Published in final edited form as:
Science. 2009 May 29; 324(5931): 1199–1202. doi:10.1126/science.1172005.

Synthetic Gene Networks that Count*

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

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*

Abstract

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 *E. coli* that can count up to three induction events, the first comprised of a riboregulated transcriptional cascade and the second of 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 thresholding effect in which some critical molecule number or density must be reached for the observed phenotypic change.

In this study, we first develop a counter, termed the Riboregulated Transcriptional Cascade (RTC) Counter, which is based on a transcriptional cascade with additional translational regulation. Figs. 1A and 1C illustrate two such cascades that can count up to 2 and 3, respectively (hence, the designations RTC 2-Counter and RTC 3-Counter). For the RTC 2-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 post-transcriptional gene expression, respectively. The *cis*-repressor sequence (cr) 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 30S ribosomal subunit to the RBS, inhibiting translation. A short, *trans*-activating, noncoding RNA (taRNA) driven by the arabinose promoter P_{BAD} binds to the *cis*-repressor in *trans*, relieving

*These authors contributed equally to this work.

RBS repression and allowing 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 (see SOM for pulse definition) by expressing a different protein in response to each pulse (Fig. 1A). With *cis*-repressed T7 RNAP 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 2-Counter construct on a high copy plasmid and transformed it into *E. coli* strain K-12pro (See SOM for details). Cells containing this construct were pulsed with the inducer arabinose, and mean fluorescence over time was measured (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 3-Counter, again with GFP as the quantitative readout. It is similar to the RTC 2-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 *trans*-activation. 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 2-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 2-Counter and 3-Counter constructs. This model, with fitted parameters (see SOM Materials and Methods Section 6 for details), was able to match both the RTC 2-Counter and 3-Counter experimental results (Figs. 2A, B). We used the model to investigate the effects of pulse frequency and pulse length on the performance of the RTC 3-Counter and 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 approximately 20 to 30 minutes and pulse intervals of 10 to 40 minutes. 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 Figs. 2C and 2D. These results are consistent with the model predictions across a wide range of temporal conditions, and confirm that the RTC 3-Counter has a sizeable 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 3-Counter is unable to count properly, presumably due to the 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, is built by chaining together modular DNA-based counting units (Fig. 3A). The DIC Counter utilizes recombinases, such as *cre* and flp_e (5), 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 (6,7). 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 (8), 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 which we have named a Single Invertase Memory Module (SIMM) (Figs. 3A, S7). Upon expression of recombinase by an upstream promoter, the entire SIMM is inverted between the recognition sites. Due to the inverted orientation of the recombinase gene with respect to the upstream promoter, further expression of recombinase protein ceases and DNA orientation is fixed.

We developed a single-inducer DIC 2-Counter (Fig. S8) and 3-Counter (Figs. 3A, S9), which are composed of one and two SIMMs, respectively, and placed them on pBAC plasmids that are maintained as single-copy episomes (9). 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 2-Counter shows high GFP output after two pulses of arabinose and only low GFP output after one pulse of arabinose, demonstrating that a single SIMM can be inverted to count events (Fig. S11). In the single-inducer DIC 3-Counter, some premature flipping of the *cre*-based SIMM did occur, resulting in a small amount of leakage, e.g., fluorescence increased after only two arabinose pulses (Figs. 3B, S12). However, this leakage was small compared to 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 3-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, demonstrating that the single-inducer DIC 3-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 3-Counter by replacing the P_{BAD} promoters in the single-inducer DIC 3-Counter with the inducible promoters $P_{Ltet0-1}$, P_{BAD}, and P_{A1lacO} (Fig. 4A, 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 3-Counter produced a high GFP output (Fig. 4B). No other permutations of the three inducers produced a high output, though some did exhibit a small amount of leakage (Figs. 4C, 4D). These results demonstrate that the circuit can be programmed to only record 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. Additionally, the architectures of both counters allow for the tunable output expression of different protein species of interest at any

number (up to 3 shown) in the counting 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 (see SOM Text for details). In addition to these shared qualities, each counter comes with its own set of properties. Our RTC Counters demonstrate fast activation due to transcriptional and translational regulatory elements, making 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 (10), and they are built with a novel SIMM design that retains counter state based on DNA orientation.

