Author correction

During the course of continued efforts developing this integrase-mediated differentiation system for improving the evolutionary stability of burdensome engineered functions, we discovered that the *E. coli* strain with differentiation activated expression of T7 RNAP (Figure 4A) contains a second copy of the T7 RNAP differentiation cassette integrated into its genome. The T7 RNAP differentiation cassette was integrated into the genome using the clonetegration method with the phage 186 integrase, which has both primary and secondary landing sites on the *E. coli* genome.[24] Though we verified the successful integration at the primary site with colony PCR and sequencing, we did not check the secondary site and discover the second integration until later experiments with improved versions of the differentiation system highlighted discrepancies in circuit behavior. Specifically, unpublished experiments indicate that the behavior of the T7 RNAP differentiation circuit in the experiments depicted in Figure 4E-F should show a larger performance difference between the (+)-chloramphenicol and (-)-chloramphenicol conditions. Revised versions of these experiments conducted with strains verified by whole genome sequencing will be included in an updated publication.

Tunable integrase-mediated differentiation facilitates improved output of burdensome functions in *E. coli*

Rory L. Williams[1], Richard M. Murray[1,2]

Abstract

Application of synthetic biology is limited by the capacity of cells to faithfully execute burdensome engineered functions in the face of Darwinian evolution. Division of labor, both metabolic and reproductive, are underutilized in confronting this barrier. To address this, we developed a serine-integrase based differentiation circuit that allows control of the population composition through tuning of the differentiation rate and number of cell divisions differentiated cells can undergo. We applied this system to T7 RNAP-driven expression of a fluorescent protein, and demonstrate both increased duration of circuit function and total production for high burden expression. While T7 expression systems are typically used for high-level short-term expression, this system enables longer duration production, and could be readily applied to burdensome or toxic products not readily produced in bacteria.

Introduction

As synthetic biology aims to engineer cells with the capacity to regulate and execute increasingly complex and burdensome functions, strategies which address the evolutionary potential of biology will only become more essential. The same force of Darwinian evolution which has provided incredible biological diversity does not discriminate between natural and engineered life, and consequently engineered functions may be readily lost in a population. It has long been observed that cell fitness negatively correlates with heterologous gene expression level,[1] and increased burden results in a shorter evolutionary half-life of engineered functions.[2,3]

Efforts to improve evolutionary stability of engineered functions have taken a variety of forms, including the most straightforward goals of reducing mutation rate and minimizing burden. Strategies to reduce the rate of such mutations have focused both on sequence design and host-genome engineering. At the level of sequence design, minimizing repeated sequences and parts diminishes mutations due to homologous recombination (HR) and improves circuit half-life,[2] and researchers may evaluate sequences *in silico* for such HR and repeat-mediated mutations with the EFM calculator.[4] Alternatively, strains have been engineered to globally reduce mutation rates by disrupting the cell's capacity for HR with *recA* knockout, knocking out error-prone polymerases to reduce point mutations, and removing selfish transposon elements that otherwise may insert themselves and disrupt circuit function.[5–7] Though such strategies may delay the acquisition of destructive mutations, other approaches are necessary to impact the rate at which mutations are selected in the population. The simplest solution to delay the selection of these mutations is to reduce the expression level of circuit components and

1. Department of Bioengineering, California Institute of Technology, Pasadena, CA
2. Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA

therefor the fitness difference between functional and non-functional cells.[1,2,8] Alternatively, rather than constitutively reducing expression, gene expression level may be dynamically regulated by co-opting transcriptional changes which occur during cell stress to drive negative feedback.[9]

Additional strategies for improving evolutionary stability, rather than directly addressing rates of mutations and burden of functions, have sought to alter the consequence of these mutations. A conceptually straightforward approach is to have multiple redundant copies of the synthetic construct, whereby multiple independent mutations are required to destroy function. The chemically inducible chromosomal evolution (CIChE) system was used to evolve strains with ~40 tandem copies of a circuit before deleting *recA*, resulting in expression of a polyhydroxybutyrate biosynthetic pathway for >100 generations compared to ~10 when expressed from a plasmid.[10] Importantly, this strategy removes random plasmid portioning as a mechanism for accelerating mutation propagation.[11] Alternatively, selection of destructive mutations could be limited by utilizing components whose mutation would inactivate not only the expression of the synthetic construct, but also an essential gene or selectable marker. Producing GFP with KanR on a single bicistronic transcript or as a fusion protein marginally improved evolutionary half-life when selecting with kanamycin, and using a bi-directional promoter to drive their expression separately increased half-life 4-10 fold.[2,12]

Strategies discussed so far have been limited to cell-level functions in uniform populations, and tactics which incorporate specialization and division of labor at a population level have not been addressed. With inspiration from microbial communities exhibiting metabolic division of labor and syntrophic interactions, there have been numerous successful implementations of metabolic division of labor for production of biomolecules of interest.[13–15] This design motif has numerous advantages, including reducing the number of genes and associated metabolic load in each specialized cell type, allowing independent optimization of separate pathways, and spatially separating potentially incompatible functions. While these benefits may be realized by combining in co-culture independently engineered strains or species, additional attractive properties become apparent with dynamically regulated division of labor in a population of genetically related or identical organisms. Such metabolic and reproductive division of labor is a recurring motif in microbiology,[16,17] but is underutilized in synthetic biology, particularly for addressing evolutionary constraints.

Examining specific instances of division of labor in bacteria gives insight into how we might use this motif in synthetic biology. In the cyanobacteria *Anabaena*, nitrogen deprivation induces a division of labor in which individual cells in a large filament terminally differentiate into heterocyst cells which are specialized for nitrogen fixation and incapable of reproduction.[18] This reproductive and metabolic division of labor allows the collective to realize an inclusive fitness benefit from a costly metabolic process; a process encoded by all cells, but expressed only in a fraction. If we imagine instead all cells expressing this function—nitrogen fixation—cells which mutate this function would certainly gain a fitness benefit, and would proliferate more quickly (assuming sufficient nitrogen). By instead having cells which have the genetic potential for nitrogen fixation,

but do not express it, there is no selective pressure for mutations which would inactivate expression of genes necessary for this process. Though this function is indeed essential for *Anabaena* survival under nitrogen deprivation conditions, we could imagine using this same strategy in synthetic biology for the expression of burdensome functions, essential and non-essential alike.

To adopt this reproductive and metabolic division of labor into a synthetic context, we propose a circuit architecture much like that seen in *Anabaena*. The simplest form of this architecture consists of two cell types, with the first being specialized for the faithful replication of an encoded function in the absence of circuit burden, and the second— generated upon differentiation of the former— for the execution of the encoded function. Though largely orthogonal and complementary to previous approaches for improving the evolutionary stability of engineered functions, this strategy may be particularly suited for certain types of applications. Functions for which a subset of cells expressing a function is sufficient are ideal, as are functions which could be divided between cells of distinct phenotypes. Importantly, functions which are highly burdensome, toxic, or incompatible with cell proliferation enter the realm of possibility.

Results

In order to increase the duration of circuit life-time, either the rate of mutations which inactivate circuit function must be drastically decreased, or alternatively the opportunity for these mutations to be selected in a population of cells must be limited. To accomplish this, we reasoned that having a population of cells that encode the circuit function, but do not express it, would allow the genetic circuit to be replicated in the absence of selective pressure for inactivating mutations. By inducing these progenitor cells to differentiate at some rate into cells expressing the function, producer cells would be continuously replenished (Fig. 1B). However, these producer cells are still susceptible to mutations which inactivate circuit function, and the opportunity for these mutants to be selected would need to be eliminated in order to prevent circuit failure. To accomplish this, we considered an architecture that would limit the number of divisions a cell could undergo following differentiation (Fig. 1C). In this way, mutations which inactivate circuit function would have negligible opportunity to be selected.

