Iterative plug-and-play methodology for constructing and modifying synthetic gene networks

Kevin D Litcofsky[1–3,5], Raffi B Afeyan[1–3,5], Russell J Krom[1–3], Ahmad S Khalil[1–3] & James J Collins[1–4]

We present a methodology for the design, construction and modification of synthetic gene networks. This method emphasizes post-assembly modification of constructs based on network behavior, thus facilitating iterative design strategies and rapid tuning and repurposing of gene networks. The ease of post-construction modification afforded by this approach and the ever-increasing repository of components within the framework will help accelerate the development of functional genetic circuits for synthetic biology.

Synthetic biology is focused on understanding biological design principles and programming new biological behaviors through the construction of artificial circuits from well-characterized genetic components. This engineering-driven approach was inspired by the development of two small genetic circuits: a bistable toggle switch[1] and a self-sustaining oscillator[2]. Since these early demonstrations, the scale and complexity of synthetic gene networks have increased[3–9].

The growing scale and complexity of synthetic gene networks requires new methods for network design, construction and modification. Traditionally, *ad hoc* plasmid construction and modification schemes have used DNA ligase to join restriction fragments. This becomes unwieldy when the number of components and the size of the construct limit the availability of unique restriction sites. To address this difficulty, several approaches have been recently developed[10–12] that focus on standardizing the assembly of larger DNA fragments. Additive assembly, however, does not address other critical issues in the design and construction of synthetic gene networks: notably, the need for post-assembly modifications and substitutions in response to the network's observed performance.

To address these needs, we developed a flexible plug-and-play approach for constructing and modifying synthetic gene networks. Inspired by the solderless breadboards used to develop electrical circuit prototypes, our platform enables rapid and scalable assembly within the familiar molecular biology framework while facilitating post-assembly modifications. The method features a set of optimal type IIP restriction enzymes whose respective restriction sites define the multiple cloning site (MCS) in the cloning vectors (**Supplementary Table 1**). The set of enzymes was chosen according to a specific set of parameters to ensure maximal compatibility during cloning (Online Methods). The method also features compatible genetic components, which have been optimized to exclude internal instances of the reserved sites. This permits post-assembly modifications by unique double digest (**Fig. 1**).

We selected an initial set of 26 well-characterized genetic components, including 12 genes and 8 promoters (**Fig. 1a** and **Supplementary Table 2**), on the basis of common usage in previously published *Escherichia coli* synthetic gene networks. We then optimized the sequences to exclude the MCS restriction sites without altering component function through synonymous codon substitution for genes and annotation-guided or randomized mutagenesis for promoters and other regulatory elements (**Supplementary Note**). The components were then constructed by either synthesis (DNA2.0, Inc.), PCR amplification or site-directed mutagenesis of the source components. We confirmed library parts with optimized sequences for proper functionality and, when possible, also compared them with their nonoptimized counterparts (**Supplementary Figs. 1–3**).

Constructing synthetic gene networks using this cloning process is straightforward (**Fig. 1b**). Components are each assigned to a directional 'slot', a pair of adjacent restriction sites within the MCS, and cloning is performed using classical molecular biology techniques (Online Methods and **Supplementary Table 3**). To demonstrate the approach, we recapitulated the original genetic toggle switch[1] by designing, constructing and tuning a bistable *lacI*-*tetR* genetic toggle switch from optimized vector and library components (**Fig. 2a**,**b**). The bistable toggle switch can maintain its respective genetic state upon removal of the chemical inducers. Induction with anhydrotetracycline (aTc) relieves *tetR* repression, allowing for high expression of *lacI* and *GFP*, whereas induction with isopropyl-β-D-1-thiogalactopyranoside (IPTG) relieves the *lacI* repression and produces the high *tetR* and *mCherry* state. We switched the toggle between the states via the addition of the respective chemical inputs and reliably maintained the states upon removal of the inducers (**Fig. 2c** and **Supplementary Fig. 4**).

We found that multiple post-assembly modifications were required to arrive at a functional, bistable genetic toggle. Our approach accelerates characterization-driven iteration by permitting modification in lieu of complete reassembly. In this case, our

[1]Howard Hughes Medical Institute, Boston University, Boston, Massachusetts, USA. [2]Department of Biomedical Engineering, Boston University, Boston, Massachusetts, USA. [3]Center for BioDynamics, Boston University, Boston, Massachusetts, USA. [4]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts, USA. [5]These authors contributed equally to this work. Correspondence should be addressed to A.S.K. (akhalil@bu.edu) or J.J.C. (jcollins@bu.edu).

