regulators, GntH and GntR, of the GntII genes for gluconate metabolism in *Escherichia coli*. *J Mol Microbiol Biotechnol* **6**, 41-56.
  19. Tsunedomi, R., Izu, H., Kawai, T., Matsushita, K., Ferenci, T. & Yamada, M. (2003). The activator of GntII genes for gluconate metabolism, GntH, exerts negative control of GntR-regulated GntI genes in *Escherichia coli*. *J Bacteriol* **185**, 1783-95.
  20. Peekhaus, N. & Conway, T. (1998). Positive and negative transcriptional regulation of the *Escherichia coli* gluconate regulon gene *gntT* by GntR and the cyclic AMP (cAMP)-cAMP receptor protein complex. *J Bacteriol* **180**, 1777-85.
  21. Bausch, C., Peekhaus, N., Utz, C., Blais, T., Murray, E., Lowary, T. & Conway, T. (1998). Sequence analysis of the GntII (subsidiary) system for gluconate metabolism reveals a novel pathway for L-idonic acid catabolism in *Escherichia coli*. *J Bacteriol* **180**, 3704-10.
  22. Bausch, C., Ramsey, M. & Conway, T. (2004). Transcriptional organization and regulation of the L-idonic acid pathway (GntII system) in *Escherichia coli*. *J Bacteriol* **186**, 1388-97.

23. Geanacopoulos, M. & Adhya, S. (1997). Functional characterization of roles of GalR and GalS as regulators of the *gal* regulon. *J Bacteriol* **179**, 228-34.
24. Savageau, M. A. (1976). *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*, Addison-Wesley, Reading, MA.
25. Martinez-Antonio, A., Salgado, H., Gama-Castro, S., Gutierrez-Rios, R. M., Jimenez-Jacinto, V. & Collado-Vides, J. (2003). Environmental conditions and transcriptional regulation in *Escherichia coli*: a physiological integrative approach. *Biotechnol Bioeng* **84**, 743-9.
26. Rigali, S., Schlicht, M., Hoskisson, P., Nothaft, H., Merzbacher, M., Joris, B. & Titgemeyer, F. (2004). Extending the classification of bacterial transcription factors beyond the helix-turn-helix motif as an alternative approach to discover new cis/trans relationships. *Nucleic Acids Res* **32**, 3418-26.
27. Savageau, M. A. (1985). A theory of alternative designs for biochemical control systems. *Biomed Biochim Acta* **44**, 875-80.
28. Savageau, M. A. (2001). Design principles for elementary gene circuits: elements, methods and examples. *Chaos* **11**, 142-159.
29. Dunlop, M. J. & Wall, M. E. (2005). Robustness in gene circuits: clustering of functional responses. *Proceedings of the 24th American Control Conference*, In press.
30. Hochbaum, D. S. & Shmoys, D. B. (1985). A best possible approximation algorithm for the k-center problem. *Math Oper Res* **10**, 180-184.

31. Wall, M. E., Rechtsteiner, A. & Rocha, L. M. (2003). Singular value decomposition and principal component analysis. In *A Practical Approach to Microarray Data Analysis* (Berrar, D. P., Dubitzky, W. & Granzow, M., eds.), pp. 91-109. Kluwer, Norwell, MA.

## Figure Captions

**Figure 1.** Illustrations of type 2 incoherent FFL models that span a comprehensive set of signal interactions: (a) type-(-,-,\*); (b) type-(-,0,\*); (c) type-(-,+,\*); (d) type-(0,+,\*); (e) type-(+,+,\*); (f) type-(+,0,\*); (g) type-(+,-,\*); (h) type-(0,-,\*). Due to symmetry in the mathematical models, the response of a system above the line ((a)-(d)) to an increase in signal is the same as that of the system with inverted signal interactions below the line ((e)-(h)) to a decrease in signal, and *vice versa*.

**Figure 2.** Functions of type-(+,+,0) circuits. The (+,+,0) subclass of type 2 incoherent FFL circuits may be either inducible or repressible, depending upon parameter values. Repressible systems may exhibit accelerated derepression rise times and repression decay times with respect to a reference system without  $X \rightarrow Y$  regulation. The following parameters were used for the reference system:  $B_Y = B_Z = 0.1$ ;  $\alpha_Y = \alpha_Z = 0.9$ ;  $\beta_Y = \beta_Z = 1$ ;  $K_{YX} = 5$ ,  $K_{ZX} = 4$ ,  $K_{ZY} = 1$ ; and  $n_{YX} = n_{ZX} = n_{ZY} = 2$ . For the repressible type-(+,+,0) circuit, external equivalence conditions were applied (Methods). The same parameter values were used for the inducible circuit, with the following exceptions:  $B_Z = 0.04386$ ,  $K_{ZX} = 10$ , and  $K_{ZY} = 0.5$  (the values of  $K_{ZX}$  and  $K_{ZY}$  were chosen arbitrarily among those that correspond to an inducible response, and the value of  $B_Z$  was chosen to align the time courses at  $t = 0$ ).

