Synthetic Gene Networks That Count

Ari E. Friedland,[1*] Timothy K. Lu,[1,2*] Xiao Wang,[1] David Shi,[1]
George Church,[2,3] James J. Collins[1]†

Synthetic gene networks can be constructed to emulate digital circuits and devices, giving one the ability to program and design cells with some of the principles of modern computing, such as counting. A cellular counter would enable complex synthetic programming and a variety of biotechnology applications. Here, we report two complementary synthetic genetic counters in *Escherichia coli* that can count up to three induction events: the first, a riboregulated transcriptional cascade, and the second, a recombinase-based cascade of memory units. These modular devices permit counting of varied user-defined inputs over a range of frequencies and can be expanded to count higher numbers.

A counter is a key component in digital circuits and computing that retains memory of events or objects, representing each number of such as a distinct state. Counters would also be useful in cells, which often must have accurate accounting of tightly controlled processes or biomolecules to effectively maintain metabolism and growth. Counting mechanisms have been reportedly found in telomere length regulation (*1*, *2*) and cell aggregation (*3*). These system behaviors appear to be the result of a threshold effect in which some critical molecule number or density must be reached for the observed phenotypic change.

In this study, we first developed a counter, termed the riboregulated transcriptional cascade (RTC) counter, which is based on a transcriptional cascade with additional translational regulation. Two such cascades are illustrated in Fig. 1, A and C that can count up to two and three, respectively (hence, the designations RTC two-counter and RTC three-counter). For the RTC two-counter, the constitutive promoter $P_{Ltet0-1}$ drives transcription of T7 RNA polymerase (RNAP), whose protein binds the T7 promoter and transcribes the downstream gene, in this case, green fluorescent protein (GFP). Both genes are additionally regulated by riboregulators (*4*), whose cis and trans elements silence and activate posttranscriptional gene expression, respectively. The cis-repressor sequence [cr in Fig. 1] is placed between the transcription start site and the ribosome-binding site (RBS), and its complementarity with the RBS causes a stem-loop structure to form upon transcription. This secondary structure prevents binding of the 30*S* ribosomal subunit to the RBS, which inhibits translation. A short, transactivating, noncoding RNA (taRNA), driven by the arabinose promoter P_{BAD}, binds to the cis repressor in trans, which relieves RBS repression and allows translation. With this riboregulation, each node (i.e., gene) in the cascade requires both independent transcription and translation for protein expression. This cascade is able to count brief arabinose pulses [for pulse definition, see (*5*)] by expressing a different protein in response to each pulse (Fig. 1A). With cis-repressed T7 RNAP

[1]Howard Hughes Medical Institute, Department of Biomedical Engineering, Center for BioDynamics and Center for Advanced Biotechnology, Boston University, Boston, MA 02215, USA. [2]Harvard-MIT Division of Health Sciences and Technology, 77 Massachusetts Avenue, Room E25-519, Cambridge, MA 02139, USA. [3]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: jcollins@bu.edu