RECEIVED 28 FEBRUARY; ACCEPTED 12 SEPTEMBER; PUBLISHED ONLINE 7 OCTOBER 2012; DOI:10.1038/NMETH.2205

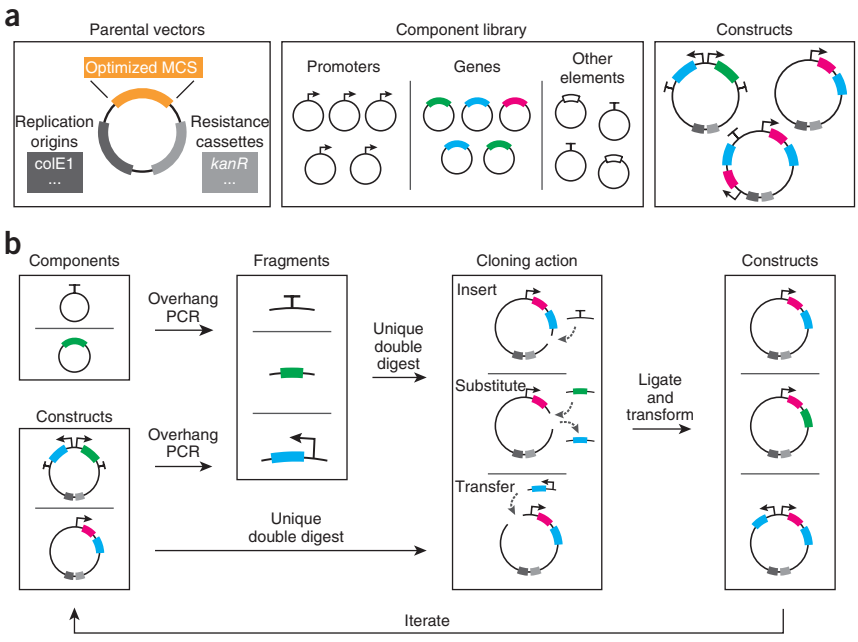

Figure 1 | Plug-and-play methodology for synthetic gene networks. (**a**) Elements comprising the framework: parental cloning vectors harboring a custom multiple cloning site (MCS) of optimal restriction enzyme sites, a library of commonly used synthetic genetic components designed to exclude the restriction sites, and a repository of assembled constructs that includes synthetic modules, intermediates and circuits. (**b**) Generalized workflow for constructing and modifying synthetic gene networks, which prioritizes and streamlines the iterative process of arriving at functional networks and modules.

initial bicistronic toggle construct (Toggle v1) did not activate in response to either of the inducers (**Fig. 2a**). To eliminate possible problems related to poor polycistronic expression and to ensure transcription of the fluorescent reporters, we added an additional instance of the appropriate promoter to each node, converting the toggle to a monocistronic design (Toggle v2). Toggle v2 exhibited aTc-induced *GFP* expression but failed to produce *mCherry* expression in response to IPTG (**Fig. 2a**). This suggested poor promoter-driven expression of both *mCherry* and *tetR*. Our component library contained multiple *lacI*-controlled promoters, so we simply swapped both instances of the promoter P_{LlacO} with $P_{trc\text{-}2}$ (**Fig. 2a**), which we found to be a stronger promoter (**Supplementary Fig. 1**). The resulting toggle (Toggle v3) activated each state upon addition of the respective inducer. However, Toggle v3 was unable to maintain the high *lacI* (*GFP*) state in the absence of aTc, displaying a bias toward the high *tetR* (*mCherry*) state. We therefore sought to balance the circuit by performing random mutagenesis of the Shine-Dalgarno sequence in the ribosome-binding site that controls the translation initiation rate of *tetR*. The fully functional, bistable toggle switch (Toggle v4) could be triggered to occupy each state by adding the respective inducer, and it successfully displayed memory of the states upon removal of the inducers (**Fig. 2c**). Of note, construction of the initial architecture of the toggle took 5 d. We completed the subsequent iterative tuning in separate 3-d efforts, guided by the network's performance (**Fig. 2a**).

