CRISPR-based Control of Biosynthetic Pathways and Genetic Biosensors, Informed by a Quantitative Investigation of Metabolic Burden

Ian Faulkner

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2021

Reading Committee:
James Carothers, Chair
Jesse Zalatan
Cole DeForest

Program Authorized to Offer Degree:
Chemical Engineering

University of Washington

Abstract

CRISPR-based Control of Biosynthetic Pathways and Genetic Biosensors, Informed by a Quantitative Investigation of Metabolic Burden

Ian Faulkner

Chair of the Supervisory Committee:
James Carothers
Department of Chemical Engineering

The heterologous, engineered genetics often used for biosynthetic chemical production share resources with host gene expression. While the host's endogenous expression has evolved along with complex regulatory networks to balance supply and demand, or cost and benefit, heterologous expression requires engineered control systems to avoid overly burdening the host. Optimizing the genetic controllers of pathway enzymes, for example, achieves greater likelihood of acceptable balance between cost and production, avoiding reductions in growth rate, population homogeneity, and production titers or circuit performance. Here, we first investigate the metabolic burden of genetic control systems implemented with CRISPR transcriptional activation and T7 RNA polymerase in *E. coli*, finding that the overhead cost of CRISPR-based control is relatively high, but that expansion of the circuit to several control nodes that share the overhead cost is an inexpensive addition. We next analyze a set of feedback control circuits suggesting that closed circuits can limit excessive heterologous gene expression at a metabolic cost comparable to simpler, and less stable, open circuitry. Finally, we build metabolic pathways producing valuable chemicals—the aromatic amino acid derivative *p*-aminocinnamic acid and the human milk oligosaccharide lacto-*N*-tetraose—making use of CRISPR-based control's expandability into broad, open

circuits controlling not only the pathway enzymes, but an potentially large array of endogenously-targeted metabolic regulators. As part of the CRISPR circuit for LNT production, we develop a novel *in vivo* biosensor, based on a ligand-responsive scRNA switch, to report on production levels, and use it alongside traditional analysis to query a library of pathway expression variants. These varied proofs-of-concept illustrate the impressive and versatile capabilities of CRISPR-based genetic control when applied to biosynthetic pathways, and are informed by an enhanced understanding of heterologous contributors to metabolic burden in biosynthetic hosts.

Table of Contents

Table of Contents **5**

List of Figures **7**

Introduction **10**

Chapter 1: Metabolic Burden of Heterologous Expression Control **13**

1.1: Introduction 14

1.2: Metabolic burden measurement scheme 16

1.3: Burden of controller components 19

1.4: Burden-output relationship is linear in *E. coli* 25

1.5: Comparison of output efficiency 26

1.6: Conclusions 30

1.7: Methods 31

1.8: Supplementary Figures 34

Chapter 2: Circuit control of heterologous expression **41**

2.1: Introduction 42

2.2: Protein-based feedback control 43

2.3: Construction and characterization of mRNA–ribozyme–gRNA devices 47

2.4: Hybrid T7/CRISPRi feedback controller 51

2.5: CRISPRa/i feedback controller 53

2.6: CRISPRa/i incoherent feedforward controller 57

2.7: Conclusions 62

2.8: Methods 63

2.9: Supplementary Figures 66

Chapter 3: Broad-circuit control of *p*-AF/*p*-ACA production in *Pseudomonas putida* **70**

3.1: Introduction 71

3.2: Heterologous genetics 73

3.3: Output copy number, burden, and integration 76
3.4: Pathway engineering for *p*-ACA production improvement 79
3.5: Endogenous CRISPRa/i for *p*-ACA production improvement 83
3.6: Circuit size considerations 84
3.7: Conclusions 84
3.8: Methods 85
3.9: Supplementary Figures 87

Chapter 4: Production and detection of LNT by a ligand-responsive scRNA 97
4.1: Introduction 98
4.2: Heterologous genetics 100
4.3: Library screen and production 101
4.4: Endogenous CRISPRa/i for LNT production improvement 102
4.5: Circuit size considerations and intracellular sensor for screening 104
4.6: Sensor development: generating an LNT-binding aptamer 105
4.7: Sensor development: design of aptamers into ligand-responsive scRNAs 109
4.8: Sensor development: a new CRISPRa promoter controlling RFP output 109
4.9: Screening strategy and results: shared activator 111
4.10: Screening strategy: orthogonal activator 113
4.11: Detection of intracellular LNT by a scRNA biosensor 113
4.12: Conclusions 117
4.13: Methods 118
4.14: Supplementary Figures 121

Chapter 5: Conclusions 132

References 135

Acknowledgements 146

List of Figures

Chapter 1

1.1. Scheme for measurement of met. burden due to expression of heterologous genetics 18
1.2. Off-target reference strains for heterologous expression controllers 21
1.3. Cost of expression of individual CRISPRa control components 22
1.4. Cost of expression of a T7 RNA polymerase controller 24
1.5. Comparisons of burden-per-output functions 27

Chapter 2

2.1. T7 lysozyme feedback on T7 RNAP control 45
2.2. LacI feedback on T7 RNAP control 46
2.3. Function of an mRNA–ribozyme–gRNA device 48
2.4. T7-controlled, mRFP1-targeted MRG repression 50
2.5. CRISPRi feedback on T7 RNAP control 52
2.6. Open-circuit CRISPRa reaching metabolic limits 54
2.7. MRG-CRISPRi feedback on CRISPRa control limits instability 56
2.8. CRISPRa/i-based IFFL yields low, pulsatile output expression 58
2.9. Burden per output of the CRISPR-based IFFL 59
2.10. High-output CRISPR-based IFFL 60
2.11. Burden per output of the high-output CRISPR-based IFFL 61

Chapter 3

3.1. Connection of the *p*-ACA production pathway to central metabolism of *P. putida* through chorismate 72
3.2. Desired product *p*-ACA is toxic to *E. coli* but not *P. putida* 74
3.3. CRISPRa-controlled, plasmid-based *p*-AF production in *P. putida* 75
3.4. Effect of plasmid burden on kinetic growth of *P. putida* 76
3.5. Effects of second-plasmid burden, and genomic copy number, on *p*-AF production 77
3.6. Production of *p*-ACA from the two-plasmid system in *P. putida* 79
3.7. *p*-ACA production using *R. glutinis* PAL 80
3.8. Summary of one-plasmid *p*-ACA production using *Rg*-PAL 81
3.9. Side products generated by *papABC/aroGL* activation or PAL activation 83

Chapter 4

4.1. The enzymatic portion of the LNT production pathway 99

4.2. Two-plasmid system for expressing the LNT production pathway 100
4.3. Orthogonal, tunable scRNA-promoter pairs constituting a combinatorial library of pathway gene expression 101
4.4. Combinatorial expression library highlights LNT-producing strains 102
4.5. Function of a ligand-responsive scRNA switch 105
4.6. Selection of RNA aptamers by SELEX 106
4.7. Sequencing-based investigation of aptamer candidate enrichment during SELEX 107
4.8. MFE structures of aptamer candidates 108
4.9. Control scRNA with no aptamer region activating J_{sen} promoter 110
4.10. Inducible CRISPRa control of LNT production 111
4.11. Scheme for sequential timer pool screen 112
4.12. Detection of intracellular LNT by a scRNA biosensor 114
4.13. LNTc2 Variant 1 biosensor response to production variants 115

Supplementary Figures can be found at the end of each chapter:

Chapter 1……34
1.S1. Output expression is unaffected by capacity monitor expression level, regardless of controller
1.S2. Induction responses of tet-controlled capacity monitor
1.S3. mRFP1 output has a stronger effect on capacity than mTag-BFP
1.S4. SDS-PAGE showing T7 RNAP levels
1.S5. CRISPRa components expressed in excess
1.S6. Weaker output RBS increases burden of T7 RNAP control
Chapter 2……66
2.S1. MRG expression burden under T7 RNAP control
2.S2. MRG expression burden under *E. coli* RNAP control
2.S3. Truncations of repressive gRNA fail to tune T7-MRG circuit
2.S4. Burden per output unaffected by gRNA truncations
2.S5. Expression cost of additional RNA elements within the CRISPR IFFL
Chapter 3……87
3.S1. Tet-inducible *p*-AF production in *E. coli*.
3.S2. CRISPRa-controlled *p*-AF production in *E. coli*.

3.S3. A self-contained CRISPRi plasmid to suppress strong-promoter output
3.S4. RFP activation from integrated synthetic CRISPRa promoters with varying strength
3.S5. Two-plasmid *p*-ACA production using *At*-PAL2
3.S6. Overlaid HPLC chromatograms of *p*-ACA production and low-concentration spike-in
3.S7. Lack of *p*-ACA consumption by either knockout or intact strain
3.S8. Lack of *p*-ACA toxicity in the knockout strain
3.S9. Weak integrated strain nonetheless activates side products
3.S10. Side product generated by different PAL enzymes
3.S11. One-plasmid *p*-ACA production using *Rg*-PAL
3.S12. Typical standard curves for HPLC quantification of *p*-AF, *p*-ACA, and *p*-ABA
Chapter 4……...121
4.S1. scRNA spacer sequence truncations tune activation of output
4.S2. Combinatorial expression library highlights LNT II-producing strains
4.S3. CRISPRi of endogenous targets in MG1655 and JM109
4.S4. In-line probing assay indicates aptamer-LNT binding by differential degradation
4.S5. Induction response to aTc in two-plasmid system in JM109
4.S6. LNT-responsive scRNA candidates
4.S7. Scheme for simultaneous timer pool screen
4.S8. Verification of LNT production in LNTc2 sensor strains
4.S9. RFP output levels of biosensor candidates
4.S10. Production variation in LNTc2 Variant 1 sensor strain
4.S11. Typical standard curves for HPLC quantification of LNT and LNT II
Table 4.S1. Revised selection conditions

Introduction

The field of synthetic biology has matured considerably in recent decades, with one of its main applications being metabolic engineering for the production of valuable chemicals from microbial hosts[1]. Academic studies detailing the construction of biosynthetic pathways, however, often fail to achieve industrially competitive product titers or yields. Moving beyond this small, proof-of-concept scale not only takes prodigious amounts of funding and development, it requires attention at all project phases to at least two immensely important concepts.

The first consideration is the metabolic burden imposed on the host strain by the heterologous pathway, meaning that the host's naturally-evolved balances of metabolic and gene expression resources, both energetic and material, have been upset by the new demands, possibly impacting growth rate, phenotype, genetic stability, and ultimately product titer[2]. Reasonable limitation of this burden, through strain selection, pathway design, control circuit design, and growth conditions, can avoid low or unpredictable production. The second consideration is the interface between the heterologous pathway and endogenous metabolism, in terms of direct use of host resources by the intended pathway[3]. Endogenous substrates, for example, could be required by pathway enzymes and could be limiting; or nebulous endogenous regulatory networks could be adjusted to boost metabolic flux toward the pathway. The natural complexity, redundancy, and autoregulation of such networks often requires widespread manipulation to achieve higher pathway flux, requiring substantially-sized heterologous genetic control circuits.

At the same time, an explosion of CRISPR-based technologies has enabled new strategies in metabolic engineering[4–7]. Namely, gene expression—even of endogenous genes—can be manipulated in a predictable and programmable way, simply by synthesizing the sequence of a guide RNA (gRNA, for CRISPR interference) or scaffold RNA (scRNA, for CRISPR activation) while following a number of design rules[8]. In our lab and collaborating labs, we have recently developed an orthogonal set of synthetic promoters, known as the J-series, that provide CRISPR activation to a weak constitutive

promoter of the Anderson series (parts.igem.org), for use in heterologous gene expression[9]. The modular nature of CRISPRa's protein components and especially its scRNAs is theoretically well-suited to the sprawling control circuits demanded by large biosynthetic pathways and widespread genomic manipulation, including master regulation, shared control elements, and arbitrarily large arrays of circuit nodes.

The protein components of CRISPR complexes are large, however, leading to concerns about their use as shared genetic controllers in production hosts already experiencing metabolic burden from heterologous pathways[10–13]. In Chapter 1, we quantify the burden arising from expression of the CRISPR components, compare it to other controllers, and separate it from the more extensive burden caused by expression of the output gene. We show that while the protein components are indeed expensive, the expansion of CRISPR control to many nodes is only a matter of expressing additional RNAs, which are relatively inexpensive. This finding is part of a larger investigation of metabolic burden in *E. coli*, showing that heterologous output expression contributes much more to burden than its expression controller does. Retroactive effects[10] on controller proteins, then, results more from total output burden than of circuit size *per se*. This output burden remains a linear response to heterologous expression, up to a point where the cell's capacity to express genes declines catastrophically[2].

It seems that this point of steep decline is easier to reach when using a synthetic expression controller that provides transcriptional activation, because its activity exists outside the bounds of the host's tightly-controlled and interconnected network of expression control (which is intended in part to regulate expression resource allocation). Therefore, limiting heterologous expression's capability to push the cells to such a situation can safeguard strain stability and therefore overall product titer. This limitation could be achieved by designing autoregulatory motifs into heterologous control circuits, to limit problematic output expression before resource exhaustion does. Exploring simple examples of these motifs as they apply to synthetic controllers, and their impacts on metabolic burden and strain stability, is the focus of Chapter 2.

Chapters 3 and 4 describe the use of CRISPR circuits controlling the microbial production of the valuable chemicals *p*-aminocinnamic acid (*p*-ACA) and lacto-*N*-tetraose (LNT), respectively, with the intent of including in each case an array of control circuit nodes aimed at manipulation of endogenous metabolism to boost production. This is a potentially widespread use case of the broad, single-layer circuit design—including a shared master regulator across numerous nodes—to which CRISPR expression control seems well-suited, so implementation in these pathways is a valuable test case. In Chapter 3, we implement the *p*-ACA pathway in the non-model bacterium *Pseudomonas putida*. In Chapter 4, we design a CRISPR-compatible genetic biosensor for *in vivo* detection and reporting of LNT production in an *E. coli* host, an extension of the overall control circuit design with myriad configurations useful to rapid strain improvement[14].

Chapter 1: Metabolic Burden of Heterologous Expression Control

Abstract

It is well-understood that the timing and levels of gene expression constituting engineered pathways and circuits must be carefully controlled to balance the supply and demand of cellular resources. Heterologous genetics share these resources with native processes, so the mechanisms of controlling heterologous gene expression are very important for optimizing the balance between cost and production. Relatively little is known about the material and energy costs of these control systems, despite their important role in engineering biology. Here, we investigate the metabolic burden of genetic control systems implemented with CRISPR transcriptional activation and T7 RNA polymerase in *E. coli*. After quantifying the costs of individual CRISPRa components, as well as separating the cost of the control systems from their outputs, we find that a single node of CRISPR-based control costs significantly more than a node of T7 control (29% of available expression capacity vs. 5%). This is with dCas9 and scRNA overexpressed to some degree, so the CRISPRa cost could potentially be lowered while retaining functionality, as capacity cost scales linearly with burden. Additionally, because the protein components of CRISPRa represent an overhead cost shared across nodes, and because they make up the bulk of CRISPR-based control cost, additional nodes can be added to CRISPR circuits relatively cheaply (< 5%). This suggests an exciting expandability of CRISPR-based circuits that is limited more by the cost of the output proteins it controls than by the cost of the control circuit itself.

1.1: Introduction

Heterologous genes, for example expressing the enzymes constituting biosynthetic pathways, share expression machinery and resources with the host cell's endogenous expression and metabolism. While the host can usually accommodate the extra expression, in doing so it experiences metabolic burden, in which other expression throughout the cell is reduced by heterologous consumption of energetic or material resources[2,15] or by heterologous occupancy of machinery such as RNA polymerase or ribosomes[12,16–18]. This burden can affect the host phenotype, from general perturbations like decreased growth rate[19] to specific ones like decreased levels of an endogenous substrate required by the heterologous pathway.

Expected levels of production from a heterologous pathway, or proper function of a genetic circuit controlling it, then, depends in part on the burden that it or other heterologous genetics imposes on the host. This burden should be minimized to avoid negative impacts on strain performance, which of course can include decreased production due to lower enzyme levels, but also more dramatic problems like bimodality across the population[20,21], ribosome loss[22], or instability of the heterologous genetics[23–25]. Critically, this instability leads to subpopulations of cells that have quit heterologous expression—and, being unburdened and fast-growing relative to the intact cells, quickly fill the overall culture with a high proportion of non-producing "cheater" cells, a common mode of failed biosynthesis cultures[26].

In an effort to better understand metabolic burden and refine strategies for its mitigation, we compared an emerging synthetic transcriptional control system (CRISPRa, related in the overall introduction) and a classic one: the RNA polymerase from T7 bacteriophage[27]. Though this latter controller theoretically is insulated from host RNAP availability due to its orthogonality, it is still sensitive to the finite expression resources of the host, because those resources are shared. When using either of these systems to control gene expression, then, we observed sometimes dramatic reductions in other, ostensibly independent expression, including of measureable products such as fluorescent protein[12].

This effect on a measurable protein is basically the capacity monitor approach[28], in which measurable decreases in constitutive expression are used to quantify metabolic burden in a more sensitive way than other measures, like reduction of growth rate. We adapted this rate-of-production capacity measurement to an endpoint total protein measurement—to accommodate varying growth rates between strains—and found that it was still a sensitive measurement of the metabolic burden imposed by heterologous expression. Categorizing this expression, then, into controllers, outputs, and the capacity monitor (Figure 1.1), we used features of the CRISPRa and T7 controllers to understand and quantify the contributions to burden from—and within—each of those categories.

In this approach, careful selection of control strains as "burdenless" reference strains allows the quantification of many elements of heterologous expression in isolation, because of the availability of off-target mismatches between controllers and their outputs. In the case of CRISPRa, disrupting the complementarity between the scRNA spacer and its target in the promoter serves to disconnect controller from output, as does the mismatched use of orthogonal promoter-scRNA pairs[8]. In the case of T7, published sets of orthogonal RNAP variants[29,30] and their cognate promoters can be mismatched in much the same way. We used these off-target controls to separately measure the expression burden supplied by an output, usually a fluorescent protein, and that supplied by the expression of the controller itself. Further, because the bacterial CRISPRa complex is composed of dCas9, MCP-SoxS as a transcriptional activator, and scRNA, we measured the individual burden contributions of each of these components, at their typical expression levels, through promoter deletions. For the output gene, because its expression level is variable depending on the action of the controller, it was more fitting to treat burden as a function: that is, to quantify the amount of burden supplied by each unit of output. Our observation that those functions were linear in *E. coli*[31] and similar in slope regardless of controller, together with the component cost data, informed later efforts using CRISPRa as a controller of biosynthetic pathway expression.

1.2: Metabolic burden measurement scheme

As a way of assessing the metabolic burden of heterologous expression, particularly its effect on the global expression capacity of the host organism, we quantified relative expression levels of an unrelated, constitutively-expressed fluorescent protein[28], usually sfGFP but occasionally mRFP1. These expression levels were compared in a test strain experiencing the burden to be measured versus a reference strain in which that burden was absent. Effects that are capable of being measured in this scheme, therefore, are limited to those which can be isolated by the use of control strains as reference. For example, it is possible to measure the burden of only the output RFP expression of CRISPR-activated expression control, while ignoring the burden of expressing the CRISPRa machinery itself, by using a reference strain in which all of the CRISPRa machinery is expressed, but with an off-target scRNA. However, it is not possible to measure the relative contributions of transcription versus translation, for example, of that same output RFP expression using that same reference strain. The relevant metric generated by this analysis can be thought of as a percentage of the host's original excess expression capacity that is still available for further expression (Figure 1.1). It is worth noting, however, that the burden imposed by extremely high levels of heterologous expression can catastrophically exhaust certain expression resources (see Chapter 2), and that the phenotypes thus generated are often not well-measured by this capacity monitor approach. It is safe to say that to avoid these metabolic cliff phenotypes, we should design expression systems to stay well away from capacity measurements approaching zero, and that milder levels of burden can be used to illuminate principles of resource allocation using a capacity monitor approach.

The usual outcome of this analysis is a somewhat lower expression level of the capacity monitor in the more-burdened experimental strain than in the less-burdened reference strain, because the extra burden decreases global expression capacity, presumably affecting some endogenous genetics in a similar way as the heterologous capacity monitor. The key metric—the amount of expression capacity that remains to

the cell after accounting for this burden—is calculated from the endpoint-measured capacity monitor expression as follows:

$$Remaining\ Expression\ Capacity = \frac{\frac{GFP}{OD_{600}}\ (experimental)}{\frac{GFP}{OD_{600}}\ (reference)} \qquad \text{Eq. 1.1}$$

Fluorescence values are always normalized to cell density as measured by OD_{600} in an effort to account for differences in growth; because we measure at endpoint, however, these differences are often minimal. (In the case of extreme burden causing genetic instability, non-producing subpopulations can arise to a high proportion of the culture, contributing density but not fluorescence: one possible mode by which this approach can fail to capture effects of extreme burden.) Optical density normalization is a feature of measurements using a plate reader, representing almost all of the data in Chapters 1 and 2, but is not necessary for those few measurements performed on a flow cytometer, in which case median values of the population are reported. This expression capacity metric represents the cost of producing whatever difference exists between the reference strain and the experimental strain.

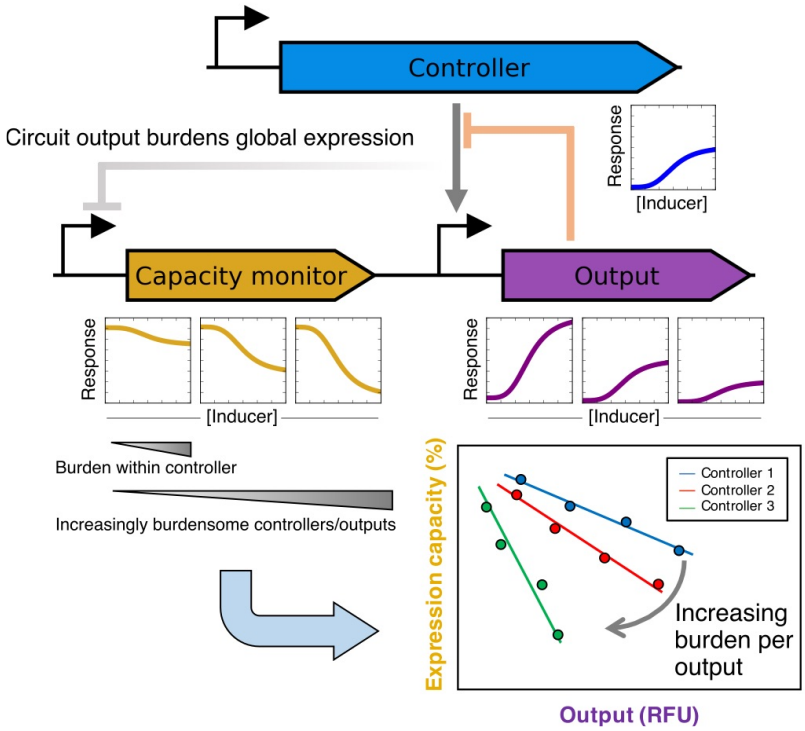


Figure 1.1. Scheme for measurement of metabolic burden due to expression of heterologous genetics. A system of medium- or low-copy plasmids pairs a controller of gene expression with its fluorescent output gene (including optional autoregulation in orange), while a fluorescent capacity monitor gene is independently controlled. Increasing levels of output result in decreased levels of overall expression capacity (grey), a fluorescent monitor of which is used here as a quantitative measure of expression burden. In *E. coli*, the effect on capacity within each genotype is a linear function of output.