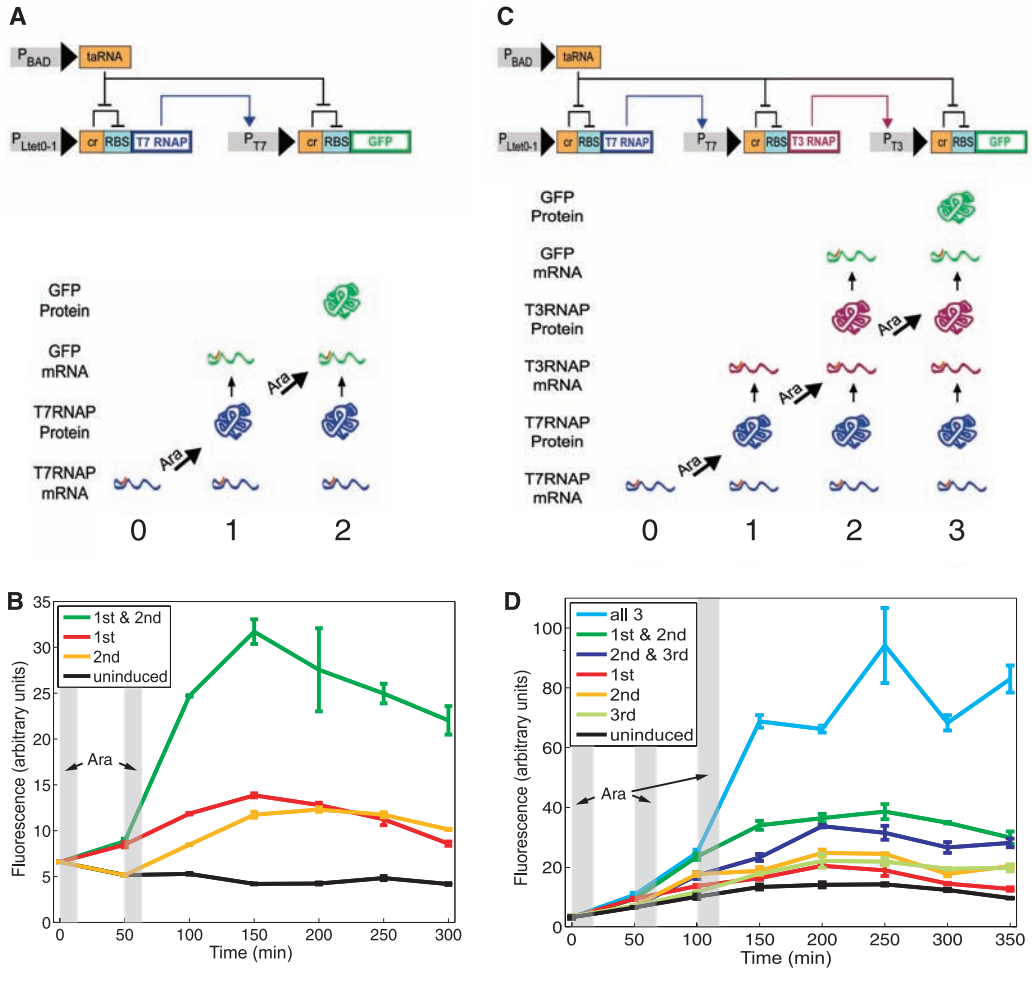

Fig. 1. The RTC two-counter and RTC three-counter construct designs and results. (**A**) The RTC two-counter is a transcriptional cascade with two nodes. Shown at the bottom are expected expression profiles after zero, one, and two arabinose (Ara) pulses. (**B**) Mean fluorescence of three replicates of RTC two-counter cell populations over time, measured by a flow cytometer. Shaded areas represent arabinose pulse duration. (**C**) The RTC three-counter is a transcriptional cascade with three nodes. Shown at the bottom are expected expression profiles after zero, one, two, and three arabinose pulses. (**D**) Mean fluorescence of three replicates of RTC three-counter cell populations over time, measured by a flow cytometer. Shaded areas represent arabinose pulse duration.

mRNAs in the cell, the first pulse of arabinose drives a short burst of taRNA production and, consequently, expression of T7 RNAP proteins. After the pulse is delivered, arabinose is removed from the cell environment, intracellular arabinose and taRNA are metabolized, and expression of T7 RNAP protein halts. The T7 RNAP proteins that have been translated then transcribe cis-repressed GFP transcripts, but few GFP proteins are made until the next arabinose pulse is delivered and translation is once again activated.

