nature biotechnology

Diversity-based, model-guided construction of synthetic gene networks with predicted functions

Tom Ellis[1,2], Xiao Wang[1,2] & James J Collins[1]

Engineering artificial gene networks from modular components is a major goal of synthetic biology. However, the construction of gene networks with predictable functions remains hampered by a lack of suitable components and the fact that assembled networks often require extensive, iterative retrofitting to work as intended. Here we present an approach that couples libraries of diversified components (synthesized with randomized nonessential sequence) with *in silico* modeling to guide predictable gene network construction without the need for *post hoc* tweaking. We demonstrate our approach in *Saccharomyces cerevisiae* by synthesizing regulatory promoter libraries and using them to construct feed-forward loop networks with different predicted input-output characteristics. We then expand our method to produce a synthetic gene network acting as a predictable timer, modifiable by component choice. We use this network to control the timing of yeast sedimentation, illustrating how the plug-and-play nature of our design can be readily applied to biotechnology.

Synthetic biology promises to transform biotechnology by applying engineering principles to biological systems[1]. In less than a decade, this field has already yielded technological applications, providing new avenues for drug manufacture[2,3], biofabrication[4] and therapeutics[5,6], while also showing promise in alternative energy[7]. A major focus of the field is the synthesis of gene networks with predictable behavior[8–10], either to endow cells with novel functions[11–15] or to study analogous natural systems[8,16–19]. Despite a booming community and notable successes, the basic task of assembling a predictable gene network from biomolecular parts remains a considerable challenge and often takes many months before a desired network is realized[20]. If synthetic biology is to advance, it is essential to identify techniques that increase the predictability of gene network engineering and decrease the amount of hands-on molecular biology required to get a functional network up and running.

Current approaches to gene network construction typically use a small set of components taken from different natural systems, which are then assembled and tested *in vivo*, often without guidance from a priori mathematical modeling[13,21]. Networks rarely behave as intended the first time, usually because chosen parts have the correct function but lack the specific quantitative properties required. Even for those few synthetic biology studies that do involve computational assistance[22–25], *in silico* results have been mainly used for data interpretation, not for guiding design and assembly. Instead, in most projects, an initial failed network is usually resolved over months of iterative retrofitting[20], often by fine-tuning imperfect parts by mutation, identifying alternative parts or adding extra features to counterbalance the problems. Directed evolution has been shown to provide a short-cut through this phase[21] but is complicated by the additional work needed to couple networks to selective pressures.

This time-consuming *post hoc* tweaking phase stems in part from having to work with a limited set of imperfect components. Although this lack of reliable parts is being addressed by community efforts[26], it remains an acute problem because most of the available components are inadequately characterized. For example, many promoters are simply characterized as being 'weak' or 'strong'. What is needed to resolve this problem and fast-track synthetic biology is an approach that creates libraries of components ahead of any assembly. Then, by starting with a finer granular range of choices for each component, modeling can be used to quickly pick out the correct part needed to generate the intended network function. This approach offers the added attraction of allowing substantially different network outcomes to be chosen in advance, simply by selecting functionally equivalent components with slightly different properties. This exploits a feature common to many types of finely balanced networks, where small changes to one component can have a large impact on the behavior of the entire system.

Using regulated promoters as our example, we describe here how a simple synthesis technique can be used to rapidly create and characterize component libraries for synthetic biology. Working in *S. cerevisiae*, we demonstrate how such libraries can be teamed with predictive modeling to rationally guide the construction of gene networks that have diverse outputs. We also illustrate a plug-and-play application for one of our network designs by using it to control the timing of yeast sedimentation.

[1]Howard Hughes Medical Institute and the Department of Biomedical Engineering, Center for BioDynamics and Center for Advanced Biotechnology, Boston University, Boston, Massachusetts, USA. [2]These authors contributed equally to this work. Correspondence should be addressed to J.J.C. (jcollins@bu.edu).

Received 26 January; accepted 26 March; published online 19 April 2009; doi:10.1038/nbt.1536

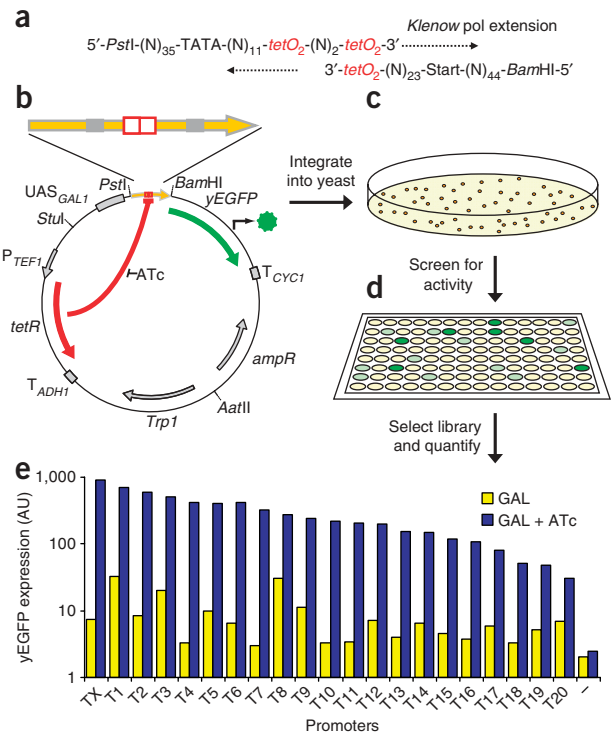

Figure 1 Regulatory promoter library synthesis, screening and characterization. Schematic design for TetR-regulated promoter synthesis. (**a**) Promoters are constructed by Klenow polymerase extension from two overlapping 110-mer oligonucleotides synthesized with unspecified nucleotides (N) between defined motifs. (**b**) Promoters are ligated between UAS_{GAL1} and the yEGFP coding sequence in the characterization vector, which also expresses TetR from the *TEF1* promoter (P_{TEF1}). (**c**) Transformation of yeast yields thousands of colonies with the genome-integrated vector. (**d**) Individual colonies are screened in 96-well format by measuring fluorescence in induction conditions after 22 h growth (media supplemented with 2% galactose + 250 ng/ml ATc). (**e**) A library of 20 regulated promoters covering a range of expression levels is selected from screening data and quantitatively characterized using flow cytometry measurement of yEGFP expression after 22 h growth in medium supplemented with 2% galactose (GAL) and with 250 ng/ml ATc (GAL + ATc). TX, control promoter; T1–T20, library promoters; –, null promoter.