Our method also enables large-scale reconstruction and design changes. For example, existing modules and circuits

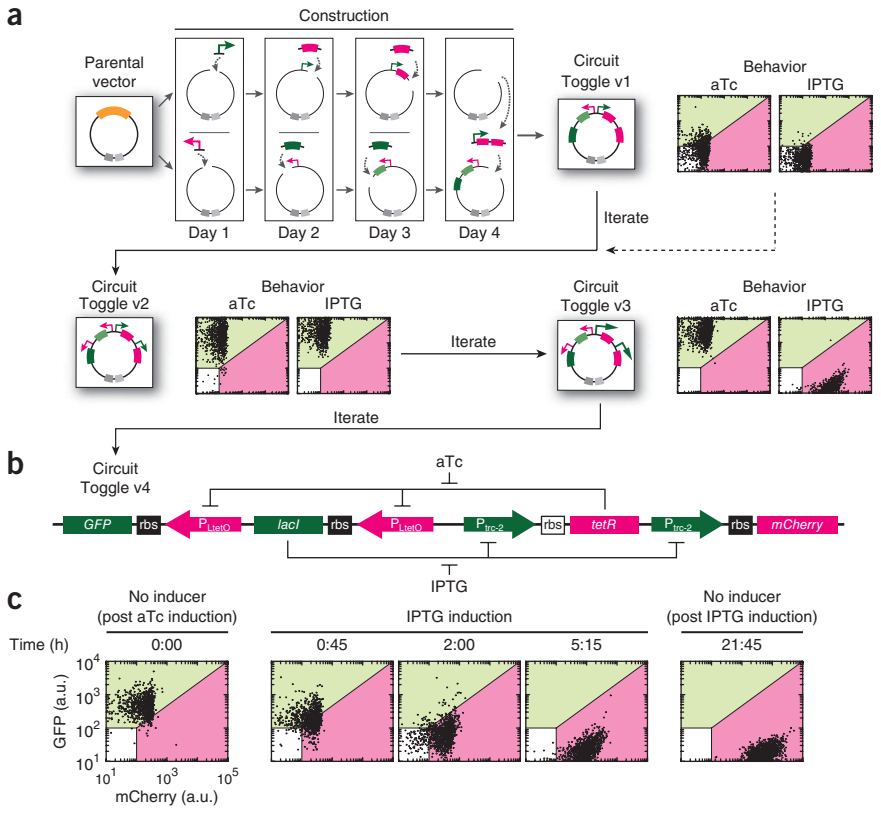

Figure 2 | Construction and tuning of a bistable genetic toggle switch. (**a**) Representation of the construction and characterization-driven tuning of a genetic toggle switch. Each of the intermediate toggle constructs was induced overnight with either aTc or IPTG, and cells were assayed for expression of fluorescent proteins (GFP and mCherry) by flow cytometry. The parental vector contains an antibiotic-resistance gene (dark gray) and an origin of replication (light gray). (**b**) Schematic of the final bistable toggle switch. (**c**) IPTG-induced switching and subsequent maintenance of the genetic toggle switch. A time course is shown for cells that harbored the circuit switching from the *GFP* state (0 h) to the *mCherry* state (0–5:15 h) through IPTG induction. These cells maintained the *mCherry* state when diluted into the no-inducer condition and grown overnight (21:45 h). Data were obtained by flow cytometry at the indicated times; $n = 10{,}000$ events per experiment. rbs, ribosome-binding site; a.u., arbitrary units.

