

# A Synthetic Recombinase-Based Feedback Loop Results in Robust Expression

Thomas Folliard,<sup>\*,†</sup> Harrison Steel,<sup>‡</sup> Thomas P. Prescott,<sup>‡</sup> George Wadhams,<sup>†</sup> Lynn J. Rothschild,<sup>¶</sup> and Antonis Papachristodoulou<sup>\*,‡,§</sup>

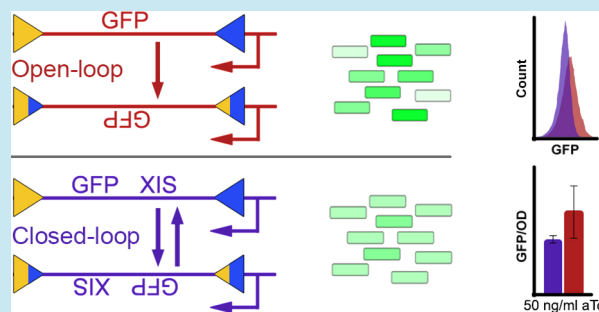
<sup>†</sup>Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, U.K.

<sup>‡</sup>Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, U.K.

<sup>¶</sup>National Aeronautics and Space Administration Ames Research Center, Moffett Field, California 94035, United States

## Supporting Information

**ABSTRACT:** Accurate control of a biological process is essential for many critical functions in biology, from the cell cycle to proteome regulation. To achieve this, negative feedback is frequently employed to provide a highly robust and reliable output. Feedback is found throughout biology and technology, but due to challenges posed by its implementation, it is yet to be widely adopted in synthetic biology. In this paper we design a synthetic feedback network using a class of recombinase proteins called integrases, which can be re-engineered to flip the orientation of DNA segments in a digital manner. This system is highly orthogonal, and demonstrates a strong capability for regulating and reducing the expression variability of genes being transcribed under its control. An excisionase protein provides the negative feedback signal to *close the loop* in this system, by flipping DNA segments in the reverse direction. Our integrase/excisionase negative feedback system thus provides a modular architecture that can be tuned to suit applications throughout synthetic biology and biomanufacturing that require a highly robust and orthogonally controlled output.



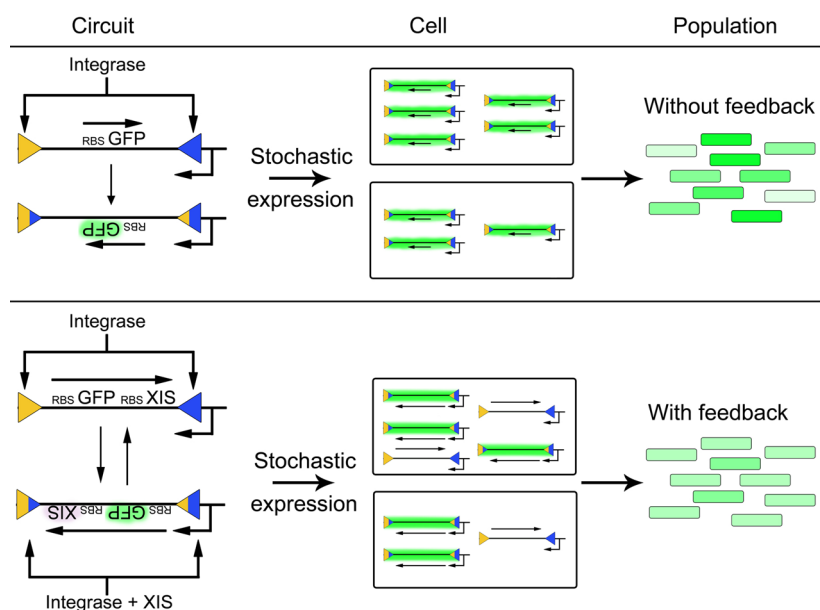
Many important biological processes, such as cell cycle control, use negative feedback to ensure highly robust and reliable operation. Negative feedback is employed extensively in engineered systems that need regulation: by using measurements of a system's output to influence its input (known as *closing the loop*) it is possible to reduce a system's response time, its input–output gain, and the dependence of its response on certain system parameters and external disturbances, which makes its overall performance more robust to fluctuations in both the system's properties and its environment.<sup>1–3</sup> The advantages of negative feedback in natural biological systems and in engineering have motivated researchers to implement similar features in synthetic biological systems.<sup>4–8</sup> The use of transcription factors to regulate expression dynamics is a long-established approach to implementing feedback;<sup>9–15</sup> however, this approach has potential limitations: the reliance of a closed-loop system on endogenous host proteins can result in problems with crosstalk, which arise from interference between its constituent regulators and other cellular processes.<sup>16</sup> Furthermore, the resulting closed-loop system may have very low gain (depending on the repression strength of the regulating transcription factor), which is in some systems undesirable and can make tuning of downstream processes challenging. One approach to overcoming the low gain of negative feedback systems is to redesign protein expression rates to compensate for the

reduced steady state;<sup>10</sup> however, this leads to further design challenges and uncertainties. Thus, it can often be very difficult to engineer a predictable orthogonal output using genetic regulation. The challenges posed by the implementation of the negative feedback systems discussed result in an underutilization of this otherwise clearly advantageous principle, widespread in natural and engineered systems, and as such there is significant need for robust, orthogonal, and easily applicable biological negative feedback architectures.

Serine integrases and their associated excisionases are a class of viral proteins that mediate integration and excision of the phage genome during the lysogenic and lytic phases of the phage life cycle.<sup>17–20</sup> Integrases allow phages to integrate their genome into the host cell and enter the lysogenic phase. They catalyze the insertion of DNA through the recognition of specific ~50 base-pair-long binding sites referred to as attB and attP (attachment Bacteria and attachment Phage, respectively). Integrases belonging to many phages, such as TP901-1, Bxb1, and  $\phi$ C31, recognize orthogonal attB and attP recognition sites via a helix-turn-helix domain.<sup>21</sup> The process of integration is undertaken via strand cleavage, inversion, and ligation to form attL and attR (attachment Left and attachment Right,

Received: April 23, 2017

Published: June 12, 2017



**Figure 1.** Schematic architecture of the open-loop (top) and closed-loop (bottom) expression systems developed in this paper. The correlated expression of excisionase and GFP closes the loop around the integrase's switching process, allowing stochastic switching between ON and OFF states, thereby implementing negative feedback. Thus, while all DNA registers in the open-loop system eventually switch to the ON position, those in the closed-loop system are regulated by excisionase concentration leading to smaller variability in the number of ON registers per cell: If GFP concentration is high (low), excisionase concentration will be likewise, resulting in more OFF (ON) registers and down (up) regulation of GFP expression. Variability introduced by fluctuations in plasmid copy number, as illustrated in the central panel of this figure, is one of many noise sources for which negative feedback can compensate. The yellow and blue triangles represent attP and attB recognition sites, while half-colored triangles represent attL and attR.