We built the RTC two-counter construct on a high-copy plasmid and transformed it into *Escherichia coli* strain K-12pro (*5*). Cells containing this construct were pulsed with the inducer arabinose, and fluorescence was measured over time (Fig. 1B). Uninduced cells show no increase in mean fluorescence, whereas cells that received either the first or second pulse show only small increases, indicating some degree of leakage—an effect in which the intended protein is expressed in each arabinose pulse, but also some unintended, downstream proteins are expressed as well. Cells that received both arabinose pulses show a substantial increase in fluorescence when the second pulse is delivered, precisely when the cells are expected to express GFP proteins.

To extend the RTC counter's capability to count to three, we built a second synthetic construct, the RTC three-counter, again with GFP as the quantitative readout. It is similar to the RTC two-counter but has three nodes in the cascade instead of two (Fig. 1C). T7 RNAP is the gene at the first node driving transcription of T3 RNAP, which ultimately drives transcription of GFP. All transcripts are likewise cis-repressed with the same riboregulator sequence. When pulsed with arabinose, this counter should primarily produce T7 RNAP proteins during the first pulse, T3 RNAP proteins during the second pulse, and GFP proteins during the third pulse (Fig. 1C).

Experimental results demonstrate that fluorescence increases substantially only when all three arabinose pulses are delivered (Fig. 1D). Flow cytometry measurements show this increase beginning at precisely the time of the third pulse, and the considerable slope at this juncture suggests that cells contain a high concentration of cis-repressed GFP transcripts ready for transactivation. The data also reveal slight leakage in cells that are pulsed only once or twice, but their fluorescence remains comparatively low. This result, in combination with the RTC two-counter evidence, shows that the temporal progression of RNA and protein species logically predicted by the counter network architecture is indeed responsible for the observed effect.

To further support these results, we constructed and analyzed a mathematical model based on the design of the RTC two-counter and three-counter constructs. This model, with fitted parameters [see section 6 of (*5*)], was able to match both the RTC two-counter and three-counter experimental results (Fig. 2, A and B). We used the model to investigate the effects of pulse frequency and pulse length on the performance of the RTC three-counter and to guide our experimental search for optimal combinations. The mathematical model predictions, shown as contour lines in Fig. 2C, indicate that maximum expression occurs with pulse lengths of ~20 to 30 min and pulse intervals of 10 to 40 min. The absolute difference in fluorescence after three pulses and two pulses is shown in Fig. 2D, with optimal counting behavior requiring similar pulse length and interval combinations noted above.

Experimentally, we sampled various pulse lengths and intervals, plotting these results as circles in Fig. 2, C and D. These results are consistent with the model predictions across a wide range of temporal conditions, and they confirm that the RTC three-counter has a sizable temporal region in which its counting behavior is robust. Within this region, the counter is also capable of counting irregular pulses; for example, it is able to distinguish between two short pulses followed by a long pulse and two long pulses, as predicted by the model (fig. S5). However, as indicated in Fig. 2, when pulse length or frequency is either too high or low, the RTC three-counter is unable to count properly, presumably because of the