To implement this differentiation circuit, we turn to bacteriophage serine integrases, a class of proteins capable of unidirectional DNA recombination between specific sequences of DNA.[19] With strategic placement of integrase attachment sites on the genome, a single integrase-mediated recombination event can simultaneously activate and inactivate the expression of desired genes (Fig. 3A). In order to tune the rate of differentiation, we rely on the inherent stochasticity of this process at low intracellular concentrations of integrase proteins, with higher expression increasing the probability that any given cell in the population will undergo the recombination event.

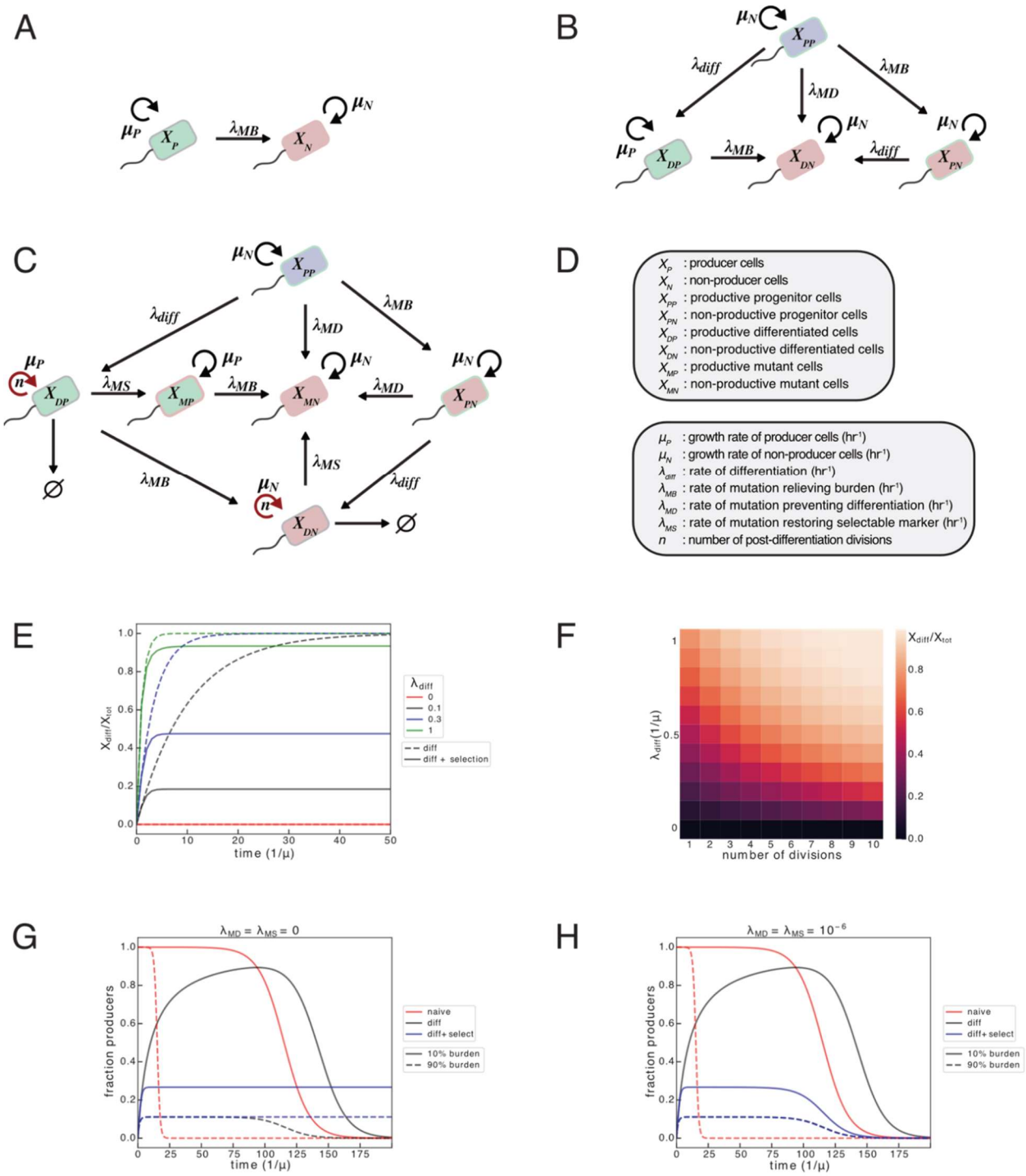

Figure 1: Architectures for implementation of differentiation circuits
(A-D) Schematics for a naive expression (A), differentiation-activated expression (differentiation) (B), and differentiation-activated expression in which the number of cell divisions following differentiation is limited (differentiation with selection) (C). (E) Deterministic ODE modeling in exponential growth conditions of differentiation circuits with and without selection ($\mu_N = \mu_P = 1$; $\lambda_{MB} = \lambda_{MD} = \lambda_{MS} = 0$). (F) Modeling as in (E) for differentiation with selection varying differentiation rate and number of divisions. (G-H) Comparison of naïve circuit with differentiation with and without selection (G: $\lambda_{MB} = 10^{-6}$, $\lambda_{MD} = \lambda_{MS} = 0$; H: $\lambda_{MB} = \lambda_{MD} = \lambda_{MS} = 10^{-6}$).

In order to allow differentiation to control diverse expression programs, the genes regulated by the recombination event may encode proteins which control the expression of numerous genes, as in the case of transcription factors, sigma factors, or orthogonal RNA polymerases. Further, we may limit the proliferation of differentiated cells by using this recombination to inactivate the expression of an essential, or conditionally essential gene. To allow tuning of the duration of differentiated cell proliferation, we take advantage of the reliance of R6K plasmid replication on the π protein encoded by the *pir* gene.[20] By varying the expression of the *pir* gene using an inducible promoter, and using the recombination event to inactivate its expression, the π protein abundance, R6K plasmid copy number, and therefor the number of cell divisions differentiated cells undergo before losing the R6K plasmid and its associated antibiotic resistance gene may be tuned (Fig. 3A).

Deterministic modeling of integrase-mediated differentiation

To gain intuition regarding the behavior of these proposed differentiation architectures, and specifically if and when differentiation-based circuits would be advantageous for improving the duration of circuit lifetime and/or total output achieved by an engineered function, we modeled the behavior deterministically using systems of ordinary differential equations. For comparison, we model a naïve expression circuit (Fig. 1A).

In the absence of any burden difference between the progenitor and differentiated cells, the differentiation architecture with unrestricted cell division in the differentiated cells results in all cells in population being differentiated if the differentiation rate is non-zero (Fig. 1E). However, when the number of cell divisions is limited, the population achieves a steady-state fraction of differentiated cells that can be tuned by both the differentiation rate and the number of cell divisions allowed by differentiated cells (Fig. 1E-F).

Considering the case when differentiated cells are performing some burdensome function and have a slower growth rate than the progenitor cells, we observe populations performing differentiation with and without limited cell divisions approaching or achieving a steady state fraction of differentiated producer cells (Fig. 1G). However, as differentiated cells are able to incur mutations inactivating the production and relieving the associated burden, populations are eventually dominated by non-productive differentiated cells at a rate that increases with burden. Conversely, when the number of divisions of differentiated cells is limited, non-productive differentiated cells do not have the opportunity to be selected, and a steady-state fraction of producers is achieved that will be disrupted only by the incredibly slow accumulation of mutations in the progenitor population (Fig. 1G).

While differentiation with selection appears to permit indefinite circuit function, we have not yet considered additional classes of mutation that become possible when implementing these differentiation circuit architectures. In the case of differentiation alone, we posited a class of mutations in progenitor cells which would destroy the ability of the cell to undergo differentiation (Fig. 1B-C). Additionally, when limiting the number of divisions a differentiated cell can undergo, a second new class of mutations becomes

apparent which would restore the ability of a differentiated cell to proliferate indefinitely (Fig. 1C). When we account for these two additional classes of mutations, we observe circuit failure even in the case where differentiated cells have limited capacity for cell division (Fig. 1H). Though all circuit architectures—naïve, differentiation, and differentiation with restricted cell division—are imperfect and ultimately fail due to mutation and natural selection, we reasoned that the best architecture would depend on a variety of factors, including the burden imposed by the engineered function, relative mutation rates (burden, differentiation, and selection mutations), as well as the needs of the specific application, such as maximizing total production or duration of circuit function.