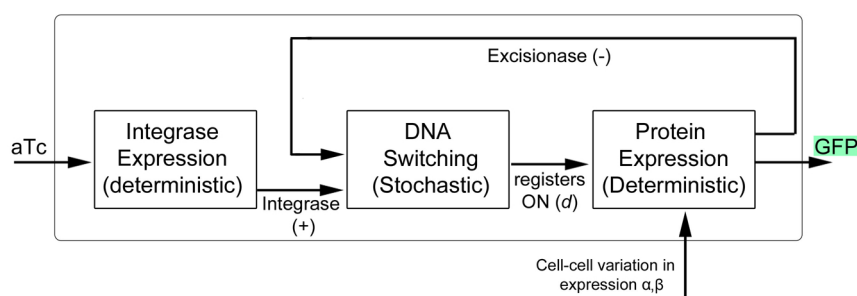
respectively) sites.<sup>22</sup> A central catalytic domain with a conserved serine residue catalyzes this reaction.<sup>22</sup>

Excisionases, via interaction with integrases, allow phages to enter the lytic phase of their life cycle. Passage from lysogenic to lytic phase is usually controlled by the cleavage or phosphorylation of a transcription factor, which flips the bias of a high gain bistable switch toward the expression of the system's integrase and excisionase proteins. The expression of an excisionase cofactor allows the integrase-excisionase complex to recognize attL and attR sites and subsequently catalyze the reverse reaction back to attP and attB.<sup>23,24</sup> Research has shown a potentially symbiotic nature in some such phage–bacterium interactions: For example, in A118-like prophage the phage regions are excised, remain nonintegrated yet also nonlytic, and may subsequently be reintegrated into the host genome to regulate host genes depending on the state of the cell and its environmental conditions.<sup>25</sup>

Integrases and excisionases have been shown to mediate site-specific integration, excision, and inversion depending on the orientation of their binding sites,<sup>22</sup> which has led to a particular focus on their characterization and application in the field of Synthetic Biology.<sup>26,27</sup> By permanently rearranging DNA, integrase/excisionase-based systems can maintain their states over many cellular generations, providing a level of evolutionary robustness that is difficult to achieve using transcription factor-based systems.<sup>32</sup> This class of proteins confers the full range of logic at the genetic level<sup>28,29</sup> (XOR, AND, OR, XNOR, etc.) and allows for a wide range of applications, such as highly sensitive medical assays in healthcare and memory storage,<sup>30–33</sup> as well as metabolic pathway assembly.<sup>34</sup> Given the digital and high gain nature of integrase devices, their use in Synthetic Biology tends to require careful tuning of system parameters to avoid a high level of output uncertainty. This arises due to the

inherent noise in cellular processes and environments, as well as variability in the integrase-mediated flipping of DNA regions, both between cells and between cell populations.<sup>32</sup> As a consequence the expression level of a particular protein of interest in an integrase circuit is often suboptimal and substantially variable, a problem currently limiting the wide application of integrases as predictable and robust biological components for biomanufacturing, healthcare, or environmental remediation systems.

Here we present an engineered synthetic negative feedback loop based on an integrase/excisionase system which could be applied in a diversity of biological settings. We redesign a previously reported DNA register<sup>28</sup> (which expressed Green Fluorescent Protein (GFP) in response to DNA inversion) on a multiple-copy plasmid to coexpress excisionase with the protein of interest, in our case a fluorescent reporter, as illustrated in Figure 1. The excisionase closes the loop around the high gain integrase system by enabling registers in an ON state to stochastically and transiently switch to an OFF state. Over a sufficiently long time period the distribution of registers in either state stabilizes, which due to the coupling of excisionase and GFP production results in a reduced variation in GFP expression between cells in the closed-loop system (both in absolute terms, and when normalized by mean expression levels) when compared to an open-loop system (which shows substantial variability due to gene expression noise). In both cases variability also arises due to cell-to-cell heterogeneity, for example due to variations in plasmid copy number or differences in proteome. In the closed-loop case, as will be demonstrated later, the correlation of excisionase and GFP production results in a reduced variation in GFP expression when compared to the open-loop system.



**Figure 2.** Block diagram for the modeled feedback system. The inducer (aTc) acts as an input to the deterministic expression of integrase via its interaction with the regulatory protein TetR. Integrase (and excisionase) concentration determines the transition probabilities of the random walk of the integer-valued state  $d$  representing the number of ON registers. Excisionase acts as negative feedback, because it biases the random walk toward transitions  $d \rightarrow d - 1$ . The state  $d$  then acts as a random input to the deterministic protein expression dynamics, together with randomly selected constant scaling parameters  $\alpha$  and  $\beta$  to represent variability between populations and cells.

## RESULTS

**Assembly of the Integrase/Excisionase System.** We implemented two synthetic systems based on the integrase DNA switching mechanism, depicted in Figure 1. Both systems placed a DNA coding region between the attachment sites attB and attP, counter-aligned from the constitutive promoter such that the enclosed proteins would not be initially expressed; we call this state the OFF state. The presence of integrase, expressed from its own plasmid (under the regulation of TetR) in response to anhydrotetracycline (aTc) inducer, inverts the DNA coding region to align with the constitutive promoter, thereby enabling expression; we call this state the ON state. In the *open-loop* system, that is, the system *without feedback*, the invertible DNA region codes only for GFP reporter, while in the *closed-loop* system (that *with feedback*) it codes for both GFP and the cognate excisionase. This means that in the closed-loop system the presence of excisionase with the available integrase enables the reversion of the coding region back to its initial OFF state. Thus, negative feedback is achieved by using excisionase as a proxy for fluorescence; when GFP concentration is large, the correlated concentration of excisionase acts to bias the random walk toward lower numbers of ON registers, thereby reducing GFP expression. To prevent the closed-loop system from heavily biasing toward the OFF state we also incorporated a strong *ssrA* degradation tag (already reported,<sup>32</sup> see Table S12 for sequence) on the excisionase to ensure that it was degraded quickly. To ensure the proposed feedback mechanism was working as intended we performed control experiments (see Figure S1) by mutating the excisionase gene, and by removing the *ssrA* tag, demonstrating that these choices were important for our system's function.

**System Modeling and Simulation.** To investigate the performance of our integrase/excisionase-based implementation of negative feedback we developed a mathematical model that combines both stochastic processes and differential equations, for which a block diagram is presented in Figure 2 (for further details see the Methods section). Representative model parameters were chosen in line with the assumptions made during the design of our system in order to define the relative rates of different processes. For example, excisionase degradation was made fast due to the presence of an *ssrA* tag, integrase/excisionase flipping rates were defined such that this process would occur over a multiple-hour time-scale, and plasmid copy numbers were defined based on known properties of the plasmid backbones used. Other free parameters in the model were subsequently tuned so that the model's behavior

would qualitatively capture that exhibited in the experimental data.