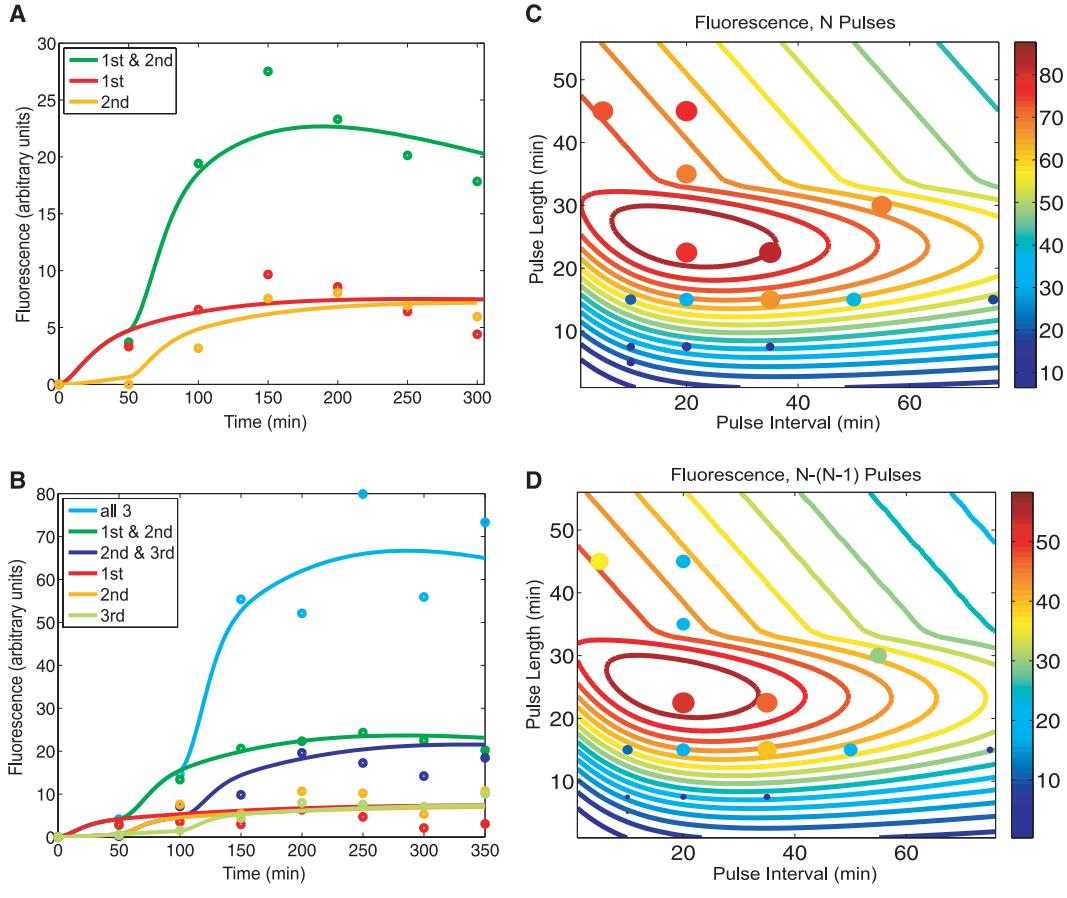

Fig. 2. Modeling predictions and RTC three-counter experimental characterization. (**A**) A model with fitted parameters captures the salient features of the normalized fluorescence results of the RTC two-counter. (**B**) An expanded model, again with fitted parameters, matches the normalized fluorescence results of the RTC three-counter. (**C**) On the basis of parameters fitted in (B), the model predicts expression output of the RTC three-counter (*N*) across a range of pulse lengths and intervals, and these calculations were used to generate the colored contour lines. Solid circles represent experimental results, with both color and size of circles indicating the level of expression. (**D**) Similar to (C), except that values shown are the difference in expression output after three (N) pulses and two (*N* – 1) pulses.

intrinsic kinetic limits of the biochemical processes involved, such as transcription and mRNA degradation.

Our second counter design, termed the DNA invertase cascade (DIC) counter, was built by chaining together modular DNA-based counting units (Fig. 3A). The DIC counter uses recombinases, such as Cre and *flp*$_e$ (*6*), which can invert DNA between two oppositely oriented cognate recognition sites, such as *loxP* and *flp*$_e$-recombination target (*FRT*) sites, respectively. Recombinases have been used for numerous applications, including the creation of gene knockouts and inducible expression systems (*7*, *8*). In our counter design, each recombinase gene (*rec*) is downstream of an inverted promoter (P_{inv}), fused to an *ssrA* tag that causes rapid protein degradation (*9*), and followed by a transcriptional terminator (Term) (fig. S7). The P_{inv}-*rec*-*ssrA*-Term DNA sequences are placed between forward and reverse recombinase recognition sites (R_f and R_r) (fig. S7), forming a single counting unit that we have named a single invertase memory module (SIMM) (Fig. 3A and fig. S7). Upon expression of recombinase by an upstream promoter, the entire SIMM is inverted between the recognition sites. Because the recombinase gene is inverted with respect to the upstream promoter, further expression of recombinase protein ceases, and DNA orientation is fixed.