To understand when each of these architectures would be best suited, we modeled each across varying production burdens, and mutation rates specific to burden, differentiation, and selection mutations. As the behavior of the two architectures involving differentiation are also impacted by the differentiation rate, we optimized this parameter in each case to maximize either the total production (Fig. 2A) or the duration of population function (Fig. 2B). In order for modeling results to be qualitatively comparable to our experiments, we modelled logistic growth with repeated 50X dilutions when the population reached 95% of the carrying capacity. Simulations were terminated when the production from an individual growth cycle was below a low threshold (Fig. 2A), or when fewer than 10% of cells were producers (Fig. 2B). Due to the mechanism of differentiation implemented experimentally—integrase-mediated recombination—differentiation is not coupled to cell division, and is modeled with a first-order rate constant coupled to time. As well, though the differentiation rate may indeed vary with the growth state of cells (i.e. exponential vs. stationary phase), we neglected this and assumed a constant differentiation rate regardless of growth phase.

Modeling the naïve implementation of production reveals matching intuition that both total production and circuit lifetime decrease with increased burden and burden mutation rate, with burden being the dominant factor. The relative benefit of differentiation on total production and circuit lifetime depends both on the burden, as well as the relative rates of burden and differentiation mutations. For both total production and circuit lifetime, this benefit increases with increased burden, but decreases as the differentiation mutation rate increases relative to the burden mutation rate.

In the case of differentiation with selection, we see largely the same trends as with differentiation alone, however with a few key differences. While there is an increased benefit for total production relative to the other two architectures as burden increases, this strategy is counterproductive at low burdens, and decreases production particularly with higher differentiation mutation rates. Further, selection allows the population to be less susceptible to the burden mutation, demonstrated by increased production and duration relative to both differentiation and naïve implementations as the burden mutation rate increases. Finally, comparing the second and third rows of Figure 2A and B reveals that the impact of the rate of the selection mutation is only revealed with low

rates of the differentiation mutation, and becomes more apparent with increased burden mutation rates.

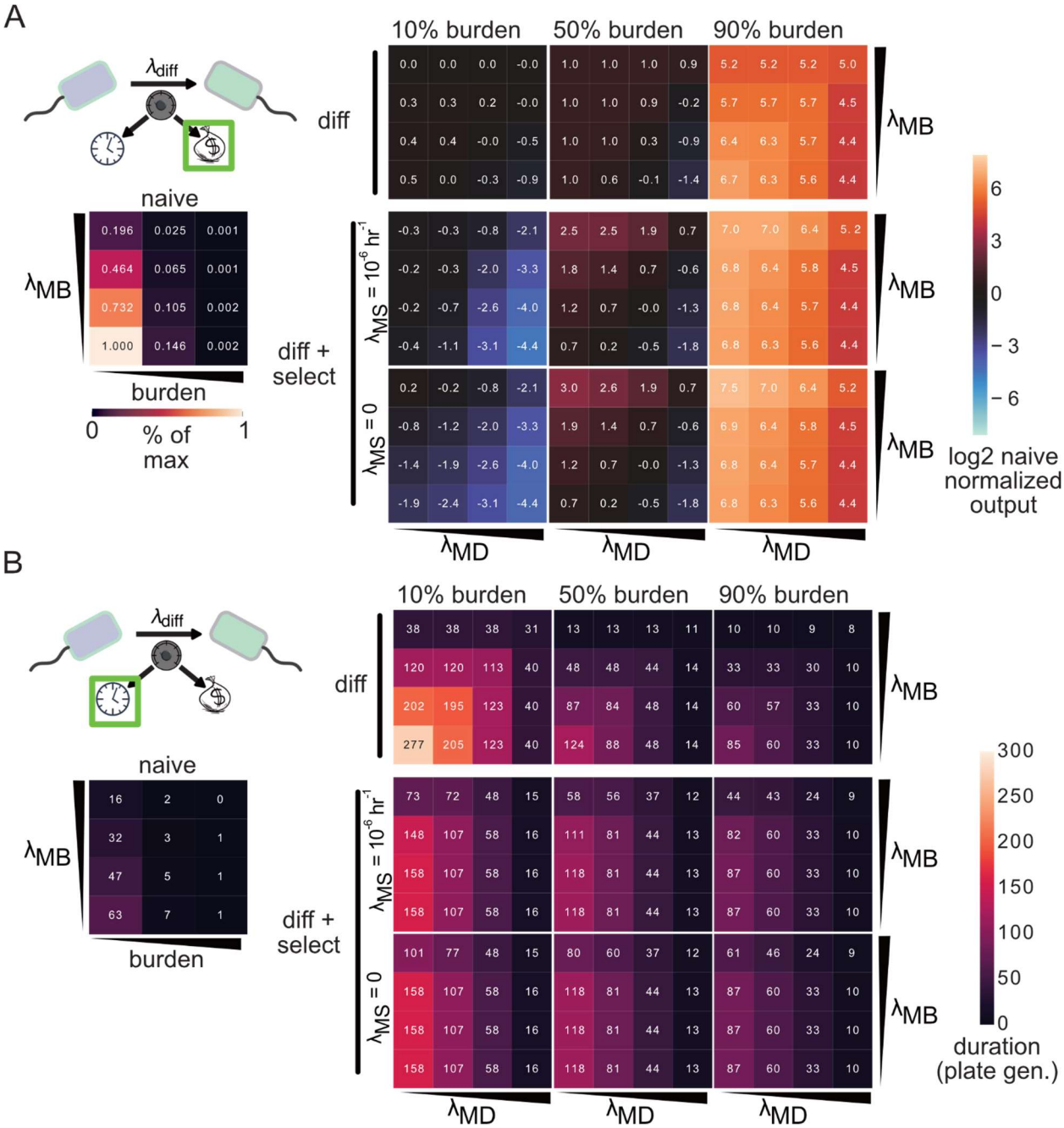

Figure 2: Differentiation architectures improve duration and output for high burden circuits
Deterministic modeling of naïve and differentiation architectures with varying burden levels (10%, 50%, and 90%) and mutation rates (λ_{MB}, λ_{MD}: 10^{-12}, 10^{-9}, 10^{-6}, and 10^{-3} h^{-1}). Simulations are of repeated 50X dilutions with logistic growth, with dilutions occurring when the population reaches 95% of the carrying capacity (K). For differentiation and differentiation with selection, the differentiation rate was optimized to maximize total production (A) and duration (B). For all simulations, $K = 10^9$ cells, $n = 4$, $\mu_P = 1.5$ h^{-1}. (A) Production rate is modeled as proportional to growth rate and varies over time, and production rate is equal across burden levels. Heatmaps present (A) total production, and (B) number of consecutive growths in which the ending fraction of producer cells is >10% of the population. (A) Heatmap (left) shows

total production normalized maximum case (λ_{MB} = 10^{-12} h^{-1} , burden=10%). Heatmaps (right) show the log2 of normalized total production (normalized to naïve production with equivalent burden and λ_{MB}).

Integrase-mediated differentiation allows tuning of population distribution

To experimentally investigate these qualitative predictions from our modeling, we first implement and characterize a differentiation architecture which allows tuning of the rate or probability of differentiation, selection against differentiated cells, and tuning of the duration of differentiated cell proliferation (Fig. 3A). In this circuit, expression of the degradation-tagged integrase Bxb1 is induced in the presence of salicylate, and catalyzes the single recombination event which terminates expression of the *pir* gene, and activates the expression of mScarletI. The rate of differentiation in the population can be tuned by varying Bxb1 expression, with the expression of mScarletI correlating with the number of differentiated cells (Fig. 3B). The expression of the *pir* gene is under the control of LasR, and the expression of *pir*, and consequently the copy number of the R6K plasmid, can be tuned by varying the concentration of Las AHL. As the R6K plasmid encodes a constitutively expressed sfGFP, the relative copy number of the R6K plasmid is inferred through sfGFP fluorescence (Fig. 3C).