Synthetic gene circuits have enlarged the molecular toolset available to bio-engineers and molecular biologists (4,11–23), enabling them to program novel cellular behaviors (24–26) and construct therapeutic agents (27,28). 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by the NIH Director's Pioneer Award Program, the NSF FIBR program, and the HHMI. Authors' contributions are in Supplementary Online Materials.

References

1. Marcand S, Gilson E, Shore D. Science 1997;275:986. [PubMed: 9020083]
2. Ray A, Runge KW. Mol Cell Biol 1999;19:31. [PubMed: 9858529]
3. Brock DA, Gomer RH. Genes Dev 1999;13:1960. [PubMed: 10444594]
4. Isaacs FJ, et al. Nat Biotechnol 2004;22:841. [PubMed: 15208640]
5. Buchholz F, Angrand PO, Stewart AF. Nat Biotechnol 1998;16:657. [PubMed: 9661200]
6. Groth AC, Calos MP. J Mol Biol 2004;335:667. [PubMed: 14687564]
7. Ham TS, Lee SK, Keasling JD, Arkin AP. Biotechnol Bioeng 2006;94:1. [PubMed: 16534780]
8. Andersen JB, et al. Appl Environ Microbiol 1998;64:2240. [PubMed: 9603842]
9. Wright DA, et al. Nat Protoc 2006;1:1637. [PubMed: 17406455]
10. Santoro SW, Schultz PG. Proc Natl Acad Sci U S A 2002;99:4185. [PubMed: 11904359]
11. Elowitz MB, Leibler S. Nature 2000;403:335. [PubMed: 10659856]
12. Gardner TS, Cantor CR, Collins JJ. Nature 2000;403:339. [PubMed: 10659857]
13. Bayer TS, Smolke CD. Nat Biotechnol 2005;23:337. [PubMed: 15723047]
14. Baker D, et al. Sci Am 2006;294:44. [PubMed: 16711359]
15. Rackham O, Chin JW. Nat Chem Biol 2005;1:159. [PubMed: 16408021]
16. Rinaudo K, et al. Nat Biotechnol 2007;25:795. [PubMed: 17515909]
17. Yokobayashi Y, Weiss R, Arnold FH. Proc Natl Acad Sci U S A 2002;99:16587. [PubMed: 12451174]
18. Guido NJ, et al. Nature 2006;439:856. [PubMed: 16482159]
19. Pedraza JM, van Oudenaarden A. Science 2005;307:1965. [PubMed: 15790857]

20. Deans TL, Cantor CR, Collins JJ. Cell 2007;130:363. [PubMed: 17662949]
21. Fung E, et al. Nature 2005;435:118. [PubMed: 15875027]
22. Kramer BP, et al. Nat Biotechnol 2004;22:867. [PubMed: 15184906]
23. Stricker J, et al. Nature 2008;456:516. [PubMed: 18971928]
24. Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R. Nature 2005;434:1130. [PubMed: 15858574]
25. Kobayashi H, et al. Proc Natl Acad Sci U S A 2004;101:8414. [PubMed: 15159530]
26. Levskaya A, et al. Nature 2005;438:441. [PubMed: 16306980]
27. Lu TK, Collins JJ. Proc Natl Acad Sci U S A 2007;104:11197. [PubMed: 17592147]
28. Lu TK, Collins JJ. Proc Natl Acad Sci U S A 2009;106:4629. [PubMed: 19255432]