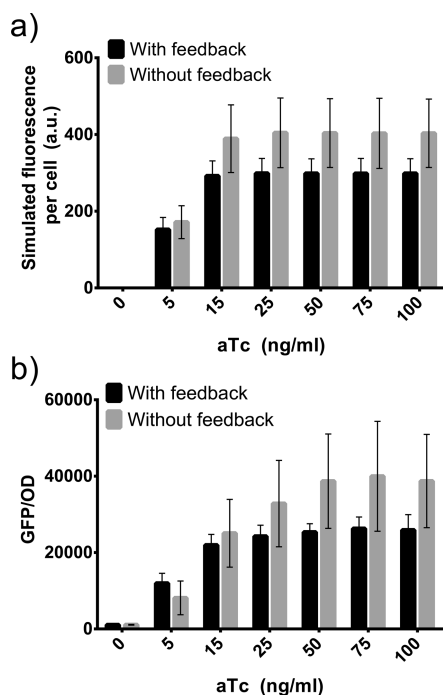
Because of the high variability found in integrase-based systems,<sup>32</sup> our model includes expression noise terms for each system component. A range of approaches to modeling gene expression noise exists,<sup>35,36</sup> of which many have demonstrated applicability to the proteins considered in this work.<sup>37</sup> Our chosen approach, the introduction of random scaling parameters that capture cell-to-cell heterogeneities that cause variability in protein production, was selected to account for both intrinsic (the inherent stochasticity of biochemical processes) and extrinsic (the fluctuations in cellular components) noise sources.<sup>38,39</sup> By utilizing an approach that accounts for variability introduced in a range of cellular processes our model can replicate the impact of major noise sources (both gene-specific and cell-wide<sup>40</sup>) without requiring excessive computational overhead. We model the time evolution of species concentrations using differential equations of the general form

$$\frac{d[y]}{dt} = \alpha \times \beta \times f_1 - f_2 \times [y] \quad (1)$$

where  $[y]$  is the concentration of a species whose noisy expression is being modeled, and  $f_1$  and  $f_2$  are general functions that may depend on system parameters and concentrations. The random parameters used throughout are  $\alpha$ , which was chosen to simulate inherited differences in behavior between populations grown from single cells, and  $\beta$ , which introduces further cell-to-cell variation within each population. Thus,  $\alpha$  accounts for some inherited noise sources (such as long-term inherited differences in proteome) as well as culture-wide variations (such as environmental differences between biological triplicates), while  $\beta$  accounts for cell-to-cell noise sources (such as variations in plasmid copy number, which can strongly impact negative feedback systems<sup>35</sup>) and small scale external noise (such as local environmental inhomogeneities). Simulated triplicates were formed by splitting cells into three populations, each with a different value for  $\alpha$ , and within each population values for  $\beta$  were sampled for individual cells. With this minimal noise model we were able to predict (from a theoretical standpoint) many of the benefits our negative feedback system would provide, which were then confirmed by experimentation as discussed in the following sections.

**Introduction of Excisionase Reduces the System's Expression Level and Output Variation.** Using our mathematical model we simulated the end-point response of

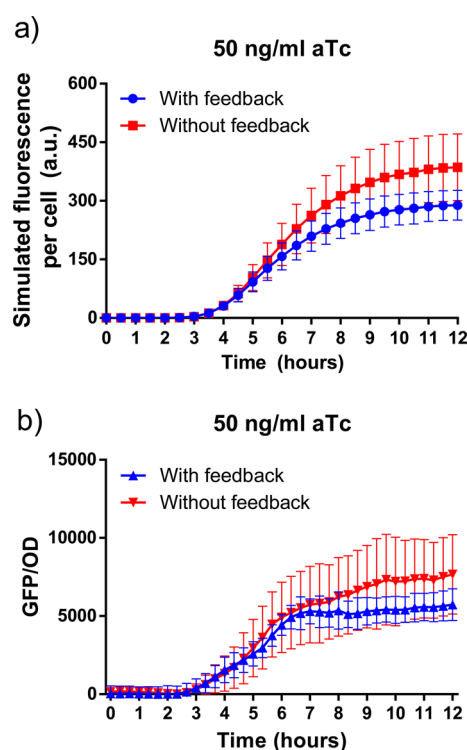
our system (defined to be the fluorescence output 24 h postinduction) to a range of anhydrotetracycline (aTc) concentrations. Figure 3a shows the steady-state expression



**Figure 3.** End-point responses for the open- and closed-loop systems over a range of aTc inputs. (a) The simulated system run for 24 h, where the mean end-point GFP concentration per cell (in arbitrary units) for 2000 sets of triplicates (each consisting of three populations of 2000 cells) is plot, with error bars indicating the mean standard deviation in fluorescence within each triplicate set. (b) Averaged end-point (taken to be 24 h) GFP expression from triplicate biological repeats, where error bars indicate the standard deviation in fluorescence of the triplicates. The y-axis is GFP expression (measured via culture fluorescence) divided by optical density (OD), in arbitrary units. For both experimental and simulated data the feedback system demonstrates a reduction in steady state fluorescence, as well as a reduced variation (when normalized by mean expression level) in expression between triplicates.

per cell, which is uniformly decreased in the closed-loop when compared to the open-loop system. This result was then confirmed experimentally, with the data (presented in Figure 3b) exhibiting a near-uniform trend toward lower expression in the closed-loop system. Figure 3 also demonstrates that over a range of aTc concentrations, in both simulation and experiment, the variation in end-point expression of the closed-loop system is lower than in the open-loop. For the end-point experimental data in Figure 3b, we find that on average inclusion of the negative feedback system reduces the coefficient of variation of fluorescence (the standard deviation of the triplicates' mean fluorescence divided by the mean fluorescence of each triplicate set) for each triplicate set (with aTc  $\geq$  15 ng/mL) by a factor of  $\sim$ 2.8.