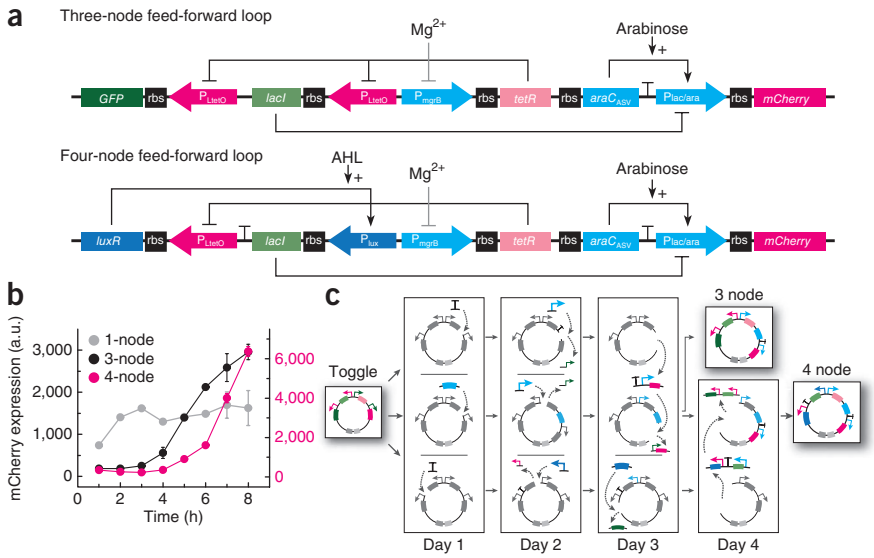

Figure 3 | Transformation of the genetic toggle switch into functional three- and four-node feed-forward loops. (**a**) Schematics of the three- and four-node feed-forward loop networks. Mg^{2+} indirectly inhibits P_{mgrB} via the PhoPQ two-component system. rbs, ribosome-binding site; AHL, *N*-acyl homoserine lactone. (**b**) Pulse response for one-, three- and four-node feed-forward loops. Cells containing the three- or four-node feed-forward loops were induced by the absence of $MgCl_2$. After a 6-h induction, $MgCl_2$ was spiked into the medium to stop induction. Flow cytometry measured *mCherry* fluorescence (a.u., arbitrary units) at 1-h intervals over an 8-h time course. Fluorescence data for one-node and three-node loops are plotted on the left (black) *y* axis; fluorescence data for the four-node loop are plotted on the right (pink) *y* axis. Points represent mean values for three experiments ± s.d.; n = 10,000 events per experiment. (**c**) Representation of the 5-d parallel subcloning process used to construct plasmids encoding three- and four-node feed-forward loops starting from a genetic toggle switch.

can be repurposed into usable cloning parts for novel synthetic gene networks that share common components and architectures. This eliminates the need for additional time-consuming *de novo* reconstruction of existing modules. To demonstrate this reconfigurability, we used the Toggle v4 as the starting point for the construction of three-node and four-node coherent feed-forward loops (FFLs)[13] (**Fig. 3a**), which are functionally distinct from the bistable toggle switch.

Coherent feed-forward loops act as sign-sensitive delays in response to step stimuli[13]. Our FFL designs are composed of a direct and indirect path to a final logical node controlling expression of a fluorescent reporter (*mCherry*). Additional nodes added to the indirect path of the FFL increase the threshold duration. In the three-node FFL (**Fig. 3a**), which is operated in a background presence of arabinose, the promoter P_{mgrB} triggers expression of the first node when induced by the absence of magnesium (Mg^{2+}) in the medium. This activates both paths, expressing *araC* in the direct path and *tetR* in the indirect path. *tetR* subsequently represses the production of *lacI* in the next node of the indirect path. The hybrid promoter $P_{lac/ara}$, which approximates an AND-like gate to express *mCherry*, then integrates the two paths at the final node. The four-node FFL (**Fig. 3a**), which is operated in a background presence of arabinose and *N*-acyl homoserine lactone (AHL), has an additional transcriptional node involving *luxR* in the indirect path.

We compared the outputs of the three- and four-node FFLs to each other and to a one-node control (**Fig. 3b**). In the control circuit, the promoter P_{mgrB} drives the expression of *mCherry* directly. We induced the three circuits for 6 h in the absence of Mg^{2+} and made fluorescence measurements every hour during an 8-h time course. The circuit responses were markedly different. The one-node control responded immediately to the inducer, whereas the three- and four-node FFLs required longer induction times before producing outputs (**Fig. 3b**).

To arrive at these new networks, we designed a parallelized construction strategy originating from Toggle v4 (**Fig. 3c**). Transformation from the toggle to the three-node FFL required two insertions and two substitutions, and the four-node FFL required a further insertion and two additional substitutions. By making these modifications in parallel, we generated both feed-forward constructs in four rounds of subcloning over the course of 5 d (**Fig. 3c**).