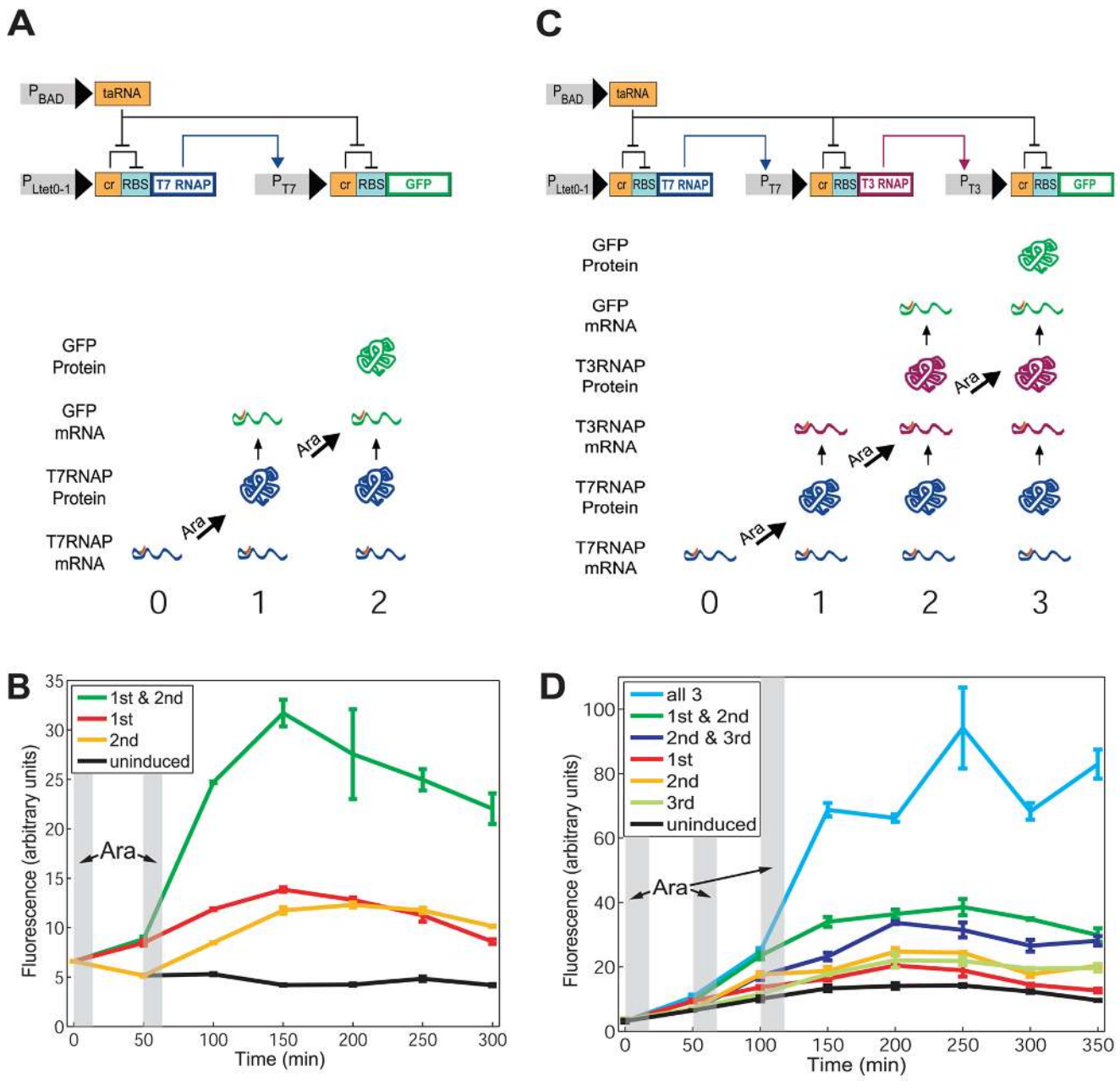

Fig. 1.

The RTC 2-Counter and RTC 3-Counter construct designs and results. (**A**) The RTC 2-Counter is a transcriptional cascade with two nodes. Shown at the bottom are expected expression profiles due to 0, 1, and 2 arabinose (Ara) pulses. (**B**) Mean fluorescence of three replicates of RTC 2-Counter cell populations over time, measured by a flow cytometer. Shaded areas represent arabinose pulse duration. (**C**) The RTC 3-Counter is a transcriptional cascade with three nodes. Shown at the bottom are expected expression profiles due to 0, 1, 2, and 3 arabinose pulses. (**D**) Mean fluorescence of three replicates of RTC 3-Counter cell populations over time, measured by a flow cytometer. Shaded areas represent arabinose pulse duration.