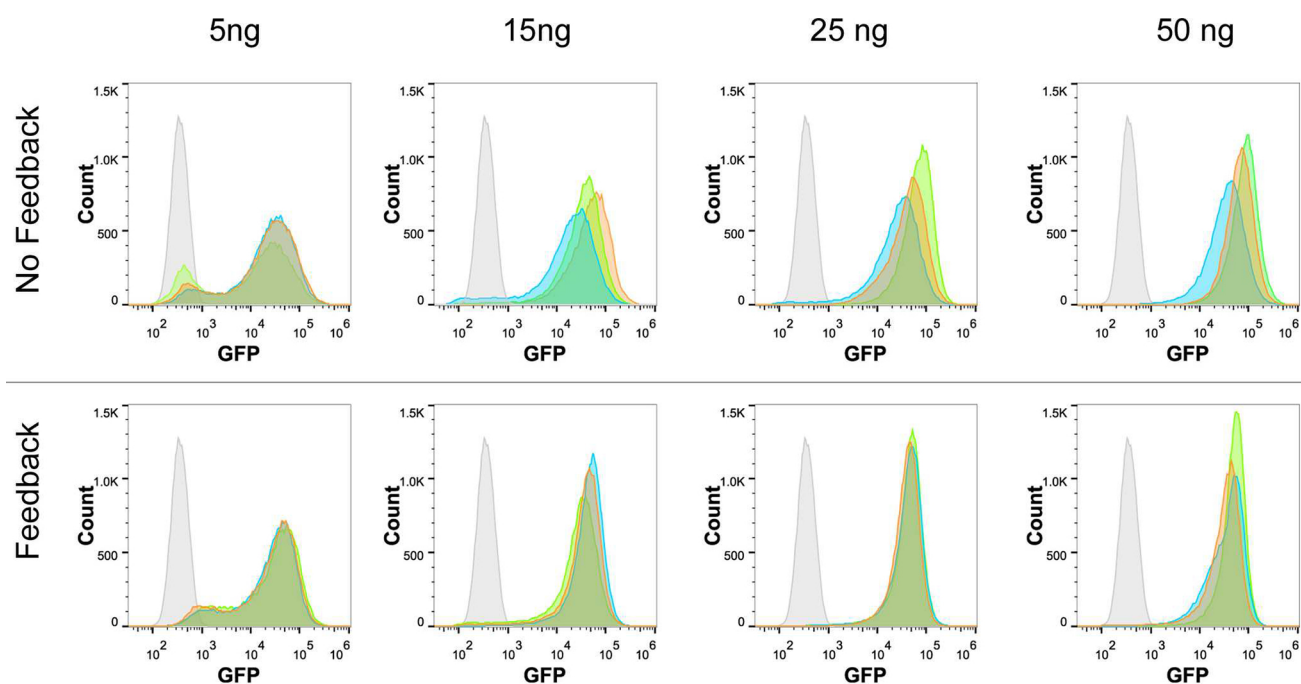
To investigate the impact of negative feedback on the system's transient response we simulated time-course data for GFP expression-per-cell with 50 ng/mL aTc concentration (Figure 4a), and then performed the corresponding experiment for comparison (Figure 4b, for other aTc concentrations see Figure S2). A control experiment was performed using the same procedure with constitutively expressed GFP to provide a



**Figure 4.** Time-course fluorescence data for the open- and closed-loop systems induced with 50 ng/mL aTc over 12 h. (a) The simulated system as in Figure 3, where the overall average of the mean and standard deviation of 2000 triplicates is plot each 30 min; (b) averaged time-course GFP expression from triplicate biological repeats, for which error bars denote the standard deviation of the individual triplicate means. The y-axis is GFP expression (measured via culture fluorescence) divided by OD, in arbitrary units. Equivalent plots for the full range of inducer concentrations are shown in Figure S2. In both the experimental and simulated data the feedback system demonstrates a reduced variation in output fluorescence levels.

time-series comparison with the integrase-based system (Figure S3). In both simulated and experimental data the closed-loop system provides a reduction in expression uncertainty: the size of the error bars, again when normalized by the mean expression level at a given point, is reduced for the closed-loop system, even early in the response when the mean expression levels do not significantly differ. We tested our system's robustness to changes in plasmid copy numbers, for both the plasmid containing the integrase gene (Figure S4), and that with the reversible DNA register containing excisionase and GFP (Figure S5). Simulations predict that the noise-reducing property of our negative feedback system is maintained over a wide range of copy numbers, except for the case (given the previous parameter values) in which the plasmid containing excisionase and GFP has very low copy (as would occur with genome integration). In this circumstance excisionase expression levels are too low to provide an adequate feedback signal, though this can be overcome (see Figure S5) by increasing the rate of excisionase expression (for example, by using a stronger RBS).

**Cell-to-Cell Noise Is Reduced by the Synthetic Feedback Loop.** Having demonstrated that our negative feedback system can provide closer alignment between population mean expressions (as depicted in Figures 3 and 4), we sought to investigate its impact on the distribution of GFP expression within a population of cells via simulations and



**Figure 5.** Single-cell flow cytometry data for triplicates of both the open- and closed-loop systems at a range of inducer concentrations. The x-axis is the level of GFP expression (measured via fluorescence) per cell, in arbitrary units. The triplicate biological repeats are overlaid and colored green, blue, and orange, and the gray distribution is a negative *E. coli* control. We observe that feedback results in higher reproducibility between replicates, both in terms of the location of the histogram's peak, as well as its overall shape. The system with feedback also results in a narrower histogram, indicating a reduction in cell-to-cell variability.

flow cytometry. We simulated triplicate colonies over the course of 24 h, recording GFP expression distribution histograms at different time points (Figure S6). We observe that expression distributions for the closed-loop system are narrower at all time points (when compared to the open-loop), and that population peaks are also more closely aligned. This prediction was validated experimentally using single cell flow cytometry to obtain the histograms in Figure 5, where each plot depicts the GFP expression distribution of three populations (biological triplicates) of an open-loop or closed-loop system at a given inducer concentration. In the presence of feedback we find a closer alignment of the triplicates' means, as seen in earlier experiments, and also observe that compared to the open-loop histograms, those of the closed-loop system have a narrower distribution of fluorescence within each population. This is made clear in Figure S7, in which the data from the closed- and open-loop cases for one of the triplicates has been overlaid. The reduction in population variation is quantized by calculating coefficients of variation (CV) for each distribution, which are on average 25% lower in the closed-loop system when compared to the open-loop (Figure S8).

**Downstream Process Is Easily Tuned to a Predictable Threshold.** Using the Salis Ribosome Binding Site (RBS) calculator<sup>41</sup> we generated a small library of GFP expression variants (that is, GFP genes with different RBSs) over an approximate 10-fold range in strength, and integrated each one into the closed-loop system. For each of these four variants we observed a reliable and distinct level of expression in the downstream process reporter (see Figure S9), which correlated with the relative theoretical RBS strengths. This demonstrates a straightforward approach via which the implementation of our feedback system can be tuned to reliably set a controlled gene's output to a desired level.

## DISCUSSION

Our results demonstrate a good correspondence between the qualitative behavior of the simulated and experimental systems, in both cases highlighting the beneficial properties of negative feedback (closed-loop) systems. Figure 3 demonstrates a reduced steady-state response in the closed-loop system, which can be justified as follows: in the open-loop system the DNA registers can only switch ON, without possibility of reversion, because of the absence of excisionase. Thus, the GFP expression of the open-loop system is maximal given the induction levels and constitutive expression rates of GFP. In contrast, for the closed-loop system the excisionase expression provides negative feedback resulting in reversion of DNA registers, and thus on average a reduced fraction will be in the ON position, leading to a lower expression level of the GFP reporter.