RESULTS

Parallel synthesis and characterization of promoter libraries

To demonstrate our library-modeling approach, we focused on regulated promoters, as they typically control gene network logic and modulate responses to stimuli. Promoter libraries have been created using DNA-shuffling/combinatorial approaches[27–29] and mutation-selection techniques[30–33], but these approaches do not allow for rational *de novo* design. We modified an efficient synthesis-with-degenerate-sequence method[32] to yield libraries of regulatory promoters that have a range of inputs and outputs. In this technique, promoters are constructed with runs of unspecified ('N') sequence-separating key motifs[32]; the fixed motif sequences ensure promoter function, and the random bases surrounding them modulate their efficiency, presumably by subtly altering local DNA conformation[34].

Our first library was designed to yield yeast promoters repressed by TetR (Tn*10*.B tetracycline repressor[35]) and inducible with the TetR-inhibitor anhydrotetracycline (ATc). We used a Klenow-based synthesis method, using inexpensive oligonucleotides[31,34], to build promoters containing the TATA box and start site from the commonly used *GAL1* promoter[36]. To introduce controlled regulation, we placed two tandem TetR operators (Tn*10* operator $tetO_2$) into the promoter at positions previously shown to provide tight repression[29].

A schematic of our library synthesis technique is shown in **Figure 1** and detailed in the Methods. To characterize the strength of every promoter, each one was cloned upstream of yeast enhanced green fluorescent protein (yEGFP)[37] and downstream of a *GAL1* upstream activation signal (UAS) in a vector that also contained the strong constitutive *TEF1* promoter[38] directing TetR expression. Constitutive expression of TetR ensures low basal levels of yEGFP, which can be relieved by adding ATc to the medium. The *GAL1* UAS allows us to avoid the effects of nucleosomes as it directs their removal from the promoter in the presence of galactose[39]. We arbitrarily chose to build a library of 20 promoters (T1–T20), covering a wide range of expression and inhibition levels. By measuring yEGFP levels with flow cytometry, we quantitatively determined every promoter's minimum (TetR repressed) output (S_{min}) and maximum (TetR unrepressed) output (S_{max}) (**Table 1**). The approach was designed to yield interchangeable promoters that are identical except for S_{min}, S_{max} and intermotif sequences. This was confirmed by comparison to TX, a control promoter retaining the *GAL1* wild-type sequence between defined motifs, as well as by DNA sequencing (**Supplementary Methods**) and dose-response curves (**Supplementary Methods**).

Table 1 Maximum and minimum output values of members of the promoter libraries

TetR-regulated promoters (P_{LibT})					LacI-regulated promoters (P_{LibL})				
Promoter	S_{max}	s.e.m.	S_{min}	s.e.m.	Promoter	S_{max}	s.e.m.	S_{min}	s.e.m.
TX	**918.00**	33.83	**7.46**	0.46	LX	**717.38**	21.06	**13.06**	0.77
T1	**694.23**	19.89	**32.79**	2.58	L1	**399.90**	25.02	**11.11**	0.60
T2	**595.79**	17.07	**8.38**	0.50	L2	**372.59**	16.87	**9.71**	0.11
T3	**506.31**	27.48	**20.22**	2.16	L3	**292.11**	11.60	**83.05**	1.09
T4	**421.78**	5.83	**3.26**	0.16	L4	**282.01**	13.61	**50.55**	1.92
T5	**408.04**	22.91	**9.87**	0.41	L5	**246.73**	6.42	**151.75**	2.77
T6	**418.60**	16.63	**6.46**	1.68	L6	**228.45**	15.37	**23.79**	0.31
T7	**319.66**	13.41	**3.04**	0.15	L7	**139.99**	8.43	**5.40**	0.35
T8	**277.75**	12.94	**30.88**	1.75	L8	**141.86**	6.23	**7.67**	0.35
T9	**244.21**	11.79	**11.34**	0.62	L9	**134.04**	9.73	**23.54**	1.55
T10	**216.99**	7.34	**3.27**	0.18	L10	**129.13**	8.04	**4.96**	0.30
T11	**203.14**	6.90	**3.41**	0.18	L11	**108.27**	4.18	**5.74**	0.45
T12	**201.76**	3.75	**7.08**	0.53	L12	**107.35**	4.73	**5.07**	0.36
T13	**154.46**	12.15	**4.01**	0.23	L13	**103.58**	9.54	**4.37**	0.29
T14	**151.03**	10.36	**6.42**	0.19	L14	**82.32**	1.50	**4.15**	0.23
T15	**118.93**	5.85	**4.62**	0.19	L15	**70.91**	4.42	**20.83**	0.96
T16	**108.22**	3.40	**3.71**	0.13	L16	**72.03**	3.05	**4.28**	0.23
T17	**81.70**	3.39	**5.91**	0.27	L17	**56.97**	1.77	**5.15**	0.36
T18	**51.75**	3.27	**3.26**	0.25	L18	**47.16**	1.33	**3.91**	0.28
T19	**48.29**	1.10	**5.13**	0.89	L19	**44.10**	2.25	**4.25**	0.20
T20	**30.69**	0.40	**6.95**	0.45	L20	**37.08**	2.12	**9.41**	0.69
TEF1	**287.38**	14.38							

Cultures grown in 2% galactose-supplemented medium for 22 h were measured for median yEGFP expression by flow cytometry; values are the means of three biological repeats with s.e.m. given. For characterization purposes, the repressor of interest (TetR or LacI) was constitutively expressed from the *TEF1* promoter, which was also measured in this study and found to remain constant in all growth conditions used. Maximum output (S_{max}) is measured with saturating concentrations of repressor inhibitor (250 ng/ml ATc for TetR, 10 mM IPTG for LacI); minimum output (S_{min}) is measured without inhibitors.