We developed a single-inducer DIC two-counter (fig. S8) and three-counter (Fig. 3A and fig. S9), which are composed of one and two SIMMs, respectively, and placed them on pBAC plasmids that are maintained as single-copy episomes (*10*). These circuits utilize P_{BAD}, so that pulses of arabinose constitute inputs to the circuit. Each pulse of arabinose results in promoter activation and expression of the next recombinase in the cascade, which then inverts the SIMM in which it is located. This allows the inverted promoter contained within that SIMM to be placed in a forward orientation to drive expression of the next SIMM stage. The single-inducer DIC two-counter shows high GFP output after two pulses of arabinose but only low GFP output after one pulse of arabinose, which shows that a single SIMM can be inverted to count events (fig. S11). In the single-inducer DIC three-counter, some premature flipping of the Cre recombinase–based SIMM did occur, which resulted in a small amount of leakage, e.g., fluorescence increased after only two arabinose pulses (Fig. 3B and fig. S12). However, this leakage was small compared with the high GFP output exhibited in response to three pulses of arabinose (Fig. 3B). In order to probe the temporal characteristics of the single-inducer DIC three-counter, we varied the pulse lengths and intervals, calculating the ratio of GFP output for cells exposed to three, versus two, pulses of arabinose (Fig. 3C). This ratio was at least 1.5 for most conditions tested, which demonstrated that the single-inducer DIC three-counter is able to successfully count pulses whose lengths and intervals range from 2 to 12 hours (Fig. 3C).

We also developed a multiple-inducer DIC three-counter by replacing the P_{BAD} promoters in the single-inducer DIC three-counter with the inducible promoters $P_{Ltet0-1}$, P_{BAD}, and P_{A1lacO} (Fig. 4A and fig. S10). These promoters respond to anhydrotetracycline (aTc), arabinose, and isopropyl β-D-1-thiogalactopyranoside (IPTG), respectively (Fig. 4A). When exposed to aTc followed by arabinose, followed by IPTG, the multiple-inducer DIC three-counter produced a high GFP output (Fig. 4B). No other permutations of the three inducers produced a high output, although some did exhibit a small amount of leakage (Fig. 4, C and D). These results demonstrate that the circuit can be programmed to record only a desired sequence of events.

We have constructed and validated two complementary designs for synthetic counters that operate across a range of time scales. These counters are both highly modular and capable of functioning with multiple inducer-promoter pairs. In addition, the architectures of both counters allow for the tunable output expression of different protein species of interest at any number (up to three are shown) in the counting

