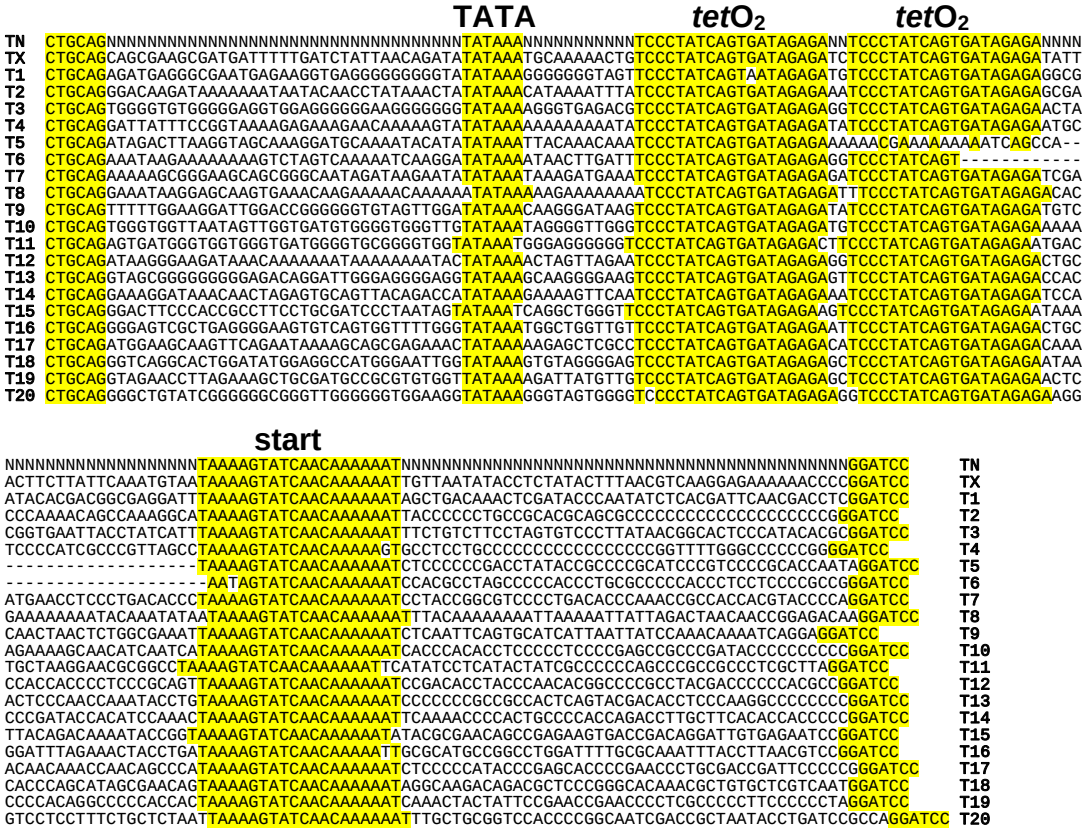
Supplementary Information:

Diversity-Based, Model-Guided Construction of Synthetic Gene Networks with Predicted Functions

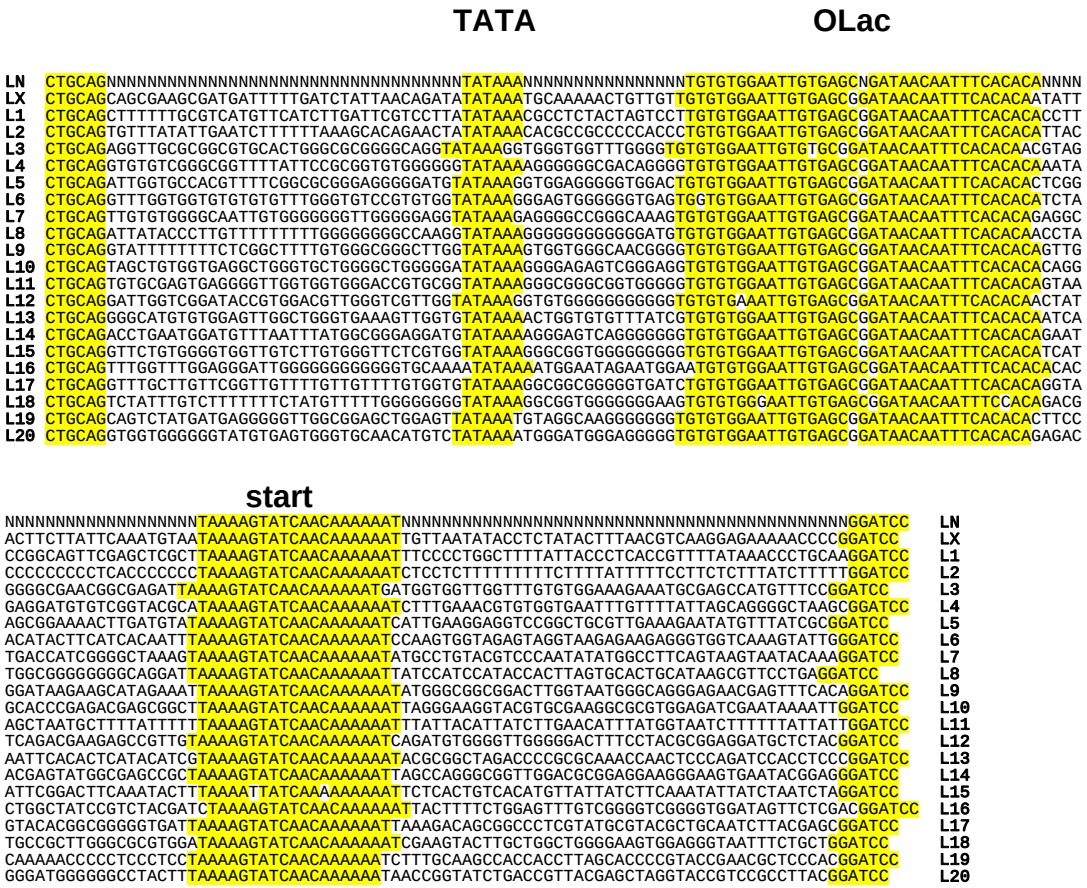
Tom Ellis*, Xiao Wang* and James J. Collins

SUPPLEMENTARY FIGURES



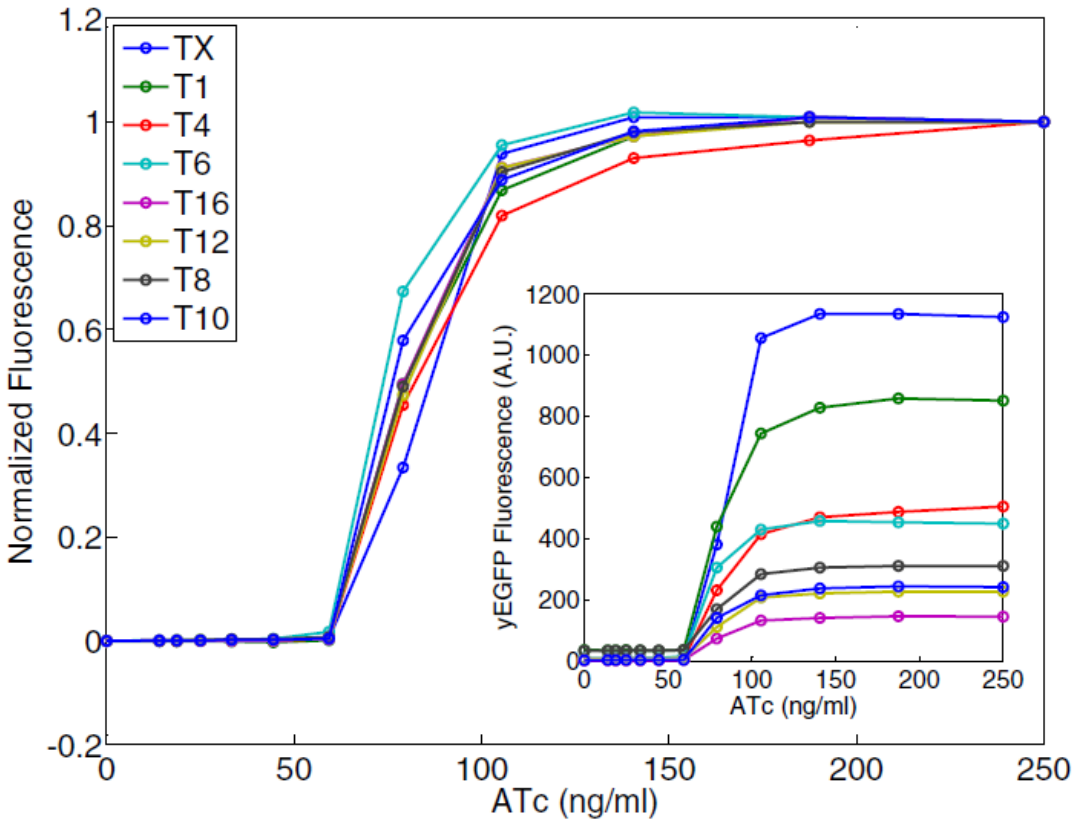
Supplementary Figure 1: Aligned sequences for all members of the TetR-regulated promoter library (P_{LibT}). The promoter sequences show conservation of defined sites (yellow), but large variations between them at sites undefined ('N') in the original design (TN).