**Figure 3.** Typical responses of type-(-,-,\*) circuits to signal (see Fig. 1(a)). The following parameter values were used for the reference system, which is a type-(-,-,0) circuit:  $B_Y = B_Z = 0.1$ ;  $\alpha_Y = \alpha_Z = 0.9$ ;  $\beta_Y = \beta_Z = 1$ ;  $K_{YX} = 5$ ,  $K_{ZX} = 4$ ,  $K_{ZY} = 1$ ; and  $n_{YX} = n_{ZX} = n_{ZY} = 2$ . For the other circuits, external equivalence conditions were applied

(Methods). a)  $T_X < T_Y$ . The type- $(-, -, 0)$  circuit exhibits a LHH behavior, type- $(-, -, +)$  exhibits a LHM behavior, and type- $(-, -, -)$  exhibits a LMH behavior. b)  $T_X > T_Y$ . The type- $(-, -, 0)$  circuit exhibits a LLH behavior, type- $(-, -, +)$  exhibits a MLH behavior, and type- $(-, -, -)$  exhibits a LMH behavior.

**Figure 4.** Responses of type- $(-, 0, *)$  circuits to the signal input illustrated in the inset of Fig. 3. a)  $T_X < T_Y$ . b)  $T_X > T_Y$ . Selection of parameter values is as in Fig. 3.

**Figure 5.** Responses of type- $(-, +, *)$  circuits to the signal input illustrated in the inset of Fig. 3. a)  $T_X < T_Y$ . b)  $T_X > T_Y$ . Selection of parameter values is as in Fig. 3.

**Figure 6.** Responses of type- $(0, +, *)$  circuits to the signal input illustrated in the inset of Fig. 3. a)  $T_X < T_Y$ . b)  $T_X > T_Y$ . c) Selection of parameter values is as in Fig. 3.

**Figure 7.** Pulsed response of a type- $(0, +, +)$  circuit, which is equivalent to a two-step cascade. The parameter values are the same as in Fig. 3, except  $B_Y = 0$ ,  $K_{YX} = 0.001$ ,  $K_{ZX} = 50$ , and  $K_{ZY} = 0.1$  to illustrate an arbitrary example of pulsing for this subclass of circuits. Two-step cascades were previously assumed to be incapable of generating pulses.<sup>5</sup>

**Figure 8.** Representative responses from 12 of 15 automatically generated clusters of time courses in response to extreme changes in the level of signal (between  $S < \min(T_X, T_Y)$  and  $S > \max(T_X, T_Y)$ ). The signal input is a two-step low-high-low pattern, the opposite to that in the inset of Fig. 2. The y-axis is the value of  $Z$  at an arbitrary scale. The x-axis is the value of time from 0 to 30 in arbitrary units. Traces are labelled

by the associated clusters A-L (Supplemental Table S1), along with the percentage of all time-courses that are in the given cluster. Responses from the remaining 3 clusters are not shown; each of them consists of fewer than 10 time courses. Response (D) appears to exhibit delayed induction; this behavior was confirmed through analysis of temporal responsiveness of type-(-,+0) circuits (Supplemental Table S2). For a more limited range of parameter values and signal interactions, the type 2 incoherent FFL was previously found to be a sign-sensitive accelerator.<sup>5</sup>

**Figure 9.** Representative responses from 20 of 30 automatically generated clusters of time courses in response to three-level changes in signal. The time dependence of the signal input is shown in an inset in Fig. 3. The y-axis is the value of  $Z$  at an arbitrary scale. The x-axis is the value of time from 0 to 50 in arbitrary units. Traces are labelled by the associated clusters A-T (Supplemental Table S3), along with the percentage of all time-courses that are in the given cluster. Responses from the remaining 10 clusters are not shown; each of them consists of 10 or fewer time-courses. The following is a classification (by eye) of the illustrated responses into the steady-state I/O behaviors defined in the text: (A) LLH; (B) LHH; (C) Null; (D) HHL; (E) HLL; (F) LHM; (G) HLM; (H) LHH; (I) LLH; (J) LMH; (K) MHL; (L) HHL; (M) HLL; (N) Null; (O) HML; (P) MLH; (Q) MLH; (R) LLH; (S) Null; (T) LHH. Only the HLH and LHL behaviors, which are scarce, did not appear among the representative responses. Response (I) exhibits delayed induction, similar to response (D) in Fig. 8.