In this way, and given sufficient modularity of the expression system, we can design an array of reference strains to tease apart many contributors to heterologous expression burden. The use of control systems that include some method of off-target lack of output is key to many of these comparisons. For example, in a CRISPRa controller, an off-target spacer sequence in the scRNA not only separates the output burden from the burden of expressing the CRISPRa components themselves, it serves as an unactivated reference strain for quantification of individual CRISPRa component burden contributions, free from the confounding influence of output burden. Similarly, the orthogonal set of full-length T7 RNA polymerases allows for off-target reference strains by mismatching the RNAP itself and the output T7 promoter variant. Finally, the function generated by plotting expression capacity versus output gives some sense of expression efficiency. In *E. coli*, we found this function to be linear, and took the line's slope as a quantitative measure of efficiency. In this work, we focused on quantifying burden contributions from controllers and their components, and on assessing the relative efficiency of different controllers in terms of burden per unit of output.

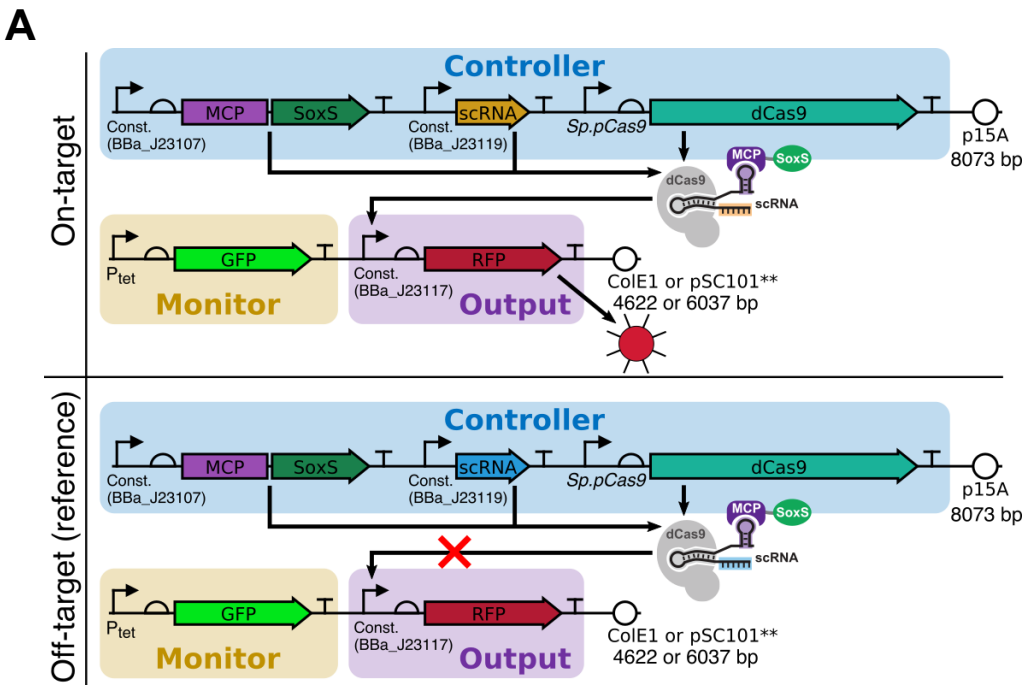
One potential issue we identified with this approach is the possibility of capacity monitor expression affecting the system, particularly output level, through expression burden of its own. If true, this effect could set up a complicated system of reciprocal burden effects that would be difficult or impossible to interpret. Fortunately, many of our implementations of this capacity monitor architecture, especially early on, included control of the capacity monitor by the aTc-responsive transcriptional repressor TetR, allowing inducible control of capacity monitor (sfGFP) expression. Comparison of different capacity monitor induction levels within the same output induction level showed very little effect on measured output expression levels (Figure 1.S1), indicating that this potential confounding effect is minimal. This finding holds true regardless of the protein

identities of the output and capacity monitor: that is, it doesn't matter which is sfGFP and which is mRFP1. Our interpretation of this finding is that relative expression levels of output and capacity monitor determine which burdens which: namely, the higher-expressed gene is relatively insensitive to fluctuations in the lower-expressed gene. A potential implication is that the concept of capacity reduction as measured by this scheme is more relevant to weakly-expressed endogenous genes than highly-expressed ones. Further, because of the tet control of the sfGFP capacity monitor, we were able to assess burden per output at many different points along the aTc induction curve (Figure 1.S2), finding that all reasonably high aTc inductions (so that sfGFP/OD_{600} is roughly > 500) gave the same expression capacity metric when normalized within that aTc induction level. Given this equivalence, in later experiments we used only 80 nM aTc for sfGFP capacity monitor induction. With these notes in mind, we can ensure proper functioning of this capacity monitor approach by calibrating capacity monitor expression to a level well below that of output expression (but still high enough to reliably quantify burden-induced reduction), regardless of which fluorescent protein is used as the monitor—and despite some observed differences in burden per output inherent to different fluorescent proteins (Figure 1.S3).

1.3: Burden of controller components

While the burden imposed by output expression is investigated in more detail in later sections, we initially noticed that it constituted the majority of heterologous expression burden: contributions from the expression of the controllers themselves were relatively small. In our efforts to quantify this controller burden, then, it was important to eliminate output as a source of confounding burden: that is, to compare a reference strain with no controller expression (and necessarily no output expression) to an experimental strain with controller expression but output expression abolished, either through its gene's absence or, preferably, an off-target mismatch between the controller and the output gene's promoter. Figure 1.2 illustrates off-target controls for two synthetic controllers thus quantified: CRISPR activation and T7 RNA polymerase.

Specifically, off-target control was achieved in a CRISPRa system by changing the spacer sequence of the scRNA so it is no longer complementary to its target in the output promoter. This approach theoretically does not affect expression burden, except by the absence of output. In this work, the on-target scRNA was J306[8] and off-target was J406, with output levels the same as the unactivated base promoter. Absence of individual controller components (dCas9, MCP-SoxS, and scRNA) was achieved by deleting the constitutive promoter driving each of them. Off-target control was achieved in a T7 RNAP system by use of a set of orthogonal full-length polymerases[29], each differing from wild-type T7 RNAP by up to 14 amino acid mutations, and each with its own cognate promoter similar to the consensus T7 promoter. In this work, we found that CGG-R12-KIRV paired with P_{CGG} (both known in shorthand as T7A) was a useful on-target pairing, though it suffered from high uninduced background under lac induction, and that T3-R5-RV (known in shorthand as T7C) paired with P_{CGG} (T7A) was a useful off-target pairing, with output levels close to zero. Additionally, for instances in which low uninduced background was critical, such as Figure 1.4, we found that CTGA-R13-AKSIRV (known in shorthand as T7B) under tet control was most useful due to the tet system's tight uninduced repression.



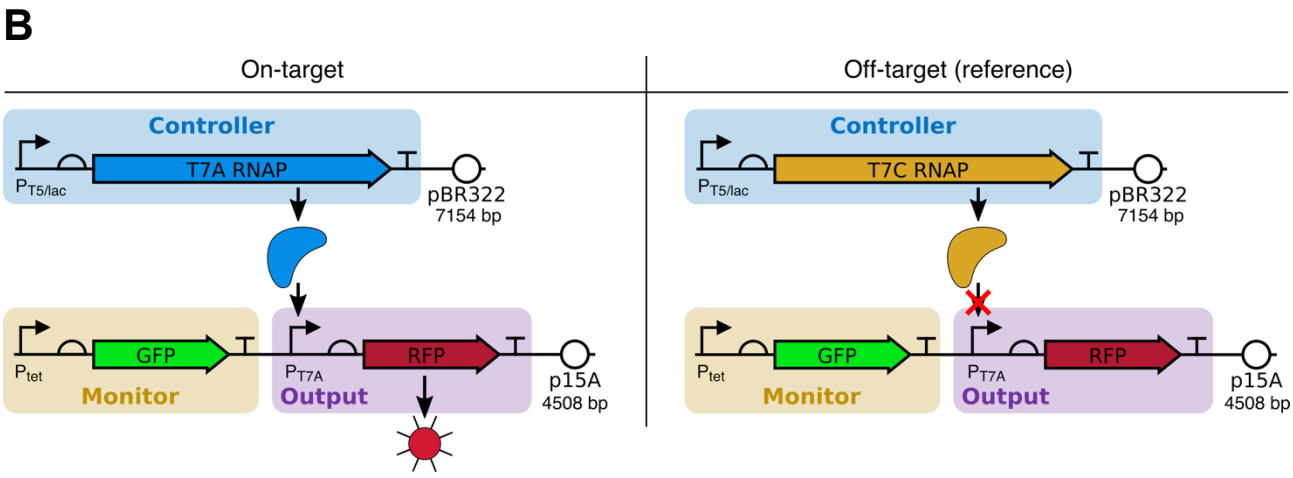


Figure 1.2. Off-target reference strains for heterologous expression controllers. The ability to separate contributors to expression burden—output from controller, or different controller components from each other—is important for understanding its nature and mitigation strategies. Here are two synthetic controllers presented within the framework of controller/output/monitor modules: CRISPR activation (A) and T7 RNA polymerase (B). In both cases, off-target mismatches between controller and output promoter allow separate measurement of the expression burden contributed by the output expression versus the expression of the controller itself.

In a CRISPRa system with a J23117 base output promoter strength and a low-copy (pSC101**) output plasmid (Figure 1.2A), we systematically eliminated expression of each of the three controller components and assessed the effects on expression burden using the tet-controlled sfGFP capacity monitor, with an off-target, intact controller as comparison. Expression was eliminated by simply deleting the corresponding promoter, which avoids confounding plasmid size effects that might be expected if the entire gene were deleted. The dCas9 promoter was its endogenous promoter from *S. pyogenes*, the MCP-SoxS promoter was the medium constitutive promoter J23107, and the scRNA promoter was the strong constitutive promoter J23119, and all components were expressed from a medium-copy p15A plasmid. Figure 1.3 shows the quantitative reduction in sfGFP capacity between a less-burdened (reference) strain with a component deleted and a more-burdened (experimental) strain with all components expressed: dCas9 expression cost the cells 14.5% of expression capacity, MCP-SoxS cost them 10.9%, and scRNA cost them 4.7%. Because of slightly different output levels between the no-MCP-SoxS culture and its off-target control

(probably due to a weak CRISPRi effect when the on-target scRNA is missing MCP-SoxS), we can use the observed burden per output function to adjust the cost of MCP-SoxS from 10.9% to about 9.6%. In each case, an on-target control with intact controller was included to demonstrate the functionality of these expression levels of the CRISPRa components—however, this does not take into account the possibility of excess component expression: for example, J23119-driven scRNA can be quite excessive for normal CRISPRa function, with recent work suggesting that J23105–scRNA and J23106–dCas9 are sufficient for most simple applications (Figure 1.S5). Regardless of excess, it appears that the cost of single-node CRISPRa control totals approximately 29% of *E. coli*'s available expression capacity.

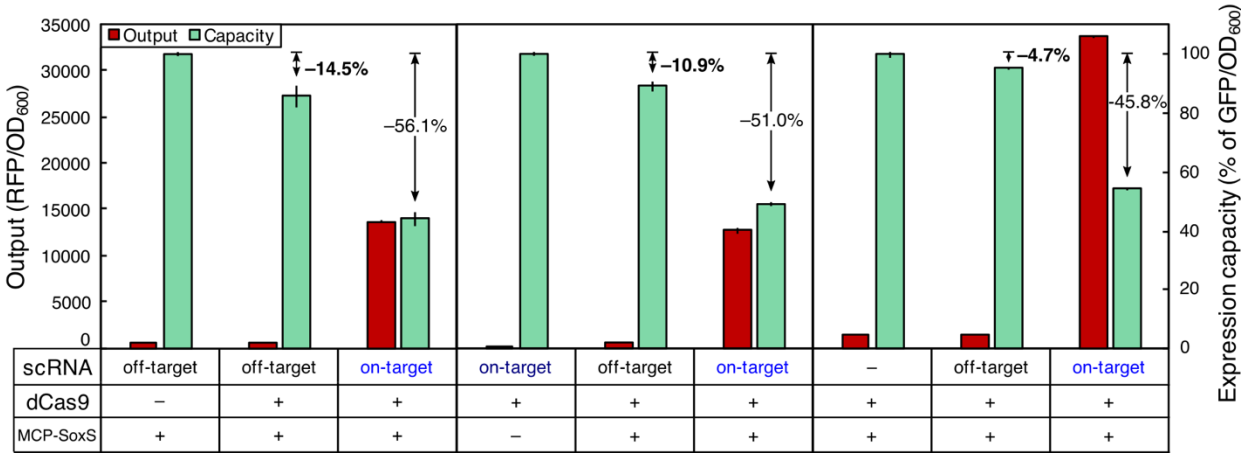
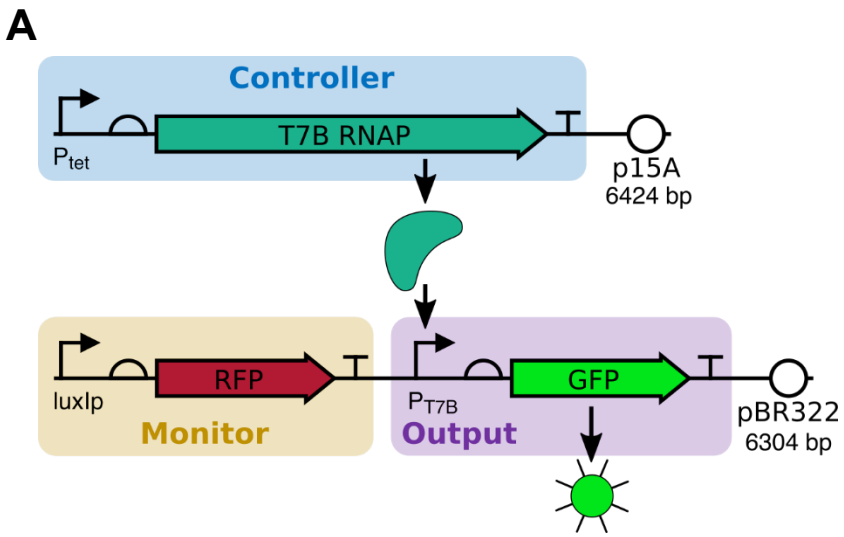


Figure 1.3. Cost of expression of individual CRISPRa control components. Here we compare off-target CRISPRa strain against reference genotypes in which various controller elements have their promoters deleted (dCas9, left panel; MCP-SoxS, middle panel; scRNA, right panel). These comparisons quantify the costs, measured in capacity reduction, required to produce each of the necessary components for CRISPRa. The cost of dCas9 expression from the *Sp*.pCas9 promoter is measured at 14.5% of expression capacity; the cost of MCP-SoxS expression from J23107 is measured at 10.9%, but accounting for a slight output difference in the reference strain adjusts the cost to 9.6%; and the cost of scRNA expression from J23119 is measured at 4.7%. All three are carried on a p15A medium-copy plasmid. These expression levels appear to result in excess dCas9 and scRNA (Figure 1.S5) and could potentially be lowered, sparing some capacity. In each panel, the high-output right column verifies capacity monitor functionality. Error bars represent standard deviation of n=3 biological replicates.

By contrast, we also quantified the cost of a T7 RNAP controller in the absence of its output, using a tet-controlled T7B RNAP, an sfGFP output gene that was absent

except in the functionality control, and an mRFP1 capacity monitor under lux control (Figure 1.4). The burden of T7B induction was quantified by comparison of uninduced (reference) versus induced (experimental) off-target control. We found that an aTc concentration that was just enough for maximal output expression when on-target—4 nM—induced an amount of T7B RNAP that, in the off-target case, was inexpensive compared to CRISPRa, at 4.5% of expression capacity. However, additional aTc increased burden drastically in the off-target case while not increasing output in the on-target case, reaching 16.6% of capacity by 10 nM aTc. This steep response suggests a much easier route to problematic overexpression of T7 RNAP compared to other controllers, with consequences possibly including genetic instability. Such a suggestion fits well in the context of the high elongation speed of T7 RNAP—and therefore its consumption rate of NTPs and production rate of transcripts that might outstrip ribosome availability—and its orthogonality to native transcriptional machinery, which places it somewhat outside of native transcriptional control networks.



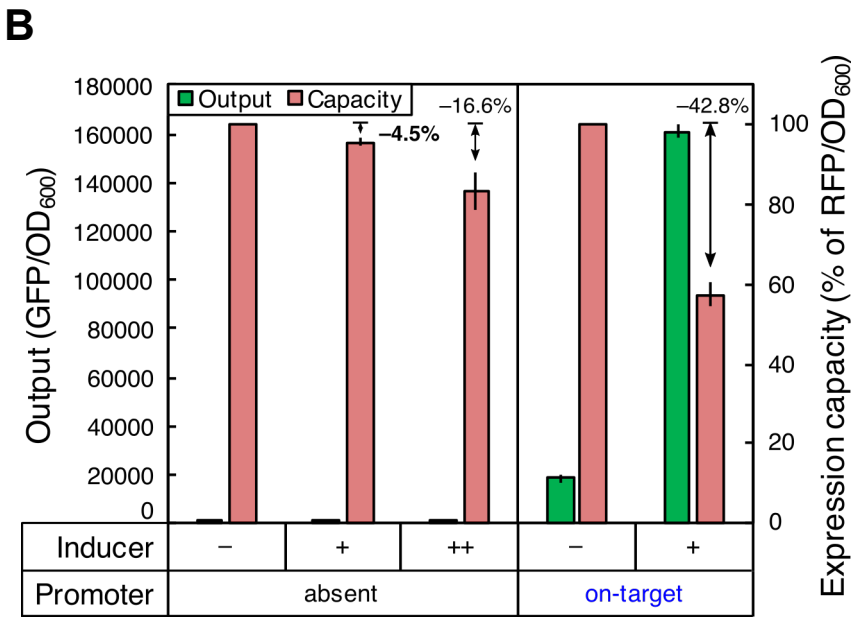


Figure 1.4. Cost of expression of a T7 RNA polymerase controller. **A)** Here we compare a functional expression level of T7B RNAP against an uninduced reference strain, without a T7B promoter in the system (to avoid output burden counfounding the measurement). Note that the controller in this case is driven by a P_{tet} promoter, the output protein is sfGFP, and the capacity monitor is luxIp-driven mRFP1, unlike in Figure 1.2B. **B)** This comparison quantifies the cost, measured in capacity reduction, required to produce enough T7 RNAP to result in high output expression (verified in the right panel using an on-target promoter). The cost of expressing that amount of T7 RNAP from the P_{tet} promoter is measured at 4.5%. However, overinduction at higher levels of aTc is possible, dramatically increasing cost to 16.6% by increasing [aTc] from 4 nM to 10 nM, without increasing output (not shown). This sensitivity could prove problematic in the context of larger circuits and their accompanying retroactive effects[10]. T7 RNAP is carried on a p15A plasmid. Error bars represent standard deviation of *n*=3 biological replicates.

Together, these results suggest that these controllers have different advantages well-suited to different applications. For expression of a single output from a single control point, or circuit node, with the goal of large amounts of output, T7 is probably the better option, as precise induction control can usually stave off overexpression, genetic instability, and culture population heterogeneity. If larger control circuits are desired, however, for example in the case of large pathway control, endogenous gene knockdown/activation, or other applications demanding numerous control nodes within the circuit, CRISPRa/i is probably well-suited, because the cost of expressing dCas9 and MCP-SoxS (or other activators) represents an “overhead cost” that is paid once for the whole circuit, after which these proteins are shared between a number of scRNAs

that represent additional circuit nodes. Because they are relatively inexpensive in terms of expression capacity, this number of scRNAs or nodes can theoretically become quite large, though in practice it is probably limited by competition for binding dCas9 and MCP-SoxS before it is limited by expression capacity. While in this experiment we only investigate the cost of one scRNA at a time, in Chapter 2 we confirm the low cost of additional scRNAs/gRNAs (up to three total).

A similar attempt to build such arbitrarily large circuits using orthogonal T7 RNAP controllers would require an additional RNAP (either full-length or partial) to represent each node of the circuit, quickly increasing controller cost. Additionally, as these nodes would start to interact with one another through their imposed expression burden, their carefully-calibrated expression levels would probably fluctuate into problematic regimes, an instance of retroactivity[10,12] with potentially high, hard-to-control consequences. It is unclear how diminished this effect could be when using a split-T7 system[30,32] for large circuits. These controller burden quantifications lead us to conclude that CRISPR-based expression exhibits an exciting expandability that makes it well-suited as a controller for large pathways and modulating endogenous expression (Chapters 3 and 4), as well as for more complicated control circuits (Chapter 2).

1.4: Burden-output relationship is linear in *E. coli*

As the burden is increased by increasing expression of output (by either increasing induction of the controller or strengthening the base CRISPRa promoter), the capacity monitor expression is proportionately decreased, as in the isocost lines previously reported[31]. We found the capacity versus output function to be linear in all cases with an *E. coli* host (Figure 1.5), suggesting a constant amount of burden contributed by each unit of output. Values of R^2 were generally > 0.95. This was true even when the fluorescent output was part of a larger operon, when the output used a stronger or weaker RBS, and when the controller was a T7 RNA polymerase. It is worth remembering that, especially for controllers like T7, it is possible to overinduce expression beyond this linear range of the output-burden relationship, resulting in "metabolic cliff" phenotypes that are difficult to measure with the capacity monitor

approach, but can result in obvious signs like severely reduced capacity, low or stagnant growth rate, and mutation of plasmids leading to non-expressing subpopulations. We suspected, and confirmed in Chapter 2, that this overinduction was possible with RNAP other than T7, but more difficult with *E. coli* RNAP due to the balance of transcriptional rate and transcriptional resources, for example, when using the host's native machinery. Even this measured linear range, however, sits in contrast to the burden response of mammalian cells[33], for example, which have more complicated resource allocation factors at play than the generally-suspected ribosome availability limitation of bacteria[12,16,18,34]. For the purposes of comparison, we take the slope of this line as a simple measure of the metabolic cost per unit of output, a metric that can be thought of as the efficiency of output driven by various controllers.