Nature Biotechnology: doi:10.1038/nbt.1536

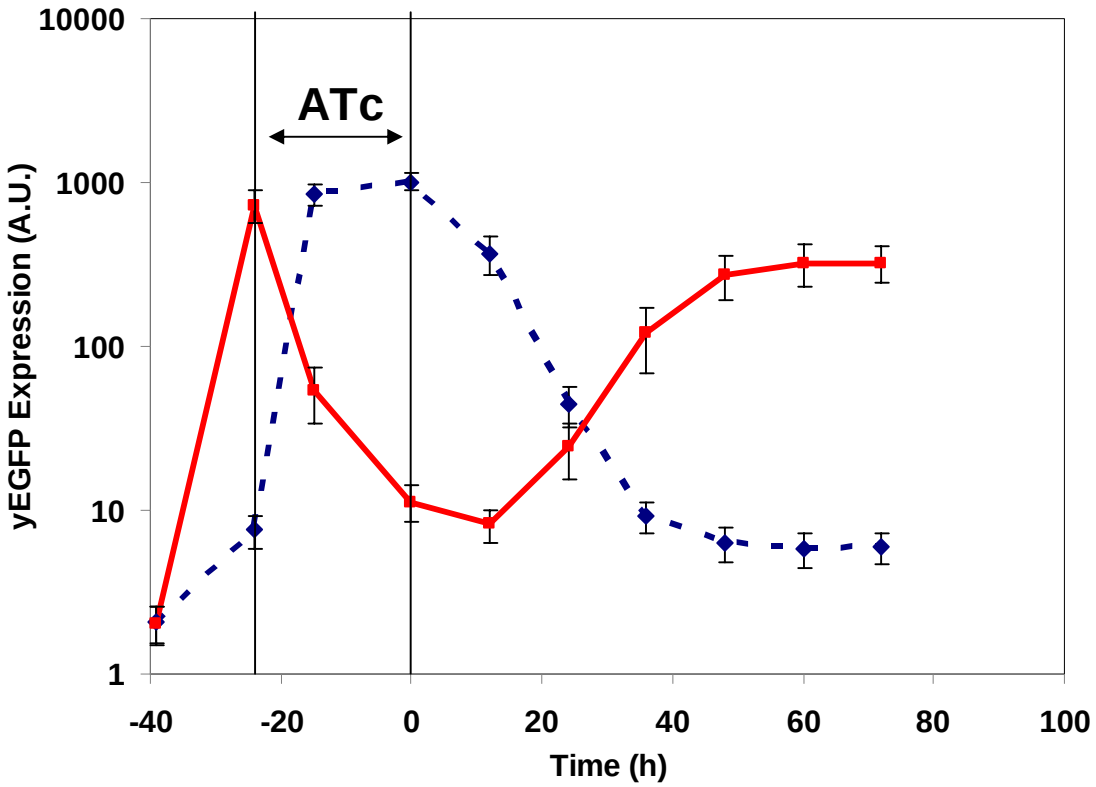


Supplementary Figure 2: Aligned sequences for all members of the LacI-regulated promoter library (P_{LibL}). The promoter sequences show conservation of defined sites (yellow), but large variations between them at sites undefined ('N') in the original design (LN).

Nature Biotechnology: doi:10.1038/nbt.1536

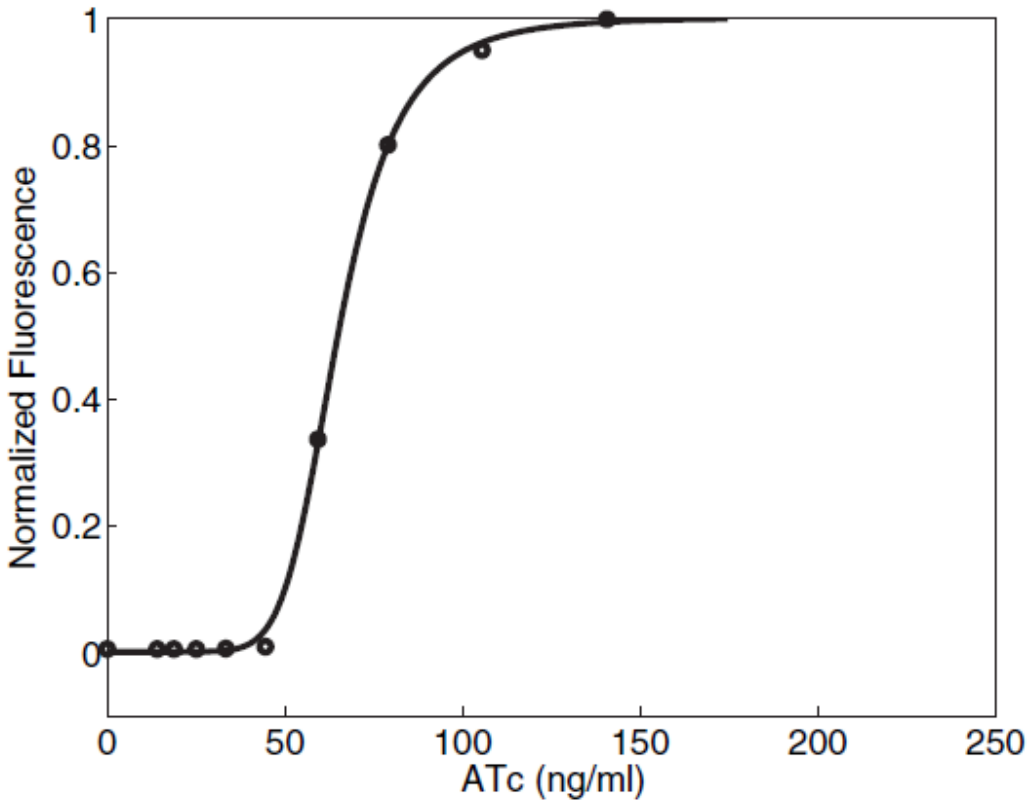


Supplementary Figure 3: Normalized dose-response curves for representative P_{LibT} promoters. All curves have very similar Hill coefficients, verifying our hypothesis that promoter mutations here only change maximum and minimum promoter output, not induction cooperativity. Inset figure shows the curves with absolute fluorescence levels.



Supplementary Figure 4: Temporal behavior of the TX-LX genetic timer network. The complete raw data for Figure 3B of the main text is shown (red) along with equivalent measurements of the reverse network, where yEGFP expression is controlled by P_{TX} and therefore directly related to LacI levels (blue discontinuous line). Cells were moved from growth in GLU-supplemented media to GAL-supplemented media at -39 hrs to activate the network. From -24 to 0 hrs, 250 ng/ml ATc was added to induce a network transition. At 0 hrs, cells were washed three times and returned to growth in GAL-supplemented media. Measurements of the median of the gated cell population and the standard deviation (error bars) were taken every 12 hours when cells were diluted into new media to maintain stable growth.