**Table 1.** Distribution of steady-state behaviors of the type 2 incoherent FFL.<sup>a</sup>

	Effect of Signal on TF Activities <sup>b</sup>			Pattern of Steady-State I/O Behavior <sup>c</sup>													Response to Change in Signal from 0 to $\infty$ <sup>d</sup>		Entropy	
	ZX	YX	ZY	Null	HLH	LHL	HHL	HLL	HML	HLM	MHL	LLH	MLH	LHH	LHM	LMH	Ind	Rep		
T <sub>x</sub> <T <sub>y</sub>	-	-	-	17	3	0	0	0	0	0	0	34	9	21	0	<u>16</u>	80	0	2.33	
	-	-	0	50	0	0	0	13	0	0	0	0	0	<u>37</u>	0	0	37	13	1.40	
	-	-	+	17	0	16	21	0	0	0	13	0	0	21	<u>13</u>	0	33	33	2.56	
	-	0	-	21	0	0	0	0	0	0	0	34	0	27	0	<u>19</u>	79	0	1.96	
	-	0	0	54	0	0	0	0	0	0	0	0	0	<u>46</u>	0	0	46	0	0.99	
	-	0	+	21	0	11	17	0	0	0	10	0	0	26	<u>14</u>	0	41	27	2.50	
	-	+	-	17	0	0	0	0	0	0	0	28	0	30	0	<u>26</u>	83	0	1.97	
	-	+	0	45	0	0	0	0	0	0	0	0	0	<u>55</u>	0	0	55	0	0.99	
	-	+	+	21	0	11	17	0	0	0	10	0	0	26	<u>14</u>	0	41	27	2.50	
	0	+	-	61	0	0	0	0	0	0	0	23	0	5	0	<u>10</u>	39	0	1.49	
	0	+	0	84	0	0	0	0	0	0	0	0	0	<u>16</u>	0	0	16	0	0.62	
	0	+	+	67	0	0	<u>33</u>	0	0	0	0	0	0	0	0	0	0	0	33	0.91
		All T <sub>x</sub> <T <sub>y</sub>			40	0	3	7	1	0	0	3	10	1	26	3	6	46	11	2.49
T <sub>x</sub> >T <sub>y</sub>	-	-	-	21	0	0	0	0	0	0	0	47	0	17	0	<u>15</u>	79	0	1.84	
	-	-	0	51	0	0	13	0	0	0	0	<u>37</u>	0	0	0	0	37	13	1.40	
	-	-	+	32	2	0	4	19	9	2	0	31	<u>2</u>	0	0	0	33	33	2.31	
	-	0	-	21	0	0	0	0	0	0	0	47	0	17	0	<u>15</u>	79	0	1.84	
	-	0	0	54	0	0	0	0	0	0	0	<u>46</u>	0	0	0	0	46	0	0.99	
	-	0	+	31	1	0	0	24	0	4	0	37	<u>4</u>	0	0	0	41	27	1.98	
	-	+	-	17	0	0	0	0	0	0	0	44	0	21	0	<u>19</u>	83	0	1.88	
	-	+	0	45	0	0	0	0	0	0	0	<u>55</u>	0	0	0	0	55	0	0.99	
	-	+	+	27	4	0	0	18	0	10	0	34	<u>7</u>	0	0	0	41	28	2.28	
	0	+	-	60	0	0	0	0	0	0	0	0	0	<u>40</u>	0	0	40	0	0.97	
	0	+	0	84	0	0	0	0	0	0	0	<u>16</u>	0	0	0	0	16	0	0.62	
	0	+	+	61	5	0	0	23	0	<u>10</u>	0	0	0	0	0	0	0	0	34	1.49
		All T <sub>x</sub> >T <sub>y</sub>			42	1	0	1	7	1	2	0	33	1	8	0	4	46	11	2.20
All				41	1	2	4	4	0	1	1	21	1	17	2	5	46	11	2.53	

<sup>a</sup> Entries in this table correspond to  $T_x < T_y$  and  $T_x > T_y$ . The steady-state behaviors of systems in which  $T_x = T_y$  may be inferred from either the  $T_x < T_y$  or  $T_x > T_y$  entries as follows: the fraction of systems that are inducible is the entry in the Ind column; the fraction of systems that are repressible is the entry in the Rep column; and the fraction of systems that are unresponsive is the sum of the entries in the Null, LHL, and HLH columns.