1.5: Comparison of output efficiency

Further, within each experiment, we compared the slope of a given experimental controller versus a typical reference controller, usually regular *E. coli* RNAP with lac repression (not the same "reference" from which capacity values are derived). When triplicate data were available, we determined significance of this slope comparison by generating a population of slopes representing all combinations of points at different inductions, for each controller, and performed a t-test on the two populations. In a few cases in Figures 1.5 and 2.2, triplicate data were not available and a 95% prediction interval is shaded around the reference function in lieu of significance testing. Overall, we observed that efficiency depended more on output characteristics like gene size and RBS strength than on controller identity, at least within the linear range.

Comparisons of synthetic controllers to *E. coli* RNAP expression resulted in surprising similarity between their burden function slopes. For comparisons between T7 RNAP and lac-repressed *E. coli* RNAP expression, we first verified that uninduced *E. coli* output—the capacity reference strain for the induced cultures—was similar to the off-target T7 control, to avoid confounding burden contributions from the output. For comparisons between CRISPRa control and constitutive *E. coli* RNAP expression, the true comparison being made is between unactivated and CRISPR-activated output

expression, with similar (more easily comparable) output levels achieved through a combination of off-target scRNA and stronger base promoter.

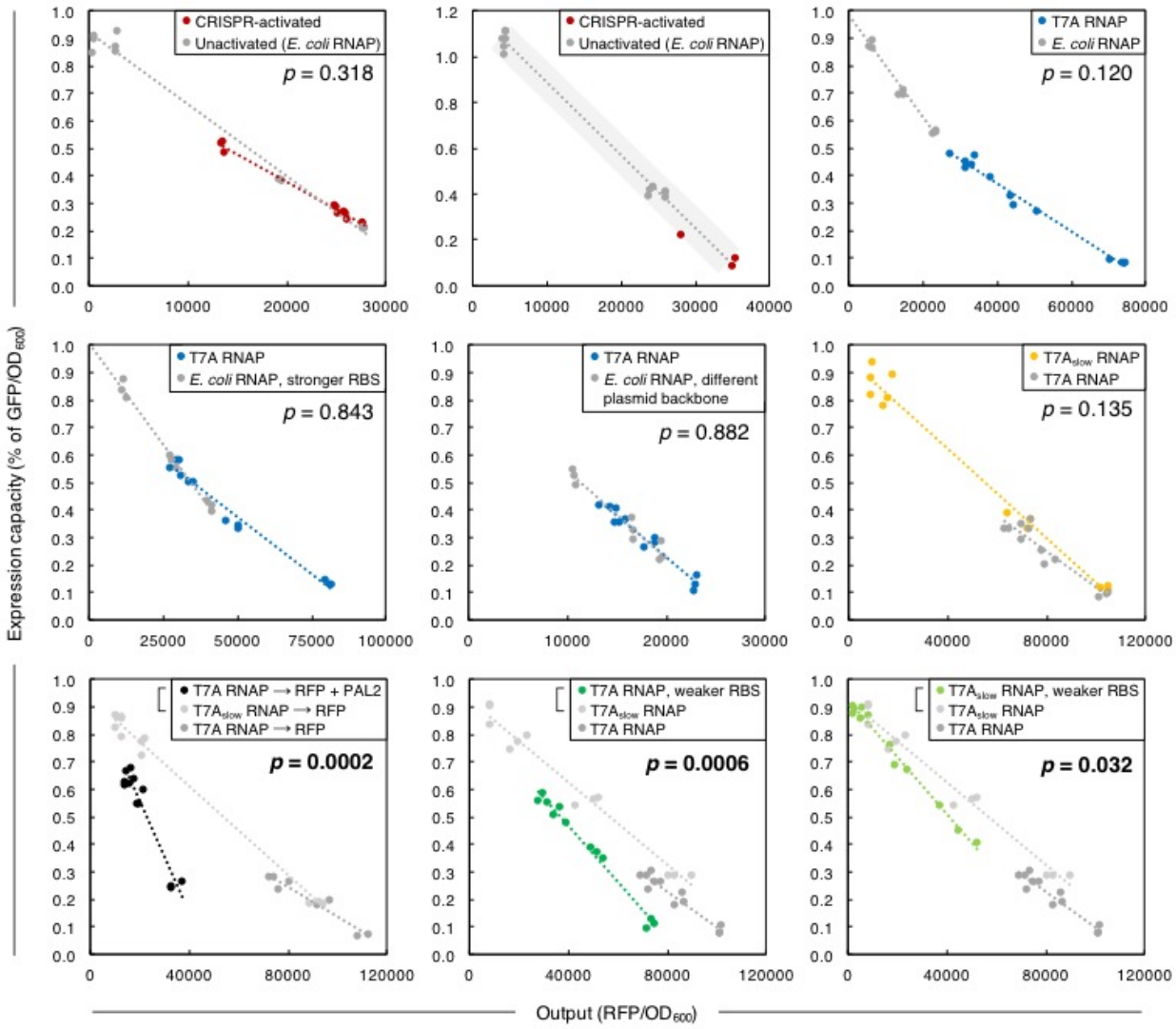


Figure 1.5. Comparisons of burden-per-output functions. Using these controllers with a given output, in *E. coli*, output levels and capacity reduction scale with each other, forming linear functions across a scale of output expression. Here, we compare slopes of these functions, as a measure of burden-per-output, or expression efficiency, between different controllers and outputs. Due to the different output promoter and UTR, CRISPRa control (panels 1 and 2, representing low- and medium-copy output, respectively) is not meant to be directly compared to T7 RNAP control, but both T7 and CRISPRa can be compared to their respective unactivated *E. coli*-controlled cognates. In those comparisons, the slope does not significantly differ when output is controlled by unactivated *E. coli* expression (regardless of the tested RBS strength or plasmid backbone), CRISPRa from a low-copy output plasmid, T7A RNAP, or T7A slow variant. (CRISPRa from a medium-copy output plasmid only produced one activated output level, with higher base promoter strengths proving unstable, but all replicates of that one point fall within a

95% prediction interval of the unactivated reference.) Only when a T7A output's RBS is altered does the slope significantly change, though the change is not drastic; or when an additional ORF (non-fluorescent PAL2, 717aa) is added to the output operon. Replicates are plotted individually.

For a given output—namely, mRFP1—with the same transcript characteristics, we observed no significant differences in output efficiency between the controllers tested, using this method of slope comparison. This similarity of efficiency suggests that the output burden is imposed mostly by resource limitation common to *E. coli* expression in general, rather than by any intrinsic controller characteristic. The limitation is often thought to be related to ribosome availability, but it is interesting to note that neither *E. coli*-expressed RBS strength was found to be significantly different than T7 control. While alternative resource limitations might include reagent pools like NTPs, or various degradation capacities, it is likely that different categories of resources are reasonably well-balanced when all control system components are derived from *E. coli*.

When using an exogenous control system, however, it is more likely to tax host resources in an unbalanced way, an effect that might result, for example, from T7 RNAP's high transcriptional elongation rate: eight-fold higher than *E. coli* transcription[35,36]. This factor might play a role in our finding that a weaker output RBS, when transcribed by T7 RNAP, resulted in lower output efficiency than a stronger RBS, and was more likely to suffer genetic instability. In the case of large amounts of transcript, as in T7 control, we might expect ribosome availability to be even more important than usual; however, we would also expect a stronger RBS to be more taxing on that availability. Because we observe the opposite in this case (and in another experimental architecture shown in Figure 1.S6), our work seems to suggest an alternative resource limitation—specifically in the case of T7 control—that is more transcriptionally oriented than ribosome availability. Possibilities include transcriptional resources, such as NTP availability, and transcription-related processes, such as mRNA degradation pathways. It is worth noting that we also observed this effect, but with less significance, when using a T7 RNAP variant with a slower elongation rate (only two-fold

higher than *E. coli* RNAP[37]), and that this variant's output efficiency was not significantly different than full-speed T7 RNAP when using the same output RBS.

The other observed change in output efficiency unsurprisingly resulted from changing the size and identity of the output protein: including the *Arabidopsis thaliana* enzyme PAL2 as a second ORF in an operon with mRFP1 significantly steepened the burden-per-output slope when controlled by T7 RNAP. Given that PAL2 is about three-fold larger than mRFP1 and is not fluorescent, it is expected that measured burden would increase and measured output would not. Attempting to account for this difference in output size might suggest that burden is still similar per mass of output, but this comparison ignores the difference in amino acid constitution between the different output proteins. In general, it should be expected that different outputs will lead to different levels of burden (Figure 1.S3), an important consideration when outputs simultaneously include, for example, all of the enzymes in a heterologous pathway. There has been some effort to develop predictive tools for output burden[38].

Specifically for T7 control, which despite its challenges remains popular for protein production and is well-suited to the single-node control circuits mentioned in Section 1.3, this work suggests a few principles of best practices when designing such systems. Namely, low levels of transcription should be implemented through low copy number of the T7 RNAP gene, or through low concentration of a chemical inducer, if such a low expression level can be kept stable despite retroactive burden effects. The RBS of the output gene should be strong, to encourage transcript stabilization through ribosome binding. The slow-elongation T7 variant should be used, as we observed drastically lower background activity. And a finding from Chapter 2: T7 lysozyme should be expressed rather strongly, possibly in a feedback architecture, again to help lower background activity.

Overall, this output efficiency study suggests that, for a given output gene, it is difficult to improve a heterologous expression system by focusing on expression efficiency per output, which seems relatively hard-coded by resource allocation. It would be more fruitful to focus such efforts on erecting boundaries to heterologous expression, to prevent overtaxation of those host resources. The linear range of

burden-per-output we have observed—in some cases reducing expression capacity to very low percentages—seems like a good target to avoid exceeding, as the ill-measured effects beyond this level of expression tend to have consequences rooted in genetic instability, which can end up reducing overall production of a hypothetical product, due to large non-expressing subpopulations arising within the culture. Put another way, output expression can either be limited by synthetic guardrails programmed into the controller design, or it can be limited by exhaustion of global expression capacity, which also affects many other processes in the host. The practitioner of heterologous expression, then, should consider the use of simple circuits to limit the extent of output activity—a focus of Chapter 2.

1.6: Conclusions

In this chapter, we established an adapted method for quantifying the metabolic burden that results from heterologous expression, and used it to assess both the "overhead cost" and efficiency of a few synthetic controllers and implementations of output genes.

Quantification of both output and capacity in these experiments has pointed to a seemingly universal (across different controllers) pattern of burden per output, but we and others[2,15] have noticed that some expression controllers are easier to drive out of a functional range than others, resulting in phenotypes reminiscent of the "metabolic cliff." For example, induction response is shifted toward lower inducer levels under T7 control than under *E. coli* control, as expected due to inherent amplification, but also has a smaller window of maximal-output inducer concentration before suffering metabolic cliff phenotypes. Similarly, the addition of CRISPR activation to regular E. coli expression can result in mutation of even medium-promoter-strength output genes (Figure 2.6). A possible mechanism for this cliff is limitation of one or many key resources of gene expression, with the classical culprit thought to be ribosome availability on top of the cliff—though our T7 work suggests that transcriptional resources could be more limiting than translational ones in that case—and energetic limitations like ATP exhaustion below the cliff. It can be difficult to pinpoint a single source of resource limitation, and it

is possible that many sources contribute, but it makes sense that the limitation is easier to reach using a synthetic controller like T7 or CRISPRa, rather than the native gene expression machinery that the host's resource pools evolved along with.

Assuming that these various synthetic controllers can be kept within this functional range of operation, there is evidently not much efficiency downside to selecting any given controller when building an expression circuit. That is, a controller should be freely selected to fit the purpose of the circuit, and due to our investigation of controller component costs, we suggest that controllers like T7 RNAP are better-suited to single-node expression (for example, for protein purification), and that CRISPRa is better-suited to large control circuits with an arbitrary or expandable number of nodes, the added control of which justifies the extra cost of the controller's protein components. Further, circuit design making use of synthetic controllers should focus on avoiding metabolic cliff phenotypes rather than pushing seemingly hard-wired efficiency limits.

With this in mind, for Chapter 2 we built controller architectures including simple feedback inhibition, with the aim of automatically reducing controller activity before its burden globally reduced all expression in the cell. We hoped therefore to spare the host's housekeeping activities from being reduced down to unhealthy levels: this in turn would hopefully expand the uses of synthetic controllers by forcing their activity levels away from the metabolic cliff.

1.7: Methods

Plasmid construction

Plasmids were constructed using either circular polymerase extension cloning (CPEC), Infusion (Takara Bio), or restriction cloning. All cloning was performed in *E. coli* DH10b chemically competent cells (Invitrogen), and DNA oligonucleotides were purchased from Integrated DNA technologies. Sanger sequencing to verify constructs was performed by Eurofins Genomics or Genewiz. DNA encoding T7 orthogonal RNA polymerases, as well as pBR322-origin-containing plasmid backbones and the lacIq gene, were gifts from Andrew Ellington (Addgene plasmids 63668, 63627, and 63628). Plasmid backbones containing p15A or ColE1 origins were taken from BglBrick plasmids[39], gifts from Jay Keasling (Addgene plasmids 35308, 35280, 35316, 35322, and 35279). DNA encoding CRISPRi/a components was obtained from multiple

plasmids from Jason Fontana[7,40]. RBS variants were selected from a previous report[41] and written into CPEC or Infusion primer overlaps. DNA encoding the enzyme PAL2 from *Arabidopsis thaliana* was as previously described[42]. DNA encoding T7 lysozyme K128Y was obtained from the pLysY plasmid contained in T7 Express cells (New England Biolabs). Point mutations in this gene, T7 RNAP genes, and promoters were written into CPEC or Infusion primer overlaps.

Experimental culture conditions

All experiments were conducted in *E. coli* MG1655 ΔrecA ΔendA cells[43], MG1655 ΔrecA ΔendA (DE3) cells, or BL21 cells with pLysY added before generating competent cells (New England Biolabs), as indicated in the text. Transformations were accomplished by heat-shock, and most experiments were conducted in two-plasmid systems. Generally, three transformed colonies were picked into replicate overnight cultures in LB (Miller) liquid media (Teknova) supplemented with appropriate antibiotics. After overnight growth, overnight cultures were diluted 1:100 into fresh LB media, with antibiotics, in a single culture tube or flask for each replicate.

For inducible cultures, particularly T7-controlled systems, after growth up to OD_{600} 0.3, as assessed using 200 uL culture in a flat-bottom, clear-bottom, 96-well plate (Corning, cat. no. 3631) in a Biotek Synergy HTX plate reader, cultures were split into 96-deep-well culture plate wells (Axygen, cat. no. P-DW-20-C-S) at 300 uL/well and induced by adding a small volume, usually 3 uL, of inducer. For constitutive cultures, particularly CRISPRa-controlled systems, overnight cultures were diluted 1:100 into fresh media containing antibiotics and inducers (aiming to match induction timing of constitutive controller and inducible capacity monitor) and grown in individual wells in the 96-deep-well plate. Density and fluorescence data were gathered in subsequent timepoints by reading 200 uL of culture, in a clear-bottom Corning plate, with OD_{600}, x485/m528 for sfGFP, and x540/m600 for mRFP1, each with a gain of 40. Raw fluorescence values at 40 gain were higher after a light source change on the plate reader, but values across bulbs were normalized when needed using maximal expression from the same expression systems.

Expression capacity measurement

Capacity monitor values were collected using a Biotek Synergy HTX plate reader, as above, and normalized by OD_{600} values. Baseline, uninduced values of Ptet-driven sfGFP/OD_{600} values were subtracted from experimental values, and expression capacity was calculated as the percent of capacity monitor expression remaining after comparing an experimental case to a reference case (e.g. from off-target to on-target), or sfGFP/OD_{600} (experimental) / sfGFP/OD_{600}

(reference). When these separate cases resulted from different levels of controller induction, within the same strain, the capacity calculation was performed within each replicate and averaged afterward, because replicate colonies were matched across induction levels. When the separate cases resulted from different strains, as in promoter deletions, for example, reference strain replicates were averaged before capacity calculation, because each replicate came from a unique colony.

Curve fitting, normalization, and statistical testing

Raw fluorescence values collected using the plate reader were always divided by OD_{600} values to normalize for cell density. Hill functions were fit using a previously described[44] Python script. Linear regressions in the capacity versus output scatter plots were fit using Microsoft Excel. For graphical purposes, averages of three biological replicates were used as points for a single linear regression, while the individual points, not their averages, were displayed on the scatter plots. For statistical purposes, a separate linear regression was calculated for all possible combinations of two points on the scatter plot (individual replicate points, not average points), excluding pairs within the same controller induction level. The resulting population of slopes was compared between strains or conditions using a two-tailed Student's t-test with unequal variance. In cases where multiple controller induction levels were not possible, for example medium-copy-output CRISPRa, or in cases where biological triplicates were not possible for some strains, the burden-per-output scatter plot includes a 95% prediction interval shaded around the reference strain (grey).

Flow cytometry

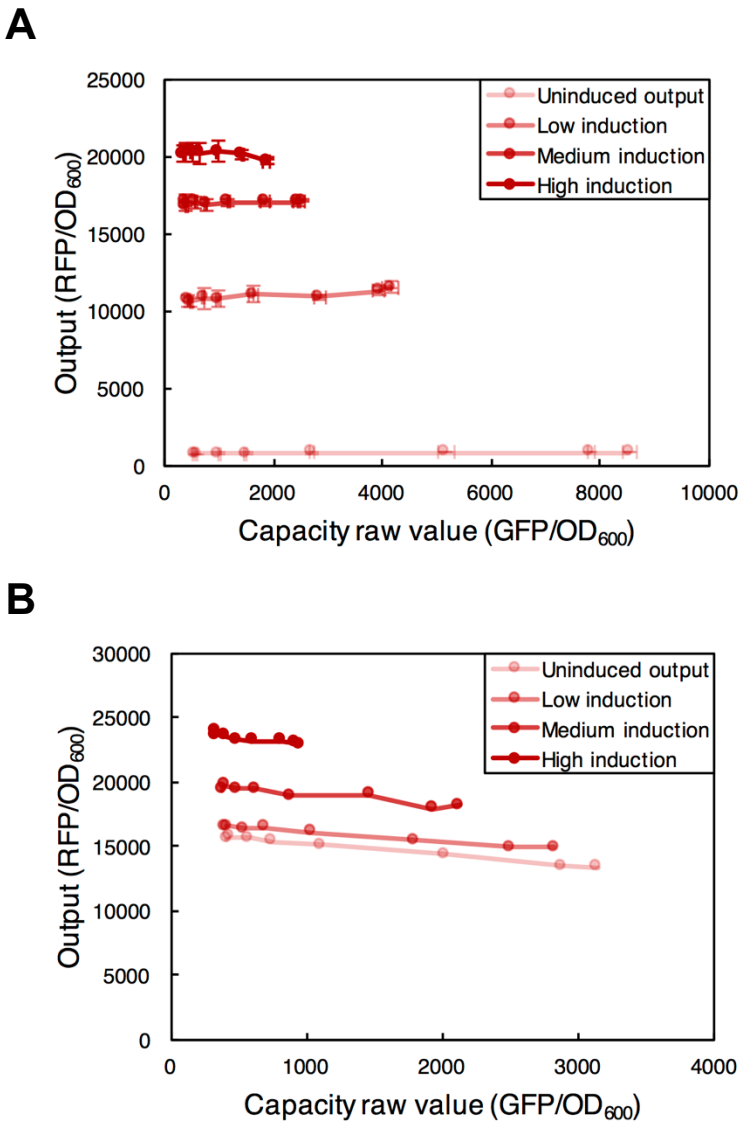
Overnight cultures were diluted 1:100 in PBS (Hyclone) and measured using a Miltenyi MACSQuant YVB flow cytometer with low flow rate; 10000 events; channels including FSC at log2 630 V, SSC as main trigger at log2 450V, Y2 at log4 350V (300V for CRISPRa systems), and B1 at log4 350V; and a trigger threshold of 15. Gating was accomplished by eliminating fringes on a SSC-A vs. FSC-A plot, then gating for linearity on a SSC-A vs. SSC-H plot. Analysis included mostly visual observation of histograms aiming to identify bimodal distributions indicating heterogeneity, but median values were sometimes used in cases of unimodal distribution.

Gel electrophoresis

All DNA was run on 1% agarose gels in TBE buffer, and cloning fragments were usually gel-purified after observation of SYBRsafe dye (Invitrogen) using a non-mutagenic wavelength. SDS-denatured proteins and RNA were run on polyacrylamide gels of various

percentage, and SDS-PAGE was run using premade gels (Bio-Rad): 7.5% for T7 RNAP and 12-20% for sfGFP.

1.8: Supplementary Figures



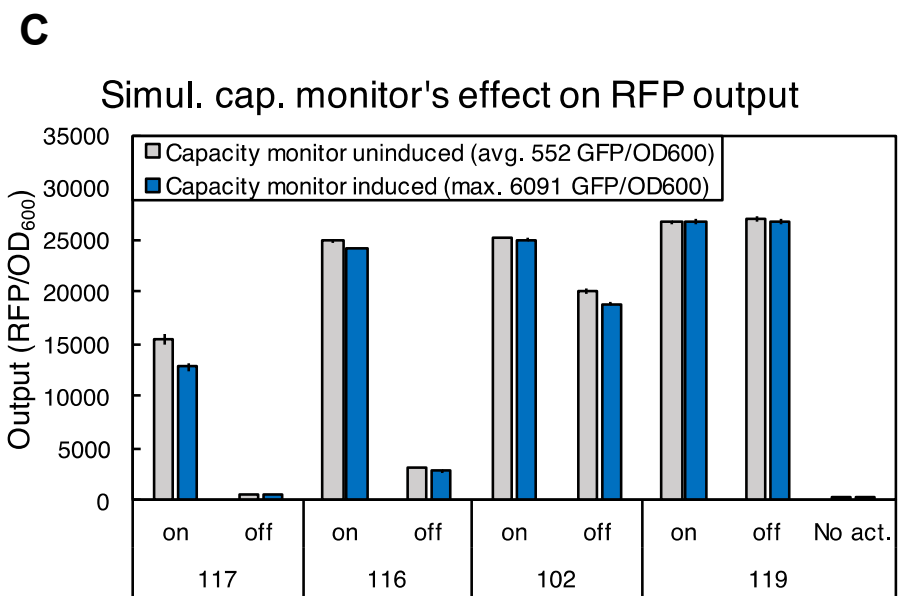


Figure 1.S1. Output expression is unaffected by capacity monitor expression level, regardless of controller. **A)** *E. coli* controller with lac induction. Within different output induction levels (the separate lines), different levels of capacity monitor induction (measured as the *x*-axis values) do not affect measured output levels (*y*-axis values), resulting in lines with roughly zero slope. Therefore, though output expression affects capacity monitor expression (Figure 1.5, also the shrinking of line length at higher output inductions in this figure), capacity monitor expression does not affect output expression, probably due to relative expression levels. This is true even if the output and capacity monitor proteins are swapped. **B)** T7 controller with lac induction. Lines similar to panel (A) are observed, with slopes close to zero. **C)** CRISPRa controller, which is effectively a low-output *E. coli* controller with engineered activation. Because this system's output is constitutive and not inducible, we did not explore an induction range, but nonetheless saw no output difference depending on binary induction of the capacity monitor (induced level was 80 nM aTc). The difference between paired grey and blue bars is the effect of capacity monitor induction. The (J23)117, 116, 102, and 119 refer to the output CRISPRa promoter's base promoter strength, as in Figure 2.6. Error bars represent standard deviation of *n*=3 biological replicates.