We characterized the long-term behavior of this differentiation circuit with varying differentiation rates and *pir* expression, using flow cytometry to determine the population fraction of progenitor and differentiated cells. Progenitor cells are identified as the sfGFP-positive/mScarletI-negative population, while differentiated cells are identified as sfGFP-negative/mScarletI-positive population, with the activation of mScarletI expression occurring before loss of sfGFP (Fig. 3E). Across all concentrations of Las AHL, the population proceeds towards 100% differentiated mScarletI-positive cells in the absence of chloramphenicol when integrase is induced, with this occurring more quickly at higher induction of the integrase (Fig. 3F). However, when selecting with chloramphenicol, the population appears to approach a steady state population distribution containing both progenitor and differentiated cells, with the relative abundance depending on the induction level of both integrase and *pir* (Fig 3F). With 5μM salicylate, differentiated cells comprise 59.2% +/- 0.03 (mean+/- SD of two replicates), 69.9% +/- 0.004, and 70.3% +/- 0.05% of the population after four plate generations with 0.3μM, 1μM, and 3μM Las AHL respectively. With 7.5μM salicylate differentiated cells comprise 88.9% +/- 0.02, 94.6% +/- 0.02%, and 96.4% +/- 0.01% of the population after four plate generations with 0.3μM, 1μM, and 3μM Las AHL respectively. These distributions qualitatively align with our deterministic modeling, namely that a higher differentiation rate and a larger number of divisions allowed by differentiated cells both increase the steady state fraction of differentiated cells.

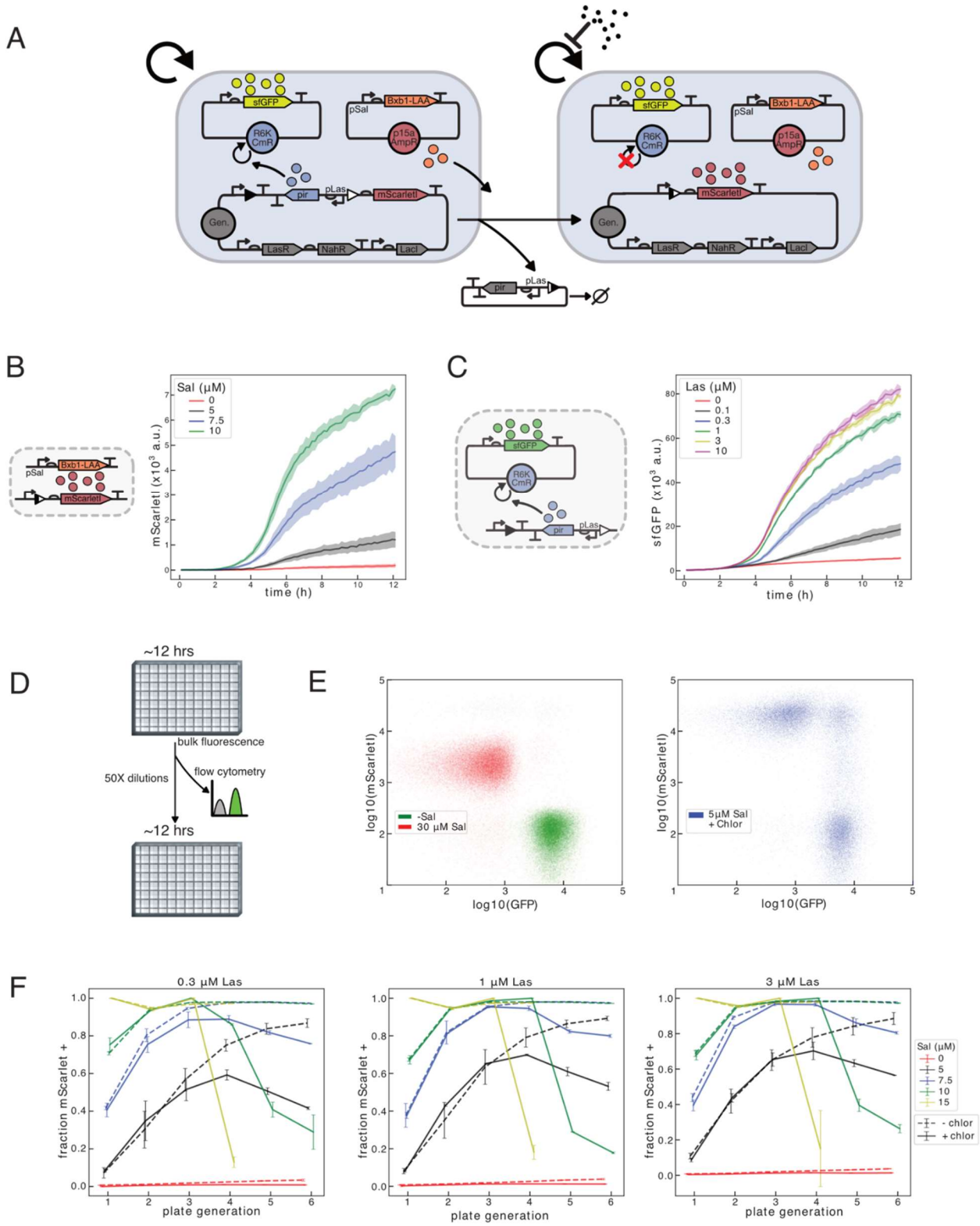

Figure 3: **Implementation of a tunable integrase mediated differentiation circuit**
(A) Schematic of a tunable integrase differentiation circuit. Las AHL induces expression of pir protein, which is needed for replication of the R6K plasmid. Salicylate induces the expression of degradation-tagged Bxb1 integrase, which excises the pir expression cassette from the genome, and activates the expression of mScarletI. Production of pir ceases after differentiation, and pir protein/R6K plasmid are

diluted by cell division, resulting in susceptibility to chloramphenicol (black dots). (B-C) Batch culture experiments of JS006 with circuit depicted in (A) grown in M9CA media. (B) mScarletI fluorescence with varying induction levels of salicylate in M9CA + carbenicillin. (C) sfGFP fluorescence with varying induction levels of Las AHL in M9CA + carbenicillin/chloramphenicol. (B-C) Means +/- standard deviation of three replicates. (D-F) Cells are grown in 300μL M9CA media with varying inducer concentrations, and diluted 50X every 12 hours into the same media conditions. Samples are taken for flow cytometry after each growth. (E) Flow cytometry results after the third 12-hour growth for cells grown in 0.3 μM Las AHL without chloramphenicol (left), and cells grown in 0.3 μM Las AHL, 5 μM salicylate, and 34 μg/mL chloramphenicol (right). (F) Results of flow cytometry analysis of cells grown for six consecutive 12-hour growths in varying inducer concentrations. Average fraction mScarletI positive (differentiated cells) for two replicate wells +/- standard deviation.

However, at higher inductions of salicylate, we clearly see circuit failure when selecting with chloramphenicol. This is revealed at all Las AHL concentrations in flow-cytometry of the fifth plate generation for 10μM salicylate, and in the fourth plate generation for 15μM salicylate (Fig 3F). Here, instead of achieving a steady state distribution comprised largely of differentiated cells, a population which is sfGFP-positive/mScarletI-negative comes to dominate the population. Though these resemble progenitor cells in gene-expression, they are no longer able to undergo differentiation, and have likely incurred a mutation analogous to the differentiation mutation we proposed in our model.

Differentiation-activated T7 expression improves burdensome function performance

We next apply our synthetic differentiation system to T7 RNAP-driven production, experimentally investigating if and when it is advantageous. To eliminate leaky expression of T7 RNAP in the absence of the differentiation event, we relied on previous research splitting T7 RNAP into functional domains to rationally choose a split site.[21] Using this strategy, production of functional full-length T7 RNAP (containing the recombined attL site and additional bases to retain the correct reading frame) occurs only after recombination joins the two fragments of the coding sequence. In the absence of integrase induction, no production of sfYFP is observed from the T7 promoter, while the production of T7 RNAP/sfYFP is tuned by induction with IPTG when integrase is induced (Fig. 4B). This is also observed in the growth phenotype, with significant growth defect occurring only with addition of both salicylate and 30 or 100μM IPTG (Fig. 4B).