<sup>b</sup> Each row corresponds to a different subclass of type 2 incoherent FFL circuits (the last row is the summary for all subclasses). The symbol entries under “ZX” indicate the effect of signal on regulation of Z by X. Symbols in the other two columns are similarly defined.

<sup>c</sup> Entries correspond to the % of 681,472 parameter combinations that result in the steady-state behavior indicated by the three-letter code. Underlined entries indicate the types that are illustrated in Figs. 3-6.

<sup>d</sup> The entries under “Ind” indicate the % of systems that are inducible; for these systems, the output at high signal is higher than the output at low signal. The entries under “Rep” indicate the fraction of systems that are repressible; for these systems, the output at high signal is lower than the output at low signal.

**Table 2.** Controlled mathematical comparison of temporal responsiveness between the (+,+,0) subclass of type 2 incoherent FFL circuits and circuits without  $X \rightarrow Y$  regulation.<sup>a</sup>

	$K_{ZY} < 1$			$K_{ZY} \geq 1$		
	Faster	Similar	Slower	Faster	Similar	Slower
Derepression Rise Time	0.55	0.45	0.00	0.16	0.84	0.00
Derepression Settling Time	0.12	0.46	0.42	0.13	0.84	0.03
Repression Decay Time	0.19	0.81	0.00	0.05	0.95	0.00
Repression Settling Time	0.11	0.81	0.08	0.04	0.95	0.01

<sup>a</sup> Each entry is the fraction of 681,472 parameter combinations (Methods) that cause the FFL to have responsiveness measures that are faster than, the same as, or slower than equivalent circuits without  $X \rightarrow Y$  regulation. Two responsiveness measures are considered to be similar if their ratio equals  $1 \pm 0.05$ .