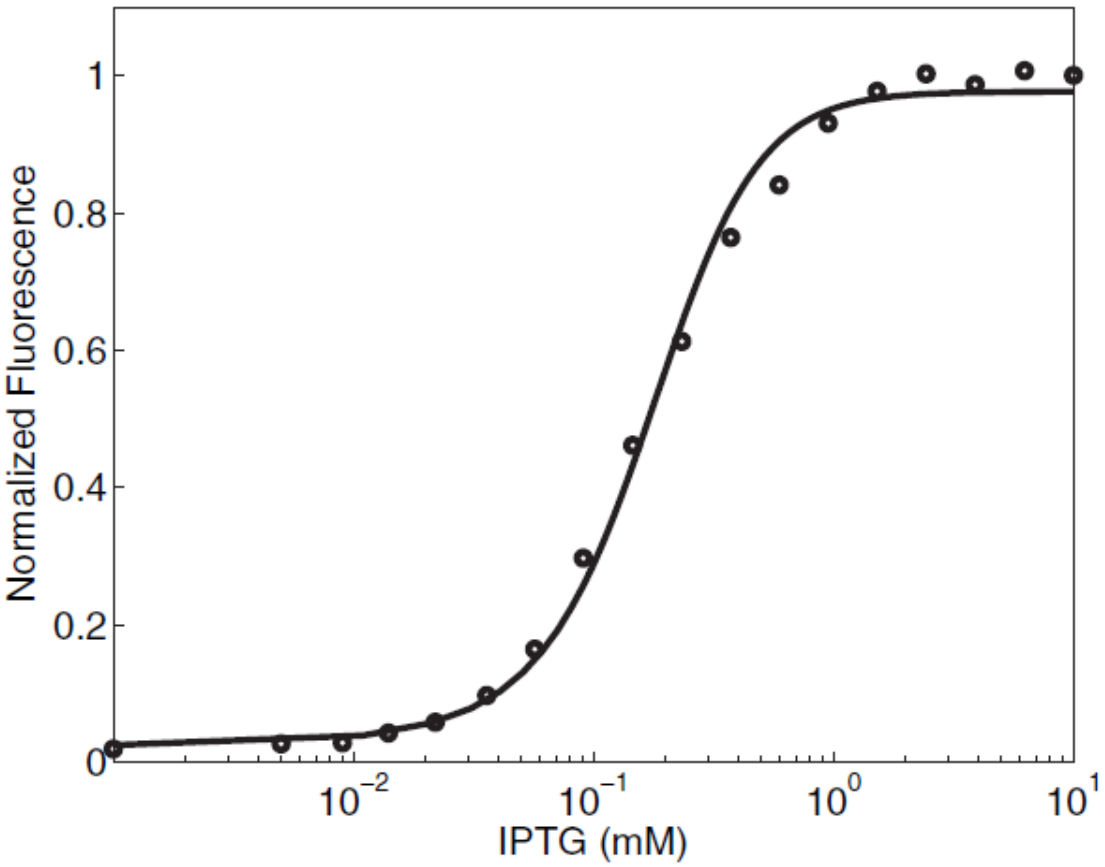
Nature Biotechnology: doi:10.1038/nbt.1536



Supplementary Figure 5: Dose response curve for the TetR-inducible switch.

Median yEGFP fluorescence from the TX promoter is normalized to the maximum value with full induction (250 ng/ml) of ATc. Our model can accurately capture the high nonlinearity and low leakage of the system, and provide input for later modeling of negative feedforward loop networks and genetic timer networks.

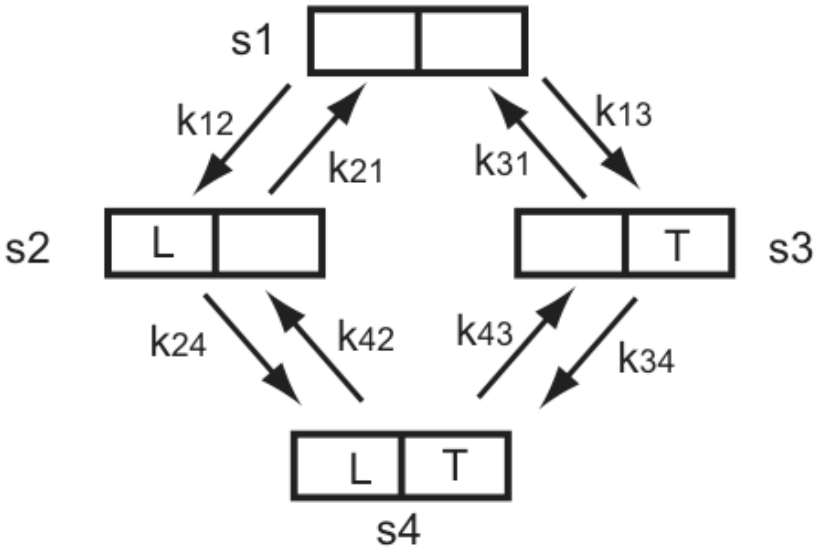
Nature Biotechnology: doi:10.1038/nbt.1536



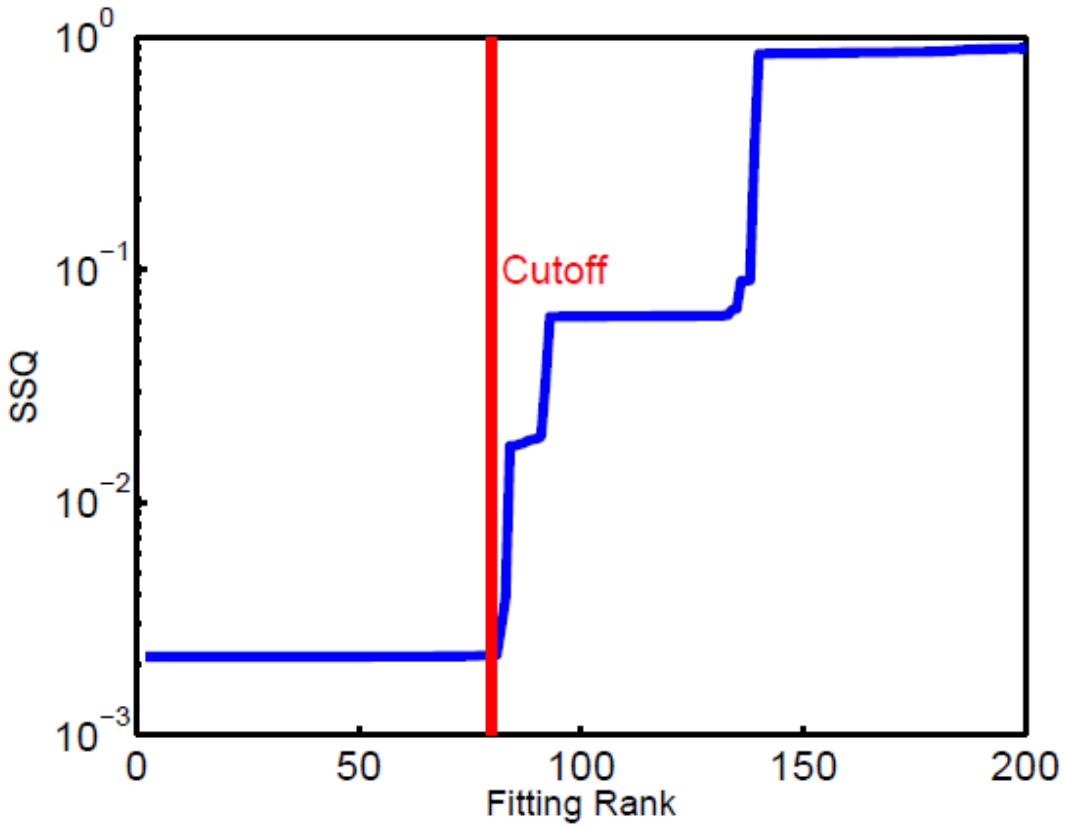
Supplementary Figure 6: Dose response curve for the LacI-inducible switch.