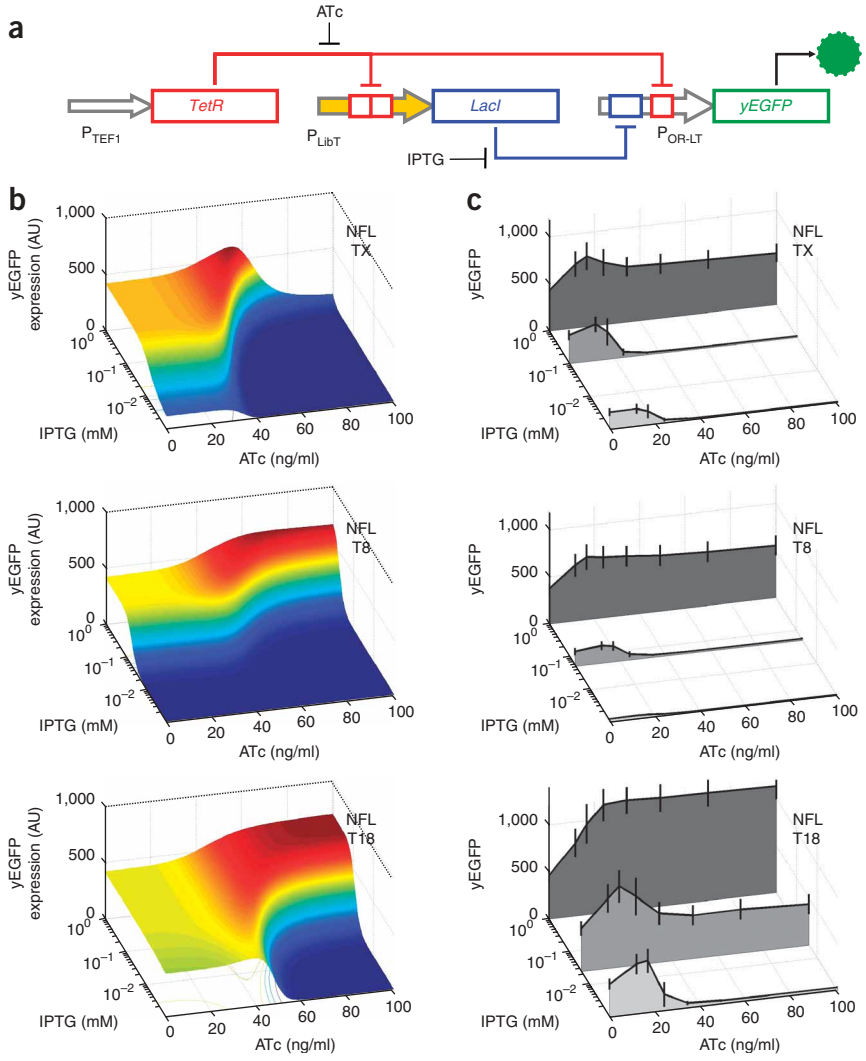

Figure 2 Modeling and synthesis of feed-forward loop networks using a promoter library. TetR-regulated promoter library data were used in conjunction with *in silico* modeling to construct negative feed-forward loop (NFL) gene networks with different predicted input-output functions. (**a**) Schematic of the network, where P_{TEF1} is *TEF1* promoter; P_{LibT} is promoter from TetR-regulated promoter library; and $P_{OR\text{-}LT}$ is LacI-TetR dual-regulated OR-gate promoter. (**b**) *In silico* modeling of the network from component properties predicts yEGFP expression (output) in response to varied concentrations of ATc and IPTG (inputs) when three different TetR-regulated promoters are used. (**c**) The three networks were assembled in *S. cerevisiae,* and median yEGFP expression was measured by flow cytometry after 22 h growth of cells in medium supplemented with 2% galactose plus varying concentrations of ATc and IPTG. Error bars show the s.d. of the gated cell population.

Feed-forward loop networks

To demonstrate how our approach can be applied in a gene network, we used the TetR-regulated promoter library with *in silico* modeling to investigate the incoherent type II negative feed-forward loop network. This is a genetic motif found in *S. cerevisiae* and mammalian cells that consists of an output gene regulated by two repressor genes, one of which is also inhibited by the other[40].

For our network, we used TetR and a eukaryotic-optimized version of the *Escherichia coli* Lac inhibitor[41] (*lacI*) as repressors (**Fig. 2a**). Each protein repressed yEGFP expression by regulating a yeast hybrid *GAL1*-based promoter ($P_{OR\text{-}LT}$) containing both the K-12 *E. coli* O_1 Lac operator (*lacO*) and a *tetO*$_2$ site. Thus, *GAL1*-driven yEGFP is downregulated in the presence of either TetR or LacI (that is, the synthetic promoter acts as an 'OR' gate). TetR expression was constitutive from the *TEF1* promoter, whereas LacI expression was driven by a promoter (P_{LibT}) selected from our library of TetR-regulated promoters. By varying the concentrations of two inputs—ATc and the LacI inhibitor isopropyl β-D-1-thiogalactopyranoside (IPTG)—the repressive effects of TetR and LacI could be tuned, modulating yEGFP expression output.

Before any network assembly, we used component properties from our experimental characterization steps to build a mathematical model to predict how network output would change when input levels (ATc/IPTG) and promoter properties were varied. The model served as a guide, predicting which components from the library could be selected to yield different network outcomes, and what dosage of ATc and IPTG would be most experimentally informative. In the model, the experimentally determined S_{max} and S_{min} values for the promoters (**Table 1**) were used; generic values were assumed for other parameters (see **Supplementary Methods** for modeling details).