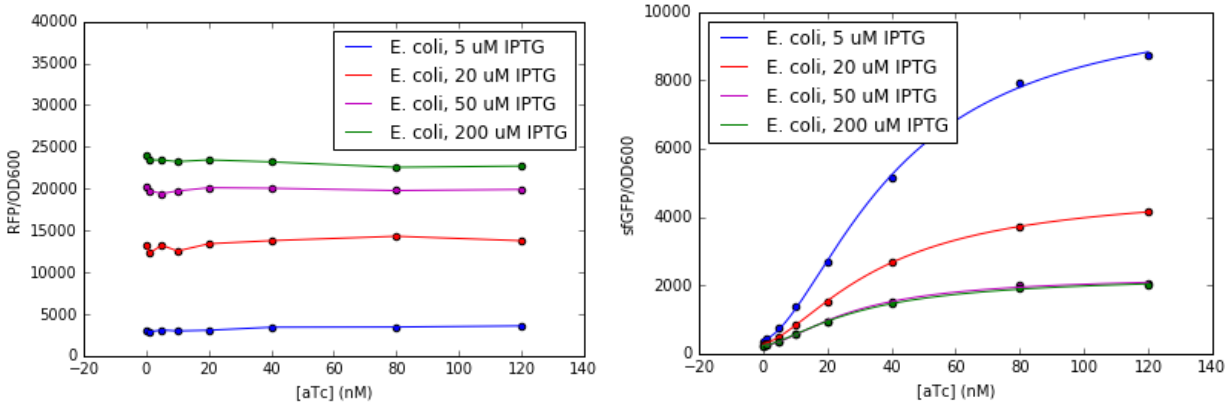


Figure 1.S2. Induction responses of tet-controlled capacity monitor seen at right. Different colors represent different output levels, the burden of which shrinks the capacity monitor's tet response vertically (right), but does not affect curve shape or sensitivity, nor RFP output (left). At any aTc concentration above 10 nM, the expression capacity normalization gives the same result; thus any of those aTc concentrations may be used to induce the capacity monitor. We primarily used 80 nM.

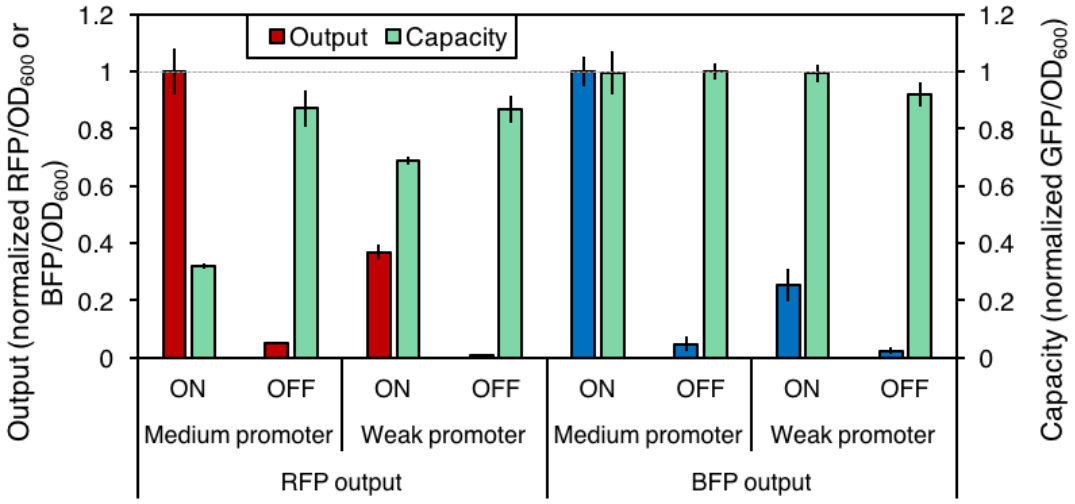


Figure 1.S3. mRFP1 output has a stronger effect on capacity than mTag-BFP. Reduction of capacity, measured by sfGFP, is drastic for high levels of RFP output, but barely measured when the output is mTag-BFP. Both outputs were expressed in the same way—CRISPRa of the J3 promoter, with the indicated base promoter strength (medium=J23116, weak=J23117), from a pSC101** plasmid—and their RBSs and ORFs were matched. It is possible that spectral overlap between sfGFP and mTag-BFP emissions somewhat inflate ON-state capacity measurements in the case of BFP output. Due to filterset limitations, BFP output levels were

determined by flow cytometry, and median values were autofluorescence-subtracted and normalized to the maximum value. Error bars represent standard deviation of *n*=3 biological replicates.

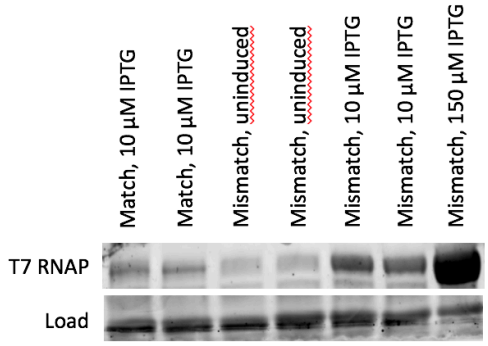
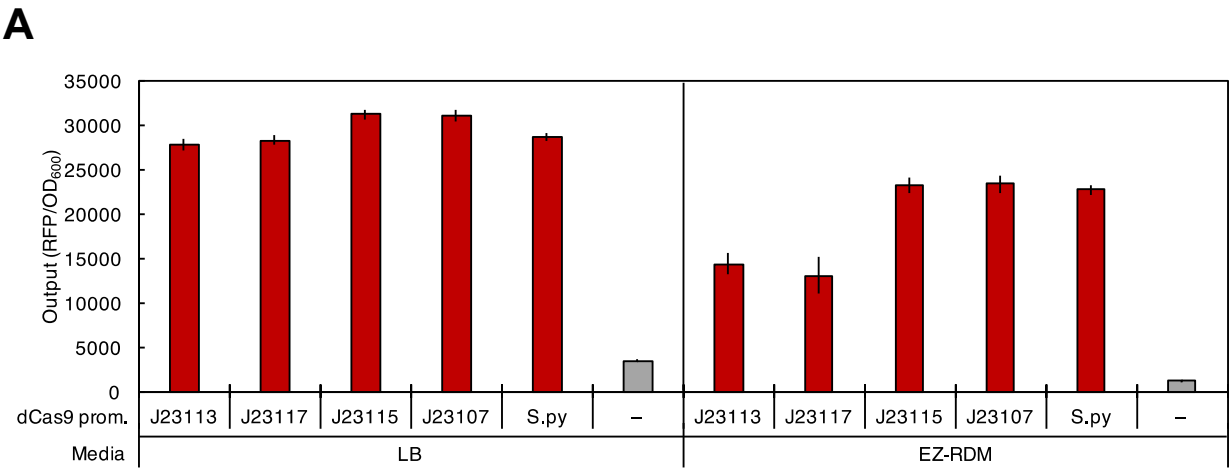


Figure 1.S4. SDS-PAGE showing T7 RNAP levels. T7 band size is indicated by the overexpressed control (right). Matched-case bands correspond to the high induction level (24000 RFP/OD_{600}) in Figure 1.S1B. Mismatched-case induction results in significant T7 RNAP expression, indicating that the off-target promoter is the reason for the lack of output; indeed, the mismatched case produces more T7 RNAP than the matched case, indicating that global expression reduction affects the T7 RNAP level as well, an instance of retroactivity within the control circuit. The "load" band is not a specific load normalization, but an endogenous band, also susceptible to expression capacity reduction.



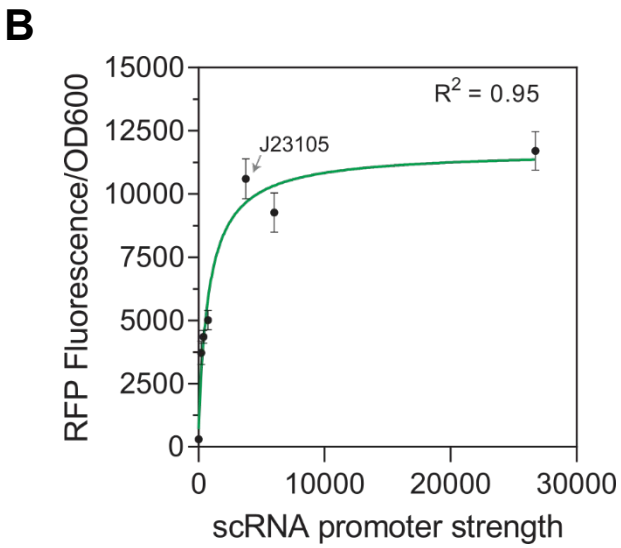
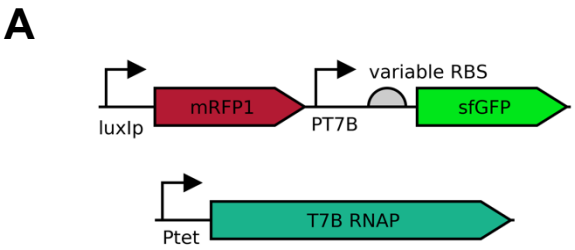


Figure 1.S5. CRISPRa components expressed in excess. **A)** Titration of dCas9 expression levels from Anderson series promoters still allows CRISPRa function at all tested promoter strengths. In LB, output is basically unchanged across the titration, indicating that even lower expression levels might be possible. In EZ-RDM, however, the more stringent growth conditions result in a decrease in output below the J23115 expression level. For proper CRISPRa function in most conditions, then, dCas9 expression level should be at least at J23115 level. It is unclear how the *Sp*.pCas9 promoter strength compares to that of the Anderson series. We were not able to detect a difference in expression capacity between these strains. **B)** Titration of scRNA expression levels, also using Anderson series promoters, allows similar output levels even at ~7-fold lower expression (J23105) than the level tested in Figure 1.3 (J23119). Error bars represent standard deviation of *n*=3 biological replicates.



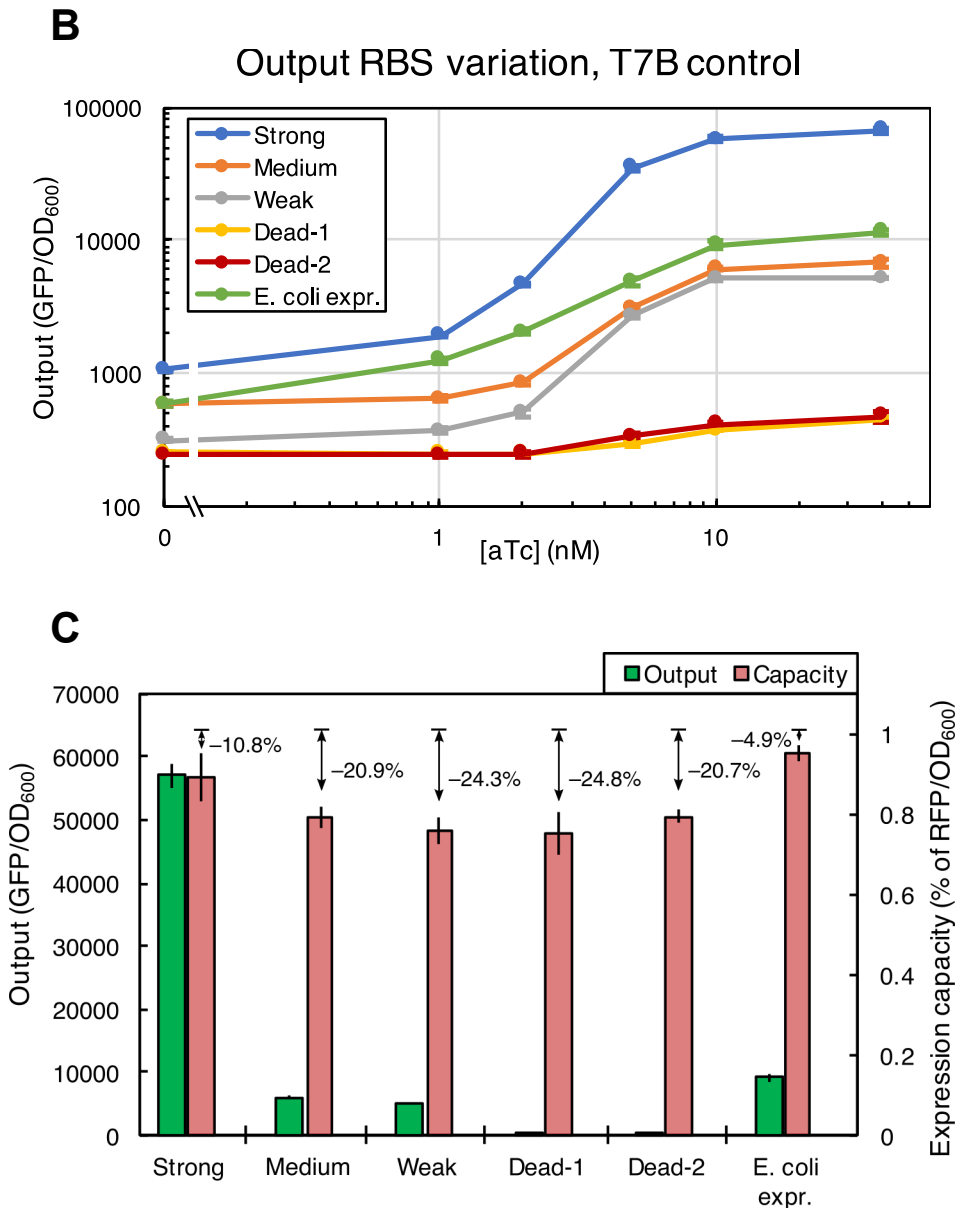


Figure 1.S6. Weaker output RBS increases burden of T7 RNAP control. **A)** With the T7B RNAP controlling sfGFP output, and mRFP1 acting as capacity monitor, we built a small library of RBS strengths controlling sfGFP translation. **B)** Effects of those RBS strengths on output responses to tet induction are as expected, though the weak RBS is stronger than predicted and the dead RBSs both retain some minimal activity. *E. coli* expression strain simply expresses sfGFP in place of T7B RNAP, with the same strong RBS. **C)** Capacity measurements resulting from maximal induction of these strains show that, even when not normalized to output levels, weaker output RBSs impose more burden on the host than stronger RBSs. This is presumably

due to the accumulation of excess transcript that is unbound by ribosomes, suggesting that it is an effect peculiar to T7 RNAP transcription, with its high elongation rate. This is counter to the expectation that ribosome availability is the primary effector of expression burden in *E. coli*. One potential limitation of this study is the possibility that nonfluorescent sfGFP fragments are translated from the weak- or dead-RBS versions, due to ribosomes acting on RBS-like sequences further down the transcript—a possibility that would likely affect capacity measurement but not output measurement. Error bars represent standard deviation of n=3 biological replicates.

Chapter 2: Circuit control of heterologous expression

Abstract

As we look toward ever-more-complex heterologous biosynthetic pathways, large size and complex features of expression control circuits could be necessary to boost metabolic flux through the pathway to economically feasible levels. Intriguing directions for this circuit expansion could include inducer-free delays of pathway expression, multi-layered circuits incorporating logical operators, or metabolite sensing for dynamic control of pathway enzymes. Chapter 1 suggested an inexpensive expandability of CRISPR-based control circuits, and also that limits on output magnitude imposed by the circuit architecture can make the strain more stable, potentially boosting production or circuit performance. As control circuits and pathways get more complex, this use of circuit architecture to enforce genetic stability could become very important to preserve production titer by limiting heterologous cultures including “cheater” non-producing cells. In this chapter, we build a set of feedback and feedforward control circuits, finding that output efficiency is generally unchanged compared to open-circuit designs, and that CRISPR-based negative regulators behave more predictably than the protein-based negative regulators we used. Further, these closed circuits should impose limits on output expression levels that have important implications for genetic stability and, in turn, production titer or circuit performance.

2.1: Introduction

In Chapter 1, we emphasized the importance of avoiding so-called metabolic cliff phenotypes and their oft-associated output bimodality, ribosome loss, and genetic instability, problems connected with heterogeneity of the culture population, itself leading dramatically to low product titer or circuit performance[20–26]. For designing robust expression control circuits, minimizing this risk of instability seems a more important focus than the efficiency of expression, or slope of the burden-per-output function, which is hard to disconnect from the host's physiology (Figure 1.5). In the controller-output-monitor system we developed, intentionally burdensome constructs are often carried by medium-copy plasmids, making genetic instability of those plasmids a common manifestation of the problematic burden of controller overexpression, beyond the cliff. The instability usually takes the form of recombination between sequences at different sites on the plasmid, resulting in rearrangement that prevents output expression; or sometimes direct novel mutation of an output promoter into nonfunctionality. Regardless of the mechanism, the most common mutations we see focus on breaking the most burdensome part of the circuit—the output—and such mutants gain enough of a selective advantage to grow to relevant proportions of the endpoint cultures.

When observed in this system, an initial hallmark of this instability is wide-ranging replicates of bulk culture measurements, as of fluorescent protein output on a plate reader. This occurs because the random timing of instability events results in different start times of nonproducer "inoculation" across the replicates, in turn resulting in different proportions of nonproducers in the endpoint cultures. Subsequent analysis of these cultures by flow cytometry reveals the distribution of output phenotypes within the culture—often forming a bimodal distribution with high numbers of low-output, non-producing cells and variable numbers of high-output, intact cells, depending on the scale of the burden.

Taking these observations as our "primary endpoint" to reduce through controller design, we aimed to incorporate autoregulatory motifs[45] into control circuits, eventually relying on the inexpensive CRISPR circuit expandability we hypothesized in Chapter 1,

as well as precedent for CRISPR-based logic circuits[44,46]. The goal was to impose an artificial limit on output level, inherent to the controller, that would hold output expression and burden back above the edge of the metabolic cliff. At the same time, we aimed to smooth stochastic fluctuations amplified by retroactivity that could push individual cells over the edge[10,12,47]. Simple motifs of autoregulatory feedback[21,47–49] and feedforward[33,50–52] would be well-suited to provide limits to expression level, and the ability to tune the strength of the negative interaction would allow tuning of output levels or other elements under the control of the same circuit.

In this chapter, we build these negative regulatory motifs into circuits controlled by CRISPRa and T7 RNAP with a measurable fluorescent protein output. The negative effectors are initially protein-based—a viral lysozyme and a transcriptional repressor—and are later standardized as CRISPRi gRNAs, for the sake of consistency, predictability, and tunability. We assess these circuits' abilities to limit output and prevent instability, and we include the same capacity monitor strategy from Chapter 1, to observe the effect on expression efficiency of including these motifs in the control circuit. Building this autoregulation into the control of heterologous pathways or circuits, we reason, is an important safeguard of the housekeeping activities the host cell needs for proper growth and production, in cases where potential instability renders it necessary.

2.2: Protein-based feedback control

A logical step for achieving feedback inhibition of one of Chapter 1's synthetic controllers was to include in an output operon the K128Y variant of T7 lysozyme[53–55], which is expressed constitutively in commercial T7 expression strains[56]. This variant, its lysozyme activity abolished by the mutation, diminishes T7 RNAP activity by binding and sequestering the RNAP protein, with the goal of reducing background T7 expression. Thinking that the burden of T7 RNAP activity when expressed from a medium-copy plasmid was enough to decrease weakly constitutive lysozyme expression to irrelevant levels, we instead included the lysozyme gene as part of an output operon along with mRFP1 (Figure 2.1A). Finding that including an extra protein

as part of an output operon reduced output, we included a control strain with a lysozyme bearing three additional mutations (H17R, H36R, Y46D), reported to abolish sequestration activity as well[53]. Uninduced background mRFP1 levels, which were high without feedback due to amplification of promoter leak, decreased 39% just by including that extra protein in the operon, and from that level decreased another 37% upon addition of feedback by the sequestration activity of the K128Y variant (Figure 2.1B, left). This probably indicates that the promoter leak was indeed overwhelming constitutive lysozyme expression, resulting in even more T7 activity and burden, and therefore that feedback inhibition can be useful for keeping T7 activity within reasonable levels. For this reason, in instances of simple, single-node T7 control of a desired output, problematic expression can probably be neatly dealt with by inclusion of an additional genetic element—either as an operon or under its own T7 promoter—consisting of a T7-expressed T7 lysozyme.