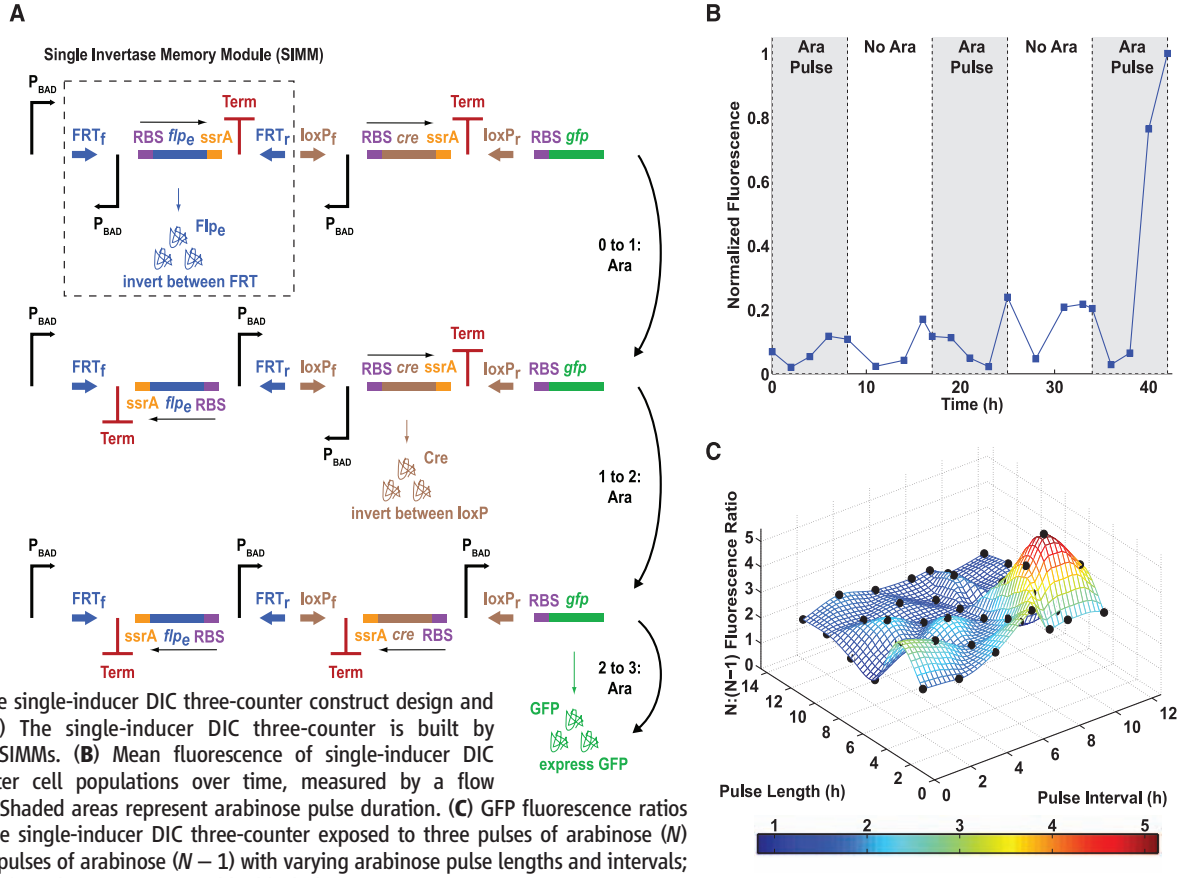

Fig. 3. The single-inducer DIC three-counter construct design and results. (**A**) The single-inducer DIC three-counter is built by cascading SIMMs. (**B**) Mean fluorescence of single-inducer DIC three-counter cell populations over time, measured by a flow cytometer. Shaded areas represent arabinose pulse duration. (**C**) GFP fluorescence ratios between the single-inducer DIC three-counter exposed to three pulses of arabinose (*N*) versus two pulses of arabinose (*N* − 1) with varying arabinose pulse lengths and intervals; experimental results are represented by black dots.