Median yEGFP fluorescence from the LX promoter is normalized to the maximum value with full induction (10 mM) of IPTG. The low nonlinearity is captured by our model and provides input for determining *ni* in the modeling related to LacI.

Nature Biotechnology: doi:10.1038/nbt.1536

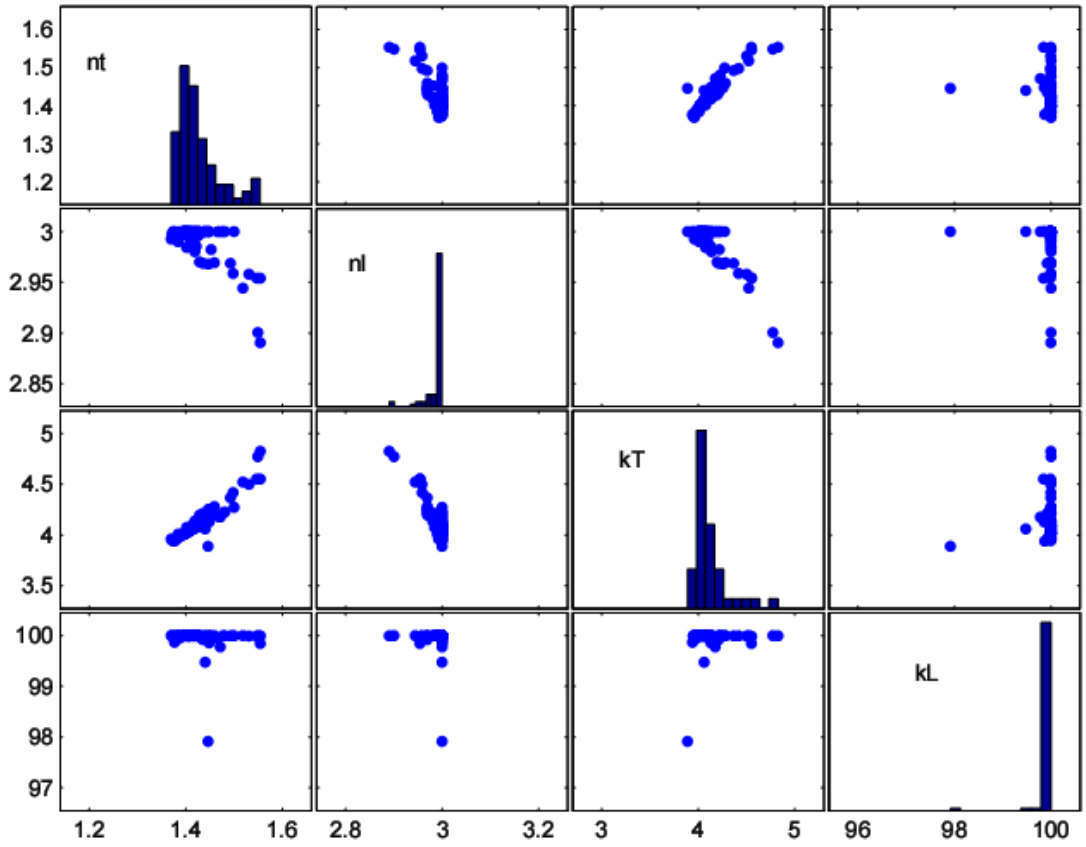


Supplementary Figure 7: Diagram of four possible states of the OR-gate promoter in the negative feedforward loop network. s1 represents the state where the promoter has no repressor binding, s2 represents the state where only LacI is binding with the promoter, s3 represents the state where only TetR is binding with the promoter, and s4 represents the state where both LacI and TetR are binding with the promoter. Transitioning rates between these states are determined by both repressor abundance and repressor-DNA association equilibrium.

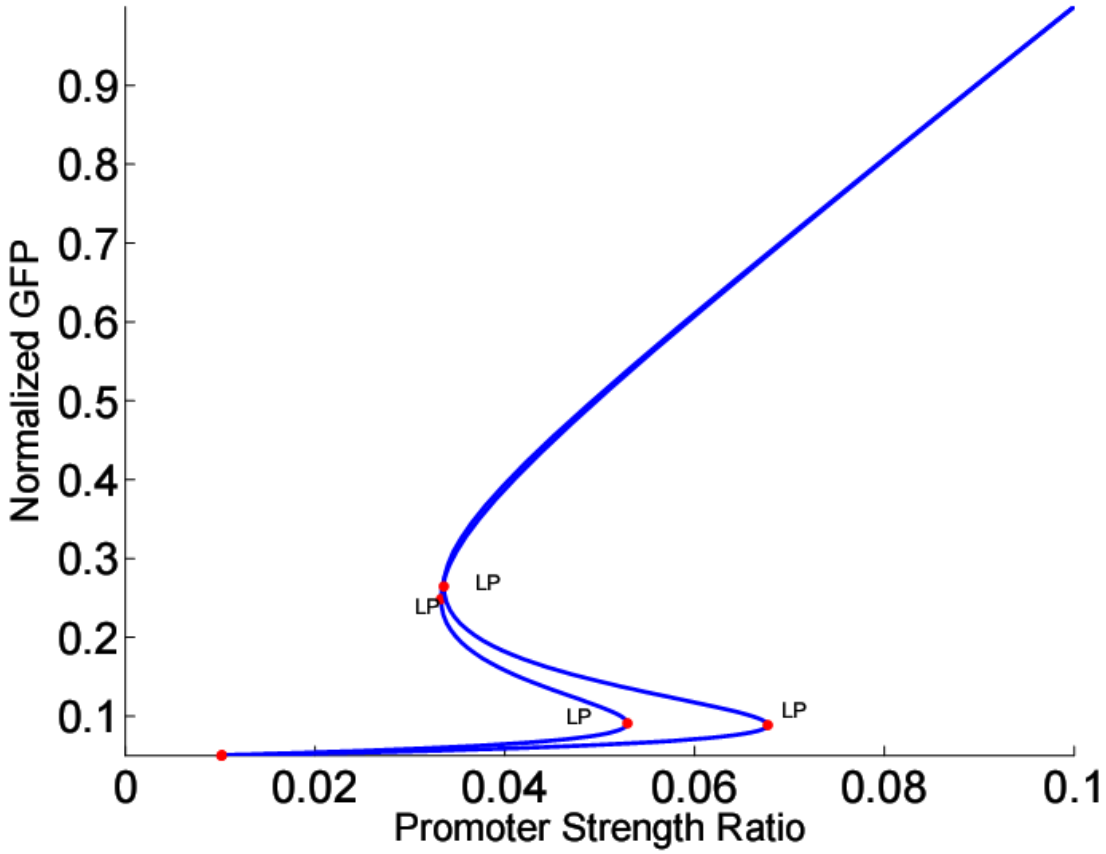


Supplementary Figure 8: Distribution of sum of square (SSQ) of the residuals for all 200 fittings. TX-LX construct is fitted using *lsqcurvefit* in Matlab 200 different times, each run with a unique and randomized initial guess. All fittings are sorted based on the SSQ value. This figure plots the SSQ against the ranking for each fitting. It is clear that over a certain threshold (0.0022) the fittings start to have much larger error, indicating that a solution is not converging to a reasonable local minimum. The fittings with SSQ larger than 0.0022 (indicated on the right of the red cutoff line) were discarded; there were 79 fittings left as candidates for further analysis.