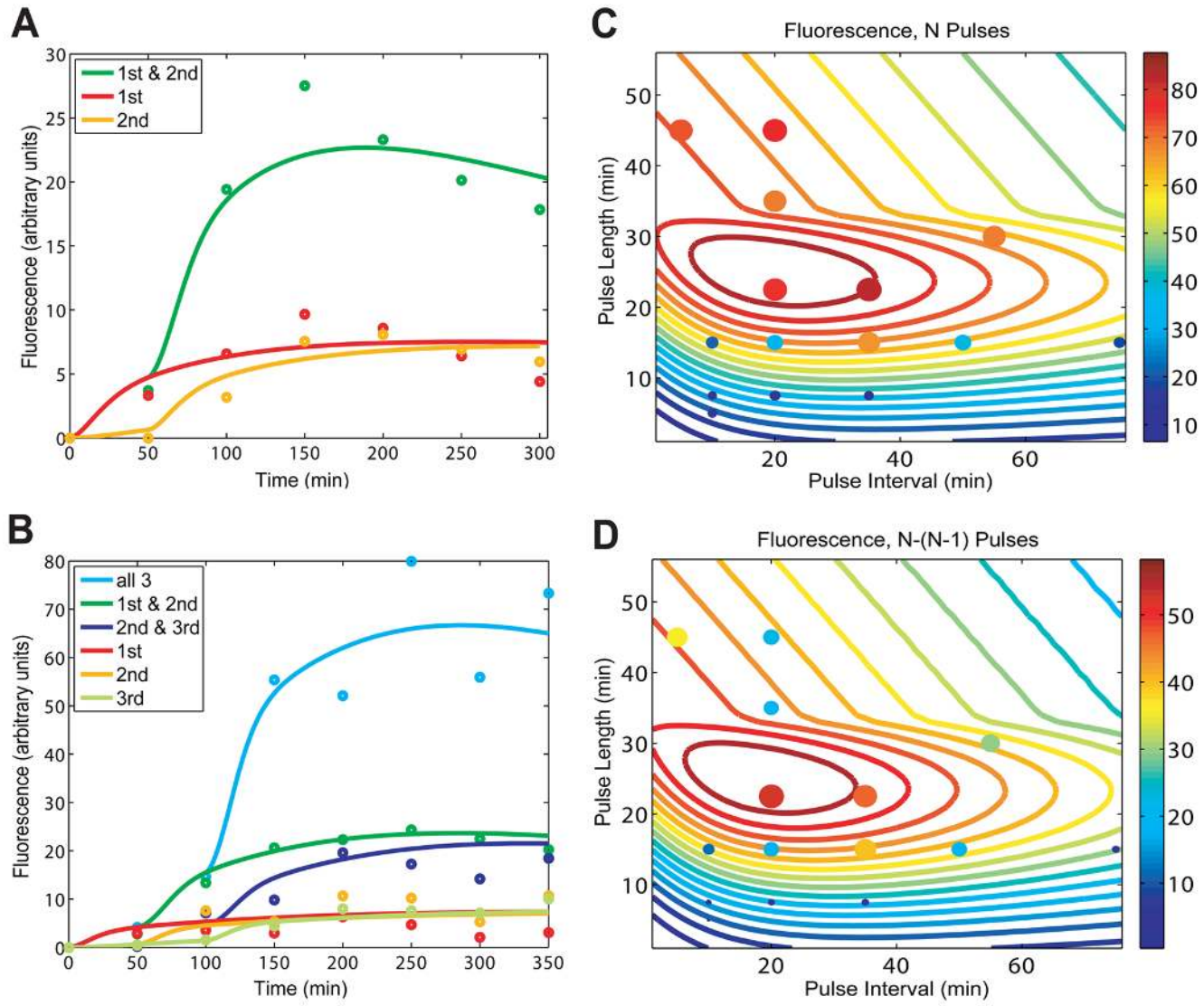

Fig 2.