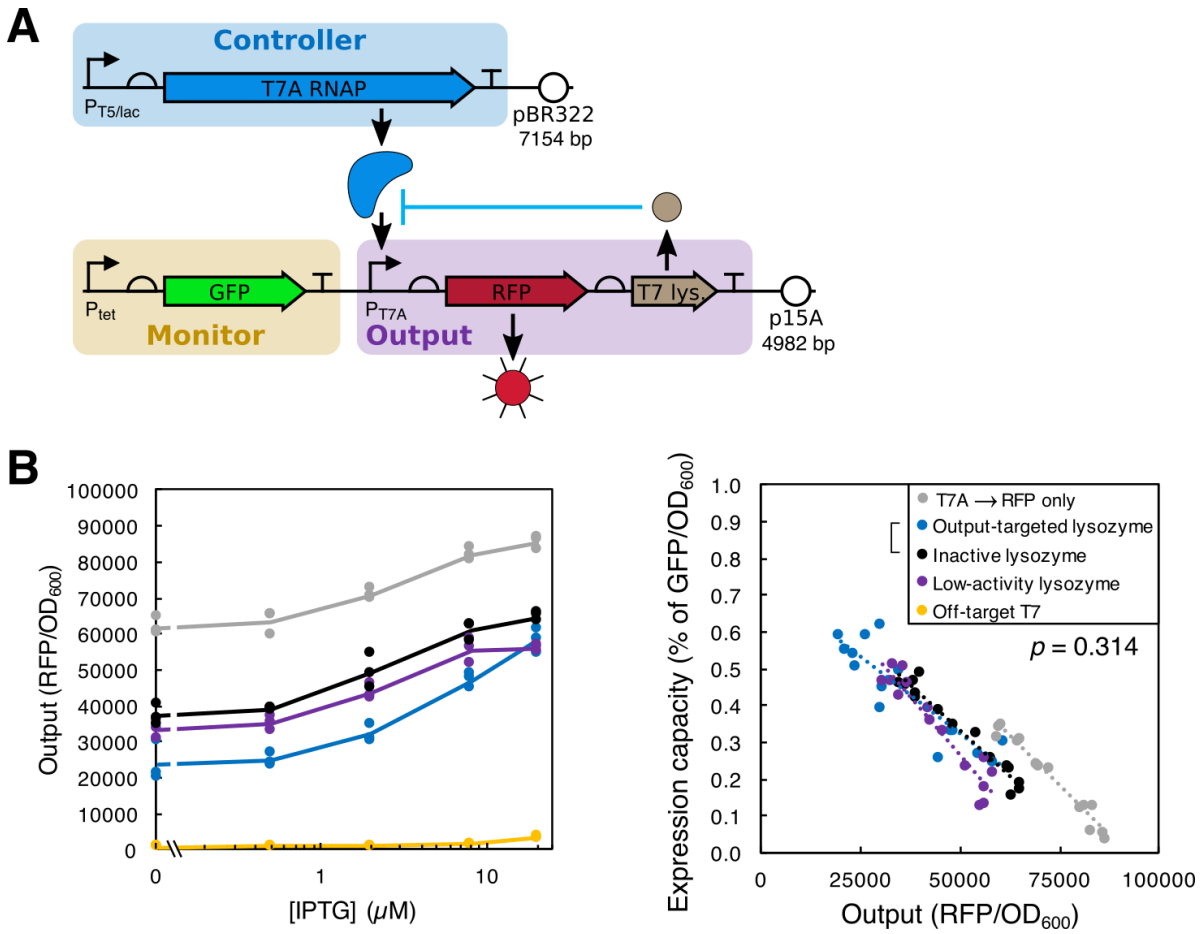


Figure 2.1. T7 lysozyme feedback on T7 RNAP control. **A)** Medium-copy T7A RNAP expresses an output operon containing mRFP1 and T7 lysozyme K128Y—which binds to and sequesters T7 RNAP—resulting in a protein-based feedback architecture. A reference strain with nonfunctional lysozyme (H17R, H36R, Y46D, K128Y) accounts for the burden required to produce the extra output protein. **B)** Output levels are decreased by adding even nonfunctional lysozyme to the output operon (left, black versus grey), probably due to the longer transcript's effect on steady-state transcript levels. Beyond this effect, additional output decrease can be attributed to the feedback effect (blue versus black). Despite the decrease in output, burden-per-output functions (used as in Figure 1.5) do not significantly differ between strains, suggesting similar efficiency between feedback-controlled and open-circuit architectures (right). Interestingly, expansion of the output operon to include any lysozyme variant did not decrease efficiency to nearly the same degree as in Figure 1.5, lower left. Replicates are plotted individually.

The lysozyme strategy of feedback, however, suffers from a lack of portability to different controllers because it is inherently paired with a T7 RNAP controller, and is therefore probably worth avoiding in cases of larger control circuits. As an alternative protein-based controller, in the style of the TetR system previously reported[57], we built the same architecture of feedback inhibition using an additional *lacI* gene in an output operon (Figure 2.2A). In this system, some portability could be achieved by including a lacO site at desired control points, in addition to the site already present in the T5/lacO promoter driving T7 RNAP. In both architectures of this LacI feedback—targeted to the T7 only and targeted to both T7 and output—we observed a high risk of genetic instability, affecting two out of three replicates in Figure 2.2B and indicating that burden to the system was greater than indicated even by the severely increased slopes of the burden-per-output functions. Apparently, the identity of the feedback element in these systems effects a large difference in the system's behavior, and among protein components, that identity can vary quite a bit, resulting in a wide range of circuit behaviors from unusable (LacI) to quite helpful (lysozyme), even within this small study. A feedback element that would be more standardized and portable could take the form of an RNA element, a possibility explored in Section 2.3 and beyond.

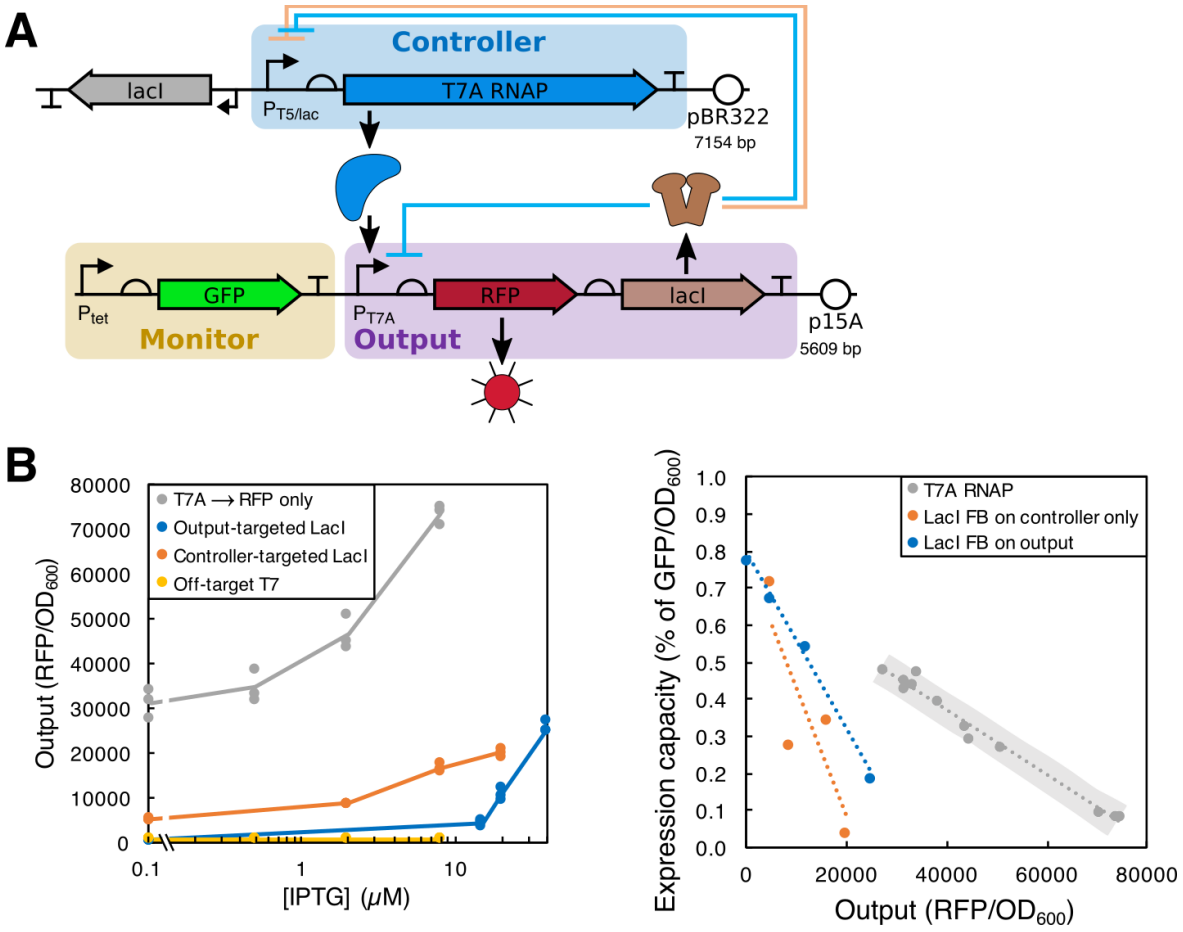


Figure 2.2. LacI feedback on T7 RNAP control. **A)** Medium-copy T7A expresses an output operon containing mRFP1 and LacI, the extra dose of which inhibits expression of additional T7A (orange) and sometimes the output operon as well (blue), depending on the inclusion of a lac operator site there. The result is a similar protein-based feedback architecture as in Figure 2.1. **B)** Output levels are severely decreased by adding LacI expression and function to the system (left), with the output-targeted version decreasing the sensitivity of the induction response as well. In contrast to the lysozyme feedback system, LacI feedback drastically decreases burden-per-output efficiency (right), highlighting the variable behavior of protein-based feedback effectors. Additionally, fluorescence data indicated a strong effect of genetic instability in the LacI feedback strains, a behavior not well-captured by the burden-per-output function slopes, but nonetheless an important indicator of poor circuit function. Grey shading indicates a 95% prediction interval around the open-circuit reference strain. Replicates are plotted individually.

Even if T7 lysozyme is a more viable feedback control strategy than LacI, its sequestration characteristics are determined by its inherent biochemistry, making it hard to tune. We considered tuning of the feedback strength to be important for allowing significant amounts of inducible output expression while still preventing overly-burdensome amounts of T7 activity: however, this system's feedback strength is

effectively only tunable through the RBS strength of the lysozyme gene. (Additional mutations can probably modulate its sequestration activity, but perhaps the easiest tuning is at the RBS at its position in the output operon.) This is a cumbersome cloning task that has the risk of unintended burden consequences (due to changes in output translation), so we sought a more programmable—and less burdensome—mode of feedback by including a CRISPRi gRNA as part of the output. Such a strategy, which uses an RNA element as the effector of feedback control, avoids the pitfall of protein-based feedback controllers: the inclusion of another protein in an output operon, which we saw in Chapter 1 could be very costly as the output of a synthetic controller. The relative inexpense of a gRNA would possibly allow for more robust circuit behavior, even if expression of dCas9 is also required.

2.3: Construction and characterization of mRNA–ribozyme–gRNA devices

To devise a system combining T7 RNAP expression with CRISPRi feedback inhibition, we relied on genomically-integrated T7 to make room in a two-plasmid system for dCas9 expression, rather than relying on a potentially unstable three-plasmid system. Not only does this CRISPRi approach standardize the feedback effector from a burden perspective, CRISPRi's programmability of repression targeting also allowed for different feedback architectures, with the feedback targeting either expression of the T7 RNAP itself or expression of its output (sfGFP in this case). Aiming to use CRISPRi to achieve these architectures, we devised a system that links gRNA levels to T7 activity by transcribing the gRNA along with the output mRNA, with the two functional parts of that single RNA cleaved posttranscriptionally by an ASBV ribozyme[44,58] (Figure 2.3). Because of the sequence of mRNA-ribozyme-gRNA, we refer to this method of gRNA expression as "MRG."

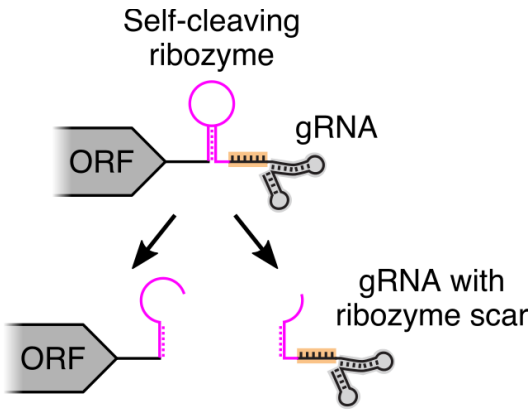


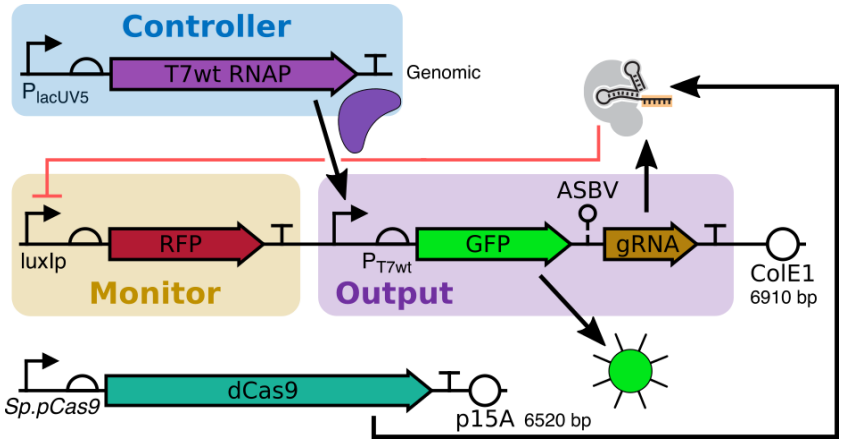
Figure 2.3. Function of an mRNA–ribozyme–gRNA device. A self-cleaving ribozyme frees a gRNA from the rest of an output mRNA, coupling gRNA levels and output levels. ASBV was chosen as the ribozyme due to its cleavage site being toward the 3' end of its sequence, minimizing the size of the scar (17 nt) that remains on the 5' end of the gRNA after cleavage.

After cleavage, the mRNA undergoes translation while the gRNA binds constitutively-expressed dCas9, forming a complex that inhibits the transcription of additional T7 RNAP mRNA or output sfGFP. This inhibition occurs after a delay accounting for gRNA expression, folding, binding the dCas9, and interrogating the DNA, theoretically resulting in an initial burst of expression before inhibition. We hypothesized that this burst of expression would result in more efficient expression control (less burden per output) if the inhibition was targeted at the T7 RNAP, rather than the sfGFP, because whatever T7 RNAP that was expressed during the burst could continue transcribing sfGFP until its degradation. It is worth noting, however, the potential for this system to behave unexpectedly compared to previous reports, possibly because of the fast elongation rate of T7 RNAP, but more likely because the ribozyme cleavage leaves a 17 nt scar at the 5' end of the gRNA, adjacent to the spacer sequence. Either of these differences could affect gRNA folding, therefore dCas9 binding or DNA binding, and therefore inhibition function.

To assess the function of this CRISPRi implementation before using it for the feedback architectures, we initially targeted the gRNA to the mRFP1 gene (RR2 spacer sequence), which is independent from the control circuit and was intended for later use as the capacity monitor for this system (Figure 2.4). As implemented here under lux control, mRFP1 had significant uninduced leak, but was inducible to medium and high

levels with addition of HSL. At low levels of mRFP1, repression levels—relative to mRFP1 expression with an off-target, MRG-expressed gRNA—from the T7-controlled MRG approach those of constitutively-expressed gRNA (expressed by *E. coli* RNAP, with no scar) (~94% versus ~96%). In both gRNA expression formats, inducing higher levels of mRFP1 by adding HSL can overwhelm the CRISPRi, reducing repression levels, but induction of additional gRNA by adding T7 inducer (IPTG) in the MRG case rescues most of the repression. This indicates that gRNA level is a limiting factor for repression especially in the MRG format, which may have a limited proportion of the gRNA ensemble folding properly, due to the ribozyme scar. This in turn suggests that CRISPRi-MRG would be responsive to varying input signals in a circuit containing feedback inhibition, and thus would be a useful guardrail against overexpression.

A



B

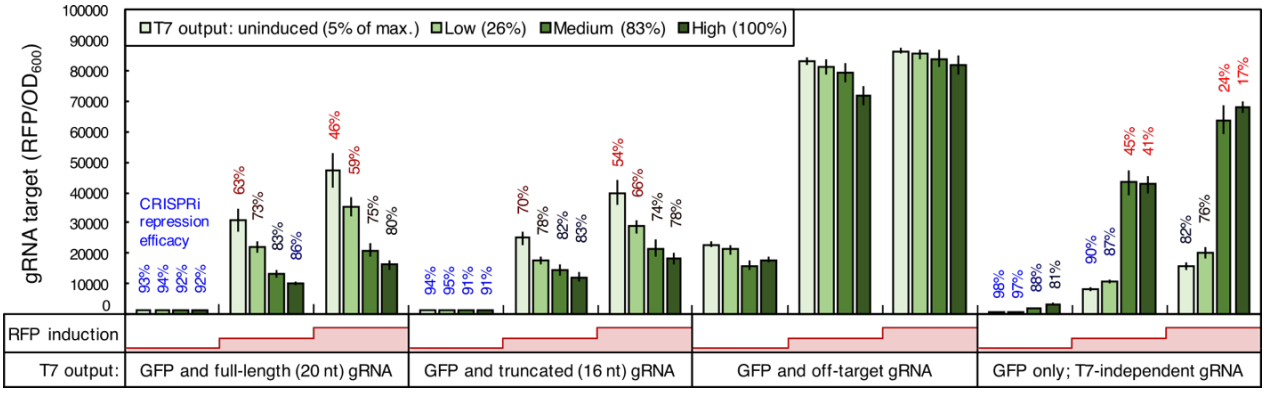


Figure 2.4. T7-controlled, mRFP1-targeted MRG repression. **A)** As an investigation of MRG functionality, we set up a system in which genomic T7 RNAP controls an sfGFP output with an MRG gRNA targeted to mRFP1. The leaky luxIp expression of mRFP1 is treated as low constitutive expression, but can be induced to higher levels by addition of HSL. **B)** Levels of the CRISPRi target mRFP1 are shown on the *y*-axis, while T7-controlled sfGFP expression is reflected in the coloring of the bars. Darker bars means more measured sfGFP, which means more gRNA when the two outputs are linked by MRG. The first three panels include this MRG link, while the fourth shows a T7 output of only sfGFP, with gRNA separately and constitutively expressed. The first two panels are mRFP1-targeted gRNAs of different spacer lengths, while the third panel is an off-target gRNA. From these unrepressed levels, repression efficacy is calculated as a percentage and shown above every on-target bar. Within each panel there are three HSL-induced levels of the CRISPRi target, mRFP1. While high levels of this target can overwhelm the CRISPRi repression, induction of additional gRNA expression through the T7-MRG system (first two panels) recovers some of the CRISPRi efficacy. On the other hand, when gRNA is uncoupled from T7 activity (fourth panel), induction of additional T7 expression (greener bars) reduces CRISPRi efficacy, indicating a burden-driven decrease in gRNA expression capacity. In general, if this repression is targeted to any burdensome output, its repression efficacy could spiral downward as each decrease in efficacy is accompanied by more output and thus an increase in burden—illustrating the importance of coupling negative regulators to burdensome outputs in a closed circuit design. gRNA spacer truncations (second panel) are discussed in Section 2.4. Error bars represent standard deviation of n=3 biological replicates.

One of the most noteworthy findings of this MRG characterization was that, while inducing additional T7 activity (indicated by sfGFP levels, but resulting also in additional gRNA, at least in the MRG strains) boosts CRISPRi efficacy in MRG architectures, it does just the opposite in the case where gRNA is constitutively expressed. The reason is that constitutive gRNA expression is sensitive to the added burden of expressing T7-controlled sfGFP (which does not include an MRG in that control case), and that burden reduces the expression of nominally-independent gRNA to less-effective levels (down to 17% in one case). The MRG format has no such sensitivity, since gRNA levels are paired with T7 activity, the main source of burden in the system. This illustrates the importance of interconnection of circuit components and controllers, so that levels of each are correlated to each other. Then, even if burden diminishes the overall magnitude of circuit expression, individual components are kept stoichiometrically linked, theoretically broadening the circuit's robustness to various metabolic burden. This is especially important for any negative regulators that are included in a control circuit, because diminishing negative regulation through burden is likely to lead to

additional burden, completing a positive feedback effect that can amplify unintended circuit behavior.

Finally, we used the mRFP1 capacity monitor to assess the burden that an off-target gRNA adds to the T7 output (sfGFP-MRG) compared to an output of sfGFP alone. We noticed two clear findings: inclusion of MRG reduces output by between 22 and 37% (or 48% at uninduced background levels) (Figure 2.S3), presumably due to differences in transcript stability; and, per unit of output, MRG expression caused significantly more burden than a non-MRG transcript (Figure 2.S1). This added burden could be due to the difference in transcript 3' end—that is, whether it was generated by transcriptional termination or by ribozyme cleavage—or simply by inclusion of extra transcription performed by T7 RNAP, with its high elongation rate.

2.4: Hybrid T7/CRISPRi feedback controller

After verification of mRFP1 repression using MRG, we targeted gRNAs either to T7 RNAP for controller feedback, sfGFP for output feedback, or a genomic site (T2) for an off-target control with no feedback (Figure 2.5A). Not only should both feedback architectures differ in feedback timing, but they could differ in strength due to the different target sequences, though both are located in their targets' ORFs. It is also worth noting that the controller-targeted feedback circuit's architecture corresponds to the LacI feedback system above, while that of the output-targeted feedback circuit corresponds to the lysozyme system. Like the protein-mediated feedback, output levels were reduced by up to 75% by addition of feedback (Figure 2.5B, left), with a concordant reduction in burden as measured by the mRFP1 capacity monitor. Indeed, burden function slope comparisons show no significant difference between the open architecture and either feedback case (Figure 2.5B, right). Interestingly, output-targeted feedback was not significantly less efficient than the controller-targeted case, suggesting that repressing the output without addressing the rest of the circuit does not waste significant expression resources. Though contrary to our initial hypothesis, this is not entirely unexpected, given the low cost of expressing a single T7 controller measured in Section 1.3.