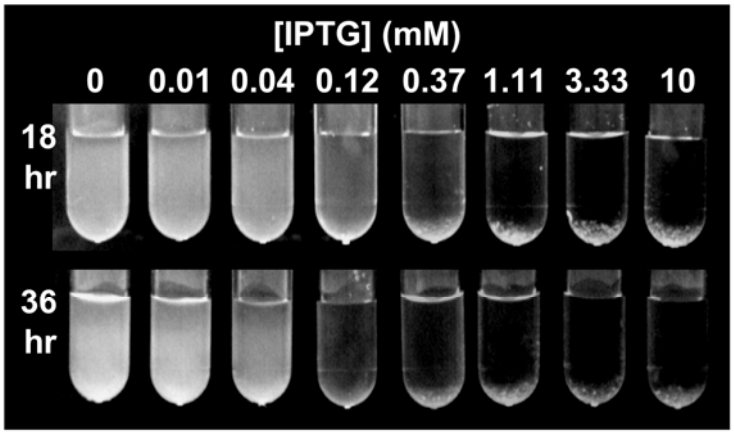
Nature Biotechnology: doi:10.1038/nbt.1536



Supplementary Figure 9. Histogram and pair-wise scatter plot of 79 candidates of the four parameter combinations. Figures on the diagonal are the histograms of 79 candidates for each of the four estimated parameters. It is shown that *nt* and *kT* are less constrained than *nl* and *kL*.



Supplementary Figure 10: Bifurcation diagrams of the mutual inhibitory network with parameters used in Figure 3B of the main text. Two bifurcation diagrams are shown utilizing the parameters used in Figure 3B of the main text. The bifurcation points are indicated as red dots with labels (LP). After the system is forced into its non-default state, it goes through a bottleneck to reset back to its default state after the external stimulus is relieved. The further the ratio is away from 0.068, the faster the system resets to its default state. The segments between two bifurcation points represent the unstable steady states, and the segments outside the region represent the stable steady states.



Supplementary Figure 11: Characterization of flocculation in L7:*Flo1* strains.

The threshold of Flo1 expression required for culture sedimentation due to flocculation was determined for L7:*Flo1 S. cerevisiae* exhibiting constitutive LacI expression. Cultures growing at high OD_{600} were induced for 36 hours with increasing concentrations of IPTG. Images shown here are 1ml cultures at 18 hour intervals, 10 minutes following brief vortexing.

Nature Biotechnology: doi:10.1038/nbt.1536

SUPPLEMENTARY METHODS

1. SEQUENCING OF PROMOTER LIBRARIES

Promoter sequences for all clones present in the TetR- and LacI-regulated promoter libraries were determined by amplifying genomic DNA from the integration locus in each clone by PCR and sequencing the promoter section of this DNA using primers annealing to opposing strands upstream and downstream of promoter. In each case, sequencing reads from the two strands were merged to give the promoter sequence for each clone and these sequences are shown aligned in **Supplementary Figures 1 and 2**. The sequencing confirmed massive variation in DNA sequence between defined sites, while the fixed motifs and separating distances were largely maintained as intended.

To confirm that introducing this variation only affects S_{min} and S_{max} properties with our library design technique, we performed a 96-well ATc dose-response experiment against eight of the promoters in the TetR-regulated library (**Supplementary Fig. 3)**. The dose-response curves of each promoter's output with increasing ATc displayed remarkably similar Hill coefficients, indicating that ATc-TetR titration is independent of sequence variation. Additionally, an attempt to find a relationship between sequence and promoter output did not produce any obvious correlations, just as in previous studies[1, 2].

2. GENE NETWORK ASSEMBLY

For library-based network assembly, the required promoter in each case was PCR-amplified from harvested genomic DNA from the appropriate library strain identified in

Nature Biotechnology: doi:10.1038/nbt.1536

screening. Promoter DNA was then inserted into the appropriate vector by standard cloning techniques.

Negative feedforward loop network (**Fig. 2A of main text**): the TetR-regulated promoters were inserted into pTVLI (a modification of pTVGI with yEGFP replaced by LacI), which was integrated into the yeast genome with tryptophan selection. To complete the network, the OR-LT promoter was inserted into pUVGI (a modification of pTVGI with *TetR* removed and *TRP1* replaced by *HIS3*) which was integrated into the genome with histidine selection.

Timer network (**Fig. 3A of main text**): yEGFP was replaced by TetR in pTVGI to obtain pTVTI, into which LacI-regulated promoters were inserted. TetR-regulated promoters were inserted into pTVLI. To create the mutually-repressive network plasmid, the *TEF1* promoter, *TetR*, A*DH1* terminator and *TRP1* were excised from pTVLI and replaced by the entire region of pTVTI spanning from the *GAL1* UAS to the *CYC1* terminator and the *LEU2* selectable marker. The resulting plasmid (pTWLI) was integrated into the genome at the *ura3-52* locus with leucine selection. To include yEGFP under the control of P_{LX} and thereby complete the network, the LacI coding sequence was removed from pLVGI-LX and this was integrated into the genome within the already-integrated pTWLI using tryptophan selection. To produce the reverse network tested in **Supplementary Figure 4**, TetR was removed from pTVGI-TX and this was integrated instead of the LacI-deleted pLVGI-LX.

Nature Biotechnology: doi:10.1038/nbt.1536

Flocculation network (**Fig. 4A of main text**): the three pTWLI constructs used in the **Figures 3B,E,F of the main text** were integrated into the genome at the *ura3-52* locus of L7:*FLO1* transformed yeast to produce networks controlling flocculation expression.

3. MODELING METHODS

Mathematical modeling is the key tool we used to integrate library component characteristics and guide the construction of synthetic gene networks such as negative feedforward loops and genetic timer networks. The library components are experimentally characterized using an inducible switch network, which includes a constitutively expressed repressor (LacI/TetR), a diversified hybrid promoter with inserted repressor binding sites and a downstream fluorescence reporter gene (**Fig. 1B, main text**). Simple experimental measurements of fluorescence levels with and without inducers provide essential quantities (e.g., parameters) for the mathematical models, which use these quantities to extract and quantify various properties of the hybrid promoters (library components) as well as predict the behavior of more complex gene networks. The details of the modeling methods used will be provided in the following sections. Even though gene expression stochasticity in the synthetic gene networks results in some interesting experimental observations, we mainly focus on the population average behaviors, both temporal and at steady state, to demonstrate our diversity-based synthetic biology approach; the mathematical models are thus based on deterministic ordinary differential equations.

Nature Biotechnology: doi:10.1038/nbt.1536

3.1. Basic Inducible Switch for Library Promoters

To characterize the properties of the hybrid promoter libraries, we first construct the mathematical formula to describe the steady-state reporter expression based on biochemical kinetics. Since LacI-regulated and TetR-regulated hybrid promoter libraries are characterized similarly, we explain the technical details in parallel for both libraries.

First we try to capture the quantitative relationship between inducer concentration and the active repressor concentration, because we used the inducer (IPTG/ATc) to inactivate the repressors (LacI/TetR) as a perturbation to the synthetic inducible switches. A general Hill function is used to quantify the fraction of inactive repressors. For example, $f_{IPTG} = [IPTG]^{ni} / (K_{IPTG}{}^{ni} + [IPTG]^{ni})$ represents the fraction of inactive LacI monomers, i.e., those bound with IPTG. Here K_{IPTG} is related to the dissociation constant between LacI and IPTG, and *ni* represents the cooperativity between IPTG and LacI. So with a constant K_{IPTG}, the bigger the [IPTG] is, the bigger f_{IPTG} is, indicating that more LacI molecules are inactive. Similarly, $f_{ATc} = [ATc]^{na} / (K_{ATc}{}^{na} + [ATc]^{na})$ represents the fraction of inactive TetR monomers, i.e., those bound with ATc, where K_{ATc} is related to the dissociation constant between TetR and ATc, and *na* represents the cooperativity between ATc and TetR. Both K_{IPTG} and K_{ATc} have the same units as inducer concentration and imply the required dose of inducer to inactivate half of the repressors. Note that in both expressions the inactive repressor fraction is independent of repressor abundance; it is based on the assumption that the repressor is abundant and the interactions between inducer molecules and repressors happen much faster than other biochemical reactions.

Nature Biotechnology: doi:10.1038/nbt.1536

This assumption is made because *TEF1* is a very strong promoter for the production of LacI/TetR in these inducible switches.