Arriving at a functional synthetic gene network that behaves according to predicted phenotype is a two-part process. It requires physically constructing the gene network by assembling DNA encoding the network components and then tuning the network components and architecture to achieve desired functionality. Many new DNA assembly techniques, such as Golden Gate[10] and Gibson[11] cloning, are addressing the front end of this process. Our method provides a flexible platform that focuses on the back end of the development cycle, and it is fully compatible with various assembly techniques that can be used at the front end, provided design considerations are made to incorporate the set of reserved restriction sites and components. Also, recycling existing circuits, intermediates and components, as afforded by our approach, reduces the time required to resynthesize or recreate validated parts. As the collection of well-characterized synthetic components, modules and circuits grows, their reuse and repurposing will simplify the assembly of subsequent networks. Altogether, our platform accelerates the design-build-test cycle in synthetic biology by integrating assembly and debugging. Furthermore, our approach is well suited for integration with emerging computational platforms intended to implement rational design and automation in synthetic biology[14–20]. Ultimately, our methodology, used with appropriate vectors and components, could be expanded to other bacterial systems as well as yeast, plant and mammalian systems, facilitating the creation of complex synthetic gene networks in a wide range of organisms.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS
This work was supported by the Howard Hughes Medical Institute and the Office of Naval Research Multidisciplinary University Research Initiative.

AUTHOR CONTRIBUTIONS
K.D.L., R.B.A., A.S.K. and J.J.C. designed the study. K.D.L., R.B.A., R.J.K. and A.S.K. performed experiments. K.D.L., R.B.A., R.J.K., A.S.K. and J.J.C. wrote the manuscript. A.S.K. and J.J.C. supervised the project.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nmeth.2205.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Gardner, T.S., Cantor, C.R. & Collins, J.J. *Nature* **403**, 339–342 (2000).
2. Elowitz, M.B. & Leibler, S. *Nature* **403**, 335–338 (2000).
3. Culler, S.J., Hoff, K.G. & Smolke, C.D. *Science* **330**, 1251–1255 (2010).
4. Danino, T., Mondragon-Palomino, O., Tsimring, L. & Hasty, J. *Nature* **463**, 326–330 (2010).
5. Khalil, A.S. & Collins, J.J. *Nat. Rev. Genet.* **11**, 367–379 (2010).
6. Khalil, A.S. *et al. Cell* **150**, 647–658 (2012).
7. Purnick, P.E. & Weiss, R. *Nat. Rev. Mol. Cell Biol.* **10**, 410–422 (2009).
8. Tamsir, A., Tabor, J.J. & Voigt, C.A. *Nature* **469**, 212–215 (2011).
9. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. *Science* **333**, 1307–1311 (2011).
10. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. *PLoS ONE* **4**, e5553 (2009).
11. Gibson, D.G. *et al. Nat. Methods* **6**, 343–345 (2009).
12. Li, M.Z. & Elledge, S.J. *Nat. Methods* **4**, 251–256 (2007).
13. Mangan, S. & Alon, U. *Proc. Natl. Acad. Sci. USA* **100**, 11980–11985 (2003).
14. Cambray, G., Mutalik, V.K. & Arkin, A.P. *Curr. Opin. Microbiol.* **14**, 624–630 (2011).
15. Carothers, J.M., Goler, J.A., Juminaga, D. & Keasling, J.D. *Science* **334**, 1716–1719 (2011).
16. Ellis, T., Wang, X. & Collins, J.J. *Nat. Biotechnol.* **27**, 465–471 (2009).
17. Salis, H.M., Mirsky, E.A. & Voigt, C.A. *Nat. Biotechnol.* **27**, 946–950 (2009).
18. Beal, J. *et al. ACS Synth. Biol.* **1**, 317–331 (2012).
19. Lux, M.W., Bramlett, B.W., Ball, D.A. & Peccoud, J. *Trends Biotechnol* **30**, 120–126 (2012).
20. Chandran, D., Bergmann, F.T. & Sauro, H.M. *J. Biol. Eng.* **3**, 19 (2009).

ONLINE METHODS

Restriction enzyme selection. The restriction enzymes used in our approach have recognition sites that are nondegenerate, palindromic and 6 bp in length. To achieve consistent digest reaction conditions, the MCS restriction enzymes were also chosen to have at least 50% activity in the common buffer NEBuffer 4 (New England Biolabs) at 37 °C. The spacing of restriction sites in the MCS allows for double digests at adjacent pairs of restriction sites, enabling unique and directional insertions. Additionally, the inclusion of nonadjacent restriction sites that produce blunt or compatible cohesive ends facilitates large-scale reorganizations of the MCS. The set of high-copy cloning vectors that include this MCS was based on the pZE vector family[21] (**Fig. 1a**). Instances of reserved restriction sites were removed from the vector backbones via codon substitution or randomized site-directed mutagenesis.