Using this system, we can directly compare the total output and duration of production for our two differentiation circuit architectures with naïve inducible production at varying burden levels. Our differentiation architectures with and without restricted cell division differ only in the presence of chloramphenicol, while the naïve architecture is identical to a differentiated cell lacking the R6K plasmid, and is grown in the absence of chloramphenicol (Fig 4A-B). As with our mScarletI/sfGFP differentiation circuit, we can tune the differentiation rate with salicylate, and the population distribution of producers and non-producers over consecutive dilutions can be measured by flow cytometry. Further, we can compare the total production using end-point bulk-fluorescence measurements monitored during growth in a plate reader. For reference, JS006 cells with the naïve circuit have experimentally determined growth rates of 0.88 +/- 0.02 h^{-1} (mean/SD of two independent colonies with six total replicates), 0.87 +/- 0.01 h^{-1}, 0.77 +/- 0.03 h^{-1}, and 0.38 +/- 0.04 h^{-1} when grown in M9CA + carbenicillin with 0, 10, 30,

and 100μM IPTG respectively (data not shown). This equates to burdens of ~1.6%, 12.9%, and 57.6% with 10, 30, and 100μM IPTG relative to the uninduced case.

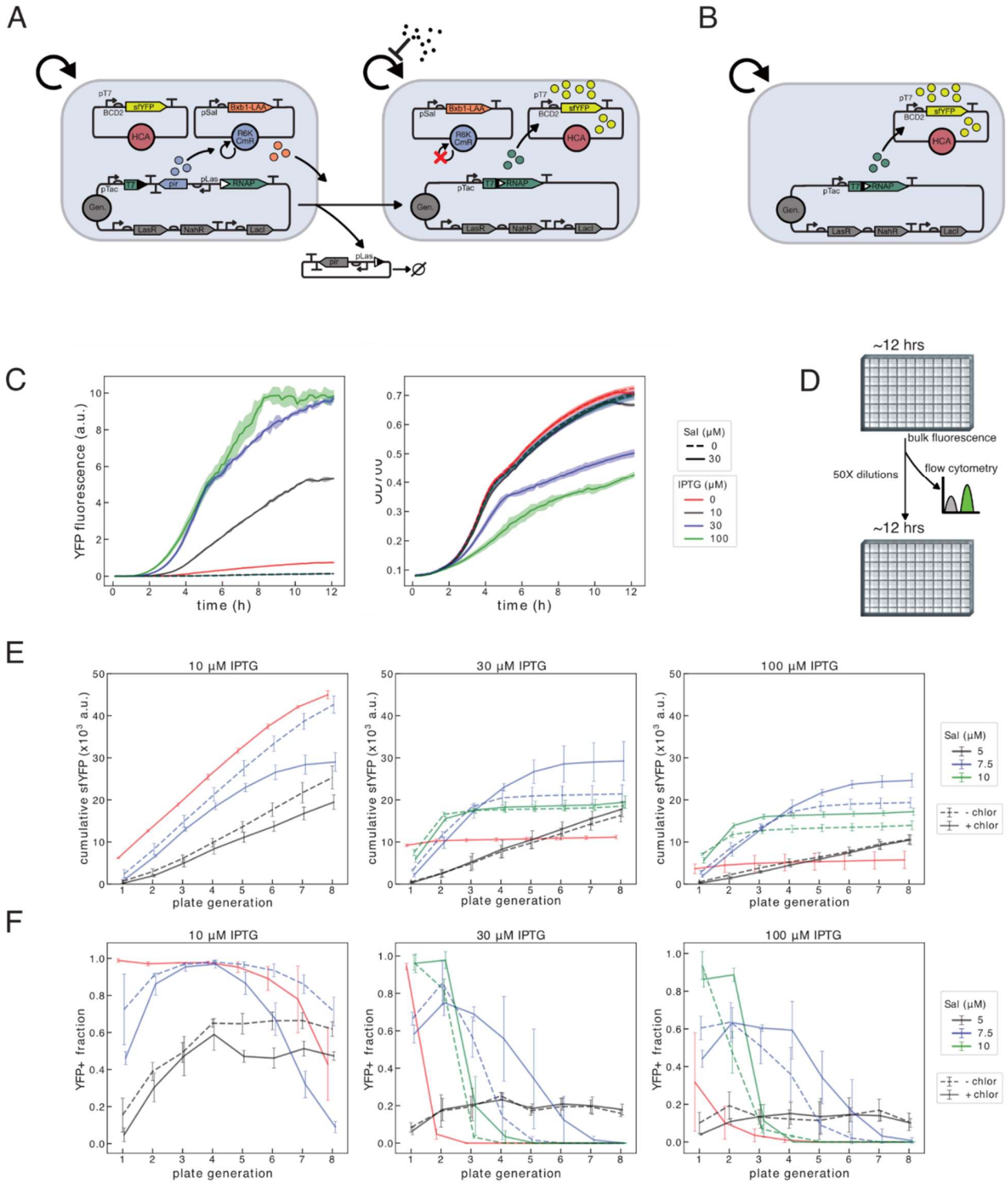

Figure 4: Differentiation improves duration and output from burdensome T7 driven expression circuits
(A) Schematic of differentiation-activated T7 RNAP driven expression. Circuit functions as in Figure 3, however the Bxb1 integrase is expressed from the R6K plasmid, and the recombination event activates expression of T7 RNAP by bringing together two fragments of the coding sequence. IPTG induces expression of full length T7 RNAP in differentiated cells, allowing expression of sfYFP from the high-copy Cole1 plasmid. (B) Schematic of naïve inducible T7 RNAP driven expression. Circuit is identical to a differentiated cell, but lacking the R6K plasmid. (C) Batch culture experiments of JS006 with circuit depicted in (A). Cells are grown in M9CA media + carbenicillin/chloramphenicol/1μM Las AHL, with or without 30μM salicylate, with varying concentrations of IPTG. Curves are means +/- standard deviation of three replicate wells. (D-F) JS006 with circuit above grown in M9CA + carbenicillin/3μM Las AHL, +/- chloramphenicol, in varying concentrations of salicylate and IPTG. Cells diluted 50X into same media conditions into 300μL total volume every ~12 h for 8 total growths. (E) Cumulative total production plotted is the sum of endpoint fluorescence values from 12 h plate reader growth. (F) Samples taken immediately after growth were analyzed by flow cytometry, and fraction sfYFP positive cells is plotted. (F-G) Means of four total replicates from two independent experiments with standard deviation error bars.

At relatively low-burden (10μM IPTG), total production increases linearly in the naïve case for the first ~6 plate generations before cells no longer producing sfYFP emerge and total production flattens to 44980 +/- 958 (mean +/- SD of four replicates). With higher burden (30 or 100μM IPTG), nearly all production occurs in the first growth, and non-producers dominate the population by the end of the second growth (total production 11167 +/- 414; 5756 +/- 2061). In the case of differentiation without selection, total production approaches that achieved by naïve production at 10μM IPTG when integrase induction is sufficiently high (>5μM salicylate). Here the delay in achieving ~100% differentiated producer cells is counteracted by a similar delay in accumulation of non-producers (Fig 4E-F).

With higher burden production, the benefit of differentiation for both circuit duration and total production becomes apparent. With a low differentiation rate (5μM salicylate) and 30μM IPTG, 15.9% +/- 0.02% are still producing sfYFP after 8 plate generations, greatly extending the duration of expression, with total production over the experiment being 16281 +/- 1508. A higher differentiation rate (7.5μM salicylate), results in a greater total production (21414 +/- 2138; ~1.9X naive) but decreased duration of expression. Increasing the differentiation rate further (10μM salicylate) pushes this rate further from an apparent optimum, decreasing total production to 18552 +/- 381 (~1.7X naïve). This benefit for total production is enhanced at the highest induction level of IPTG, with ~3.4X and ~2.4X the naïve production with 7.5 and 10μM salicylate respectively.