Figure 1

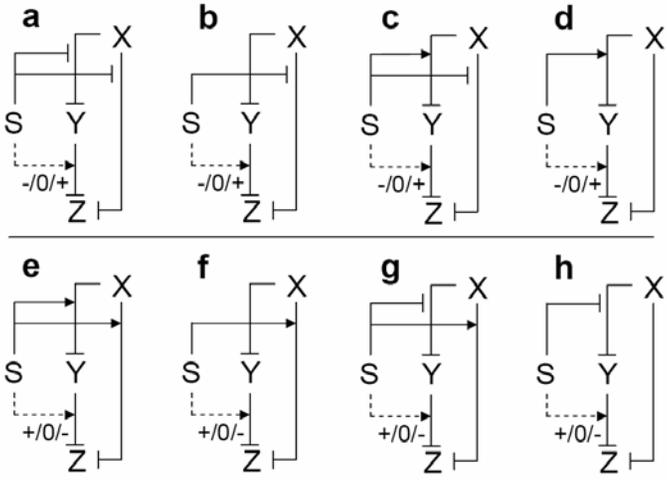


Figure 2

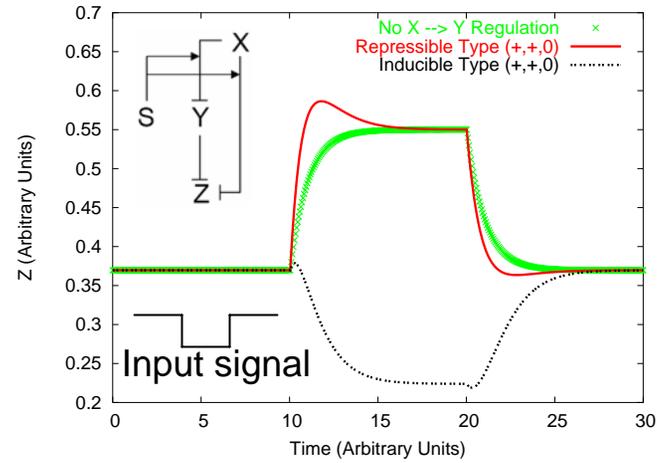


Figure 3

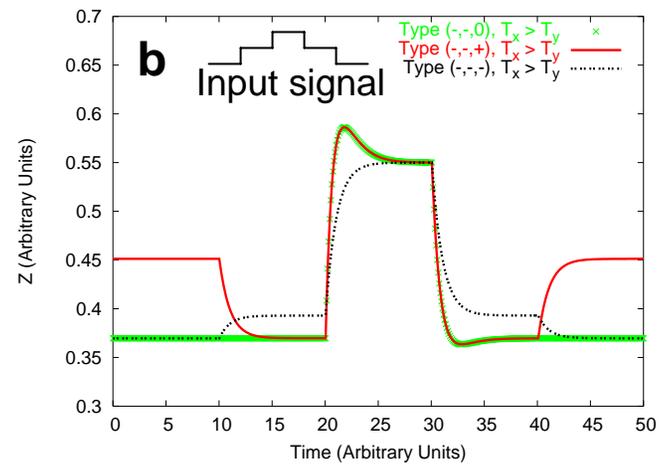
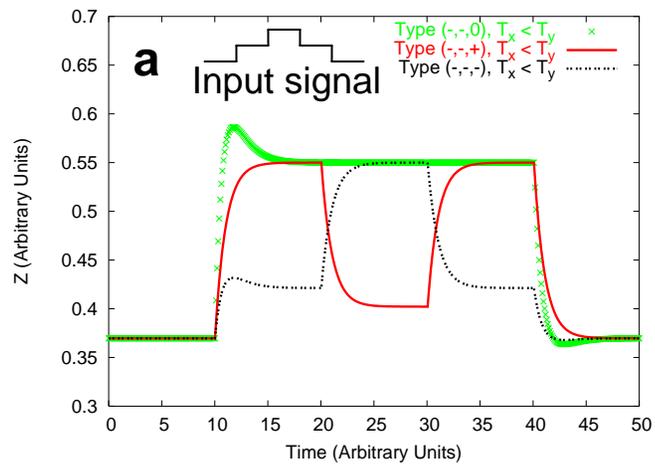


Figure 4

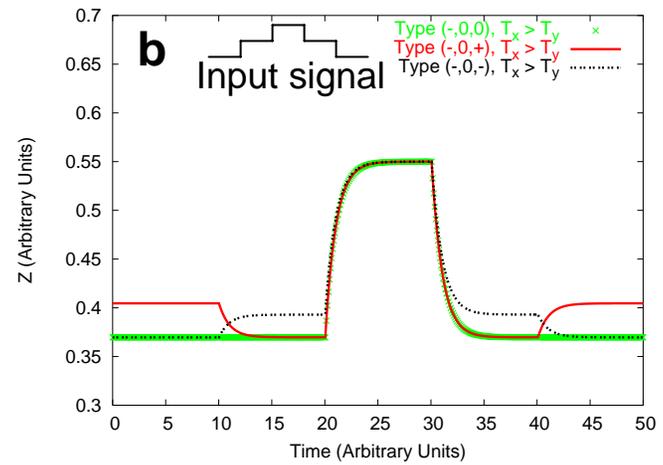
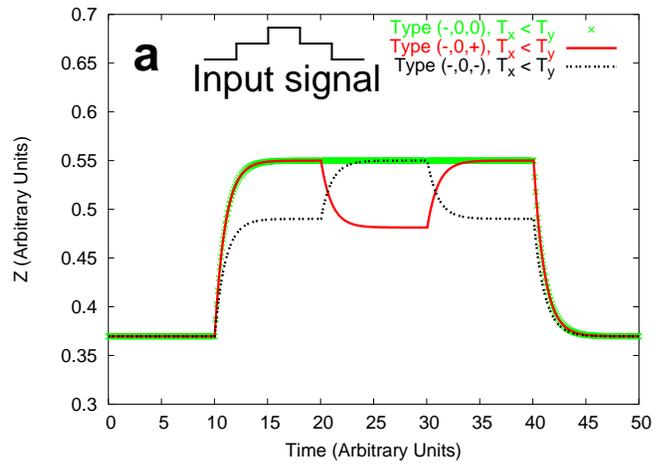