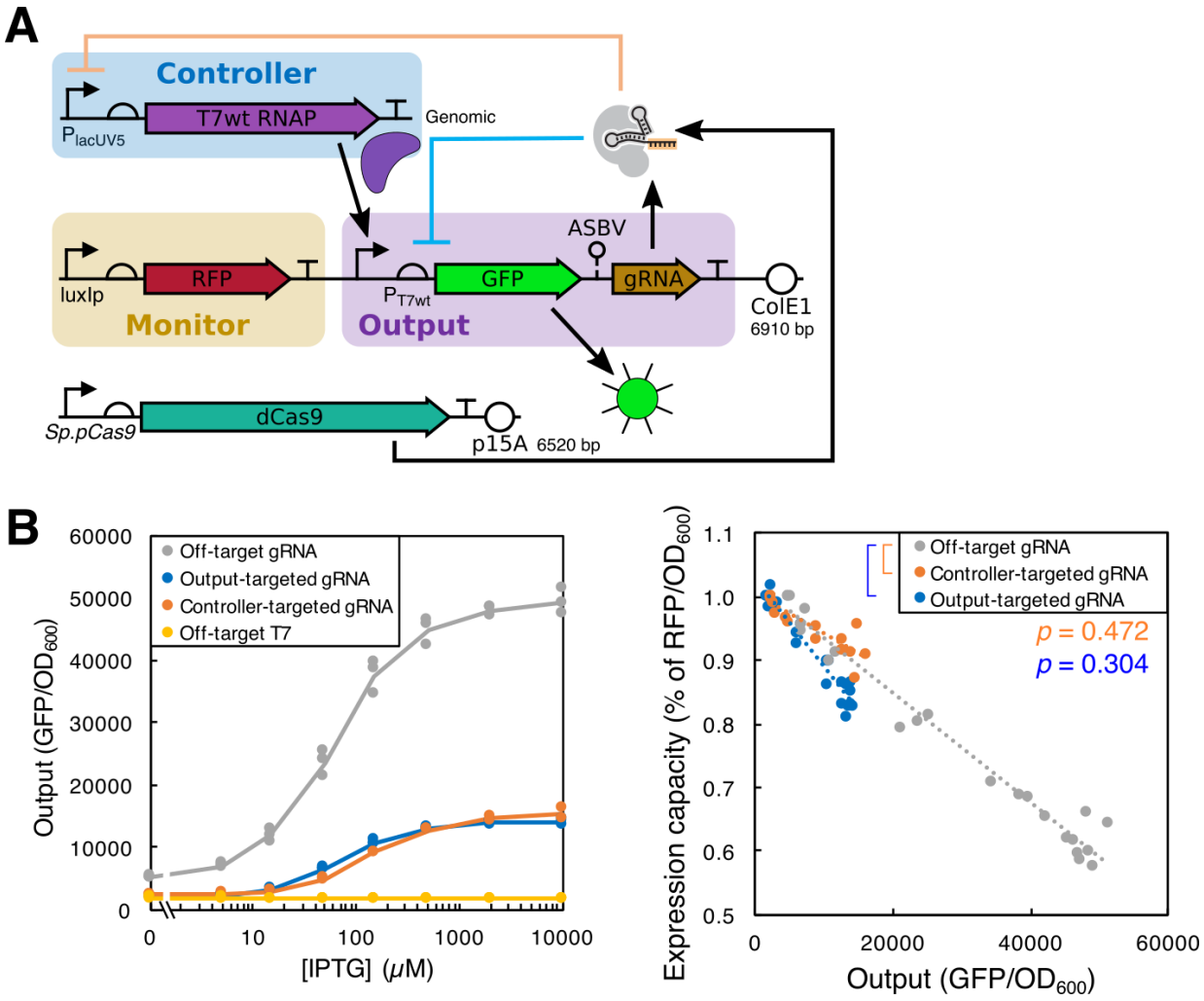


Figure 2.5. CRISPRi feedback on T7 RNAP control. **A)** To complete the T7-MRG feedback architecture, we changed the gRNA spacer sequence in the system from Figure 2.4 to target either the output sfGFP (blue) or the controller T7 RNAP (orange). In this case, mRFP1 only acts as capacity monitor. **B)** Beyond the effect of just adding the gRNA to the transcript (Figure 2.S3), output is greatly reduced by either feedback architecture, though the efficiency of output expression is not significantly different from the open-circuit architecture (grey). Changing repression strength could be a reliable way to tune output levels, though the MRG makes it difficult to do this through spacer truncations (Figure 2.S3). Replicates are plotted individually.

Wondering if we could use previously reported CRISPRi tunability through gRNA spacer truncation[4] to establish an optimal feedback strength that would preserve high levels of output while preventing excess burden from the T7 controller, we introduced truncations to the spacer sequence of the T7-targeted gRNA in order to create a small library of feedback strengths. Output levels from this library did not follow a trend according to truncation length, but nonetheless varied from 60 to 84% reduction relative

to the off-target control (Figure 2.S3). This raised the question of whether output level variation was primarily due to feedback strength variation or to differential burden imposed by circuits with varying parameters. Previous findings in Chapters 1 and 2, however, indicate that the control circuit burden is probably not fluctuating significantly relative to output burden, because output burden makes such a dominant contribution to total expression burden. Indeed, burden levels indicated by mRFP1 were constant across these feedback strengths when normalized to output levels (Figure 2.S4), indicating that feedback strength variation was the primary driver of output variation, and therefore that excess controller burden is probably minimal in all these cases.

This means, however, that attempted rational tuning of this MRG system through gRNA spacer truncations does not result in a predictable effect on output. The reason is that cleavage of the ribozyme leaves on the 5' end of the gRNA a scar of RNA sequence internal to the ribozyme—meaning that these spacer truncations are actually partial mismatches to the target. Perhaps more importantly, the inclusion of this extra sequence in slightly different positions could influence gRNA folding in such a way that affects its repression efficacy, by changing the proportion of the gRNA population that is properly folded and can bind dCas9. Indeed, we saw in the RFP-targeted CRISPRi of Section 2.3 that a truncation to 16 nt actually increased repression efficacy slightly. This difficulty of tunability is one of the main limitations of expressing gRNA through an MRG system for feedback control.

2.5: CRISPRa/i feedback controller

As part of the CRISPRa burden quantification in Section 1.3, we built a small series of base promoter strengths within the J3 synthetic promoter[8], all stronger than the typical J23117, with the idea that stronger, but unactivated, expression would represent an output-matched reference strain for burden measurement. We found that these stronger promoters became less activatable, measured by fold-increase in output, as their output ran up against a ceiling of allowable expression (Figure 2.6). However, when increasing the output plasmid copy number from low (pSC101**) to medium (ColE1), we found a strong risk of genetic instability in lieu of that output ceiling. Even

activation of J23117 was slightly unstable, as indicated by heterogeneity measured as bimodal distributions by flow cytometry, and by wide differences in bulk output between biological replicates on a plate reader. Medium-copy J23116 was tolerated when unactivated, but the additional burden of activation resulted in enough instability to decrease the bulk output measurement. Stronger promoters had even more drastic results: J23102 was unstable even when unactivated, and J23119 failed to clone.

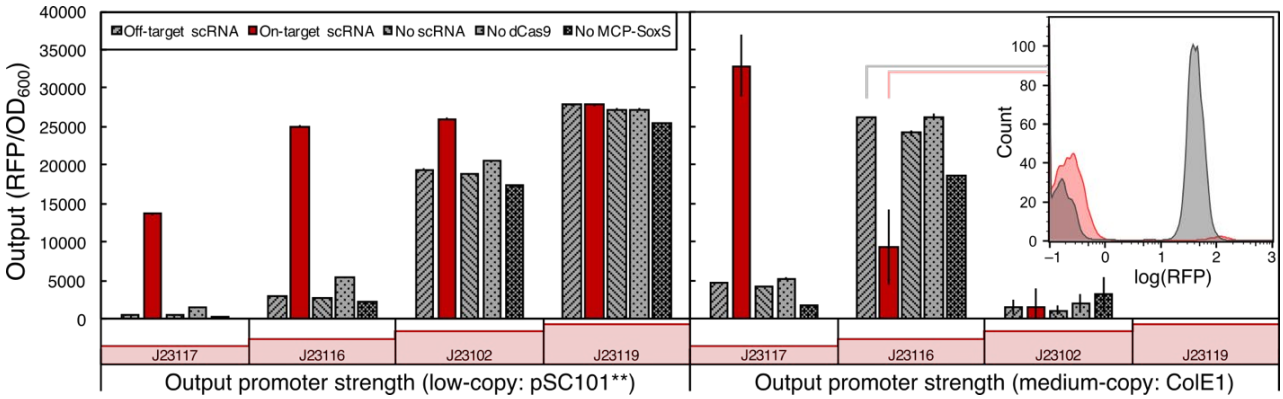


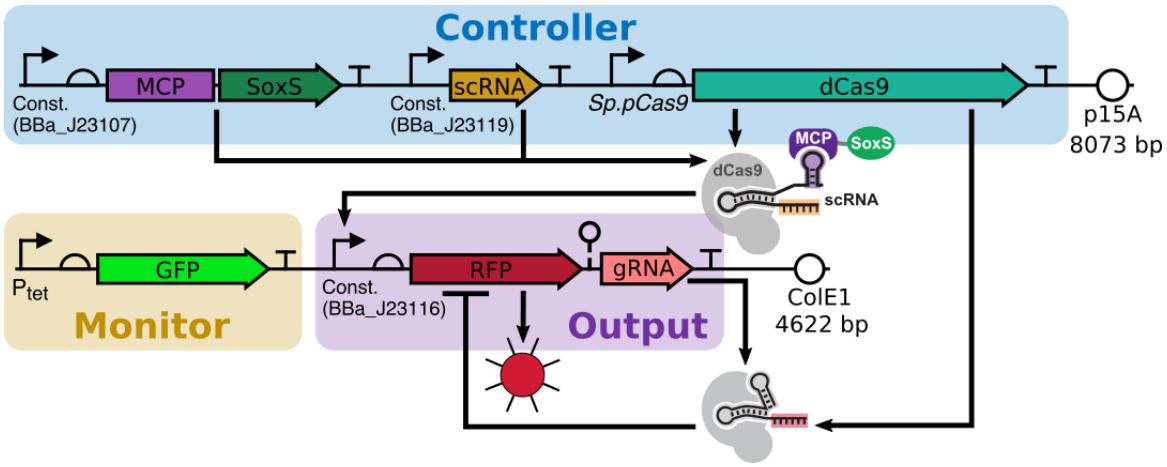
Figure 2.6. Open-circuit CRISPRa reaching metabolic limits. A two-plasmid system as in Figure 1.2A compares on-target mRFP1 output expression to various off-target or no-component controls. For both a low-copy (left) and medium-copy (right) output plasmid, the base promoter strength of the J3 promoter is increased as indicated. The low-copy plasmid remains stable but reaches a ceiling of output beyond which activation cannot increase; the medium-copy plasmid suffers instability even at activated J23117 promoter strength, resulting in large non-expressing subpopulations. The bar plot shows bulk cultures measured by a plate reader, and the inset shows typical flow cytometry histograms for heterogeneous (red) and intact (grey) cultures. Error bars represent standard deviation of *n*=3 biological replicates.

Within the medium-copy (ColE1) version of this system, we added on the same MRG construct as in Figure 2.3: an ASBV ribozyme and RR2 gRNA appended to the mRFP1 transcript, representing output-targeted feedback (Figure 2.7A). This purely-CRISPRa/i control circuit is advantageous in terms of controller burden due to its shared use of CRISPR's protein components, while still maintaining functionality of both activation and inhibition. It is possible with larger CRISPRa/i circuits for some nodes' efficacy to be reduced by competition for dCas9 or MCP binding, but it is unlikely to occur in this simple, two-node system. The MRG output, as noted earlier, avoids some of the drastic and variable burden effects of protein feedback effectors, but adds a bit of

burden relative to an output of mRFP1 alone. We hypothesized that this added burden is diminished now that the output is being transcribed by an *E. coli* RNAP rather than by T7 RNAP, as in previous sections. The answer seems to depend on the growth media, with MRG appearing less costly than mRNA alone in LB (possibly confounded due to higher-than expected output in one genotype), but showing no cost difference in EZ-RDM (Figure 2.S2). The full experiment was performed in LB.

We found the effect of the CRISPRi feedback to limit output quite severely, with activated and repressed output having a similar value to unactivated and unrepressed output (Figure 2.7B). Unactivated but repressed output was quite diminished, due to some gRNA being in the system, and activated but unrepressed output, as before (Figure 2.6), suffered from genetic instability. The intriguing result in this experiment is the relative stability of the strain undergoing activation and repression, potentially supporting our assertion that feedback control is important for preventing problematic levels of output expression—it seems that the mere addition of MRG-based feedback rescued the ColE1–J23116 culture from instability. Because output was limited all the way down to unactivated levels, however, it is hard to tell whether this stability is inherent to the feedback control, or if it simply mirrors the stability seen in the unactivated, unrepressed cultures that show a similar output level.

A



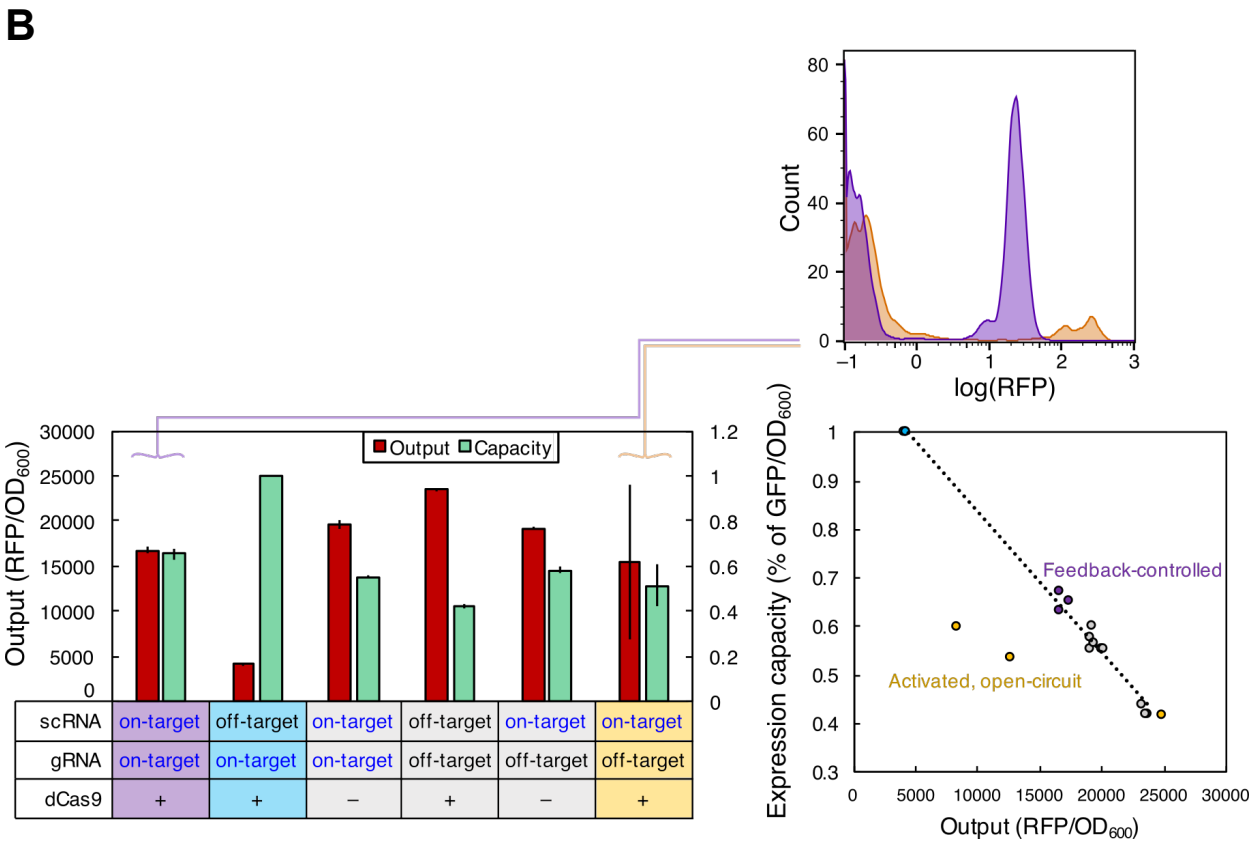


Figure 2.7. MRG-CRISPRi feedback on CRISPRa control limits instability. **A)** The same MRG gRNA expression system as in Figures 2.3-5 is applied to an mRFP1 output controlled by CRISPRa, with sfGFP acting as capacity monitor. The gRNA is targeted to mRFP1, completing the feedback. The base promoter strength of the J3 output promoter is J23116, on a ColE1 plasmid, a combination which led to instability when activated in Figure 2.6. **B)** The feedback-controlled circuit (purple) results in similar output levels as unactivated, unrepressed strains (grey), and remains much more stable and homogenous than the activated, open-circuit strain (yellow). At upper right is the comparison between the two of flow cytometry histograms, while the bar plot is bulk culture measurements from a plate reader. It is unclear whether the stability of the feedback-controlled culture is inherent to the architecture or simply a result of low output. An unactivated, repressed culture (blue) is taken as the reference for expression capacity, and the feedback-controlled cultures fall on a line between the reference and the unactivated, unrepressed cultures (right), indicating similar expression efficiency regardless of architecture. Two replicates of the activated, open-circuit strain show greatly reduced output expression, due to the non-expressing subpopulations, but also have lost enough expression capacity to deviate strongly from the overall burden-per-output function. Bar plot error bars represent standard deviation of *n*=3 biological replicates; scatter plot replicates are plotted individually.

Ideally, we would adjust the strength of either the activation or repression to shift the balance toward higher output, to see how high we can push a feedback-controlled circuit. Diminishing the repression strength of the gRNA would be the simpler way to achieve this; however, the MRG system makes spacer truncation difficult to predict, as

we saw in Section 2.4. An important future experiment to assess the usability of this fully-CRISPR feedback system would be to diminish gRNA function in another way, for example by selecting a less effective target site, though this approach is not nearly as neatly and predictably tunable as truncation.

Instead, we separated the repressive gRNA from the output mRNA by placing them under the control of separate, but both CRISPR-activatable, promoters, thus removing the ribozyme component and its problematic scar while still linking output and gRNA levels. This allowed tunability of the repression strength, and also changed the circuit topology to a type 1 incoherent feedforward loop.

2.6: CRISPRa/i incoherent feedforward controller

The result of independent output promoters for mRFP1 and gRNA was the incoherent feedforward loop reported recently[59], slightly adapted to include a constitutive sfGFP capacity monitor (Figure 2.8A). A hallmark behavior of IFFL circuits is the capability to express output as a temporal pulse[51], due to the lag time of the repressive element, so we checked for this behavior in our CRISPR-based IFFL using a continuous dilution experiment in *E. coli*. We found that output levels were very low, but that a pulse was present (Figure 2.8B). The theoretical usefulness of this pulse in control circuit design and heterologous pathway control is blunted by its miniscule output, prompting us to re-balance the activation and repression to boost output. In later experiments, we used the low-output IFFL to measure the expression costs of expressing scRNAs and gRNAs in real-world, multi-layered circuits, as a test of our CRISPR circuit expandability argument from Chapter 1. We additionally were interested to assess how the pulse dynamics influenced the measurement of burden-per-output, if at all.

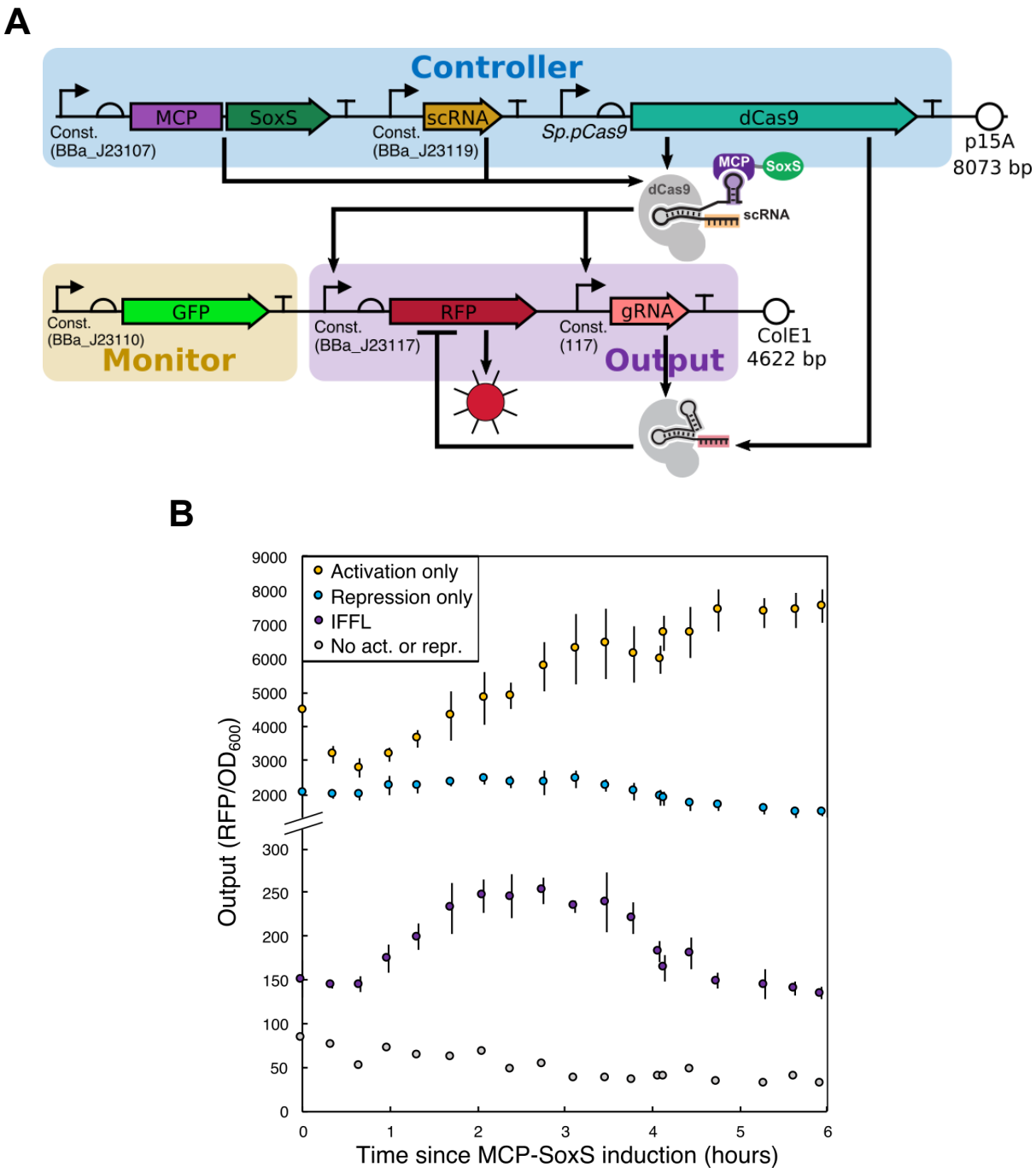


Figure 2.8. CRISPRa/i-based IFFL yields low, pulsatile output expression. **A)** Relative to Figure 2.7, independent activation of the gRNA changes the circuit architecture to a type 1 incoherent feedforward loop. The only genetic difference is control of the gRNA by the J2 promoter, and an additional scRNA (J206) expressed from the controller plasmid (omitted here). **B)** To assess the dynamic behavior of this circuit, we performed a continuous dilution experiment in which culture volume removed for measurement was discarded and replaced with fresh media, at a frequency such that culture densities were kept constant. While the activated, unrepressed culture (yellow) continually increased, the IFFL culture (purple) initially increased and then experienced repression, forming a dynamic expression pulse that indicates the circuit

is working as predicted. The IFFL culture is also limited to extremely low output, pointing toward lower CRISPRi efficacy as a way to make relevant amounts of output while still achieving pulsatile expression dynamics. The repression-only strain (blue) uses a strong promoter driving mRFP1 (J23119), to better distinguish its output from background. The cost of expressing scRNAs or gRNAs within this circuit (Figure 2.S5) were found to be similarly low as when expressed in isolation (Figure 1.3). Error bars represent standard deviation of *n*=3 biological replicates.