Modeling predictions and RTC 3-Counter experimental characterization. (**A**) A model with fitted parameters captures the salient features of the normalized fluorescence results of the RTC 2-Counter. (**B**) An expanded model, again with fitted parameters, matches the normalized fluorescence results of the RTC 3-Counter. (**C**) Based on parameters fitted in (B), the model predicts expression output of the RTC 3-Counter 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 values shown are the difference in expression output after three pulses and two pulses.

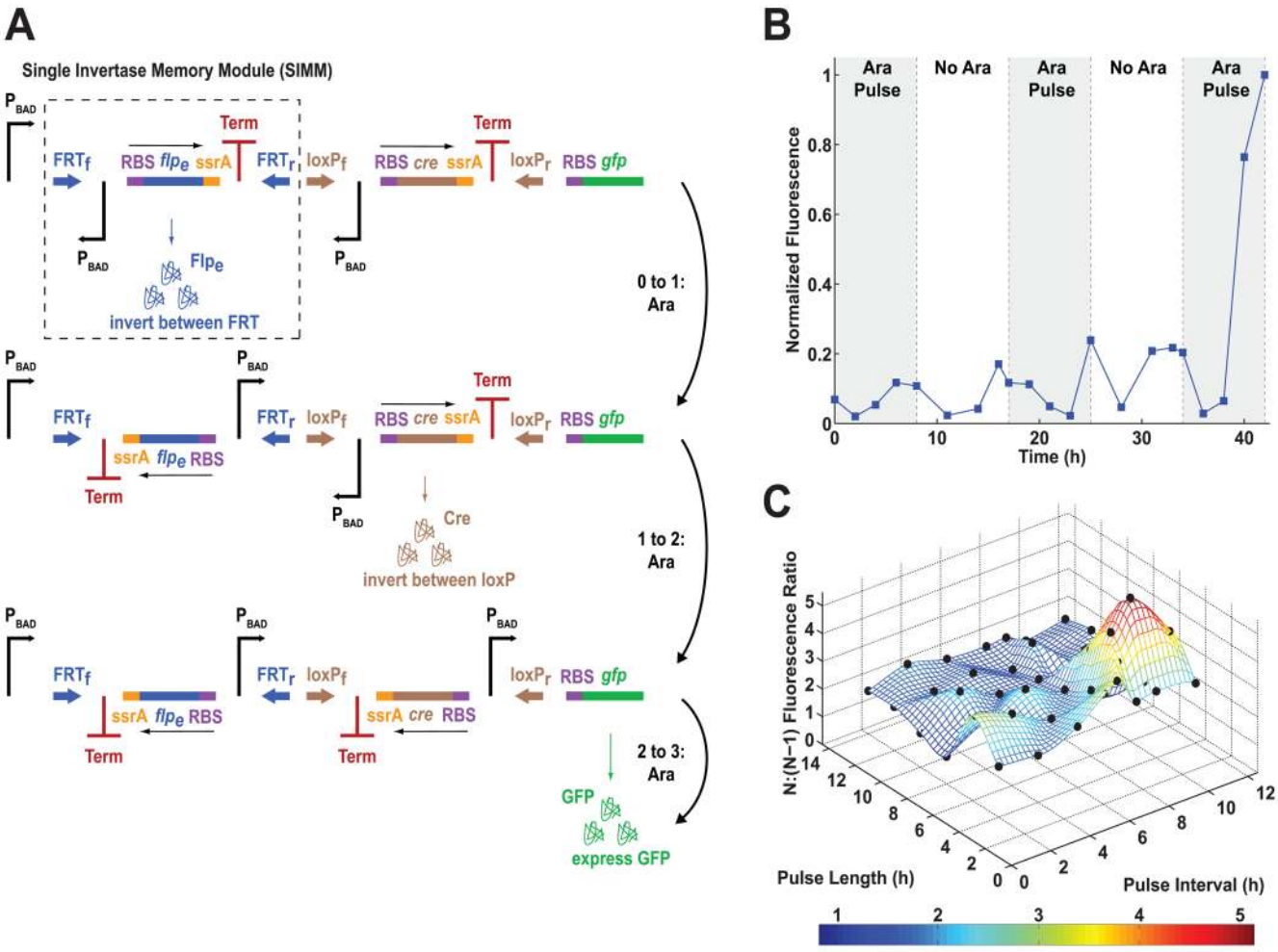

Fig. 3.