The effect of selecting against differentiated cells, as in our modeling, is dependent on both the differentiation rate and the expression burden. At low burden (10μM IPTG), this selection decreases total output in comparison to both naïve and differentiation, with total production of 29010 +/- 2243 and 17822 +/- 832 (0.64X and 0.4X naïve) with 7.5 and 10μM salicylate respectively. However, with higher burden production, differentiation with selection facilitates a total production of 29224 +/- 4629 with 7.5μM salicylate, 2.62X naïve production and 1.28X of differentiation without selection. As with differentiation alone, this benefit is exaggerated at the highest induction of IPTG, with production 4.28X naïve production, and 1.26X that of differentiation alone.

In addition to evaluating the performance of these two differentiation architectures with respect to naïve production, it is also useful to characterize the mutations which destroy circuit function for each case. For the naïve and differentiation without chloramphenicol selection cases, colonies isolated after eight plate generations contained plasmid capable of expressing sfYFP when transformed into cells expressing T7 RNAP, but these cells failed to express sfYFP when transformed with the same expression construct encoded on a pSC101-chlor plasmid (data not shown). It is apparent in both of these cases—although specific mutations were not identified—that a mutation on the genome has disrupted the expression of T7 RNAP, equivalent to the burden mutation described in our modelling. For the case of differentiation with restricted cell division, it is apparent from the continued chloramphenicol resistance, loss of sfYFP expression, and insensitivity to induction with salicylate, that a mutation equivalent to the differentiation mutation (or alternatively both a burden mutation and a selection mutation) we posited in our modeling has occurred. Though mutations disrupting the expression of the integrase—either directly on integrase expression cassette encoded on the R6K plasmid or at the level of the transcription factor NahR—would accomplish this, we first examined the differentiation cassette itself. Apart from a ~1.3kb deletion which destroyed the attP site, we also observed two independent mutations in which the integrase-recombination event appears to have inverted rather than excised the intervening sequence, destroying both integrase attachment sites (supplementary sequences). Though we observed apparent differentiation mutations in the differentiation cassette in these three cases, we did not so in several others, and additional sequencing is required to determine all possible sources of mutation.

Discussion

With inspiration from bacterial reproductive and metabolic division of labor observed in nature, we developed a synthetic differentiation system that allows fractional tuning of progenitor and differentiated cells through inducible integrase-mediated differentiation and conditionally restricted cell division in differentiated cells. We applied this system to T7 RNAP-driven expression of fluorescent protein, and demonstrated that differentiation can improve total production output and duration of production for high burden production circuits, qualitatively matching deterministic modeling results. Further, we demonstrated that limiting the capacity of differentiated cells to undergo cell division was counterproductive for both total production and duration of production at low burden, but can increase both metrics at high burden relative to naïve and differentiation architectures if the differentiation rate is appropriately tuned.

In modeling our differentiation architectures, we saw that the benefit of differentiation with and without restricted cell division relative to naïve expression depended heavily on the expression burden. Specifically, the performance (both total production and duration) of differentiation with and without restricted cell division relative to naïve production improves with increased burden. Though this trend agrees qualitatively with experimental results, it does not match exactly. While in our modeling differentiation and particularly differentiation with selection are ineffective or harmful with 10% burden, experimentally we see a strong benefit for both total production and duration of

production at 30μM IPTG, corresponding to a ~12.9% growth penalty. This incongruency may be a result of cells being diluted from stationary phase (in all but the first plate generation) directly into media containing inducer. It has previously been shown that the cost/burden of unneeded protein production is elevated during the first few cell divisions following dilution from stationary phase.[22] Given that cells will undergo 5-6 cell divisions after a 50X dilution, an increase in effective burden for the first several cell divisions extends through much of the growth and may largely explain this difference between our modeling and experiments. If we instead perform experiments using higher fold dilutions or in continuous culture, this difference may be diminished.

Apart from effects due to burden, our modeling also revealed that the performance of our two differentiation architectures depended on the relative mutation rates. In particular, differentiation with restricted cell division is less sensitive to an increased rate of burden mutation, and the benefit of restricted cell division relative to differentiation alone is apparent with higher burden mutation rates and lower differentiation rates. Though direct quantification of mutation rates is complicated by potentially variable cost/burden of production across growth phase following cell dilution, we can infer that the differentiation mutation rate is likely not orders of magnitude larger than the burden mutation rate, and may indeed by on the same order or lower. We draw this conclusion because our modeling reveals that differentiation with selection tends to perform better than differentiation alone only when the burden mutation rate is of equal or greater order of magnitude than that of the differentiation mutation. While this may not be surprising, we recognize that a differentiation mutation may be facilitated by errors during integrase-mediated recombination, the rate of which has not been quantified to our knowledge.

While here we have demonstrated that limiting cell divisions in differentiated cells through selection with chloramphenicol can provide a modest improvement to total production and duration of production with respect to differentiation alone, it may be possible to increase this benefit with minor adjustments to our circuit architecture. First, because we are using chloramphenicol—which acts on the ribosome to inhibit protein synthesis—to select against differentiated cells, we may be unnecessarily inhibiting production. By using an alternative selectable marker such as mFabI and selecting with triclosan which inhibits lipid synthesis, we may be able to remove this inhibition of production in differentiated cells while maintaining inhibition of cell proliferation, thereby increasing total production.[23] Additionally, though both differentiation architectures may benefit from reduction in the differentiation mutation rate, differentiation with restricted cell division is more sensitive to this mutation. Though this rate may be reduced by optimizing the circuit sequence following more careful analysis of mutations which occur in this case, we may alternatively reduce the effective mutation rate by requiring two or more mutations to occur to destroy differentiation potential. This may be accomplished by integrating two modified copies of the differentiation cassette, or by having two redundant differentiation mechanisms utilizing orthogonal integrases.

This first demonstration of utilizing synthetic differentiation to improve performance of burdensome functions, though reminiscent of examples of division of labor found in

bacteria, differs in important ways which may limit the benefit gained from its implementation. Most obvious is that the function we are expressing in differentiated cells—T7 RNAP and fluorescent protein—is entirely non-essential. Replacing or supplementing this unneeded metabolic load with a function that is beneficial or essential for both progenitor and differentiated cells would likely improve the evolutionary stability of such differentiation architectures, and future more successful implementations of synthetic differentiation and division of labor may be closer to natural examples in this respect. However, despite this shortcoming of our existing circuits, both differentiation architectures provide advantages not achieved by existing strategies to improve the evolutionary stability of engineered functions.

While existing strategies can aid greatly in reducing the rate of mutations in engineered circuits, the only general strategies to reduce the rate at which mutations are selected are to reduce the burden of expression,[1–3] or alternatively to integrate numerous copies of a genetic construct on the genome.[10] Reducing the burden of expression may not be a viable option for the production of toxic proteins or metabolites, or for certain industrial applications in which maximizing production is essential for economic viability. As well, though genomically integrating many copies of a construct may be effective at limiting the generation and selection of non-productive mutant cells, cells expressing particularly burdensome or toxic functions will have impaired or destroyed ability to proliferate, potentially rendering long duration or continuous production nonviable. Further, applications in which engineered cells are not growing in mono-culture in a laboratory environment, but instead must compete with cells in a complex microbial community, will require consideration of this competition rather than simply the competition between functional and non-functional engineered cells.

The differentiation architectures we describe here represent a qualitatively new strategy for addressing the constraints imposed on synthetic biology by evolutionary forces, and may be applied generally to diverse circuits and functions. With the implementation of division of labor through differentiation, we can remove selective pressure for mutations relieving burden in a subset of the population, mitigate fitness defects in these progenitor cells, and sacrifice a tunable fraction to production with little regard for the burden or toxicity of the function. Circuit architectures utilizing differentiation can allow longer duration production—potentially in continuous culture—without the uniform growth defect common to all other strategies, and show promise in improving the evolutionary stability of engineered functions.