A simulation with P_{LibT} = TX (control promoter: S_{max} 918.0, S_{min} 7.46) leads to an interesting nonmonotonic expression landscape with an output peak at intermediate inputs (**Fig. 2b**). This occurs because TetR has simultaneously opposing effects on yEGFP output—inhibiting the production of yEGFP by binding to $P_{OR\text{-}LT}$, while also relieving LacI inhibition of yEGFP by binding to P_{LibT} and repressing LacI production. This is consistent with previous *in situ* studies of naturally occurring negative feed-forward loops[42]. Our synthetic library-modeling approach enables the investigation of this motif without the hindrance of interconnected regulatory networks[42–44].

By changing the S_{min} and S_{max} values of P_{LibT} in the model, we can computationally examine how different promoters from our library affect this response landscape. The model predicts significant differences in output among the members of the promoter library. Two examples most divergent from the TX simulation are shown in **Figure 2b**. Increasing the S_{min} value of the TetR-regulated promoter removes almost all expression in low concentrations of IPTG (P_{LibT} = T8: S_{min} 30.88), whereas decreasing the S_{max} value shifts peak expression to occur only at higher concentrations of ATc (P_{LibT} = T18: S_{max} 51.75). This demonstrates quantitatively that the same external induction (ATc or IPTG) can elicit very different responses from the motif simply due to small changes in promoter strength.

To test these *in silico* predictions, we assembled three negative feed-forward loop networks (**Fig. 2b**) using corresponding components from our libraries and quantified their output responses to varied ATc/IPTG inputs using flow cytometry measurement of yEGFP. The experimental data (**Fig. 2c**) correlated well with our computational

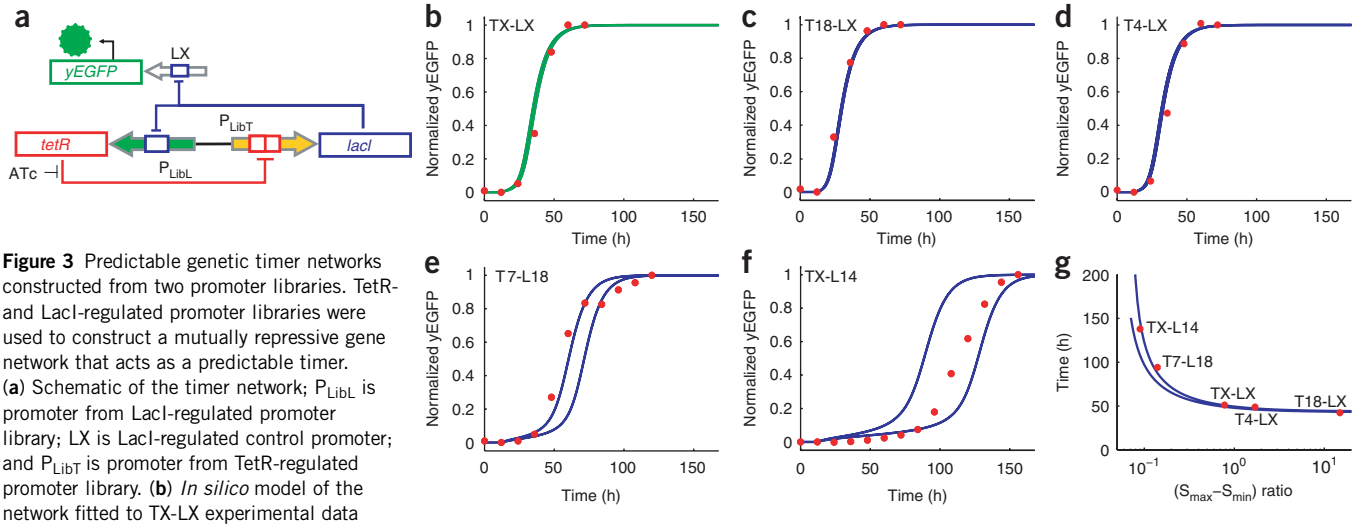

Figure 3 Predictable genetic timer networks constructed from two promoter libraries. TetR- and LacI-regulated promoter libraries were used to construct a mutually repressive gene network that acts as a predictable timer. (**a**) Schematic of the timer network; P_{LibL} is promoter from LacI-regulated promoter library; LX is LacI-regulated control promoter; and P_{LibT} is promoter from TetR-regulated promoter library. (**b**) *In silico* model of the network fitted to TX-LX experimental data (**Supplementary Methods**) shows yEGFP expression changing over time after ATc induction is removed at time 0. Yeast cells with the TX-LX network genomically integrated were grown for 36 h with 250 ng/ml ATc to induce the network, washed three times and monitored starting from time 0 until the expression state reset to a maximum. Normalized yEGFP output (red circles), which was calculated from flow cytometry measurements taken every 12 h, matches the model output (green lines) well. Both upper and lower bounds of model fittings are plotted (**Supplementary Methods**). (**c**–**f**) Median yEGFP expression (red circles) was measured by flow cytometry every 12 h for four different promoter combinations (T18-LX, T4-LX, T7-L18 and TX-L14). Cultures were induced and treated the same as in **b**. Blue lines are model predictions based on parameters inferred from **b**. Lower and upper bounds use the parameters corresponding to lower and upper bounds in **b**. (**g**) The relationship between the reset time and the ratio for all of its values is plotted. Lower and upper bounds use the parameters corresponding to lower and upper bounds in **b**. The reset time can be approximated as $T = C_1 + C_2/\sqrt{(|r - C_3|)}$, where T = reset time, C_1 = basal reset time, C_2 = scale factor and C_3 = bifurcation ratio (**Supplementary Methods**).

predictions, particularly considering the many nonfitted generic parameters used in the model. These results demonstrate how small changes in promoter strength can have dramatic consequences on network responses. These findings also show that a model built from just generic parameters and data from individual components can offer insights into a network response landscape, and that such a model, when teamed with component libraries, can serve as a useful, rapid guide for producing networks with different predictable characteristics.