Though final output levels of the closed- and open-loop systems differ in this manner, their behavior aligns closely for the first few hours of each experiment. This is due to the time required for sufficient excisionase to be produced in order to complete the feedback loop, which itself is delayed by the time required for aTc induction to prompt transcription and translation of the integrase. A useful feature of many feedback systems is their ability to reduce rise time, which has previously been demonstrated in biological systems.<sup>10</sup> However, based upon analysis of time-course data (using the method of Rosenfeld et al.<sup>10</sup>) the feedback system considered in this work did not provide a statistically significant speed-up of system response. To further study this effect future work will focus on the utilization of a turbidostat setup that is able to maintain a culture's growth in exponential phase.<sup>42</sup> Such an approach would also allow experiments to be run for longer periods of

time, and reduce the impact of changing growth-phase on time-series data.

In all experiments and simulations it was found that variability was reduced in the closed-loop system. For example, at steady state in Figure 3b the coefficient of variation for the triplicates was reduced by a factor of  $\sim 2.8$  in the closed-loop system when compared to the open-loop variant. This might be considered counterintuitive, since one could initially hypothesize that expressing excisionase introduces extra stochasticity to the expression of the reporter protein, thereby increasing the variability of its output. However, to understand why this intuitive reasoning fails, first note that the regulated output is not the number of DNA registers in the ON state, but rather the concentration of GFP. In populations of cells with faster (slower) constitutive GFP expression rates than the mean, the correlation between excisionase and GFP expression (since they are in the same operon) means that excisionase will also be produced at a correspondingly faster (slower) rate, thereby speeding (slowing) the rate at which the populations' DNA registers revert back to OFF. Thus, populations with GFP expression that tends away from the mean in open-loop due to natural cell-to-cell variation will be regulated closer to the mean in closed-loop, providing negative feedback functionality. Considering this in the context of our model, if the random scaling factor  $\alpha\beta$  is large (small), which might correspond to the GFP/excisionase plasmid having higher (lower) copy number in that cell, then for each ON register there will be a large (small) expression of both GFP and excisionase, which will reduce (increase) the number of ON registers. So, for cells with large (small)  $\alpha\beta$  the level of expression of the target protein per ON register will be high (low), but there will be fewer (more) ON registers.

The variability between populations demonstrated by flow cytometry measurements in Figure 5 (and their similarity to Figure S6) supports the approach taken to introduce random variations between cells in our model: we observe (particularly in the open-loop case) substantial variation between triplicates in terms of the center of their GFP fluorescence distribution, which corresponds to varying values for the random parameter  $\alpha$ , as well as a wide distribution of fluorescence values within each population, corresponding to variation in the parameter  $\beta$ . Both of these parameters reflect natural differences in behavior that arise between cells:  $\alpha$  accounts for differences (for example, in the proteome) that are inherited from the small number of cells used to initiate each triplicate, while  $\beta$  reflects the spread in cell properties within each triplicate during growth and experiment (for example, due to inhomogeneities in environmental conditions or plasmid copy number fluctuations). Further work could be done to refine the probability distributions from which  $\alpha$  and  $\beta$  are sampled, including experimental study to determine appropriate correlations between the  $\beta_i$ 's that were used to introduce noise into different equations.

Though the negative feedback system developed in this work has achieved a reduction in our system's interpopulation and cell-to-cell variability, further development and improvement is possible. We studied how the parameter  $D$  (which represents the copy number of the plasmid containing the DNA register) impacts the system's noise-reducing behavior (Figure S5), demonstrating that by adjusting system parameters (such as the excisionase expression or degradation rates) our closed loop system can be tuned to function at a wide range of copy numbers (including single-copy genome integration). Given

that the feedback signal is excisionase concentration, the "actuator" of GFP expression is the flipping of DNA registers to ON or OFF positions. Thus, the actuator signal controlling GFP expression is inherently quantized, since only an integer number  $d$  of the total  $D$  registers can be ON, leading to expression rates at fractions  $d/D$  of the maximum. In engineered control systems, quantized signals are known to degrade performance below the optimum.<sup>1</sup> We thus theorize that an increased number of DNA registers ( $D$ ) would reduce the effect of quantization, thereby improving the noise-attenuating properties of our system further. However, larger copy numbers would increase the burden placed on the system's host cell (something not considered in our modeling), potentially causing a subsequent loss of output signal.

To further improve the robustness of our integrase-based system future work may focus on including the expression of TetR within its own negative feedback loop, thereby compensating for variability in its (presently constitutive) expression. This would then reduce noise in the integrase expression process, making our system's response to varying inducer concentrations more reliable. Experimentally this could be achieved by placing TetR under the control of a pTet promoter (as was used in this work to regulate integrase expression), resulting in a negative autoregulatory architecture (though this may result in leaky integrase expression, which can circumvent induction systems at very low levels<sup>28,32</sup>). Similar architectures have been used extensively in the literature, demonstrating their ability to robustly reduce expression noise of downstream genes.<sup>14,15</sup>

Though existing transcription factor-based negative feedback systems provide an effective approach to regulation of gene expression noise,<sup>14</sup> the synthetic feedback system developed herein is of particular interest as it centered on serine integrases: these components have been extensively characterized in Synthetic Biology, and libraries of integrase proteins with orthogonal binding domains exist.<sup>43,44</sup> Thus, our design is likely to be easily scalable, allowing a number of integrase-excisionase feedback circuits to function within the same cell. Given that phage-derived integrases have been used in systems with a wide variety of output proteins over the last three decades,<sup>45–47</sup> there should be no significant limitations in replacing the GFP reporter in our design with an alternate target protein, making this framework widely adaptable (and tunable, as demonstrated in Figure S9) to a broad range of applications in a plug and play fashion. Furthermore, the phage proteins are non-native and have been shown to function in many different domains of life.<sup>48,49</sup> Therefore, there should be minimal crosstalk between cellular processes and the feedback circuit (as is often found in transcription-based feedback designs), which makes this framework widely transferable between different hosts and applications. It would therefore be especially useful in systems that require highly accurate and reproducible protein expression, such as for the regulation of degradase expression for degradation of orthogonally tagged proteins, or to uniformly express a complemented gene in knock-in studies.