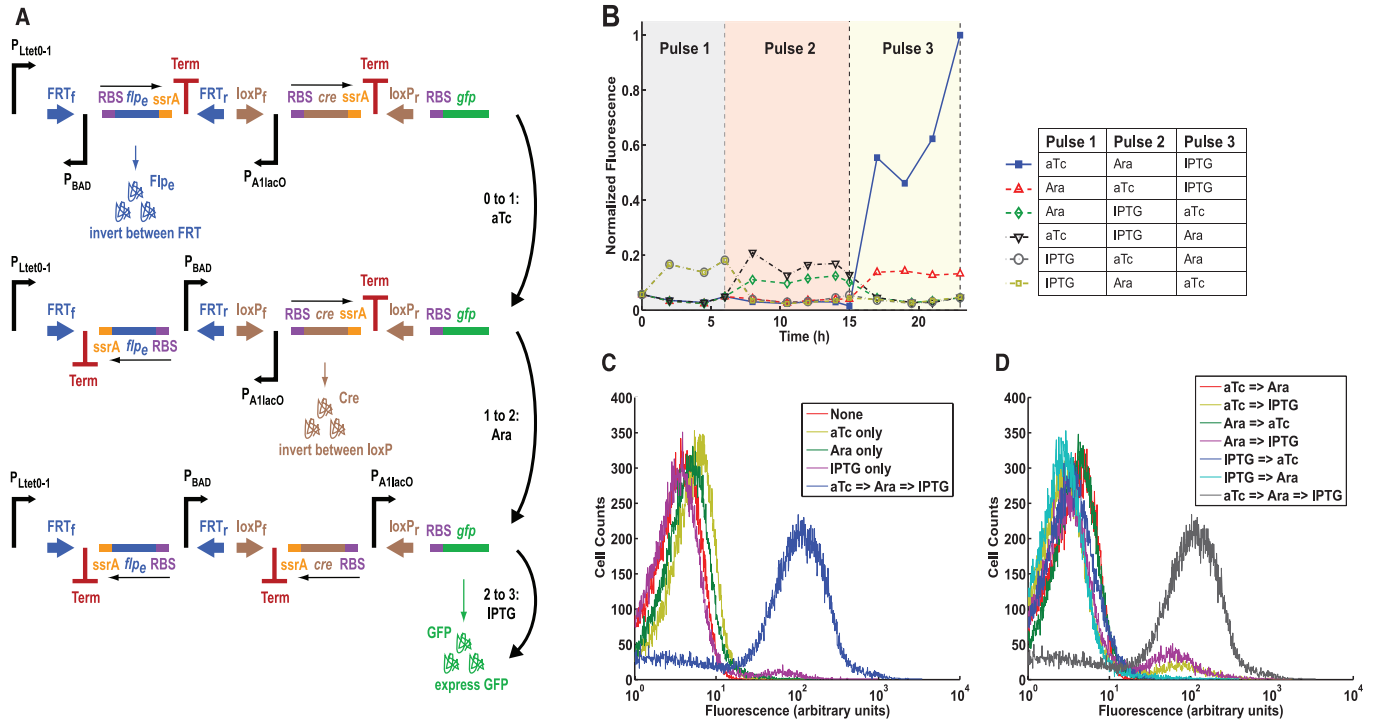

Fig. 4. The multiple-inducer DIC three-counter construct design and results. (**A**) The multiple-inducer DIC three-counter is similar to the single-inducer DIC three-counter in Fig. 3, except that each promoter is a unique inducible promoter: $P_{Ltet0\text{-}1}$, P_{BAD}, and P_{A1lacO}. These promoters respond to aTc, arabinose, and IPTG, respectively. (**B**) Mean fluorescence of multiple-inducer DIC three-counter cell populations over time, measured by a flow cytometer. Colored areas represent the durations of consecutive inducer pulses. (**C**) Flow cytometry population data showing the multiple-inducer DIC three-counter when exposed to the desired sequence of three inducers and to single inducers only. (**D**) Flow cytometry population data showing the multiple-inducer DIC three-counter when exposed to the desired sequence of three inducers and to all pairwise permutations of inducers.

process. Our constructs were built to count up to three events, but they both should be extensible with the use of other unique polymerases or recombinases, of which many are known (*5*). In addition to these shared qualities, each counter comes with its own set of properties. Our RTC counters demonstrate fast activation because of transcriptional and translational regulatory elements, which makes them useful for counting cellular events on the time scale of cell division. The DIC counters operate on time scales of hours (fig. S13) as a result of DNA recombination dynamics (*11*), and they are built with a novel SIMM design that retains counter state based on DNA orientation.