Genetic timer networks

Having demonstrated our approach in a relatively simple network using one promoter library, we next used two promoter libraries in a more complex network with a richer set of dynamics. Using the mutual-repression motif of the genetic toggle switch[14], we set out to produce predictable genetic 'timers'. These timers exploit the finely balanced nature of a mutual inhibitory network[14], where changes in opposing repressor levels can disrupt bistability, and memory of induction can be lost as the system resets to its original default state. These timers are effectively genetic toggle switches operating in a monostable regime, and their rate of resetting is directly related to relative expression levels of the two repressors—the further they are from the balanced values required for bistability[14], the more rapidly memory of induction is lost.

For yeast timers, we used LacI and TetR as the two mutually repressive gene products (**Fig. 3a**). LacI is expressed from a TetR-regulated promoter (P_{LibT}) selected from our component library described above, and TetR is expressed from a LacI-regulated promoter (P_{LibL}) taken from the second component library. This second library of promoters (L1–L20 plus control LX) was synthesized and characterized as before, but with the Lac operator (*lacO*) in place of the tandem Tet operators (see **Table 1** for promoter data and **Supplementary Methods** for sequences). To follow the expression state of the timers, we placed yEGFP under the control of the LX promoter, giving an expression read-out directly correlated to TetR expression (**Fig. 3a**).

An initial mathematical model based on the properties of components from both libraries gave us qualitative insights into how timers can be set by means of imbalanced mutual inhibition (**Supplementary Methods**). It revealed that changing the ratio of expression from the two opposing promoters affects the reset time. However, due to the complexity of the biochemical reactions involved, a model built solely from component data cannot capture important quantitative features of the network. In particular, the model was not able to accurately predict by how much the reset time would change. The temporal dynamics of this system cannot be quantitatively predicted without first seeing a system in action to dissect some of the lumped parameters that remain fixed for all possible timers (**Supplementary Methods**). To address this, we assembled and tested a single example timer using the two control promoters (TX-LX); we then used the experimental data (**Fig. 3b**) from this system to calibrate a quantitatively predictive model for the other 440 possible timers afforded by our libraries.

The quantitative model gave us predictions as to how reset time could be varied by promoter selection, specifically by adjusting the ratio of relative expression from opposing promoters (using $S_{max} - S_{min}$ to determine relative expression for each). We used the model to quantitatively predict the reset behavior of two timers (T18-LX and T4-LX) with ratios greater than that for TX-LX and two timers (T7-L18 and TX-L14) with smaller ratios. These timers were assembled and tested in yeast. The experimental data for all four networks fell within the upper and lower bounds of the model predictions, validating our approach and ability to make quantitative predictions (**Fig. 3c–f**).

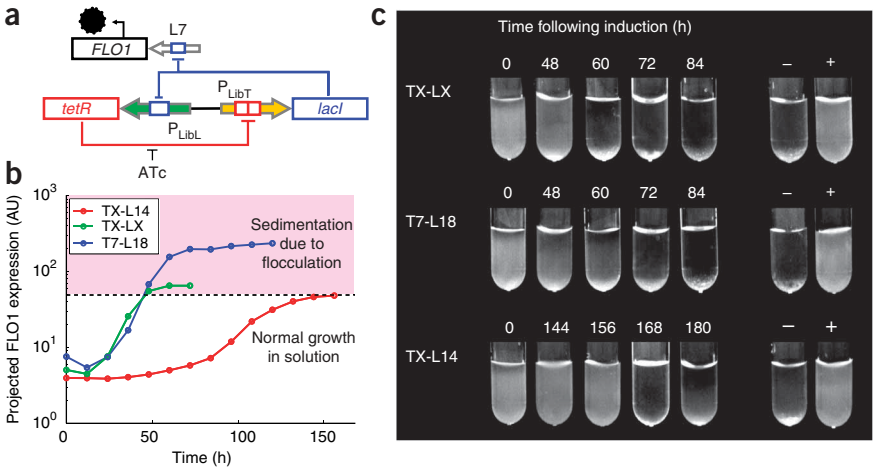

Figure 4 Controlling the timing of yeast sedimentation using a predictable gene network. The synthetic networks tested in **Figure 3b**,**e**,**f** were used to control the timing of yeast sedimentation caused by flocculation. (**a**) Schematic of flocculation gene networks. Flocculation is regulated by replacing yEGFP and LX in the gene network shown in **Figure 3a** with *FLO1* under the control of the promoter L7. (**b**) Rescaled yEGFP data from **Figure 3b**,**e**,**f** were used to project temporal FLO1 expression levels and predict the timing of cell sedimentation due to flocculation (**Supplementary Methods**). (**c**) The timing of sedimentation from the three synthetic networks. Cultures induced by growth with 250 ng/ml ATc for 36 h were washed twice and grown at high OD_{600} with shaking and diluted into fresh medium every 12 h, until sedimentation cleared the suspension. Images shown here are 1 ml cultures at 12-h intervals, 10 min after brief vortexing. Controls: –, growth in 10 mM IPTG; +, growth in 250 ng/ml ATc.

The model with calibrated parameters provides us with a 'confidence interval' of reset times for all ratios (**Fig. 3g**). Closer inspection of the model reveals that reset times of different timer networks with low S_{min} values are approximately proportional to the reciprocal of the square root of the distance between the $S_{max} - S_{min}$ ratio and the bifurcation ratio (the ideal $S_{max} - S_{min}$ ratio for bistability). Mathematically, this is due to a temporal lag in resetting caused by the network passing through a 'bottleneck' as it leaves bistability[45,46]. This direct relationship allows timers with any reset time between 50 and 150 h to be chosen simply based on the strengths of the two promoters selected from the respective libraries (**Fig. 3g**).