## METHODS

**Plasmids.** The target plasmid<sup>28</sup> containing GFP flanked by inverted Bxb1 recognition sites in the opposite orientation to the P7 constitutive promoter<sup>50</sup> was used as the control (open-loop) system without feedback (P15A origin, 15–20 copies, Ampicillin resistance,<sup>32</sup> see Figure S10 for schematic). The

excisionase gene<sup>32</sup> with a tuned *ssrA* degradation tag was synthesized by IDT and inserted after the GFP output by Gibson assembly<sup>51,52</sup> (see Figure S11 for schematic). A second plasmid<sup>28</sup> (ColE1, 50–60 copies, Chloramphenicol resistance) containing Bxb1 integrase under control of the pTet promoter was cotransformed in each experiment to provide Bxb1 integrase expression. Sequences for ribosome binding sites and *ssrA* tags are provided in Table S12.

**Cell Culture and Experimental Conditions.** All experiments were performed using *E. coli* DH5 $\alpha$ Z1.<sup>53</sup> For each experiment three colonies were picked from freshly transformed plates into LB media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl)<sup>54</sup> with appropriate antibiotics and grown overnight, each was then diluted 1/1000 with reported concentrations of inducer and appropriate antibiotics. For end point experiments (Figures 3 and 5) cultures were grown in a dark incubator for 24 h at 37 °C and 225 rpm in 1.5 mL opaque deep well plates (Fisher). For time course experiments 200  $\mu$ L microcultures were prepared in 300  $\mu$ L black well, clear bottom plates (Fisher) and run for 12 h in a dark BMG clariostar plater reader with a gain of 1000 at 37 °C and shaken at 200 rpm. The antibiotics used were carbenicillin (100  $\mu$ g/mL) and chloramphenicol (25  $\mu$ g/mL) (Sigma). Anhydrotetracycline (Sigma) was used in reported concentrations.

**Measurement and Data Analysis.** End point experiments were centrifuged and resuspended in PBS.<sup>55</sup> Individual GFP and optical density (OD) OD600 measurements were taken using a Clariostar plate reader with a gain of 1000 before the data was normalized by OD600. Time course data was collected using sequential measurement of GFP and OD600. Flow cytometry was performed on an Attune flow cytometer; for each sample 50000 cells were recorded. GFP intensities were quantified by removing any data points below the negative control population, to remove cell fragments and other artifacts. All experimental data presented herein is available online.<sup>56</sup>

**Mathematical Modeling.** Considering our system at the single-cell level, we assumed that each cell contains  $D$  copies of the plasmid containing the DNA register depicted in Figure 1, where the attB and attP integrase recognition sites are separated by either a reporter coding region (i.e., open-loop), or by both reporter and excisionase coding regions (i.e., closed-loop).

Integrase is expressed on a different plasmid of copy number  $N$ , under the control of the tetracycline transcription factor (TetR). We modeled the concentration of active (repressing) TetR ( $[TetR]$ ) by assuming there is a constant total TetR concentration per cell ( $[TetR]_0$ ) that is reversibly converted to its inactive (nonrepressing) form at a rate proportional to aTc concentration  $[aTc]$ . This results in the deterministic differential equation

$$\frac{d[TetR]}{dt} = k_1(\alpha\beta_1[TetR]_0 - [TetR]) - k_{-1}[TetR][aTc]$$

where  $k_1$  and  $k_{-1}$  are the forward and reverse reaction rates, respectively. The combined parameter  $\alpha\beta_1$  is a cell-specific scaling factor for expression rates, chosen as discussed in the following section, which models the cell-to-cell uncertainty in protein expression (in this case the equilibrium TetR concentration per cell). We modeled the dynamics of integrase concentration ( $[I]$ ) as the net balance of its expression under the regulation of TetR with degradation and dilution. This results in the deterministic differential equation

$$\frac{d[I]}{dt} = N\mu_1\alpha\beta_2\frac{K_1}{K_1 + [TetR]} - \delta_I[I]$$

where  $\mu_1$  and  $\delta_I$  are parameters representing the maximal integrase expression per plasmid, and the degradation/dilution rates of integrase molecules, respectively. The parameter  $K_1$  is the dissociation constant for TetR repressing integrase expression, and in this case  $\alpha\beta_2$  models the cell-to-cell uncertainty in integrase expression around the nominal rate  $\mu_1$ .

Similarly to integrase expression above, when  $d$  of the  $D$  plasmids have flipped to the ON state, we have expression of GFP and (in the closed-loop case) excisionase, and also degradation of each. We thus model the concentrations of GFP ( $[GFP]$ ) and excisionase ( $[X]$ ) using the differential equations

$$\begin{aligned}\frac{d[X]}{dt} &= d\alpha\beta_3\mu_X - \delta_X[X], \\ \frac{d[GFP]}{dt} &= d\alpha\beta_3\mu_G - \delta_G[GFP]\end{aligned}\quad (2)$$

where the scaling  $d$  of the expression rates  $\mu_X$  and  $\mu_G$  reflects an assumption that the expression of excisionase and GFP is linearly proportional to the number of ON registers. In this case the combined parameter  $\alpha\beta_3$  models the cell-to-cell uncertainty in protein expression around the nominal rates  $\mu_{3X}$  and  $\mu_{3G}$ , which is set equal for  $[X]$  and  $[GFP]$  since they are expressed from the same operon on the same plasmid. Note that setting  $\mu_X = 0$  models the open-loop behavior, while  $\mu_X > 0$  models closed-loop behavior.

Consider now the closed-loop circuit. The presence of integrase causes the  $D$  registers to begin to flip from OFF to ON. When  $d$  of the  $D$  registers have flipped ON, both GFP and excisionase are expressed. Integrase and excisionase together cause some of the  $d$  ON registers to flip back to OFF, thus slowing expression of GFP and excisionase. Hence, in addition to the linear deterministic models described above, we also need to model the stochastic, digital behavior of the  $D$  DNA registers. The variable  $d(t)$  representing the number of ON registers at time  $t$  takes a random walk on the set of integers  $d \in 0, 1, \dots, D$ . Given  $d(t)$ , the rate at which  $d$  jumps to  $d + 1$  is modeled as

$$r_{ON}(D - d)\frac{([I])^4}{K_{ON}^4 + ([I])^4}$$

This rate is constructed as a scalar parameter  $r_{ON}$ , representing a switching rate, multiplying the expected number of OFF registers that have two integrase dimers bound, thereby being in the configuration able to stochastically switch to the ON state. In particular, the integrase concentration-dependent fraction in this rate is a Hill function with parameter 4, representing the fact that a total of four integrase molecules needs to bind to the DNA to be in the correct configuration to switch the register to the ON position. The parameter  $K_{ON}$  is a combined parameter that is a function of the dissociation constant of integrase dimerization and the dissociation constant of integrase dimer binding to each DNA recognition site. More complex models of DNA recombination exist<sup>57,58</sup> and may be used to expand our model in the future. However, at present we have utilized a simple hill function to model recombination due to the small number of free parameters it introduces, and to reduce computational complexity.