Figure 5

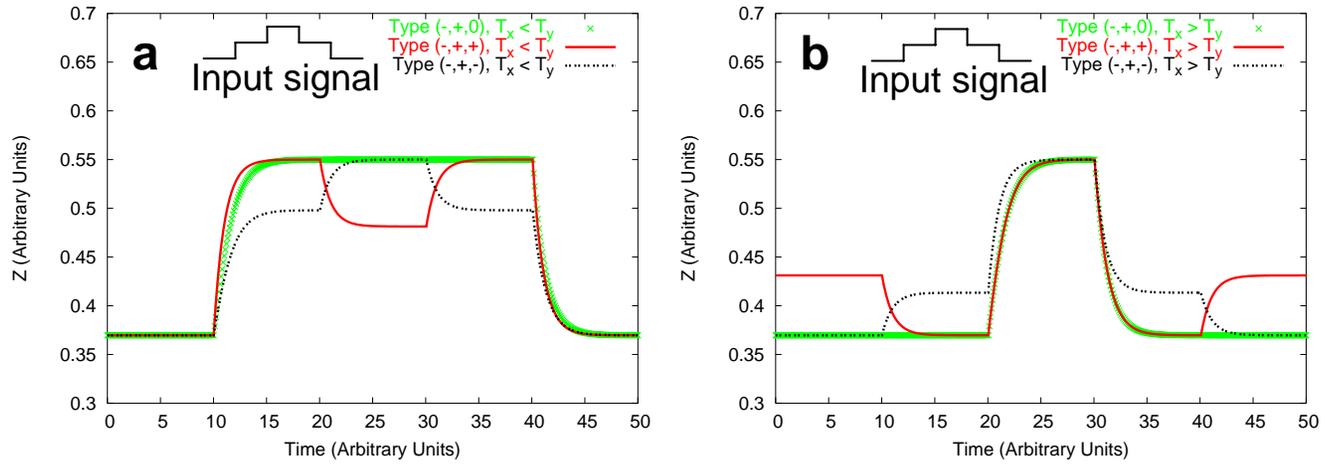


Figure 6

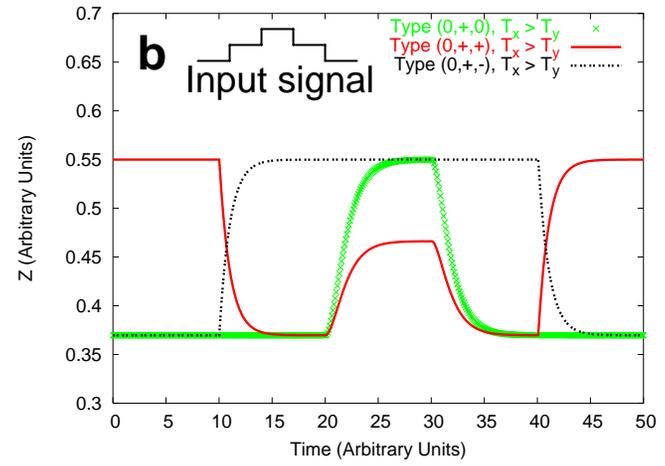
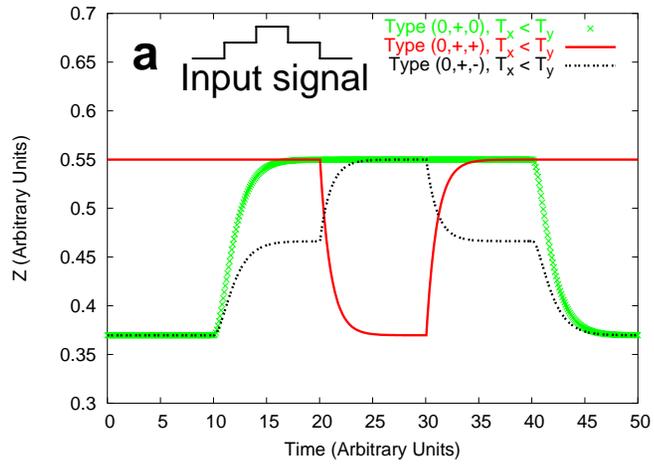