Control of yeast sedimentation timing

To demonstrate how our approach can be readily applied in a biotechnology scenario, we used the plug-and-play nature of our timer networks to control the flocculation of yeast. Flocculation occurs when yeast cells express *FLO1*, which functions as a yeast-specific adhesin that causes cells to clump together and sediment from the medium[47–49]. The phenotype is crucial in industrial beer, wine and bioethanol fermentation, as it allows for easy removal of yeast sediment after all sugars have been converted to ethanol[48].

Because the reset times of T4-LX and T18-LX are very close to that of TX-LX, we chose TX-LX, T7-L18 and TX-L14 to test the application of genetic timers. Using these three networks, we controlled the timing of sedimentation by replacing yEGFP with the *FLO1* gene (**Fig. 4a**). In our laboratory yeast strain, *FLO1* is not expressed, but replacing its native promoter with a strong promoter reactivates flocculation, causing sedimentation to occur when a threshold of *FLO1* expression is passed. The timing of sedimentation can therefore be tied to the resetting of each timer network by choosing an appropriate regulated promoter from the libraries. With the 441 possible timers, and a choice from two library sets of promoters for controlling *FLO1* expression, we had the potential to produce >17,000 different flocculating networks. We selected the L7 promoter, which has relatively high S_{max} and very low S_{min}, to give a wide dynamic range. When LacI was abundant, the minimal expression from this promoter did not elicit sedimentation, allowing yeast to grow in suspension.

We experimentally determined the threshold of *FLO1* expression that causes sedimentation (**Supplementary Methods**). Rescaling the timer data in **Figure 3b**,**e**,**f** to match the S_{max} and S_{min} values for the L7 promoter allowed us to plot estimates as to when this threshold would be passed for our networks (**Fig. 4b**). We tested these predictions by assembling the networks in yeast, and then growing the yeast cultures until sedimentation, which occurred days after the initial induction (ATc) was removed (**Fig. 4c**). For the TX-LX, T7-L18 and TX-L14 networks, sedimentation was seen at 60, 60 and 168 h, respectively. These findings closely matched the predicted times assuming a 12-h lag, presumably due to a longer phenotype maturation for flocculation compared to yEGFP fluorescence. This experiment demonstrates that we can quickly apply predictable networks built with our approach to control an industrially relevant phenotype. Such accurate control of flocculation timing provides a wide window of opportunity to harvest fermentation product from cells and could be applied to improve biomass recycling in the biofuels industry.

DISCUSSION

This work establishes a strategy to rapidly increase the number of network components as well as decrease the time and effort required to engineer gene networks with desired functions. Our approach is compatible with plug-and-play synthetic biology and facilitates gene network construction. Here we focused on generating, characterizing and using component libraries of promoters, but our method is applicable to other biomolecular components, as diversity in nonessential sequence also affects functional efficiency in proteins and RNA.

Although screening of mutated parts is not a new technique, our approach represents an advance over previous methods by coupling qualitative and quantitative modeling with library diversity to guide the construction of synthetic gene networks with predictable functions. In robust networks like our feed-forward loop, models built entirely from component property sets are sufficient to guide the choice of parts required to elicit specific network phenotypes, such as high sensitivity to inputs or low maximum output. Although previous studies have shown that it is possible, in some cases, to accurately predict network behaviors based solely on component properties[8,50], we found that this conclusion cannot be generalized to more complex, finely balanced networks, such as the timers described here. Instead, we found that one must first assemble and experimentally characterize a single exemplary network of interest to create a generalizable model with quantitative predictive capabilities. The experimental work needed to construct and test this one network is quickly offset by the yield of accurate quantitative predictions for all other possibilities, and the benefit of this is especially significant when

one considers that each additional component library incorporated increases the number of potential networks exponentially.

Our strategy effectively moves component 'tweaking' to the front-end of gene network engineering. This arrangement is instinctively more rational than network retrofitting and is made feasible by the coupling with mathematical modeling. As component libraries are produced in parallel at the same point in the process that individual parts are typically characterized for modeling, they require little extra effort in return for significant reward. Projects undertaken with this approach will help accelerate synthetic biology by yielding many more components for the community, and as library-synthesized components are designed to show variation only in intended properties, the need for extensive characterization of each component is eliminated or substantially reduced. Our work also provides an accessible method for introducing predictable, controlled variability to networks, a feature that is increasingly becoming desirable as synthetic biology enters its second decade[18,19]. With advances in modern DNA synthesis technologies, the range of our approach will expand as synthesis becomes faster and cheaper, and as longer regions of biomolecules are able to be specifically varied in a systematic fashion.

METHODS

Strains and media. *S. cerevisiae* strain YPH500 (*α, ura3-52, lys2-801, ade2-101, trp1Δ63, his3Δ200, leu2Δ1*) (Stratagene) was used in all experiments, and all genomic integrations were specifically targeted to the redundant *ura3-52* locus. Culturing, genetic transformation and verification of transformation were done as previously described[29], using either the *TRP1*, *HIS3* or *LEU2* genes as selectable markers.

Plasmid construction. The TetR-regulated promoter library characterization vector (pTVGI, **Fig. 1b**) was adapted from the previously described yeast integrative plasmid pRS4D1[29], removing the *GAL1/GAL10* promoter region and replacing it with the *S. cerevisiae TEF1* promoter directing TetR expression and the *GAL1* UAS region plus a synthesized library promoter directing yEGFP expression. A 489-bp span of arbitrary sequence from the gene *ura3* was included between these promoters to buffer any cross-talk between them and to allow the vector to site-specifically integrate into the *ura3-52* locus. For the LacI-regulated promoter library characterization vector (pLVGI), TetR was replaced by a synthetic codon-optimized LacI[41] that had been altered to remove both an internal *Pst*I restriction site (without changing the codon sequence or efficiency) and the hyper-strong SV40 nuclear localization signal. The control TX promoter was amplified directly from pRS4D1, whereas the LX promoter and OR-LT promoters were generated, as previously described, by standard oligonucleotide PCR mutation methods from the TX promoter and T123 pRS4D1 promoter, respectively[29]. All plasmids were constructed and used to transform *E. coli* to harvest DNA for yeast transformations, as previously described[29].