Synthetic gene circuits have enlarged the molecular tool set available to bioengineers and molecular biologists (*4*, *12*–*24*) and have enabled them to program novel cellular behaviors (*25*–*27*) and to construct therapeutic agents (*28*, *29*). Our synthetic counters represent complementary designs that can be used in different settings for a variety of purposes across a range of time scales. For example, if inputs to our RTC counter were coupled to the cell cycle, one might program cell death to occur after a user-defined number of cell divisions as a safety mechanism in engineered strains used for biosensing, bioremediation, or medical purposes. In addition, the multiple-inducer DIC counter might be used to study sequential events that occur in settings such as developmental biology and gene cascades; the single-inducer DIC counter could record events encountered in its environment (e.g., for biosensing); and our SIMM design could be used in other synthetic circuits to maintain genetic memory of low-frequency events, for therapeutic or other applications, such as studying neural circuits.

References and Notes

1. S. Marcand, E. Gilson, D. Shore, *Science* **275**, 986 (1997).
2. A. Ray, K. W. Runge, *Mol. Cell. Biol.* **19**, 31 (1999).
3. D. A. Brock, R. H. Gomer, *Genes Dev.* **13**, 1960 (1999).
4. F. J. Isaacs *et al.*, *Nat. Biotechnol.* **22**, 841 (2004).
5. Materials and methods are available as supporting material on *Science* Online.
6. F. Buchholz, P. O. Angrand, A. F. Stewart, *Nat. Biotechnol.* **16**, 657 (1998).
7. A. C. Groth, M. P. Calos, *J. Mol. Biol.* **335**, 667 (2004).
8. T. S. Ham, S. K. Lee, J. D. Keasling, A. P. Arkin, *Biotechnol. Bioeng.* **94**, 1 (2006).
9. J. B. Andersen *et al.*, *Appl. Environ. Microbiol.* **64**, 2240 (1998).
10. D. A. Wright *et al.*, *Nat. Protocols* **1**, 1637 (2006).
11. S. W. Santoro, P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4185 (2002).
12. M. B. Elowitz, S. Leibler, *Nature* **403**, 335 (2000).
13. T. S. Gardner, C. R. Cantor, J. J. Collins, *Nature* **403**, 339 (2000).
14. T. S. Bayer, C. D. Smolke, *Nat. Biotechnol.* **23**, 337 (2005).
15. D. Baker *et al.*, *Sci. Am.* **294**, 44 (2006).
16. O. Rackham, J. W. Chin, *Nat. Chem. Biol.* **1**, 159 (2005).
17. K. Rinaudo *et al.*, *Nat. Biotechnol.* **25**, 795 (2007).
18. Y. Yokobayashi, R. Weiss, F. H. Arnold, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16587 (2002).
19. N. J. Guido *et al.*, *Nature* **439**, 856 (2006).
20. J. M. Pedraza, A. van Oudenaarden, *Science* **307**, 1965 (2005).
21. T. L. Deans, C. R. Cantor, J. J. Collins, *Cell* **130**, 363 (2007).
22. E. Fung *et al.*, *Nature* **435**, 118 (2005).
23. B. P. Kramer *et al.*, *Nat. Biotechnol.* **22**, 867 (2004).
24. J. Stricker *et al.*, *Nature* **456**, 516 (2008).
25. S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, R. Weiss, *Nature* **434**, 1130 (2005).
26. H. Kobayashi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8414 (2004).
27. A. Levskaya *et al.*, *Nature* **438**, 441 (2005).
28. T. K. Lu, J. J. Collins, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11197 (2007).
29. T. K. Lu, J. J. Collins, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4629 (2009).

30. Supported by the NIH Director's Pioneer Award Program, the NSF Frontiers in Integrative Biological Research program, and the Howard Hughes Medical Institute. Authors' contributions are in supporting online materials.

Supporting Online Material

www.sciencemag.org/cgi/content/full/324/5931/1199/DC1
Materials and Methods
SOM Text
Figs. S1 to S13
Tables S1 and S2
References

9 February 2009; accepted 31 March 2009
10.1126/science.1172005