Materials and Methods

Strains and constructs

JS006 strain *E. coli* were used for all experiments, and constructs were assembled using 3G assembly. Constructs were genomically integrated with clonetegration using pOSIP-KO and pOSIP-CH.[24] pOSIP plasmids were double digested with BamHI and

SpeI, and PCR purified. For assembly, 3G was used,[25] however P1 and PX adapters were used in place of UNS1 and UNSX to allow compatibility with the pOSIP backbone. Modified adapters were used to generate the bicistronic transcriptional unit for LasR/NahR (UNS3_D/UNS3_B), and inverted *pir* transcriptional unit (UNS3_E*/UNS3_A*). Generation of modified MoClo[26] compatible T7 RNAP parts, as well as degradation tagged Bxb1 integrase and UNS1-UNSX R6K-chlor plasmid backbone, is summarized in the table below. Sequences for Bxb1 integrase attachment sites attB and attP were obtained from Ghosh, et. al.[27] Sequences of all parts, primers, and final constructs can be found in the supplementary information.

Part	Template	FOR	REV	note
U1-R6K-chlor-UX	pOSIP-CH	RW.U1.R6kbb.F	RW.UX.R6kbb.R	
T7(attL)RNAP	C95m	RW.T7RNAP.split2.FOR	RW.T7RNAP.split2.REV	1-piece Gibson
T7LEFT_attB	T7(attL)RNAP	pseqF	RW.T7split.attB.bsaID.R	
attP_T7RIGHT	T7(attL)RNAP	RW.T7split.attP.bsaIB.F	pseqR	
Bxb1LAA	Bxb1	pseqF	RW.bxb1.LAA.rev1/ RW.bxb1.LAA.rev2	

Deterministic modeling of differentiation circuits

Circuits depicted in Figure 1A-C were modeled deterministically using systems of coupled ordinary differential equations. In all simulations, rates for differentiation and mutations are first-order with respect to cell number, and do not depend on growth rate. Exponential growth without carrying capacity was assumed for simulations in Fig. 1 E-H. For Fig. 2, circuits were simulated with logistic growth with a carrying capacity of 10^9 cells, with cells being diluted 50X when cells reached 95% of the carrying capacity. Production was modeled as being proportional to the ratio of specific growth rate (actual growth rate after accounting for effect due to carrying capacity) to maximum growth rate for the specific cell type, and production rate was 1 for all simulations regardless of burden. For differentiation with selection, a cell with *n* remaining cell divisions divides to generate two cells with *n-1* remaining cell divisions. Cells with one remaining cell division therefor divide into two cells which do not divide. This was equated to cell death, and these cells have zero production and do not count towards the carrying capacity. For comparing total production (Fig. 2A) and duration of circuit function with >10% of cells being producers (Fig. 2B), the differentiation rate was selected using optimization (three independent starting values: 0.001,0.1,1) to maximize total production or duration. Jupyter notebooks describing and running all simulations are available on the Github repository listed in supplementary information.

Differentiation experiments

For experiments with differentiation cells, cells were grown from glycerol stock in 3mL culture of M9CA glucose (Teknova M8010) with 34μg/mL chloramphenicol, 100μg/mL carbenicillin, and 1μM Las-AHL. Overnight cultures were diluted 1:100 into the same media and grown ~2-3 hours to OD 0.2-0.4. To avoid cross-over of antibiotics and inducers, cells were pelleted (3500g for 10min) before resuspending in M9CA with

appropriate antibiotics (carb for differentiation, carb + chlor for differentiation with selection) to OD ~0.1. Control cells with naïve inducible expression of T7 RNAP from the genome were treated as above but grown in M9CA glucose + carbenicillin. Cells at OD ~0.1 were diluted 1:10 into a total volume of 300μL containing appropriate antibiotics (carbenicillin +/- chloramphenicol) and various inducer concentrations (IPTG, salicylate, Las AHL). Cells were grown in 96-well square-well plate (Brooks MGB096-1-2-LG-L) at 37°C with maximum-speed linear shaking in a BioTek Synergy H1m. OD700, sfYFP fluorescence (503/540 nm excitation/emission; gain 61 and 100), sfGFP fluorescence (485/515 nm excitation/emission; gain 61 and 100), and mScarletI fluorescence (565/595 nm excitation/emission; gain 100) were measured at 10 minute intervals as appropriate. For long-term experiments, cells were diluted 1:50 after ~12h growth into the same media conditions into a replicate plate. All data and Jupyter notebooks are available on the Github repository.

Flow-cytometry

Immediately after the conclusion of a 12h growth, cells were diluted 1:50 into 100μL 1X PBS for analysis with flow cytometry. Samples were run on a Miltenyi Biotech MACSQuant VYB Flow Cytometer equipped with Violet 405nm, Blue 488nm, and Yellow 561nm lasers. sfYFP was measured with the 488nm laser with 525/50nm filter, sfGFP with the 405nm laser with 525/50nm filter, and mScarletI with the 561nm laser with 661/20nm filter. 50,000 ungated events were recorded for each sample, and results were analyzed with custom python code available in the Github repository listed supplementary information. Briefly, peak locations were determined from KDE fits of ungated flow data, gaussian mixture models used to assign cells to peaks, and cells within peaks were designated positive or negative for the respective fluorescent protein using a chosen threshold for peak mean. For T7 RNAP differentiation experiments, peaks with mean log10(sfYFP) > 2.5 were designated 'on'. For mScarletI/sfGFP differentiation experiments, peaks with mean log10(mScarletI) > 3 were designated as 'differentiated'.

Identification of mutations in differentiation-activated T7 RNAP expression

Cells from the eighth plate generation were struck for single colonies. For the case of differentiation with selection (+chloramphenicol), cells from two independent wells from 7.5μM salicylate/0μM IPTG and 7.5μM salicylate/30μM IPTG were plated on LB agar + chloramphenicol, carbenicillin, and Las AHL. Colony PCR using p4_186_primary_FOR and pOSIP_insert_REV was performed on four colonies from each, two of which were sequenced with RW.pir.int.R and RW.pir.int.F2, and sequences were mapped to 'differentiation cassette split T7 RNAP'. This was done similarly for naïve (0μM and 30μM IPTG) and differentiation without selection (7.5μM salicylate/0μM IPTG and 7.5μM salicylate/30μM IPTG), however PCRs were unsuccessful. The source of mutation was determined by (1) isolating plasmid DNA from two isolated colonies from each plate and transforming into cells with genomically integrated inducible T7 RNAP (naïve production cells lacking the pT7-sfYFP construct) and (2) transforming a pSC101-chlor-pT7-BCD2-sfYFP-T2m construct into competent cells prepared from these same cells.

Acknowledgements

We would like to thank Andy Halleran, Anandh Swaminathan, and Andrey Shur for productive conversations; Andy Halleran for providing code for analysis of flow cytometry data; Samuel Clamons for providing code for tidying and analyzing Biotek data; and Andrey Shur and Andy Halleran for providing cloning resources. pSal, pLas, and pTac and their associated evolved transcription factors were kind gifts from Adam Meyer. The CIDAR MoClo Parts Kit was a gift from Douglas Densmore (Addgene kit # 1000000059).[26] This research is supported by the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 and cooperative agreement W911NF-19-2-0026 from the U.S. Army Research Office. The content of the information on this page does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

(1) Glick, B. R. Metabolic Load and Heterlougos Gene Expression. *Biotechnol. Adv.* **1995**, *13* (2), 247–261.

(2) Sleight, S. C.; Bartley, B. A.; Lieviant, J. A.; Sauro, H. M. Designing and Engineering Evolutionary Robust Genetic Circuits. *J. Biol. Eng.* **2010**, *4* (1), 12. https://doi.org/10.1186/1754-1611-4-12.