Figure 7

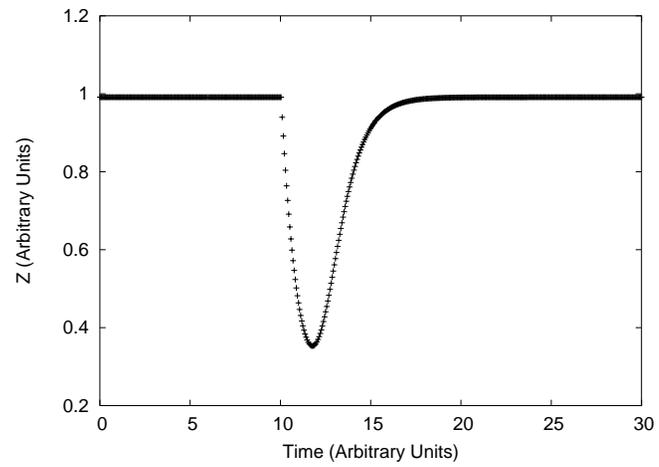


Figure 8

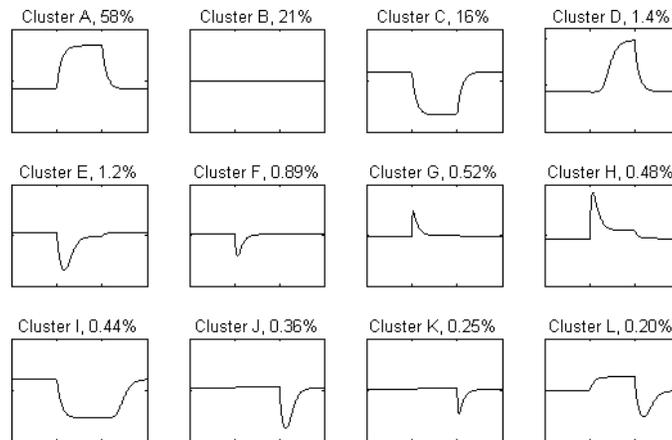
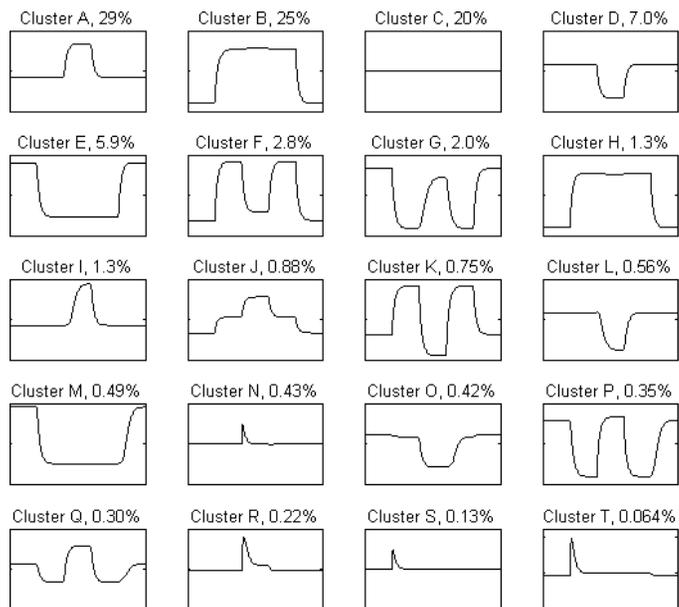


Figure 9



# **Multiple Functions of a Feed-Forward-Loop Gene Circuit**

## ***Supplementary Material***

Michael E. Wall<sup>1,2\*</sup>, Mary J. Dunlop<sup>3</sup>, William S. Hlavacek<sup>4</sup>

<sup>1</sup>*Computer and Computational Sciences Division and* <sup>2</sup>*Bioscience Division, Mail Stop B256, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

<sup>3</sup>*Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA*

<sup>4</sup>*Theoretical Division, Mail Stop K710, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

*\* Corresponding author*

## Supplementary Tables

Table S1. Distribution of time courses among 12 of 15 clusters for the response of the type 2 incoherent FFL to changes in signal between extremely high and low levels.<sup>a</sup>

Effect of Signal on TF Activities			% of Time Courses in Clusters A-L												Entropy
ZX	YX	ZY	A	B	C	D	E	F	G	H	I	J	K	L	
-	-	-	85	5	0	0	0	0	0	0	0	4	3	2	0.89
-	-	0	46	22	20	0	0	0	3	3	5	0	0	0	2.04
-	-	+	38	9	47	0	0	0	3	3	0	0	0	0	1.70
-	0	-	93	7	0	0	0	0	0	0	0	0	0	0	0.37
-	0	0	69	31	0	0	0	0	0	0	0	0	0	0	0.90
-	0	+	49	11	40	0	0	0	0	0	0	0	0	0	1.38
-	+	-	95	5	0	0	0	0	0	0	0	0	0	0	0.27
-	+	0	71	21	0	8	0	0	0	0	0	0	0	0	1.11
-	+	+	43	9	36	4	4	4	0	0	0	0	0	0	1.95
0	+	-	66	34	0	0	0	0	0	0	0	0	0	0	0.92
0	+	0	38	57	0	5	0	0	0	0	0	0	0	0	1.22
0	+	+	0	37	47	0	10	6	0	0	0	0	0	0	1.62
<b>All</b>			58	21	16	1	1	1	1	0	0	0	0	0	1.76

<sup>a</sup> The organization of this table is similar to that of Table 1. Systems are subjected to a two-level signal input with levels  $S < \min(T_X, T_Y)$  and  $S > \max(T_X, T_Y)$ . The signal input has a low-high-low time dependence, which is the opposite of that illustrated in an inset of Fig. 2. The three clusters with fewer than 10 members are not included in the table (they were used to calculate the entropies, however). Representative time courses for the behavior types (labelled A–L) are illustrated in Fig. 8.