Promoter library synthesis and screening. Promoters were created from partially overlapping pairs of 110-mer PAGE-purified oligonucleotides (**Supplementary Methods**), which were custom synthesized by Sigma-Genosys. Second strand DNA synthesis by Klenow polymerase was followed by agarose gel electrophoresis purification to obtain fragments ready for insertion into characterization vectors[34]. Upon ligation, *DH5α E. coli* (New England Biolabs) were transformed with the plasmid vectors, and clones were selected by ampicillin resistance. We pooled 10^4–10^5 colonies directly from Luria-Bertani agar plates and harvested them for plasmid using the QIAprep Spin Miniprep Kit (Qiagen). The promoter plasmid library was then used to transform yeast as previously described[29], scaling up by a factor of 10 and plating on 250 mm × 250 mm plates to yield ~3,500 individual colonies. Single colonies (192 total) were transferred to two 96-well plates (300 µl of media supplemented with 2% galactose and 250 ng/ml ATc, per well) and grown for 22 h. Cell fluorescence was measured at 450 nm using a SpectraFluor Plate Reader (Tecan). Approximately one-quarter of clones produced a detectable level of expression when TetR was inhibited by ATc, and roughly three-quarters of these responded to the removal of ATc with a drop in expression, indicating controlled regulation. Expression was undetectable in glucose. Colonies selected to create our 20-member library were PCR-tested for single genomic integration, and then characterized by flow cytometry in the presence and absence of 250 ng/ml ATc. For LacI-regulated libraries, ATc was replaced with 10 mM IPTG in the screening and characterization stages.

Flow cytometry and data analysis. Flow cytometry measurements were carried out as previously described[8], running samples on a medium flow rate until 20,000 cells had been collected within a small forward and side scatter gate to reduce extrinsic noise. Data files were analyzed using MatLab (The MathWorks), linearizing log-binned fluorescence intensity values and then calculating the median and s.d. of the gated population. For both promoter library data (**Table 1**) and control ATc and IPTG induction curves, 3 ml cultures were grown for 20 h to an optical density at 600 nm (OD_{600}) of 1.00 at 30 °C with orbital shaking before measurement. For the negative feed-forward loop and genetic timer data, 300 µl of cells were grown to OD_{600} of 1.00 at 30 °C in 96-well format. For the negative feed-forward loop data, cells were grown for 22 h before measurement. For the genetic timer data, cells were grown for 12 h, a sample was taken for measurement, and then a fraction of the remaining cells was diluted in fresh medium for the next 12 h of growth.

Flocculation. To obtain flocculating strains, the L7 library promoter was inserted into pFA6a-KanMX6-pGAL1 in place of the *GAL1* promoter sequences, and PCR amplification from primer pairs FL1 and FL2 was used to integrate these in place of the wild-type *FLO1* promoter as described previously[47]. To measure flocculation over time, 1.2 ml cultures were grown for 12 h to $OD_{600} = 1.50$ at 30 °C with orbital shaking. We removed 1 ml for measurement and replaced with 1 ml of fresh culture medium to continue growth. For measurement, 1 ml of culture was vortexed for 5 s before sitting for 10 min in a clear 3 ml culture tube. Cultures were photographed with a light box behind, and the image inverted and auto-contrasted using Picasa imaging software (Google).

For details of modeling, predictions, sequencing and gene network assembly, see the **Supplementary Methods**.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We thank P.R. Jensen for advice on promoter library synthesis methods, K. Verstrepen for guidance and materials relating to yeast flocculation and H.H. Lee for valuable ideas in genetic device construction. This work was supported by the National Institutes of Health (NIH) through the NIH Director's Pioneer Award Program, grant number DP1 OD003644, the National Science Foundation Frontiers in Integrative Biological Research program and the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

T.E., X.W. and J.J.C. designed the study; T.E. performed the experiments; X.W. conducted the modeling work; T.E., X.W. and J.J.C. analyzed the data and wrote the paper.

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Andrianantoandro, E., Basu, S., Karig, D.K. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.* **2**, 2006–0028 (2006).
2. Withers, S.T. & Keasling, J.D. Biosynthesis and engineering of isoprenoid small molecules. *Appl. Microbiol. Biotechnol.* **73**, 980–990 (2007).
3. Weber, W. *et al.* A synthetic mammalian gene circuit reveals antituberculosis compounds. *Proc. Natl. Acad. Sci. USA* **105**, 9994–9998 (2008).
4. Basu, S., Gerchman, Y., Collins, C.H., Arnold, F.H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–1134 (2005).
5. Lu, T.K. & Collins, J.J. Dispersing biofilms with engineered enzymatic bacteriophage. *Proc. Natl. Acad. Sci. USA* **104**, 11197–11202 (2007).
6. Lu, T.K. & Collins, J.J. Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. *Proc. Natl. Acad. Sci. USA* **106**, 4629–4634 (2009).
7. Lee, S.K., Chou, H., Ham, T.S., Lee, T.S. & Keasling, J.D. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr. Opin. Biotechnol.* **19**, 556–563 (2008).