An intriguing development in the low-output IFFL experiment was that expression capacity, measured at endpoint, seemed to increase drastically relative to the open-circuit reference (Figure 2.9), even if output isn't increasing much above unactivated background. We think this could be due to the dynamics of expression: because the output expression is more pulsatile in the IFFL architecture, there is more capacity for the cell to make the capacity monitor (sfGFP) by the time that output is repressed. However, this effect was absent in the later, high-output IFFL experiment, so perhaps it is peculiar to the comparison between low-output IFFL and open-circuit control.

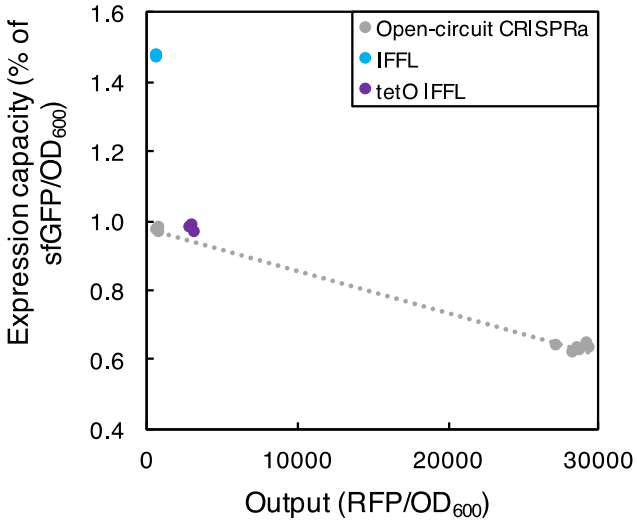


Figure 2.9. Burden per output of the CRISPR-based IFFL. By comparing open-circuit CRISPRa with no repression at different base promoter strengths, we can establish a reference burden-per-output function (grey) against which to compare IFFL output and capacity (measured by constitutive sfGFP). Surprisingly, capacity measurements increased in the IFFL cultures: an effect that could be due to the difference in output expression dynamics, with a pulsatile output expression allowing for extra capacity monitor expression after the end of the pulse. This effect is absent in the tetO IFFL (purple), which is an attempt at boosting output by using tet repression to tighten leaky expression of the repressive gRNA. Replicates are plotted individually.

Because the output level of the initial IFFL was so impractically low, we made use of the independent output promoters to truncate the gRNA (down to 11nt) and weaken the repression, an intervention not possible in the similarly low-output CRISPRa MRG feedback architecture of Section 2.5. Output increased drastically (Figure 2.10), rendering the circuit more practically useful and giving us a better chance of reliably measuring burden. In contrast to the low-output case, the burden imposed by this higher output was quite in line with a non-IFFL control (Figure 2.11), so perhaps expression capacity measurement is less dependent on timing than initially indicated by the previous experiment. It is worth noting that a failure of the open-circuit culture forced us to use an alternative non-IFFL reference strain in this experiment: CRISPRi repressing a strong (J23119) mRFP1 promoter in the absence of activation.

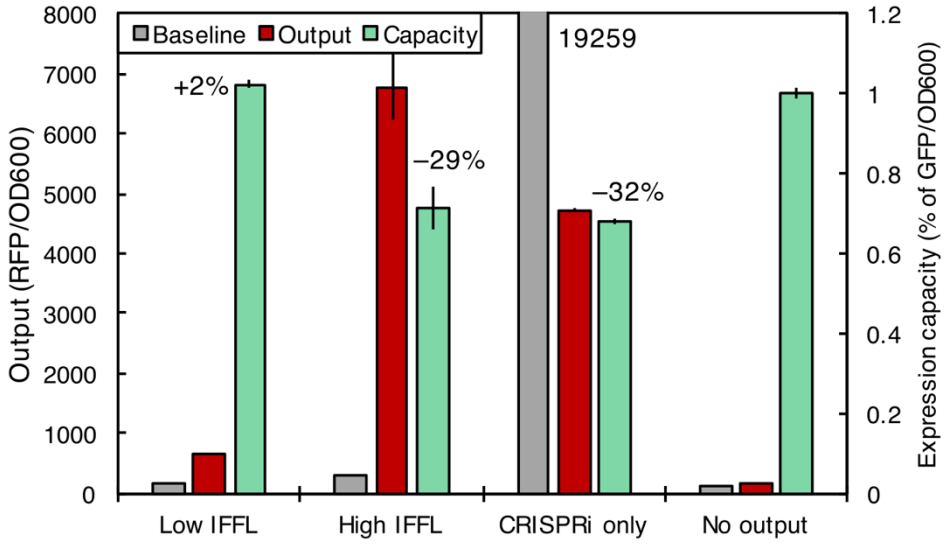


Figure 2.10. High-output CRISPR-based IFFL. Aiming for a circuit with reasonably useful output expression levels, and also better comparability of capacity data, we increased output of the IFFL in Figure 2.8 by truncating the gRNA spacer sequence, reducing CRISPRi efficacy. The result is the IFFL balance being shifted toward CRISPRa, boosting output. As in Figure 2.8, the CRISPRi-only control uses a J23119 output promoter, to better distinguish its output from background. Error bars represent standard deviation of *n*=3 biological replicates.

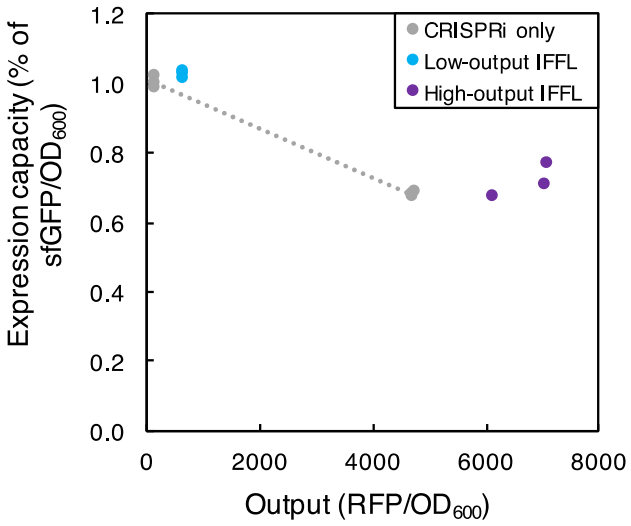


Figure 2.11. Burden per output of the high-output CRISPR-based IFFL. As in Figure 2.9, we establish a reference burden-per-output function (grey) against which to compare IFFL output and capacity (measured by constitutive sfGFP), but due to a failure of the CRISPRa-only strain we were forced to use the CRISPRi-only strain. Interestingly, the capacity increase seen in Figure 2.9 is absent from this experiment, even in the low-output IFFL (blue). More reliably (due to its similarity in output level to the CRISPRi-only strain), the high-output IFFL (purple) is located close to the reference burden-per-output function, indicating similar expression efficiency. Output of this IFFL is increased by reducing CRISPRi efficacy through gRNA spacer truncation. Replicates are plotted individually.

Even if this IFFL provides no efficiency advantage over open-circuit designs, it is remarkably a CRISPR-based control circuit with tunable output levels, it theoretically prevents output expression levels running away into problematically high regimes, and provides intriguing expression dynamics that could prove useful for certain problems of pathway control.

Finally, this particular circuit, with its modularity and numerous opportunities to easily clone off-target scRNAs or gRNAs, represented a useful testbed for determining scRNA or gRNA expression cost in the context of a larger, multi-layered circuit—an important feature of the argument from Chapter 1 that CRISPR-based circuits are well-suited to arbitrary expandability. This is also a more real-world test than the promoter deletion of Figure 1.3, and it turns out that the results are very similar: to add a J23119-expressed scRNA costs 4.4% of expression capacity in this measurement (Figure 2.S5), while to use that scRNA to CRISPR-activate a gRNA driven by the J2 synthetic

promoter costs seemingly less, with measurements actually negative in this experiment, though somewhat confounded by slightly different output levels in the reference strain. With these expression costs being so low relative to that of the shared CRISPR protein components, it indeed seems that large control circuits—shaped either broadly or deeply—are achievable using CRISPR-based expression, depending on the burden of the outputs and the interactions between circuit layers.

2.7: Conclusions

In our efforts to keep output expression limited to reasonable levels, we built control circuits, based on the controllers investigated in Chapter 1, that incorporate closed-loop architectures mostly consisting of negative feedback. We observed various behaviors inherent to different negative regulators: with protein-based regulators exhibiting highly variable burden effects; MRG CRISPRi achieving stable burden contribution but lacking tunability; and non-MRG CRISPRi overpowering CRISPRa until balanced by truncation, boosting output. In these circuits, we did not see a case where a closed-loop, genetically stable output level exceeded an open-loop, genetically unstable one, as hypothesized, but it is still true that such closed-loop control decreases the likelihood of burden-based genetic instability and other problematic phenotypes, even if it does so only by keeping output levels low and stable. This stability leads to homogeneity within the culture, which can end up being more productive than a higher-output but unstable and heterogeneous culture.

A note of caution about boosting output levels by decreasing repression efficacy in these circuits is that high outputs, close to the threshold of problematic expression, result from the weakest repression efficacies. This weak repression also diminishes the output-stabilizing effect of the negative feedback or feedforward motif, increasing the spread of the output distribution and therefore the chance that individual cells could undergo genetic instability. Close to the metabolic cliff is where tight control of output stochasticity is most important, so this coincidence of high output and high stochastic output fluctuation resulting from weak repression might call for exploration of alternative output-tuning strategies, such as increasing activation strength under strong repression.

In observing the burden consistency of CRISPRi negative regulators, and reaffirming the low cost of additional nodes of CRISPR circuit control, we have validated the acceptability of using CRISPR to express large control circuits, which may or may not involve negative autoregulation. Theoretically, these circuits can achieve large size through breadth or depth or both, but in biological practice, deep cascades of logical operations tend to suffer from signal decay[44,60,61]. The following chapters explore examples of "real-world" implementations of CRISPR circuits controlling metabolic pathways for the production of valuable chemicals from bacterial hosts. We additionally explore the expandability of these circuits, focusing on breadth (with extra nodes used for endogenous CRISPRa/i or for biosensors), with the understanding that if necessary—that is, if instability is observed during circuit expansion—an additional node encoding CRISPRi feedback can always be added, potentially rescuing the circuit's stability and allowing further expansion[49].

2.8: Methods

Plasmid construction

Plasmids were constructed using either circular polymerase extension cloning (CPEC), Infusion (Takara Bio), or restriction cloning. All cloning was performed in *E. coli* DH10b chemically competent cells (Invitrogen), and DNA oligonucleotides were purchased from Integrated DNA technologies. Sanger sequencing to verify constructs was performed by Eurofins Genomics or Genewiz. DNA encoding T7 orthogonal RNA polymerases, as well as pBR322-origin-containing plasmid backbones and the lacIq gene, were gifts from Andrew Ellington (Addgene plasmids 63668, 63627, and 63628). Plasmid backbones containing p15A or ColE1 origins were taken from BglBrick plasmids[39], gifts from Jay Keasling (Addgene plasmids 35308, 35280, 35316, 35322, and 35279). DNA encoding CRISPRi/a components was obtained from multiple plasmids from Jason Fontana[7,40]. RBS variants were selected from a previous report[41] and written into CPEC or Infusion primer overlaps. DNA encoding T7 lysozyme K128Y was obtained from the pLysY plasmid contained in T7 Express cells (New England Biolabs). Point mutations in this gene, T7 RNAP genes, and promoters were written into CPEC or Infusion primer overlaps.

Experimental culture conditions

All experiments were conducted in *E. coli* MG1655 ΔrecA ΔendA cells[43], MG1655 ΔrecA ΔendA (DE3) cells, or BL21 cells with pLysY added before generating competent cells (New England Biolabs), as indicated in the text. Transformations were accomplished by heat-shock, and most experiments were conducted in two-plasmid systems. Generally, three transformed colonies were picked into replicate overnight cultures in LB (Miller) liquid media (Teknova) supplemented with appropriate antibiotics. After overnight growth, cultures were diluted 1:100 into fresh LB media, with antibiotics, in a single culture tube or flask for each replicate.

For inducible cultures, particularly T7-controlled systems, after growth up to OD_{600} 0.3, as assessed using 200 uL culture in a flat-bottom, clear-bottom, 96-well plate (Corning, cat. no. 3631) in a Biotek Synergy HTX plate reader, cultures were split into 96-deep-well culture plate wells (Axygen, cat. no. P-DW-20-C-S) at 300 uL/well and induced by adding a small volume, usually 3 uL, of inducer. For constitutive cultures, particularly CRISPRa-controlled systems, overnight cultures were diluted 1:100 into fresh media containing antibiotics and inducers (aiming to match induction timing of constitutive controller and inducible capacity monitor) and grown in individual wells in the 96-deep-well plate. Density and fluorescence data were gathered in subsequent timepoints by reading 200 uL of culture, in a clear-bottom Corning plate, with OD_{600}, x485/m528 for sfGFP, and x540/m600 for mRFP1, each with a gain of 40. Raw fluorescence values at 40 gain were higher after a light source change on the plate reader, but values across bulbs were normalized when needed using maximal expression from the same expression systems.

Expression capacity measurement

Capacity monitor values were collected using a Biotek Synergy HTX plate reader, as above, and normalized by OD_{600} values. Baseline, uninduced values of Ptet-driven sfGFP/OD_{600} values were subtracted from experimental values, and expression capacity was calculated as the percent of capacity monitor expression remaining after comparing an experimental case to a reference case (e.g. from off-target to on-target), or sfGFP/OD_{600} (experimental) / sfGFP/OD_{600} (reference). When these separate cases resulted from different levels of controller induction, within the same strain, the capacity calculation was performed within each replicate and averaged afterward, because replicate colonies were matched across induction levels. When the separate cases resulted from different strains, as in promoter deletions, for example, reference strain replicates were averaged before capacity calculation, because each replicate came from a unique colony.

Curve fitting, normalization, and statistical testing

Raw fluorescence values collected using the plate reader were always divided by OD_{600} values to normalize for cell density. Hill functions were fit using a previously described[44] Python script. Linear regressions in the capacity versus output scatter plots were fit using Microsoft Excel. For graphical purposes, averages of three biological replicates were used as points for a single linear regression, while the individual points, not their averages, were displayed on the scatter plots. For statistical purposes, a separate linear regression was calculated for all possible combinations of two points on the scatter plot (individual replicate points, not average points), excluding pairs within the same controller induction level. The resulting population of slopes was compared between strains or conditions using a two-tailed Student's t-test with unequal variance. In cases where multiple controller induction levels were not possible, for example medium-copy-output CRISPRa, or in cases where biological triplicates were not possible for some strains, the burden-per-output scatter plot includes a 95% prediction interval shaded around the reference strain (grey).

Flow cytometry

Overnight cultures were diluted 1:100 in PBS (Hyclone) and measured using a Miltenyi MACSQuant YVB flow cytometer with low flow rate; 10000 events; channels including FSC at log2 630 V, SSC as main trigger at log2 450V, Y2 at log4 350V (300V for CRISPRa systems), and B1 at log4 350V; and a trigger threshold of 15. Gating was accomplished by eliminating fringes on a SSC-A vs. FSC-A plot, then gating for linearity on a SSC-A vs. SSC-H plot. Analysis included mostly visual observation of histograms aiming to identify bimodal distributions indicating heterogeneity, but median values were sometimes used in cases of unimodal distribution.

Continuous dilution experiment

Three transformed colonies were picked into replicate overnight cultures in LB media supplemented with appropriate antibiotics. After overnight growth, cultures were diluted 1:50 into 300 µL fresh EZ-Rich defined media, with antibiotics, in a single deep-well plate well. After 1 hour of growth, 100 µL of culture was removed for measurement on a Biotek Synergy HTX plate reader; this culture volume was discarded and replaced with an equal volume of fresh media, at a frequency (about 20 minutes) such that culture densities, measured by OD_{600}, eventually stabilized (after about 1 hour) and were subsequently kept constant. Once OD_{600} stabilized (at about 0.25), aTc was added to a final concentration of 50 nM. Measurements continued every 20 minutes for 8 hours.

2.9: Supplementary Figures

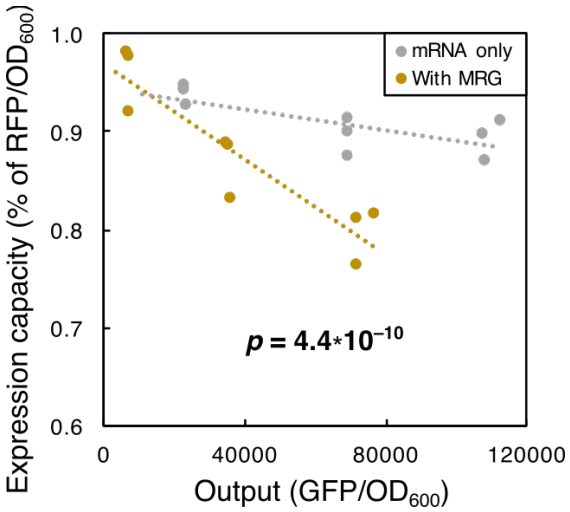


Figure 2.S1. MRG expression burden under T7 RNAP control. Output transcripts with or without MRG were used to generate burden-per-output functions, which showed significantly different slopes when expressed by T7 RNAP as in Figure 2.5A. It is a reasonable expectation that this decrease in efficiency when adding MRG is specific to T7 expression, due to its high transcriptional elongation rate. The MRG transcript's reduction in output level is apparent along the *x*-axis. The gRNA was off-target in this case. Replicates are plotted individually.

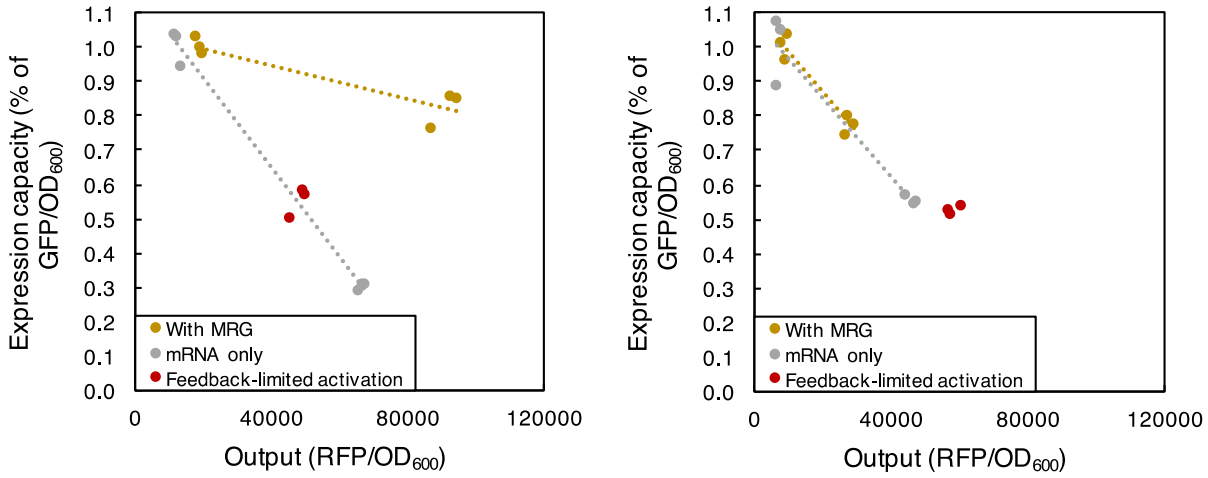


Figure 2.S2. MRG expression burden under *E. coli* RNAP control. Due to an odd media dependence, versions of this experiment in LB (left) and EZ-RDM (right) are shown here. Output transcripts with or without MRG were used to generate burden-per-output functions as in Figure 2.S1, but under *E. coli* transcription as in Figure 2.7. In EZ, coincidence of the functions

indicates similar efficiency regardless of MRG inclusion, with an activated but feedback-limited strain (red here, purple in Figure 2.7) showing similar efficiency. In LB, output is much higher than expected in one MRG point, making the MRG burden-per-output function appear dramatically more efficient than the MRG-less function: an unlikely outcome when the gRNA is off-target. A higher resolution of output levels, populating more points along the function, could help resolve this discrepancy. The gRNA was off-target in all strains except the feedback-limited activation. Replicates are plotted individually.

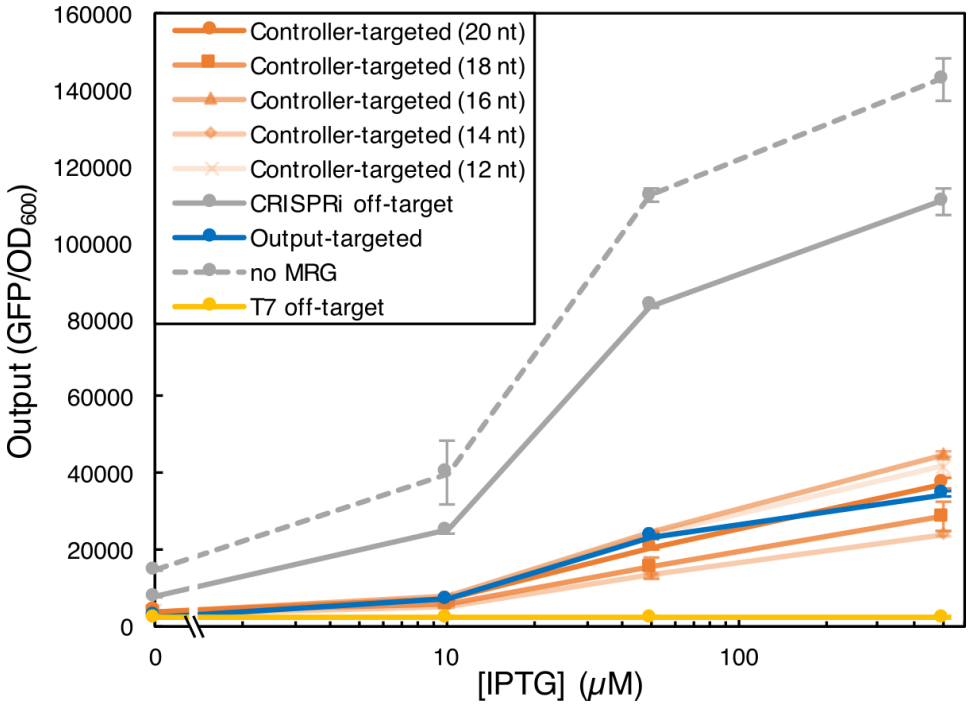


Figure 2.S3. Truncations of repressive gRNA fail to tune T7-MRG circuit. Tuning the strength of controller-targeted feedback through truncations of the gRNA spacer sequence yields no clear trend in output levels, probably due to ribozyme scars interacting differently with the various truncated sequences. Reduction in output expression as a result of MRG inclusion on the transcript is apparent by comparing the off-target (grey solid) and no-MRG (grey hashed) response functions. Error bars represent standard deviation of *n*=3 biological replicates.