The single-inducer DIC 3-Counter construct design and results. (**A**) The single-inducer DIC 3-Counter is built by cascading SIMMs. (**B**) Mean fluorescence of single-inducer DIC 3-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 3-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.

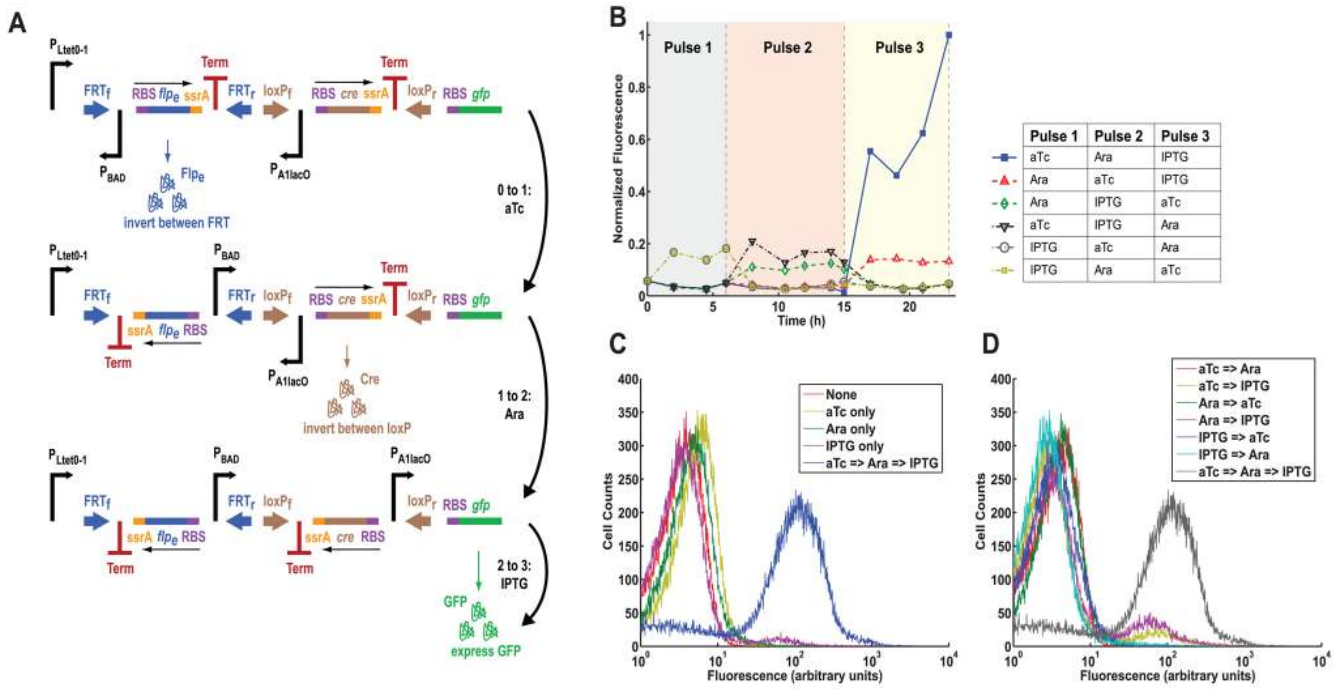

Fig. 4.

The multiple-inducer DIC 3-Counter construct design and results. (**A**) The multiple-inducer DIC 3-Counter is similar to the single-inducer DIC 3-Counter in Fig. 3 except that each promoter is a unique inducible promoter: $P_{Ltet0–1}$, P_{BAD}, and $P_{A1lacO.}$ These promoters respond to anhydrotetracycline (aTc), arabinose, and IPTG, respectively. (**B**) Mean fluorescence of multiple-inducer DIC 3-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 3-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 3-Counter when exposed to the desired sequence of three inducers and to all pairwise permutations of inducers.