Next we take repressor multimerization into account. TetR molecules form dimers to function as active repressors. The relationship between TetR dimer concentration and monomer concentration follow the expression, $[TetR_2] = [TetR]^2 / K_{DT}$, where [TetR$_2$] represents the TetR dimer concentration and K_{DT} is the dissociation constant for TetR dimerization. Similarly, LacI molecules form tetramers as active repressors. The tetramer concentration of LacI is given by $[LacI_4] = [LacI]^4 / K_{DL}{}^2 K_{TL}$, where K_{DL} and K_{TL} are dissociation constants for LacI dimerization and tetramerization.

Finally we combine the effects of induction and multimerization with repressor-DNA binding dynamics to quantify reporter expression. For example, for the LacI-regulated promoter library, we assume that IPTG binds independently to LacI monomers. The concentration of LacI tetramers without IPTG binding would be $(1-f_{IPTG})^4[LacI]^4/(K_{DL}{}^2 K_{TL})$ and the concentration of LacI tetramers with at least one IPTG molecule binding would be $(1-(1-f_{IPTG})^4)[LacI]^4/(K_{DL}{}^2 K_{TL})$, where f_{IPTG} is the fraction of active LacI tetramer repressors as described. With the assumption that LacI tetramers with 1, 2, 3 or 4 IPTG molecules binding have the same binding affinity with DNA, we can work out that the probability of the LacI-regulated promoter being empty is

$$p_{e,lac} = \frac{1}{1 + \frac{(1-f_{IPTG})^4[LacI]^4}{K_{LDNA}K_{DL}{}^2K_{TL}} + \frac{(1-(1-f_{IPTG})^4)[LacI]^4}{K_{LIDNA}K_{DL}{}^2K_{TL}}} \quad (1)$$

Nature Biotechnology: doi:10.1038/nbt.1536

where K_{LDNA} and K_{LIDNA} are the dissociation constants between LacI tetramer and DNA (without and with IPTG binding, so $K_{LDNA} << K_{LIDNA}$).

Similarly, for the inducible switch regulated by TetR, we can derive the expression for $P_{e,tet}$. Since there are two TetR operator sites, the reaction of repressor binding has the form

$$\begin{aligned} O &\underset{k2}{\overset{TR*k1}{\rightleftharpoons}} O1 \\ O1 &\underset{k4}{\overset{TR*k3}{\rightleftharpoons}} O2 \end{aligned} \tag{2}$$

where *TR* is the TetR dimer without ATc binding (we assume that the TetR dimer bound with Atc cannot bind with DNA, or at least can be ignored). So the probability of the TetR-regulated promoter being empty is

$$p_{e,tet} = \frac{1}{1 + TR^2/(K1K2) + TR/K1} \tag{3}$$

where *K1 = k2/k1* and *K2 = k4/k3* are dissociation constants for TetR dimer and DNA interactions, respectively. *TR* represents the TetR dimer without ATc binding and it equals $(1-f_{ATc})^2[TetR]^2/K_{DT}$. So Eq. (3) can be written as

$$p_{e,tet} = \frac{1}{1 + \frac{(1-f_{ATc})^4[TetR]^4}{K_{DT}{}^2 K1K2} + \frac{(1-f_{ATc})^2[TetR]^2}{K_{DT}K1}} \tag{4}$$

The probabilities of having an empty promoter [Eqs (1) and (4)] directly affect the synthesis strength of downstream genes. The steady-state protein abundance of this gene is $eff\,(p(1-cr)+cr)$, where *eff* is the transcription and translation efficiency of the gene, and *p* is the probability of this operator being empty. It can be either $p_{e,lac}$ in Eq. (1) or $p_{e,tet}$ in Eq. (4). *cr* is the relative synthesis efficiency when the operator is occupied. Eqs

Nature Biotechnology: doi:10.1038/nbt.1536

(1) and (4) are the foundation for simpler formulae in the following sections, and they remain the same for all promoters; only *eff* and *cr* vary for different library components. These approximations are made because repressor inducer interactions are independent of the sequence around the operator sites. We also assume repressor-DNA binding affinity is only marginally affected by mutations around the operator sites and hence this effect can be ignored.

3.2. Negative Feedforward Loop

In the negative feedforward loop motif (**Fig. 2A, main text**), TetR production is constitutively driven by the *TEF1* promoter and LacI production is controlled by the TetR-regulated promoters with two *tet*O_2 sites. yEGFP production is regulated by an engineered promoter with both *OLac* and *tet*O_2; thus, it is regulated by both LacI and TetR.

To first model the production of LacI, we simplify the expression of $p_{e,tet}$ in Eq (4) as

$$p_{e,tet} = \frac{1}{1 + b(a(1 - f_{ATc})^4 + (1 - f_{ATc})^2))} \tag{5}$$

where *a* and *b* are lumped nondimensional parameters representing $[TetR]^2 / (K_{DT} K2)$ and $[TetR]^2 / (K_{DT} K1)$, respectively. By substituting the expression of f_{ATc} into (5) and fitting *p(1-cr)+cr* to the dose response curve using TX as the promoter regulating LacI production (see **Supplementary Fig. 5**), we get an estimation of five parameters as follows: *na* = 3, *a*=464,*b*=768, k_{ATc}=20.4 and *cr*=0.007. Since TetR production is under the control of a constitutive promoter (P_{TEF1}), its total abundance is assumed to remain

constant in the model. Other parameters are biochemical kinetic parameters, which also remain constant for different library components. Now we can get an estimate of total LacI production in the negative feedforward loop: $TLacI=eff*(p_{e,tet}*(1-cr)+cr)$, where the values of *eff* and *cr* depend on the specific TetR-regulated promoter used (TX, T8, etc.).

Next we describe the regulation of yEGFP production mathematically. yEGFP expression in the negative feedforward loop is regulated by the hybrid OR-gate promoter which contains both TetR and LacI operator sites. Similarly, the downstream gene expression is determined by how the promoter is occupied by the repressors. The state of this promoter is modeled with four discrete states (**Supplementary Fig. 7**), specifically, no repressor binding (s1), only LacI binding (s2), only TetR binding (s3), or both LacI and TetR binding (s4). The possible transitions between these states and the corresponding kinetic rates are indicated in **Supplementary Figure 7**.

To determine how much yEGFP will be produced, we first compute the probability that the promoter is at a certain state at any given time. The state transition is essentially the evolution of a four-state markov chain. The probability of the promoter at each of the four states (P_i, $i=1,2,3,4$) can be described using equations:

$$(k_{12}+k_{13})P_1 = k_{21}P_2 + k_{31}P_3 \tag{6}$$

$$(k_{21}+k_{24})P_2 = k_{12}P_1 + k_{42}P_4 \tag{7}$$

$$(k_{42}+k_{43})P_4 = k_{24}P_2 + k_{34}P_3 \tag{8}$$

$$P_1 + P_2 + P_3 + P_4 = 1 \tag{9}$$

Nature Biotechnology: doi:10.1038/nbt.1536

where k_{42} and k_{31} are the dissociation rates between the TetR dimer and DNA. Similarly, k_{43} and k_{21} are dissociation rates between the LacI tetramer and DNA, k_{24} and k_{13} are association rates between TetR and DNA, and k_{34} and k_{12} are association rates between LacI and DNA. Eqs (6)-(8) formulate the transitions as a four-state continuous-time Markov chain. Eq. (9) ensures that the system satisfies the law of total probability. To simplify the notation, we can obtain the following relationship by assuming that LacI and TetR bind with DNA independently:

$$\begin{aligned} k_{42} &= k_{31} = k_{DT} \\ k_{43} &= k_{21} = k_{DL} \\ k_{24} &= k_{13} = k_{AT} \\ k_{34} &= k_{12} = k_{AL} \end{aligned} \tag{10}$$

where k_{DT} and k_{DL} are the dissociation rates between TetR/LacI and DNA, and k_{AT} and k_{AL} are the association rates between TetR/LacI and DNA. These parameters not only depend on the association equilibrium constant; they also depend on active TetR/LacI abundance. By defining $k_{EL}=k_{AL}/k_{DL}$ and $k_{ET}=k_{AT}/k_{DT}$ and substituting (10) into (6)-(9), we can solve the probability expression for each state

$$\begin{aligned} P_1 &= \frac{1}{k_{EL}k_{ET} + k_{EL} + k_{ET} + 1} \\ P_2 &= \frac{1}{k_{ET} + 1 + \frac{k_{ET}}{k_{EL}} + \frac{1}{k_{EL}}} \\ P_3 &= \frac{1}{k_{EL} + 1 + \frac{k_{EL}}{k_{ET}} + \frac{1}{k_{ET}}} \\ P_4 &= \frac{1}{\frac{1}{k_{ET}} + 1 + \frac{1}{k_{EL}} + \frac{1}{k_{EL}k_{ET}}} \end{aligned} \tag{11}$$