(3) Canton, B.; Labno, A.; Endy, D. Refinement and Standardization of Synthetic Biological Parts and Devices. *Nat. Biotechnol.* **2008**, *26* (7), 787–793. https://doi.org/10.1038/nbt1413.

(4) Jack, B. R.; Leonard, S. P.; Mishler, D. M.; Renda, B. A.; Leon, D.; Suárez, G. A.; Barrick, J. E. Predicting the Genetic Stability of Engineered DNA Sequences with the EFM Calculator. *ACS Synth. Biol.* **2014**, *4* (8), 939–943. https://doi.org/10.1021/acssynbio.5b00068.

(5) Renda, B. A.; Hammerling, M. J.; Barrick, J. E. Engineering Reduced Evolutionary Potential for Synthetic Biology. *Mol. Biosyst.* **2014**, *10* (7), 1668–1678. https://doi.org/10.1039/c3mb70606k.

(6) Pósfai, G.; Plunkett, G.; Fehér, T.; Frisch, D.; Keil, G. M.; Umenhoffer, K.; Kolisnychenko, V.; Stahl, B.; Sharma, S. S.; De Arruda, M.; et al. Emergent Properties of Reduced-Genome Escherichia Coli. *Science (80-.).* **2006**, *312* (5776), 1044–1046. https://doi.org/10.1126/science.1126439.

(7) Csörgo, B.; Fehér, T.; Tímár, E.; Blattner, F. R.; Pósfai, G. Low-Mutation-Rate, Reduced-Genome Escherichia Coli: An Improved Host for Faithful Maintenance of Engineered Genetic Constructs. *Microb. Cell Fact.* **2012**, *11*, 1–13. https://doi.org/10.1186/1475-2859-11-11.

(8) Sleight, S. C.; Sauro, H. M. Visualization of Evolutionary Stability Dynamics and Competitive Fitness of Escherichia Coli Engineered with Randomized Multigene Circuits. *ACS Synth. Biol.* **2013**, *2* (9), 519–528. https://doi.org/10.1021/sb400055h.

(9) Boo, A.; Stan, G.-B.; Borkowski, O.; Gorochowski, T. E.; Furini, S.; Ellis, T.; Ladak, Y. N.; Gilbert, C.; Ceroni, F.; Awan, A. R. Burden-Driven Feedback Control of Gene Expression. *Nat. Methods* **2018**, *15* (5).

https://doi.org/10.1038/nmeth.4635.
(10) Tyo, K. E. J.; Ajikumar, P. K.; Stephanopoulos, G. Stabilized Gene Duplication Enables Long-Term Selection-Free Heterologous Pathway Expression. *Nat. Biotechnol.* **2009**, *27* (8), 760–765. https://doi.org/10.1038/nbt.1555.
(11) Halleran, A. D.; Flores-bautista, E.; Murray, R. M. Quantitative Characterization of Random Partitioning in the Evolution of Plasmid-Encoded Traits. **2019**, 1–14.
(12) Yang, S.; Sleight, S. C.; Sauro, H. M. Rationally Designed Bidirectional Promoter Improves the Evolutionary Stability of Synthetic Genetic Circuits. *Nucleic Acids Res.* **2013**, *41* (1), 1–7. https://doi.org/10.1093/nar/gks972.
(13) Minty, J. J.; Singer, M. E.; Scholz, S. A.; Bae, C.-H.; Ahn, J.-H.; Foster, C. E.; Liao, J. C.; Lin, X. N. Design and Characterization of Synthetic Fungal-Bacterial Consortia for Direct Production of Isobutanol from Cellulosic Biomass. *Proc. Natl. Acad. Sci.* **2013**, *110* (36), 14592–14597. https://doi.org/10.1073/PNAS.1218447110.
(14) Zhou, K.; Qiao, K.; Edgar, S.; Stephanopoulos, G. Distributing a Metabolic Pathway among a Microbial Consortium Enhances Production of Natural Products. *Nat. Biotechnol.* **2015**, *33* (4), 377–383. https://doi.org/10.1038/nbt.3095.
(15) Roell, G. W.; Zha, J.; Carr, R. R.; Koffas, M. A.; Fong, S. S.; Tang, Y. J. Engineering Microbial Consortia by Division of Labor. *Microb. Cell Fact.* **2019**, *18* (1), 1–11. https://doi.org/10.1186/s12934-019-1083-3.
(16) West, S. A.; Cooper, G. A. Division of Labour in Microorganisms: An Evolutionary Perspective. *Nat. Rev. Microbiol.* **2016**, *14* (11), 716–723. https://doi.org/10.1038/nrmicro.2016.111.
(17) van Gestel, J.; Vlamakis, H.; Kolter, R. Division of Labor in Biofilms: The Ecology of Cell Differentiation. *Microbiol. Spectr.* **2015**, *3* (2), 1–24. https://doi.org/10.1128/microbiolspec.mb-0002-2014.
(18) Kumar, K.; Mella-Herrera, R. A.; Golden, J. W. Cyanobacterial Heterocysts. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (4), 1–19. https://doi.org/10.1101/cshperspect.a000315.
(19) Landy, A. Dynamic, Structural, and Regulatory Aspects of Lambda Site-Specific Recombination. *Annu. Rev. Biochem.* **1989**, *58* (1), 913–941. https://doi.org/10.1146/annurev.bi.58.070189.004405.
(20) Rakowski, S. A.; Filutowicz, M. Plasmid R6K Replication Control. *Plasmid* **2013**, *69* (3), 231–242. https://doi.org/10.1016/j.plasmid.2013.02.003.
(21) Segall-Shapiro, T. H.; Meyer, A. J.; Ellington, A. D.; Sontag, E. D.; Voigt, C. A. A “resource Allocator” for Transcription Based on a Highly Fragmented T7 RNA Polymerase. *Mol. Syst. Biol.* **2014**, *10* (7), 742–742. https://doi.org/10.15252/msb.20145299.
(22) Shachrai, I.; Zaslaver, A.; Alon, U.; Dekel, E. Cost of Unneeded Proteins in E. Coli Is Reduced after Several Generations in Exponential Growth. *Mol. Cell* **2010**, *38* (5), 758–767. https://doi.org/10.1016/j.molcel.2010.04.015.
(23) Jang, C. W.; Magnuson, T. A Novel Selection Marker for Efficient DNA Cloning and Recombineering in E. Coli. *PLoS One* **2013**, *8* (2), 1–7. https://doi.org/10.1371/journal.pone.0057075.
(24) St-Pierre, F.; Cui, L.; Priest, D. G.; Endy, D.; Dodd, I. B.; Shearwin, K. E. One-

Step Cloning and Chromosomal Integration of DNA. *ACS Synth. Biol.* **2013**, *2* (9), 537–541. https://doi.org/10.1021/sb400021j.
(25) Halleran, A. D.; Swaminathan, A.; Murray, R. M. Single Day Construction of Multigene Circuits with 3G Assembly. *ACS Synth. Biol.* **2018**, *7* (5), 1477–1480. https://doi.org/10.1021/acssynbio.8b00060.
(26) Iverson, S. V.; Haddock, T. L.; Beal, J.; Densmore, D. M. CIDAR MoClo: Improved MoClo Assembly Standard and New E. Coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synth. Biol.* **2016**, *5* (1), 99–103. https://doi.org/10.1021/acssynbio.5b00124.
(27) Ghosh, P.; Pannunzio, N. R.; Hatfull, G. F.; Gottesman, M. Synapsis in Phage Bxb1 Integration: Selection Mechanism for the Correct Pair of Recombination Sites. *J. Mol. Biol.* **2005**, *349* (2), 331–348. https://doi.org/10.1016/j.jmb.2005.03.043.

Supplementary Information

All data and data analysis, and code for modeling of differentiation circuits are available on the Github repository: https://github.com/rlwillia/Integrase-differentiation

Sequences of all parts, primers, and constructs, and a table with descriptions of mutations observed in experiments with T7 RNAP differentiation with selection, can be found in the supplementary excel file.