8. Guido, N.J. *et al.* A bottom-up approach to gene regulation. *Nature* **439**, 856–860 (2006).
9. Hasty, J., McMillen, D. & Collins, J.J. Engineered gene circuits. *Nature* **420**, 224–230 (2002).
10. Guet, C.C., Elowitz, M.B., Hsing, W. & Leibler, S. Combinatorial synthesis of genetic networks. *Science* **296**, 1466–1470 (2002).
11. Levskaya, A. *et al.* Synthetic biology: engineering *Escherichia coli* to see light. *Nature* **438**, 441–442 (2005).
12. Deans, T.L., Cantor, C.R. & Collins, J.J. A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells. *Cell* **130**, 363–372 (2007).
13. Kramer, B.P. *et al.* An engineered epigenetic transgene switch in mammalian cells. *Nat. Biotechnol.* **22**, 867–870 (2004).
14. Gardner, T.S., Cantor, C.R. & Collins, J.J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
15. Weber, W. *et al.* A synthetic time-delay circuit in mammalian cells and mice. *Proc. Natl. Acad. Sci. USA* **104**, 2643–2648 (2007).
16. Grilly, C., Stricker, J., Pang, W.L., Bennett, M.R. & Hasty, J. A synthetic gene network for tuning protein degradation in *Saccharomyces cerevisiae*. *Mol. Syst. Biol.* **3**, 127 (2007).
17. Elowitz, M.B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
18. Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519 (2008).
19. Tigges, M., Marquez-Lago, T.T., Stelling, J. & Fussenegger, M. A tunable synthetic mammalian oscillator. *Nature* **457**, 309–312 (2009).
20. Marguet, P., Balagadde, F., Tan, C. & You, L. Biology by design: reduction and synthesis of cellular components and behaviour. *J. R. Soc. Interface* **4**, 607–623 (2007).
21. Yokobayashi, Y., Weiss, R. & Arnold, F.H. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. USA* **99**, 16587–16591 (2002).
22. Kobayashi, H. *et al.* Programmable cells: interfacing natural and engineered gene networks. *Proc. Natl. Acad. Sci. USA* **101**, 8414–8419 (2004).
23. Acar, M., Becskei, A. & van Oudenaarden, A. Enhancement of cellular memory by reducing stochastic transitions. *Nature* **435**, 228–232 (2005).
24. Blake, W.J. *et al.* Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* **24**, 853–865 (2006).
25. Pedraza, J.M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005).
26. Canton, B., Labno, A. & Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* **26**, 787–793 (2008).
27. Cox, R.S. III, Surette, M.G. & Elowitz, M.B. Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.* **3**, 145 (2007).
28. Gertz, J., Siggia, E.D. & Cohen, B.A. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature* **457**, 215–218 (2009).
29. Murphy, K.F., Balazsi, G. & Collins, J.J. Combinatorial promoter design for engineering noisy gene expression. *Proc. Natl. Acad. Sci. USA* **104**, 12726–12731 (2007).
30. Alper, H., Moxley, J., Nevoigt, E., Fink, G.R. & Stephanopoulos, G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**, 1565–1568 (2006).
31. Jensen, P.R. & Hammer, K. Artificial promoters for metabolic optimization. *Biotechnol. Bioeng.* **58**, 191–195 (1998).
32. Hammer, K., Mijakovic, I. & Jensen, P.R. Synthetic promoter libraries–tuning of gene expression. *Trends Biotechnol.* **24**, 53–55 (2006).
33. Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. USA* **102**, 12678–12683 (2005).
34. 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).
35. Hillen, W. & Berens, C. Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annu. Rev. Microbiol.* **48**, 345–369 (1994).
36. Johnston, M. & Davis, R.W. Sequences that regulate the divergent GAL1–GAL10 promoter in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**, 1440–1448 (1984).
37. Cormack, B.P. *et al.* Yeast-enhanced green fluorescent protein (yEGFP)a reporter of gene expression in Candida albicans. *Microbiology* **143**, 303–311 (1997).
38. Schirmaier, F. & Philippsen, P. Identification of two genes coding for the translation elongation factor EF-1 alpha of *S. cerevisiae*. *EMBO J.* **3**, 3311–3315 (1984).
39. Floer, M., Bryant, G.O. & Ptashne, M. HSP90/70 chaperones are required for rapid nucleosome removal upon induction of the GAL genes of yeast. *Proc. Natl. Acad. Sci. USA* **105**, 2975–2980 (2008).
40. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA* **100**, 11980–11985 (2003).
41. Scrable, H. & Stambrook, P.J. Activation of the lac repressor in the transgenic mouse. *Genetics* **147**, 297–304 (1997).
42. Kaplan, S., Bren, A., Dekel, E. & Alon, U. The incoherent feed-forward loop can generate non-monotonic input functions for genes. *Mol. Syst. Biol.* **4**, 203 (2008).
43. Entus, R., Aufderheide, B., Herbert, M. & Sauro, M.H. Design and implementation of three incoherent feed-forward motif based biological concentration sensors. *Syst. Synth. Biol.* **1**, 119–128 (2007).
44. Tsang, J., Zhu, J. & van Oudenaarden, A. MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. *Mol. Cell* **26**, 753–767 (2007).
45. Wang, X., Hao, N., Dohlman, H.G. & Elston, T.C. Bistability, stochasticity, a nd oscillations in the mitogen-activated protein kinase cascade. *Biophys. J.* **90**, 1961–1978 (2006).
46. Strogatz, S.H. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Addison-Wesley, Reading, Massachusetts, 1994).
47. Guo, B., Styles, C.A., Feng, Q. & Fink, G.R. A *Saccharomyces* gene family involved in invasive growth, cell-cell adhesion, and mating. *Proc. Natl. Acad. Sci. USA* **97**, 12158–12163 (2000).
48. Verstrepen, K.J. & Klis, F.M. Flocculation, adhesion and biofilm formation in yeasts. *Mol. Microbiol.* **60**, 5–15 (2006).
49. Smukalla, S. *et al.* FLO1 is a variable green beard gene that drives biofilm-like cooperation in budding yeast. *Cell* **135**, 726–737 (2008).
50. Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S. & Elowitz, M.B. Accurate prediction of gene feedback circuit behavior from component properties. *Mol. Syst. Biol.* **3**, 143 (2007).