Because K_{EL} depends on LacI abundance and K_{ET} depends on TetR abundance, their expressions can be expanded as

$$K_{EL} = k_{al}((1 - f_{IPTG})LacI)^4 \quad (12)$$

$$K_{ET} = k_{at}(1 - f_{ATc})^2 \quad (13)$$

where k_{al} = 0.004 is LacI-DNA association equilibrium, and k_{at} = 150 is TetR-DNA association equilibrium times total TetR abundance. Since total TetR abundance is assumed to be constant (because of the constitutive promoter P_{TEF1}), we can lump these two parameters together as one single parameter, k_{at}. LacI abundance in Eq. (12) is explicitly calculated as described above based on Eq. (5). It is worth pointing out that the parameter k_{at} includes abundance of TetR and hence represents a different parameter from k_{al}, even though they share some similarities. These two parameters are generic parameters within a biological regime, but have different units to ensure that final probabilities are nondimensional. Also k_{iptg}=0.01 and ni=1. These are inferred from fitting LacI-inducible switches (see next section and **Supplementary Fig. 6**). Even though LacI-inducible switches use a different promoter, the parameters k_{iptg} and ni are still applicable to our negative feedforward loop system because the interaction between IPTG and LacI remains the same.

In the end, yEGFP production is described by the expression: $effg*(P_1*s_1+ P_2*s_2+ P_3*s_3+ P_4*s_4)$. Here $effg$ is yEGFP production efficiency, P_i $(i=1,2,3,4)$ is the probability of state i, and s_i (i=1,2,3,4) is the relative production strength of yEGFP when the promoter is at state i. We used generic parameters for the other constants and predict the behavior of the negative feedforward loop. The parameters we used are: k_{iptg}=0.01,

Nature Biotechnology: doi:10.1038/nbt.1536

k_{al}=0.004,k_{at}=150, s_1=1, s_2=0.02, s_3=0.6, and s_4=0.01. The reason s_2 is much smaller than s_3 is that from our experimental characterization of the OR-gate promoter we found LacI repression of yEGFP production to be much tighter than TetR. With the expression for yEGFP production, we use a grid of ATc and IPTG doses as inputs to generate the expression landscape shown in **Figure 2B of main text**.

3.3. Mutual Inhibitory Networks

To capture the temporal dynamics of the mutual inhibitory dynamics with limited information about the components, we use a further simplified form to construct the differential equations. For the mutual inhibitory networks underlying the genetic timers (**Fig. 3A, the main text**), the production of LacI (symbolized as L) and TetR (symbolized as T) can be described by the following two ordinary differential equations:

$$\begin{aligned} \frac{dL}{dt} &= crl + p_{e,tet}(cil - crl) - \delta L \\ \frac{dT}{t} &= crt + p_{e,lac}(cit - crt) - \delta T \end{aligned} \tag{14}$$

where *crl*, *cil*, *crt* and *cit* are synthesis rates for LacI and TetR, with or without repressor binding the promoter, respectively. These parameters are easily estimated from library component measurements and listed in Table1 (S_{min} and S_{max}). The first and second terms on the right of these equations represent the synthesis of repressor with or without repressor binding, respectively. The third term represents the exponential decay of the repressors. Yeast cells grown in galactose media have doubling times of about 6 hours. Because both TetR and LacI are very stable proteins, the decrease of intracellular abundance of these repressors is through cell division, so here the degradation rate is

Nature Biotechnology: doi:10.1038/nbt.1536

δ=0.002 min^{-1}. Transcription and translation are combined as one step to simplify the equations. The dynamics of yEGFP are described by the same equations as (14) except the synthesis terms are constant regardless of which library component is used; this is because yEGFP is driven by LX. From previous sections, we know that $p_{e,tet}$ is a function of TetR abundance and $p_{e,lac}$ is a function of LacI abundance. To further simplify the analysis, we used a more general form of $p_{e,tet}$ and $p_{e,lac}$, e.g.,

$$p_{e,tet} = \frac{1}{1+\left(\frac{\frac{T}{(1+\frac{ATc}{k_{ATc}T})^m}}{k_T}\right)^{nt}} = \frac{1}{1+\left(\frac{T}{kT}\right)^{nt}\left(1+\frac{ATc}{k_{ATc}T}\right)^{-mnt}} \quad (15)$$

where T represents TetR abundance, k_{ATc} represents the ATc-TetR dissociation rate and k_T represents TetR-DNA binding affinity. The powers m and nt represent binding cooperativity between inducer and repressor (m) and between repressor and DNA (nt), respectively. From fitting the dose response curves for library components, we know $m*nt$ is about 11.5. This indicates that TetR has very high cooperativity with ATc, itself or DNA. So in the later fittings, the range for nt is set between 2 and 10 (details below). In the experiments, we only use full induction of ATc to force the system to a specific state. With k_{ATc} set to be small, e.g., when ATc = 250 ng/ml as we used in the experiments, *(1+ATc/(kATc*T))^(-m*nt)* is very close to zero. On the other hand, without ATc, *(1+ATc/(kATc*T))^(-m*nt)* equals 1. So $p_{e,tet}$ can be further simplified as

$$p_{e,tet} = \frac{1}{1+\left(\frac{T}{kT}\right)^{nt}\Delta_{ATc}} \quad (16)$$

where Δ_{ATc} is a delta function of ATc. It equals 1 when there is no ATc and 0 with ATc; the parameter *kT* is to be determined. Similarly,

$$p_{e,lac} = \frac{1}{1+\left(\frac{L}{kL}\right)^{nl} \Delta_{IPTG}} \tag{17}$$

From the data for the LacI-inducible switches, we found that LacI has much smaller cooperativity with IPTG, itself or DNA. So in later fittings, the range for *nl* is set to be between 1 and 3.

3.4. Parameter Fittings

To investigate the effects of different library components on the resetting temporal dynamics of the genetic timer networks, we examine a key parameter, the expression ratio (*r* = (*cit-crt)/(cil-crl)*). By dividing all variables by (*cil-crl)*, Eqs (14) can be written as

$$\begin{aligned} \frac{dL'}{dt} &= crl' + p_{e,tet} - \delta L' \\ \frac{dT'}{t} &= crt' + r \bullet p_{e,lac} - \delta T' \end{aligned} \tag{18}$$

where all parameters and variables are divided by (*cil-crl)* (indicated by prime). In Eq. (18), parameters *crl*', *crt*' and *r* can be computed from the values in **Table 1** of the main text. The S_{min} values are very small compared to the S_{max} values. Considering the fact that the cell auto-fluorescence can be at a range of 2-4 and with the ratio being the critical parameter, we set *crl*' and *crt*' to be 0.005 for all constructs. However, parameters *nt*, *nl*, *kL* and *kT* are unknown. As implied in Eq. (15), these parameters cannot be inferred from

Nature Biotechnology: doi:10.1038/nbt.1536

the dose response curves of individual components (**Supplementary Fig. 5 and 6**). So in order to predict the behavior of genetic timer networks assembled from many different components, we first need to estimate these parameters from one construct, the control construct with engineered wild-type promoters (TX-LX).

To estimate these parameters, we need to specify the possible range of each parameter. As described in previous sections, the ranges for *nl* and *nt* are set with information from the dose response curves of library components. With δ equal to 0.002 min^{-1}, the normalized protein abundance is 500. So we set the range of both *kL* and *kT* to be between 1 and 100.