The rate at which  $d$  jumps to  $d - 1$  is modeled as

$$r_{\text{OFF}} d \frac{([I])^4}{K_{\text{ON}}^4 + ([I])^4} \frac{([X])^4}{K_{\text{OFF}}^4 + ([X])^4}$$

Similarly to the case above, this rate is constructed as a scalar parameter  $r_{\text{OFF}}$ , representing a switching rate, multiplying the expected number of ON registers that have two integrase dimer/excisionase dimer complexes bound, thereby being in the configuration able to stochastically switch to the OFF state. In this equation both the integrase concentration-dependent fraction and the excisionase concentration-dependent fraction are Hill functions with parameter 4, since four integrase and four excisionase molecules are required for an ON register to flip to OFF. Similarly to the previous case, the parameter  $K_{\text{OFF}}$  represents the dissociation constant of excisionase with integrase.

A diagram summarizing the structure of our model is depicted in Figure 2. The number  $d$  of DNA registers in the ON state follows the dynamics of a random walk, with time-varying transition rates that depend on the concentrations  $[I]$  and  $[X]$  of integrase and excisionase. In the absence of feedback,  $[X]$  is assumed to equal zero for all time, so that the switching process only increases the number of ON registers.

**Simulation Setup and Parameters.** For the open- and closed-loop cases ( $\mu_X > 0$  and  $\mu_X = 0$ , respectively) with input levels 0, 5, 15, 25, 50, 75, and 100 ng/mL, we simulated 5000 single cells using the mathematical model described above with the parameters given in Table S13. This was done with values of  $\alpha\beta_i$  sampled evenly (in probability) from the cumulative distribution function of a log-normal distribution,  $\ln \mathcal{N}(\mu - \sigma_1^2/2, \sigma_1^2)$  with  $\mu = 1$  and  $\sigma_1 = 0.5$ , which gave us a distribution of cell behaviors from which we could artificially simulate triplicates. At every half-hour during the simulation the following process was repeated 2000 times to form and analyze an equivalent to 2000 sets of biological triplicates (triplicates are indexed by  $j$ ): We sampled values  $\alpha_j \in \ln \mathcal{N}(-\sigma_2^2/2, \sigma_2^2)$  for  $j = 1, 2, 3$  with  $\sigma_2 = 0.25$ , and then randomly selected 2000 cells with  $\alpha\beta_i \in \ln \mathcal{N}(\ln(\alpha_j) - \sigma_2^2/2, \sigma_2^2)$  from the 5000 single cells simulations to form each of the three triplicate populations. Scaling factors for different processes ( $\beta_1, \beta_2, \beta_3$ ) are thus sampled independently from a log-normal distribution with mean  $\alpha_j$  for each triplicate. This process of sampling cell populations from a constant pool of simulated cells with a predefined range of scaling factors was pursued instead of simulating unique populations to form triplicates as it allows us to compute reliable statistics much more efficiently. For each simulated triplicate the mean fluorescence of each of the three populations was calculated, as was the standard deviation of these means. After the 2000 repetitions of this process the mean of the individual triplicate means was calculated, as was the mean of the triplicate standard deviations, which is the data presented in (for example) Figure 4a. To give the end-point data (as in Figure 3a) a simulation of 5000 cells was modeled for 24 h, after which the process described above was performed to generate mean triplicate statistics.

## ■ ASSOCIATED CONTENT

### ● Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.7b00131.

Figures S1–11 and Table S12–13 (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: tom@folliard.org.

\*E-mail: antonis@eng.ox.ac.uk.

### ORCID

Antonis Papachristodoulou: 0000-0002-3565-8967

### Author Contributions

A.P. designed the research; T.F. designed and performed experiments; T.P.P., A.P., and H.S. performed modeling and computational work; all authors analyzed data; all authors wrote the paper.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by EPSRC project EP/M002454/1. H.S. would like to acknowledge the General Sir John Monash foundation for partial financial support of this work.

## ■ REFERENCES

- (1) Aström, K. J., and Murray, R. M. (2010) *Feedback Systems: An Introduction for Scientists and Engineers*, Princeton University Press, New Jersey, USA.
- (2) Thattai, M., and van Oudenaarden, A. (2001) Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U. S. A.* 98, 8614–8619.
- (3) Oyarzun, S. A., Lugagne, J. B., and Stan, G. B. V. (2015) Noise propagation in synthetic gene circuits for metabolic control. *ACS Synth. Biol.* 4, 116–125.
- (4) Iglesias, P. A., and Ingalls, B. P., Eds. (2010) *Control Theory and Systems Biology*, MIT Press, Cambridge, MA, USA.
- (5) Cosentino, C., and Bates, D. (2011) *Feedback Control in Systems Biology*, CRC Press.
- (6) Shimoga, V., White, J. T., Li, Y., Sontag, E., and Bleris, L. (2013) Synthetic mammalian transgene negative autoregulation. *Mol. Syst. Biol.* 9, 1–7.
- (7) Bloom, R. J., Winkler, S. M., and Smolke, C. D. (2015) Synthetic feedback control using an RNAi-based gene-regulatory device. *J. Biol. Eng.* 9, 1–13.
- (8) Briat, C., Gupta, A., and Khammash, M. (2016) Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks. *Cell Systems* 2, 15–26.
- (9) Jusiak, B., et al. (2014) Synthetic Gene Circuits. *Reviews in Cell Biology and Molecular Medicine*, 1–56.
- (10) Rosenfeld, N., Elowitz, M. B., and Alon, U. (2002) Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.* 323, 785–793.
- (11) Madar, D., Dekel, E., Bren, A., and Alon, U. (2011) Negative auto-regulation increases the input dynamic-range of the arabinose system of *Escherichia coli*. *BMC Syst. Biol.* 5, 111.
- (12) Daniel, R., Rubens, J. R., Sarpeshkar, R., and Lu, T. K. (2013) Synthetic analog computation in living cells. *Nature* 497, 619–623.
- (13) Ferrell, J. E. (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell Biol.* 14, 140–148.
- (14) Nevozhay, D., Adams, R. M., Murphy, K. F., Josic, K., and Balázsi, G. (2009) Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5123–5128.
- (15) Nevozhay, D., Adams, R. M., and Balázsi, G. (2011) Linearizer Gene Circuits with Negative Feedback Regulation in *Yeast Genetic Networks: Methods and Protocols* (Becskei, A., Ed.) pp 81–100, Springer.
- (16) Cardinale, S., and Arkin, A. P. (2012) Contextualizing context for synthetic biology-identifying causes of failure of synthetic biological systems. *Biotechnol. J.* 7, 856–866.