Table S2. Mathematically controlled comparison of temporal responsiveness between the  $(-,+,0)$  subclass of type 2 incoherent FFL circuits and circuits without  $X \rightarrow Y$  regulation.<sup>a</sup>

	$K_{ZY} < 1$			$K_{ZY} \geq 1$		
	Faster	Similar	Slower	Faster	Similar	Slower
Induction Rise Time	0.00	0.34	0.66	0.00	0.89	0.11
Induction Settling Time	0.00	0.34	0.66	0.00	0.89	0.11
De-induction Decay Time	0.00	0.88	0.12	0.00	0.94	0.06
De-induction Settling Time	0.00	0.88	0.12	0.00	0.94	0.06

<sup>a</sup> Each entry is the fraction of the 681,472 parameter combinations (Methods) that cause the FFL to have responsiveness measures that are faster than, similar to, or slower than equivalent circuits without  $X \rightarrow Y$  regulation. Two responsiveness measures are considered to be similar if their ratio equals  $1 \pm 0.05$ .

Table S3. Distribution of time courses among 20 of 30 clusters for the response of the type the type 2 incoherent FFL to a three-level signal input.<sup>a</sup>

Effect of Signal on TF Activities			% of Time Courses in Clusters A-T																	Entropy						
ZX	YX	ZY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T				
Tx < Ty	-	-	-	48	31	5	0	0	0	3	0	0	4	0	0	0	0	6	4	0	0	0	2.00			
	-	-	0	0	33	22	0	13	0	0	14	0	0	0	12	0	0	0	0	0	3	2	2.51			
	-	-	+	0	20	4	38	0	25	0	6	0	0	6	0	0	0	0	0	0	0	0	2.21			
	-	0	-	50	40	7	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	1.45			
	-	0	0	0	70	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.89			
	-	0	+	0	29	7	32	0	21	0	6	0	0	6	0	0	0	0	0	0	0	0	0	2.26		
	-	+	-	43	50	4	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1.35		
	-	+	0	0	80	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.72		
	-	+	+	0	29	7	32	0	21	0	6	0	0	6	0	0	0	0	0	0	0	0	0	0	2.26	
	0	+	-	43	23	33	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1.58	
	0	+	0	0	44	56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.99	
	0	+	+	0	0	42	58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.98	
All Tx < Ty			15	37	20	13	1	6	0	3	0	1	2	0	1	0	0	1	0	0	0	0	0	2.57		
Tx > Ty	-	-	-	61	27	8	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	1.41			
	-	-	0	45	0	22	7	0	0	0	0	0	0	8	0	6	8	0	0	3	0	0	2.33			
	-	-	+	37	0	9	3	31	0	2	0	0	0	6	0	5	2	1	1	3	0	0	0	2.55		
	-	0	-	61	27	8	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	1.41		
	-	0	0	69	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.89		
	-	0	+	51	0	11	0	31	0	4	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1.70	
	-	+	-	57	34	6	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	1.42	
	-	+	0	64	0	21	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.29
	-	+	+	41	0	8	0	27	0	14	0	6	0	0	0	0	0	0	1	1	0	0	0	0	0	2.21
	0	+	-	0	65	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.93
	0	+	0	31	0	59	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.30
	0	+	+	0	0	35	0	39	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.56
All Tx > Ty			43	13	21	1	11	0	4	0	3	1	0	1	0	1	1	0	0	0	0	0	0	0	2.42	
All			30	27	21	8	6	3	2	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	2.77	

<sup>a</sup> The organization of this table is the similar to that of Table 1. Systems are subjected to a three-level signal input with an intermediate level. The time dependence of the signal input is illustrated in an inset in Fig. 3. Ten clusters that have 10 or fewer members are not included in the table (they were included in

the calculation of the entropies, however). Representative time courses for the behavior types (labelled A–T) are illustrated in Fig. 9.