Cloning details. The desired restriction sites are added to the 5′ and 3′ ends of each component using overhang PCR. For each insertion, both component and vector are uniquely double digested. During post-assembly modifications, the prior insert is removed by gel purification following vector digest. The fragments are then ligated and transformed, and the resulting clones are screened for the desired product. Because the restriction sites used throughout construction remain unique, assembly order is not constrained, which allows for a greater degree of flexibility in planning parallel assembly.

Component characterization. Input/output functions were obtained for all of the optimized components and, when possible, compared to the input/output functions for their unaltered counterparts. To obtain characterization data, custom test circuits were first constructed for each class of components, as detailed below. Then appropriate dose responses were generated with fluorescence measurements being performed with a BD FACSAriaII flow cytometer (BD Biosciences).

For promoters, cassettes were constructed in which the promoters drive the expression of a *GFP* reporter (**Supplementary Fig. 1**); these plasmids were transformed into MG1655 Pro cells. Dose-response experiments were performed using the appropriate inducers for each promoter (**Supplementary Fig. 1**). The promoters were compared to their commonly used, unmodified equivalents. In virtually all cases, the input/output functions matched well. Notably, the optimized P_{LlacO} promoter exhibited a larger dynamic range than the original unaltered version, with other dose-response dynamics matching closely. The P_{LtetO} promoter was not optimized, as the original did not have any reserved restriction sites in its sequence.

For transcription factors (TFs), cascade circuits were constructed in which the magnesium-responsive P_{mgrB} drives the inducible expression of the TF, which subsequently operates on a specific promoter to drive the expression of a *GFP* reporter (**Supplementary Fig. 2**). These circuits were transformed into the respective MG1655-derived knockout strains (that is, strains that have the native TF gene knocked out). Dose responses were performed on the TF-sensitive promoters in the absence (with magnesium) and presence (without magnesium) of the TF. In the case of the cI protein, the dose response was performed with magnesium as the inducer, as cI is not regulated by a chemical inducer. All of our optimized TFs were able to be expressed, and they predictably regulated their specific promoters in a dose-dependent fashion.

For our optimized terminator, two simple circuits were constructed in which P_{BAD} drove the expression of two bicistronic reporters, *GFP* and *mCherry*, with and without the terminator between the two reporters (**Supplementary Fig. 3**). These circuits were transformed into MG1655 cells. The expression of the two reporters was measured in response to induction. For the circuit with the terminator present, the downstream reporter, *mCherry*, produced no expression, thereby confirming proper function.

Two optimized reporters were also tested for fluorescence output. The test circuits consisted of P_{LtetO} driving the expression of the respective fluorescent reporter (**Supplementary Fig. 3**). The circuits were transformed into MG1655 cells. The tested reporters all exhibited dose-dependent expression.

Primer design. The PCR primers used for plasmid constriction and modification were designed according to the following algorithm. Each primer began (5′) with six bases arbitrarily selected in a manner to create primers with similar T_m, calculated using OligoCalc (http://www.basic.northwestern.edu/biotools/oligocalc.html). The next six bases (5′–3′) comprised the desired restriction enzyme recognition site. The remainder of the primer, the 3′ end, consisted of either the 'Fw. Primer Homology' or the 'Rev. Primer Homology' sequences, annotated in the component sequence entries. Thus, the final primer design was 5′ – six arbitrary bases + six bases for recognition site + fw./rev. primer homology – 3′. The primers were ordered from IDT.

Induction conditions. All toggle experiments were performed in M9 minimal medium supplemented with 0.2% glucose and 0.02% casamino acids. We induced the P_{trc-2} promoter with IPTG at a concentration of 1 mM and the P_{LtetO} promoter with 100 ng/mL aTc.

All feed-forward–loop experiments were performed in M9 minimal medium supplemented with 0.2% glucose and 0.02% casamino acids and a background of 0.01% arabinose. Experiments involving the four-node feed-forward loop were also performed in the presence of 10 μM AHL. We performed experiments on the feed-forward loops by repressing the P_{mgrB} promoter with 50 mM $MgCl_2$.

21. Lutz, R. & Bujard, H. *Nucleic Acids Res.* **25**, 1203–1210 (1997).