- (17) Smith, M. C. M., and Thorpe, H. M. (2002) Diversity in the serine recombinases. *Mol. Microbiol.* 44, 299–307.
- (18) Breüner, A., Brondsted, L., and Hammer, K. (2001) Resolvase-like recombination performed by the TP901–1 integrase. *Microbiology* 147, 2051–2063.
- (19) Keravala, A., and Calos, M. P. (2008) Site-specific chromosomal integration mediated by phiC31 integrase. *Methods Mol. Biol.* 435, 165–173.
- (20) Jain, S., and Hatfull, G. F. (2000) Transcriptional regulation and immunity in mycobacteriophage Bxb1. *Mol. Microbiol.* 38, 971–985.
- (21) Morita, K., et al. (2009) In vitro characterization of the site-specific recombination system based on actinophage TG1 integrase. *Mol. Genet. Genomics* 282, 607–616.
- (22) Grindley, N. D. F., Whiteson, K. L., and Rice, P. A. (2006) Mechanisms of site-specific recombination. *Annu. Rev. Biochem.* 75, 567–605.
- (23) Savinov, A., Pan, J., Ghosh, P., and Hatfull, G. F. (2012) The Bxb1 gp47 recombination directionality factor is required not only for prophage excision, but also for phage DNA replication. *Gene* 495, 42–48.
- (24) Ghosh, P., Wasil, L. R., and Hatfull, G. F. (2006) Control of phage Bxb1 excision by a novel recombination directionality factor. *PLoS Biol.* 4, e186.
- (25) Feiner, R., et al. (2015) A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nat. Rev. Microbiol.* 13, 641–650.
- (26) Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S., and Lu, T. K. (2016) Synthetic recombinase-based state machines in living cells. *Science* 353, aad8559.
- (27) Hsiao, V., Hori, Y., Rothmund, P. W., and Murray, R. M. (2016) A population-based temporal logic gate for timing and recording chemical events. *Mol. Syst. Biol.* 12, 869.
- (28) Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P., and Endy, D. (2013) Amplifying genetic logic gates. *Science* 340, 599–603.
- (29) Siuti, P., Yazbek, J., and Lu, T. K. (2013) Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* 31, 448–452.
- (30) Yang, L., et al. (2014) Permanent genetic memory with > 1-byte capacity. *Nat. Methods* 11, 1261–1266.
- (31) Courbet, A., Endy, D., Renard, E., Molina, F., and Bonnet, J. (2015) Detection of pathological biomarkers in human clinical samples via amplifying genetic switches and logic gates. *Sci. Transl. Med.* 7, 289ra83.
- (32) Bonnet, J., Subsoontorn, P., and Endy, D. (2012) Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. U. S. A.* 109, 8884–8889.
- (33) Farzadfar, F., and Lu, T. K. (2014) Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* 346, 1256272.
- (34) Colloms, S. D., Merrick, C. A., Olorunniji, F. J., Stark, W. M., Smith, M. C. M., Osbourn, A., Keasling, J. D., and Rosser, S. J. (2014) Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res.* 4, e23.
- (35) Dublanche, Y., Michalodimitrakis, K., Kümmerer, N., Foglierini, M., and Serrano, L. (2006) Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol. Syst. Biol.* 2, 41.
- (36) Tian, T., and Burrage, K. (2006) Stochastic models for regulatory networks of the genetic toggle switch. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8372–8377.
- (37) Murphy, K. F., Adams, R. M., Wang, X., Balazsi, G., and Collins, J. J. (2010) Tuning and controlling gene expression noise in synthetic gene networks. *Nucleic Acids Res.* 38, 2712.
- (38) Swain, P. S., Elowitz, M. B., and Siggia, E. D. (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12795–800.
- (39) Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002) Stochastic gene expression in a single cell. *Science* 297, 1183–6.
- (40) Sanchez, A., and Golding, I. (2013) Genetic determinants and cellular constraints in noisy gene expression. *Science* 342, 1188–1193.
- (41) Espah Borujeni, A., Channarasappa, A. S., and Salis, H. M. (2014) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* 42, gkt1139.
- (42) Miliadis-Argeitis, A., Rullan, M., Aoki, S. K., Buchmann, P., and Khammash, M. (2016) Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth. *Nat. Commun.* 7, 7.
- (43) Colloms, S. D., Merrick, C. A., Olorunniji, F. J., Stark, W. M., Smith, M. C., Osbourn, A., Keasling, J. D., and Rosser, S. J. (2014) Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res.* 42, e23.
- (44) Fogg, P. C., Colloms, S., Rosser, S., Stark, M., and Smith, M. C. (2014) New applications for phage integrases. *J. Mol. Biol.* 426, 2703–16.
- (45) Camilli, A., and Mekalanos, J. J. (1995) Use of recombinase gene fusions to identify *Vibrio cholerae* genes induced during infection. *Mol. Microbiol.* 18, 671–683.
- (46) Castillo, A. R., Woodruff, A. J., Connolly, L. E., Sause, W. E., and Ottemann, K. M. (2008) Recombination-based in vivo expression technology identifies *Helicobacter pylori* genes important for host colonization. *Infect. Immun.* 76, 5632–5644.
- (47) Lowe, A. M., Beattie, D. T., and Deresiewicz, R. L. (1998) Identification of novel staphylococcal virulence genes by in vivo expression technology. *Mol. Microbiol.* 27, 967–976.
- (48) Russell, J. P., Chang, D. W., Tretiakova, A., and Padidam, M. (2006) Phage Bxb1 integrase mediates highly efficient site-specific recombination in mammalian cells. *BioTechniques* 40, 460–464.
- (49) Xu, Z., and Brown, W. R. A. (2016) Comparison and optimization of ten phage encoded serine integrases for genome engineering in *Saccharomyces cerevisiae*. *BMC Biotechnol.* 16, 13.
- (50) Mutalik, V. K., et al. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* 10, 354–360.
- (51) Gibson, D. G., et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345.
- (52) Gibson, D. G., et al. (2010) Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods* 7, 901–903.
- (53) Lutz, R., and Bujard, H. (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* 25, 1203–1210.
- (54) Bertani, G. (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* 62, 293–300.
- (55) Dulbecco, R., and Vogt, M. (1954) Plaque formation and isolation of pure lines with poliomyelitis viruses. *J. Exp. Med.* 99, 167–182.
- (56) Folliard, T. (2017) A synthetic viral feedback loop results in robust expression, Oxford University Research Archive, <https://doi.org/10.5287/bodleian:wrNKzxaY2>.
- (57) Bowyer, J., Zhao, J., Subsoontorn, P., Wong, W., Rosser, S., and Bates, D. G. (2016) Mechanistic Modeling of a Rewritable Recombinase Addressable Data Module. *IEEE Transactions on Biomedical Circuits and Systems* 10, 1161–1170.
- (58) Bowyer, J., Hsiao, V., and Bates, D. G. (2016) Development and Experimental Validation of a Mechanistic Model of a Recombinase-Based Temporal Logic Gate. *Proceedings of the 12th IEEE Biomedical Circuits and Systems Conference, October 17-19, Shanghai, China.*