With the bound for each parameter to be set, we fit the model against the experimental measurements of the construct with engineered wild-type promoters (TX-LX). The experimental characterization of the temporal response of TX-LX is shown in **Supplementary Figure 4,** and the normalized resetting period is shown in **Figure 3B of the main text**. The initial decrease of yEGFP levels after removal of ATc is probably due to the stress caused by washing and tight binding between ATc and $tetR^3$, and is independent of the synthetic gene network considered; so we treat the first 12 hour results as constant and fit the data after the initial 12 hours. 200 sets of initial guesses of the four parameters are used to fit the data. The sum of square of the residuals (SSQ) is used to rank the fittings from ones with low error to ones with high error (**Supplementary Fig. 8**). **Supplementary Figure 8** shows the increase of SSQ for the fittings that did not converge to a reasonable local minimum. It is easy to identify that fittings with SSQ

greater than 0.0022 have dramatically larger error and hence be discarded, indicated by the red cutoff line in **Supplementary Figure 8**. Eventually there are 79 parameter fittings left as candidates to predict the temporal behavior of other constructs. The distribution of these parameters is plotted in **Supplementary Figure 9**.

3.5. Bottleneck Effects

We use the steady-state value of the system with various values of the ratio *r* (the bifurcation diagram) to investigate the relationship between *r* and the system's temporal dynamics. Using 79 sets of candidate parameters, we computed all 79 bifurcation diagrams. The diagrams shown in **Supplementary Figure 10** are the ones with the leftmost and rightmost lower bifurcation points, demonstrating that the right bifurcation point has a ratio between 0.05 and 0.07.

With a ratio greater than 0.07, the system has default low LacI abundance (**Supplementary Fig. 10**) and high TetR and yEGFP abundance. By adding a full dose of ATc induction, the initial state is forced to be opposite of the default state. After removal of the ATc induction, the system goes back to its default state. If the ratio is greater than but close to 0.07, the network will evolve for a period of time as if it has a ratio smaller than the bifurcation ratio. Eventually it will approach its steady state after this initial lag, which is called the bottleneck effect. The consequence of this bottleneck is that yEGFP dynamics exhibits a prolonged lag before the system starts to increase towards its default state. The time the system spends going through the lag depends on how close the actual ratio is to the bifurcation ratio.

Nature Biotechnology: doi:10.1038/nbt.1536

To determine the range of the bifurcation ratio, we use the parameter candidates from fitting TX-LX to locate the bifurcation point and hence predict the temporal behavior of other possible constructs. We chose a variety of possible combinations, and our model predicted the temporal dynamics of the other constructs with high accuracy. We used the parameters for the bifurcations diagrams in **Supplementary Figure 10** to predict the fastest and slowest possible dynamics for four other constructs with different ratios. As can be seen from **Figure 3G of the main text**, the network temporal behavior is sensitive to the ratio as it gets close to 0.07. However, the model still predicts the correct range, verified by experimental measurements.

It is possible to approximate the time spent passing the bottleneck using a formula taking into account the scaling law of bottleneck effects[4]. The formula is

$$T_{total} = C_1 + \frac{C_2}{\sqrt{|r - C_3|}} \tag{19}$$

where C_1 represents the constant time spent to degrade the leftover repressors, and the second term represents the time spent going through the bottleneck. This time obeys a square root scaling law. In the formula, C_2 is a scaling factor and C_3 is the bifurcation ratio. The bigger the difference between the actual ratio and the bifurcation ratio, the further away the system is from the bifurcation point and hence the time spent going through bottleneck is smaller. The values that can approximate the experimental data are C_1=36.51, C_2=15.2 and C_3=0.068. A simple rearrangement of Eq. (19) gives us $\log(T_{total} - C_1) + \frac{1}{2}\log(|r - C_3|) = \log C_2$. So the total reset time and promoter strength

ratios are linearly correlated at log scale. This illustrates that the temporal resolution we can achieve for our timer depends on the temporal objective. For example, for a timer with reset time over 90 hours, our library can potentially achieve a temporal resolution below 4 hours; but for a timer with reset time below 50 hours, our library can, theoretically, achieve a resolution of minutes or even seconds. However, such high resolution is unlikely due to measurements limitations and variations of experimental conditions.

4. PREDICTION OF FLOCCULATION TIMING

To test the dynamic range of flocculation, pLVGI was integrated into L7:*FLO1* strains (construction described in the main text) at the *ura3-52* locus giving constitutive *TEF1* promoter-driven expression of LacI. This strain, when cultured in increasing concentrations of IPTG, showed sedimentation due to flocculation at concentrations ≥ 0.12 mM after 36 hours (**Supplementary Fig. 11**). This IPTG concentration corresponds to a normalized fluorescence of 0.3554 from LacI-regulated promoters created in this study (**Supplementary Fig. 6**), which is equivalent to yEGFP expression of 49.75 A.U from the L7 promoter (equivalent expression is S_{max}L7 multiplied by normalized fluorescence). To predict when flocculation will occur based on the timer networks, the data shown in **Figure 3B of the main text** were rescaled to match the expression outputs of the L7 promoters using the following equation:

$$E_P = [(E_M - S_{min}M) / (S_{max}M - S_{min}M)] * (S_{max}P - S_{min}P) + S_{min}P$$

where E_M is measured expression from promoters (M) and E_P is predicted expression from promoters (P). For TX-LX, TX-L14 and T7-L18, M = LX and P = L7. Predicted

expression (E_P) for the four constructs was plotted as a function of time in **Figure 4B of the main text,** along with shading indicating 49.75 A.U.

5. OLIGONUCLEOTIDES USED IN THIS STUDY

TL1 for TetR-regulated library construction:

TCTGGGGTACTGCAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
TATAAANNNNNNNNNNNTCCCTATCAGTGATAGAGANNTCCCTATCAGTGAT
AGAGA

TL2 for TetR-regulated library construction:

ATAGGATCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNATTTTTTGTTGATACTTTTANNNNNNNNNNNNNNNNNNNNNTCTCTA
TCACTGAT

LL1 for LacI-regulated library construction:

TCTGGGGTACTGCAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
TATAAANNNNNNNNNNNNNNNNTGTGTGGAATTGTGAGCNGATAACAATTT
CACACA

LL2 for LacI-regulated library construction:

ATAGGATCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNATTTTTTGTTGATACTTTTANNNNNNNNNNNNNNNNNNNNNTGTGTG
AAATTGTT

FL1 for integration of L7 promoter in place of the genomic *FLO1* promoter:

GTGCCAGAAGTGTAAAGACTGCCAAAAACATATAGCGATGAGGCATTGTCAT

CATTTTGGGATCCTTTGTATTACTTACTGAAGG

FL2 for integration of L7 promoter in place of the genomic *FLO1* promoter:

CTTCCGGGTTCTTATTTTTAATTCTTGTCACCAGTAAACAGAACATCCAAAAG

AATTCGAGCTCGTTTAAAC

List of key genes/proteins: FLO1; Yeast Enhanced Green Fluorescent Protein (yEGFP); Tetracycline Repressor (TetR); Lac Inhibitor (LacI).

REFERENCES

1. Jensen, P.R. & Hammer, K. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Appl Environ Microbiol* **64**, 82-87 (1998).
2. De Mey, M., Maertens, J., Lequeux, G.J., Soetaert, W.K. & Vandamme, E.J. Construction and model-based analysis of a promoter library for E. coli: an indispensable tool for metabolic engineering. *BMC Biotechnol* **7**, 34 (2007).
3. Berens, C. & Hillen, W. Gene regulation by tetracyclines. Constraints of resistance regulation in bacteria shape TetR for application in eukaryotes. *Eur J Biochem* **270**, 3109-3121 (2003).
4. Strogatz, S.H. Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering. (Addison-Wesley Pub., Reading, Mass.; 1994).

Nature Biotechnology: doi:10.1038/nbt.1536