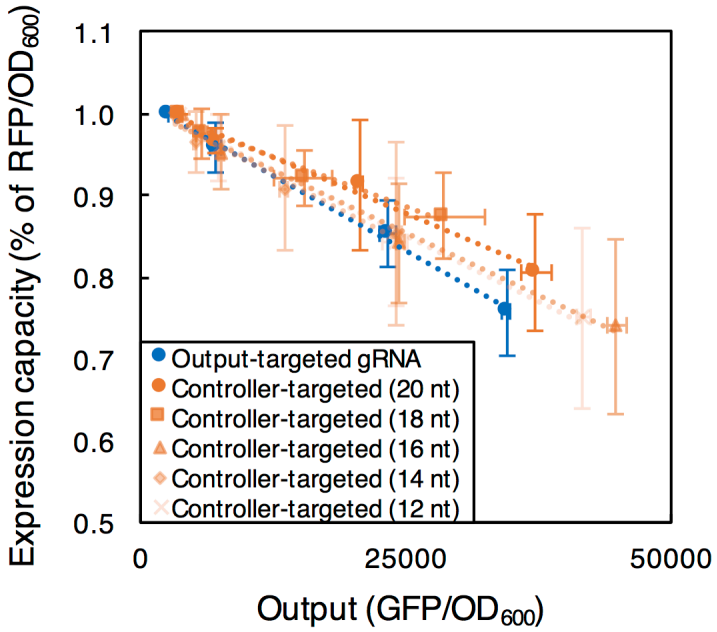
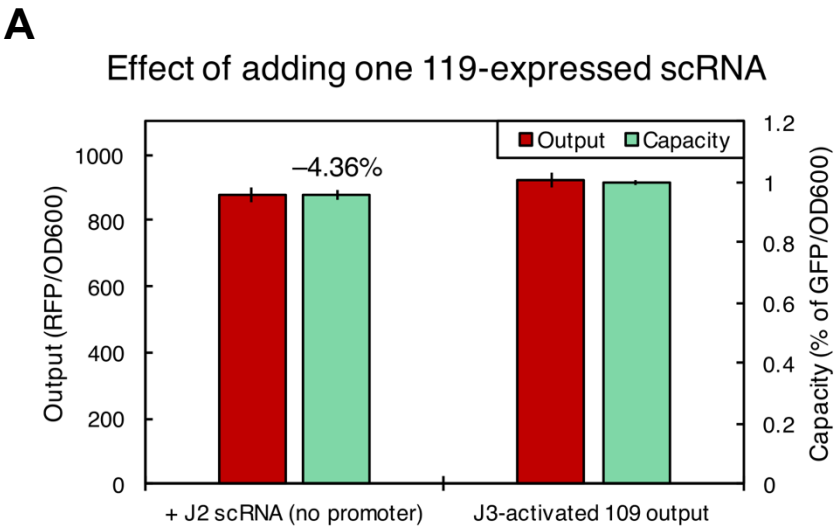


Figure 2.S4. Burden per output unaffected by gRNA truncations. The range of gRNA spacer truncations in Figure 2.S3 shows no clear trend in burden per output, just as it showed no trend in output. For clarity, replicates are not plotted individually. Points and error bars represent average and standard deviation, respectively, of *n*=3 biological replicates.



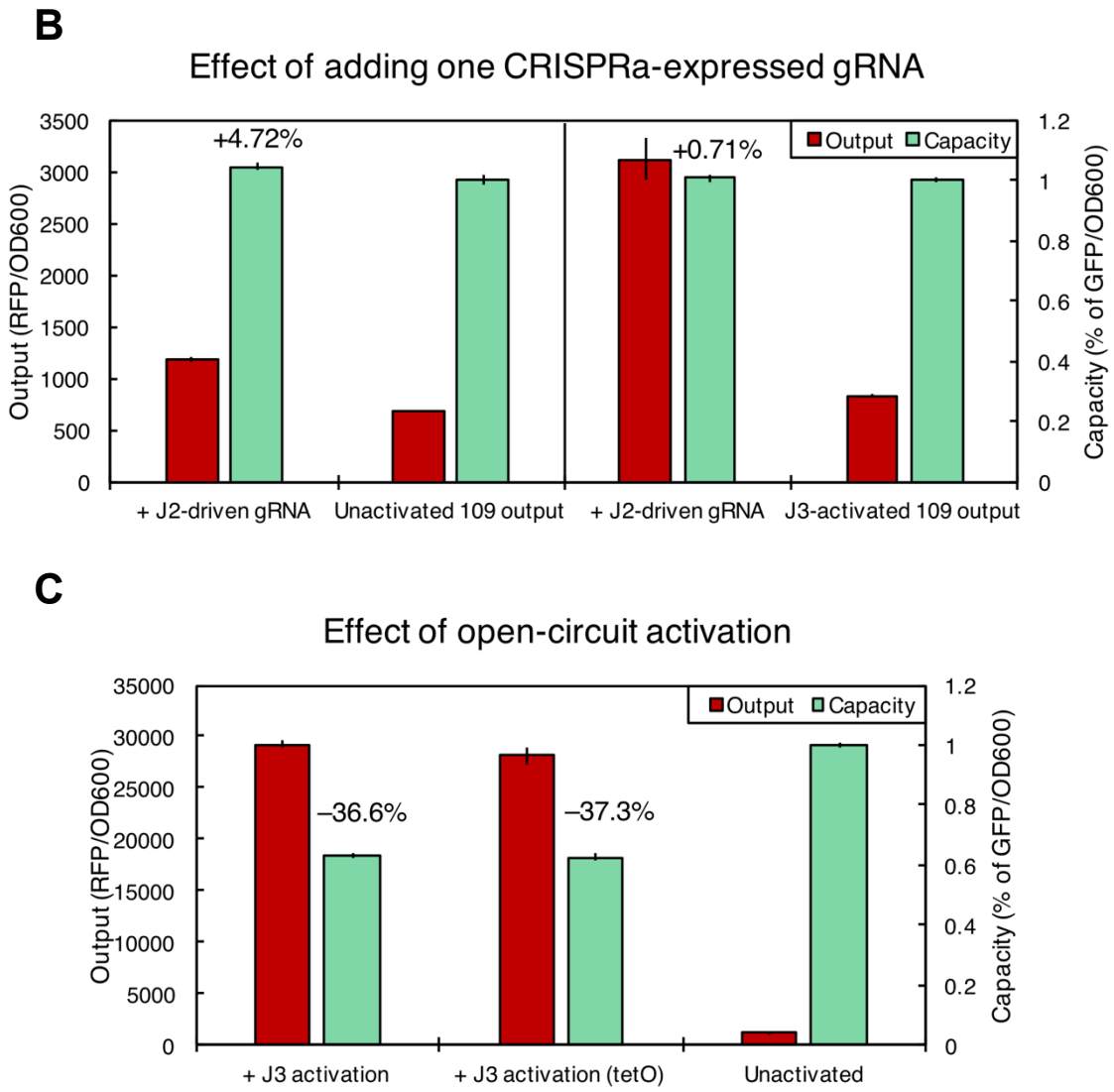


Figure 2.S5. Expression cost of additional RNA elements within the CRISPR IFFL. Reference strains are selected to highlight the difference between expressing or not expressing (A) the J206 scRNA, expressed by a J23119 promoter and lacking an output promoter to activate; (B) the off-target gRNA, expressed by the CRISPR-activated promoter J2; (C) activated RFP output without repression, to verify the functionality of the capacity monitor. scRNA and gRNA expression costs are similarly low as in Figure 1.3, even when expressed from a synthetic, CRISPR-activated promoter as part of the second layer of the IFFL circuit. In all cases, the reference strain is on the right. Error bars represent standard deviation of *n*=3 biological replicates.

Chapter 3: Broad-circuit control of *p*-AF/*p*-ACA production in *Pseudomonas putida*

Abstract

A companion feature to the expandability of CRISPR-based control circuits is their programmability. The target site of CRISPR complexes is almost arbitrarily programmable, and we are steadily learning the rules for bacterial CRISPRa of even endogenous genes. It is fitting, then, to imagine broad CRISPR circuits of many control nodes that represent a program of widespread activation and repression aimed at boosting heterologous pathway performance. In this chapter, we engineer a *Pseudomonas putida* strain capable of producing the valuable aromatic chemical *p*-aminocinnamic acid, as part of a larger effort to incorporate data-driven strain design into the incremental improvements of design-build-test-learn cycles in non-model bacterial hosts. When finished, the improved strain will be a test case in the use of large CRISPR-based control circuits to build and optimize bacterial strains for chemical production. For now, we relate the construction and characterization of the unimproved strain: in its own right a noteworthy accomplishment of bacterial production of *p*-ACA.

3.1: Introduction

Production of valuable chemical compounds using engineered biological hosts is a promising route with many chemical advantages, but accommodating, avoiding, or taking advantage of (by harnessing material and energy from) endogenous metabolism and its accompanying regulation can be a major obstacle to industrially relevant bioproduction[1]. Often, overcoming this obstacle requires wide-ranging alterations of endogenous metabolism, and new tools have emerged to understand the effects of such changes. Large-scale observation of strain engineering effects using -omics technologies[62], combined with genome-scale modeling and design-build-test-learn (DBTL) approaches enhanced by machine learning[63,64] hold great promise for rapid improvement of production strains through well-targeted changes to endogenous metabolism.

We see the combinatorial, orthogonal, and tunable features of CRISPR-based expression control as an advantage well-matched with the framework of DBTL cycles for incremental strain improvement. While additional work is needed before arbitrarily targeting any endogenous gene for CRISPRa, many targets have already shown promise, and the portability of CRISPR control allows for proper matching of host and product[65,66]. Though details might differ, the design rules, programmability, and expandability of CRISPR control circuits should remain fundamentally unchanged across a wide range of bacterial production hosts in the near future, with some better suited to certain classes of products.

Aromatic compounds are a promising but challenging class of bioproducts[67] due to their connection to the host's central metabolism through the aromatic amino acid precursor chorismate[68]. Development of aromatic-producing strains is particularly attractive when using renewable, non-edible lignocellulosic feedstocks[69]. The products and intermediates can pose challenges, however, due to toxicity and solubility concerns[70,71]. In particular, in this work we focus on *p*-aminocinnamic acid[72] (*p*-ACA) as a precursor for *p*-aminostyrene, but *p*-ACA production in *E. coli* is accompanied by toxic effects (Figure 3.2), even though its immediate precursor *p*-aminophenylalanine (*p*-AF) can safely accumulate in that host. A more suitable host, free from the toxic effects, is

Pseudomonas putida[73], the metabolism of which, including the heterologous pathway, is shown in Figure 3.1.

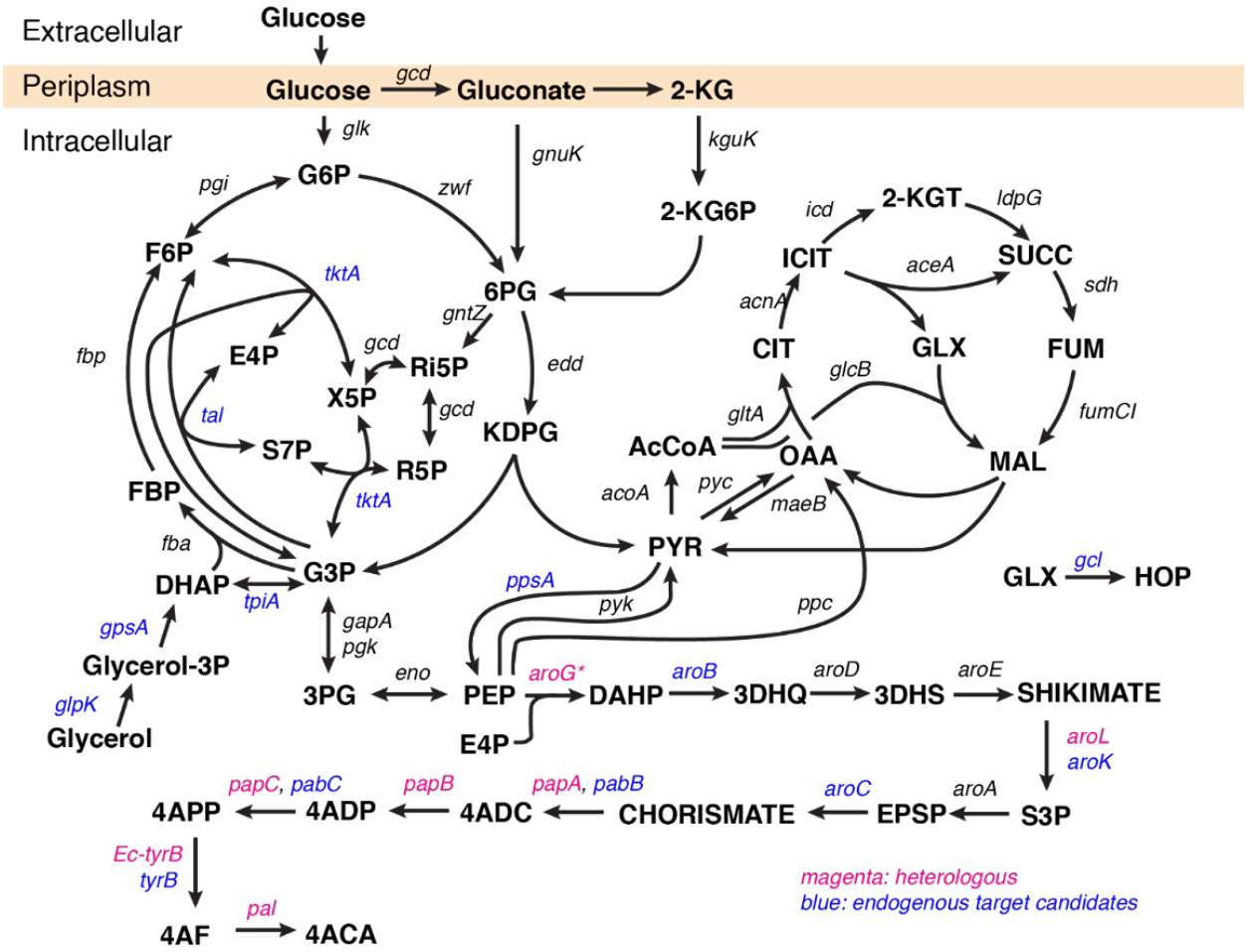


Figure 3.1. Connection of the *p*-ACA production pathway to central metabolism of *P. putida* through chorismate. To make *p*-AF or *p*-ACA, the heterologous genes shown in magenta are required, though *Ec*-tyrB is only sometimes included, because there is an endogenous copy as well. Blue text indicates possible targets for endogenous CRISPRa/i to boost metabolic flux through the heterologous pathway.

The heterologous contributions to the *p*-ACA production pathway (magenta in Figure 3.1) start with a feedback-resistant AroG[74] and overexpression of AroL aimed at boosting levels of the endogenous precursor chorismate. From there, the *Pseudomonas fluorescens* enzymes PapA, PapB, and PapC produce *p*-aminophenyl pyruvate, which becomes *p*-AF through endogenous transaminase activity, sometimes

supplemented by additional expression of *E. coli* TyrB. Finally, a phenylalanine ammonia lyase enzyme (PAL), either from *Arabidopsis thaliana*[42] or *Rhodotorula glutinis*[68,75], converts *p*-AF to *p*-ACA.

Production of *p*-ACA from this strain is likely to be enhanced by boosting metabolic flux into the pathway and by limiting loss of flux to side products[76], and we aim to rapidly design this enhancement using the machine-learning-based DBTL approach. The modularity and orthogonality of not only heterologous enzyme expression by CRISPRa, but of a whole array of endogenous CRISPRa/i interventions[77], is an ideal system to introduce the variation driving such DBTL improvement, especially in a host like *P. putida* (other hosts are planned as well). Additionally, our findings in Chapters 1 and 2 suggest substantial freedom in our ability to expand this array of scRNAs/gRNAs to arbitrary size. This expansion relates less to the autoregulatory architectures of Chapter 2 and more to a wide-ranging, single-layered control circuit—but if circuit expansion becomes overly burdensome at some point, either through expression burden or through changes to endogenous metabolism, Chapter 2 provides confidence that autoregulation can be added to the circuit as easily as adding another gRNA.

While most of this overall project is forward-looking, toward improving a basic production strain using novel and powerful techniques, this chapter describes the effort to establish that base strain as a starting point for further improvements: a non-model bacterial strain producing *p*-ACA under the control of CRISPR-based expression.

3.2: Heterologous genetics

As originally implemented in our lab for production in *E. coli*, the *p*-AF production pathway consisted of the *P. fluorescens papABC* operon[78] under tet-inducible control, along with a feedback-resistant *E. coli aroG*[74] and *aroL* in their own operon, also tet-inducible. Inducing this pathway in the DH10B strain, in particular, routinely produced up to 800 μM *p*-AF (Figure 3.S1), but trying to use an *A. thaliana* PAL2 enzyme to extend this pathway flux to the products downstream of *p*-AF proved difficult, probably due to the toxic effect of *p*-ACA on *E. coli* arresting the growth of any cells producing it.

Toxic amounts of extracellular *p*-ACA, the result of a spike into the media, are shown in Figure 3.2, with 50% reduction of *E. coli* growth occurring above 10 mM extracellular.

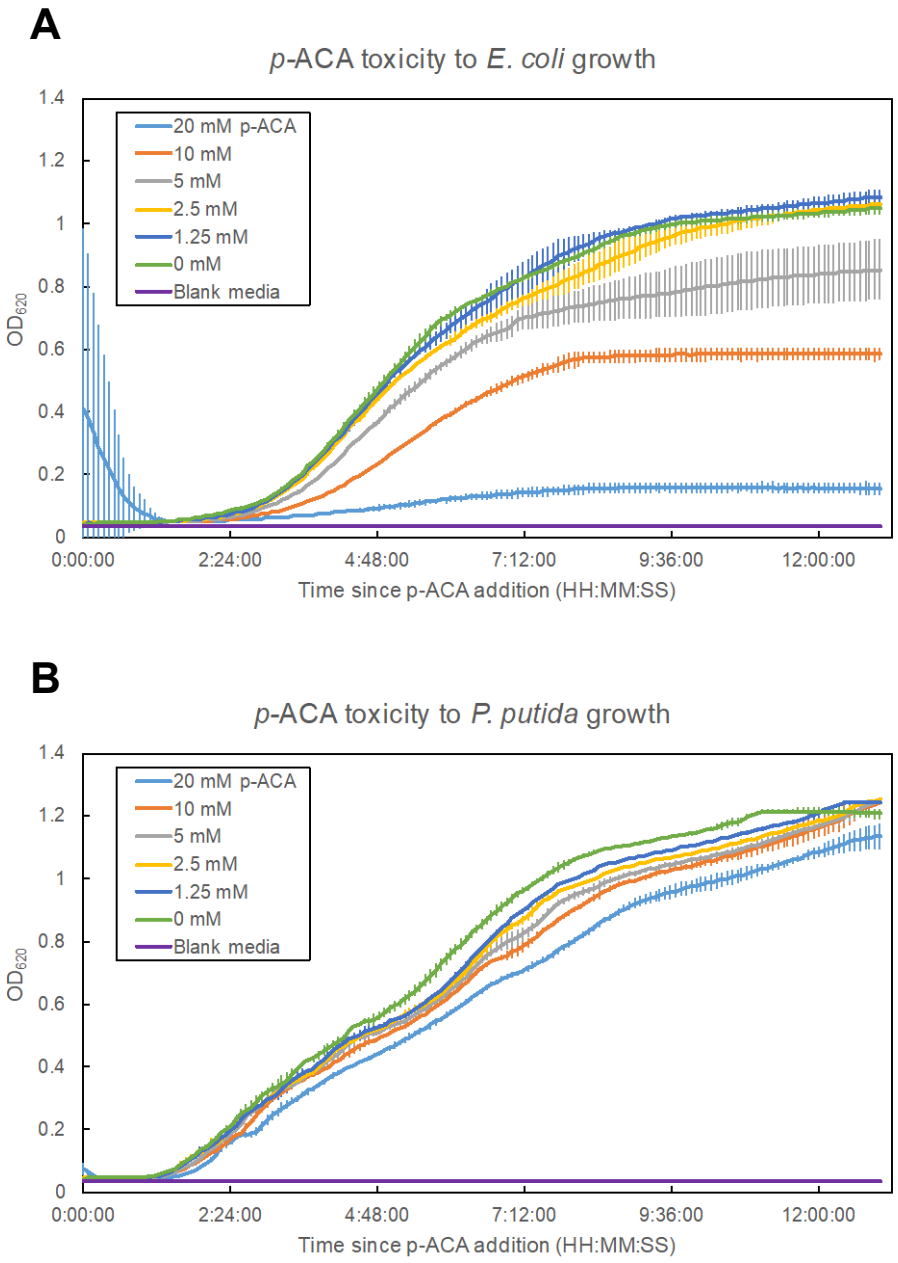


Figure 3.2. Desired product *p*-ACA is toxic to *E. coli* but not *P. putida*. In this kinetic growth experiment, extracellular *p*-ACA is supplied in the media at the indicated concentrations, and cultures are grown in a plate reader with periodic density measurements. **A)** *E. coli* growth is severely limited above 10 mM *p*-ACA, with milder effects at 5 mM. Early timepoints are sometimes obscured by *p*-ACA precipitate, which eventually resolubilizes. **B)** *P. putida* growth

is relatively unaffected, inspiring some confidence that *p*-ACA-producing cells will remain prominent in the culture. Error bars represent standard deviation of *n*=3 biological replicates.

In contrast to *E. coli*, *p*-ACA has little effect on *P. putida* growth, even up to 20 mM extracellular (close to its solubility limit). We therefore sought to port the existing *p*-AF pathway to *P. putida* and extend it to more valuable downstream aromatic products. This pathway portability, however, would benefit from a portable synthetic controller replacing tet-inducible control, which fit well with ongoing efforts to use CRISPRa as that portable controller. Before testing it fully in *P. putida*, we first verified in *E. coli* that CRISPRa control of *papABC* and *aroGL* produced *p*-AF (Figure 3.S2), then did the same on a broad-host-range plasmid in *P. putida* (Figure 3.3).

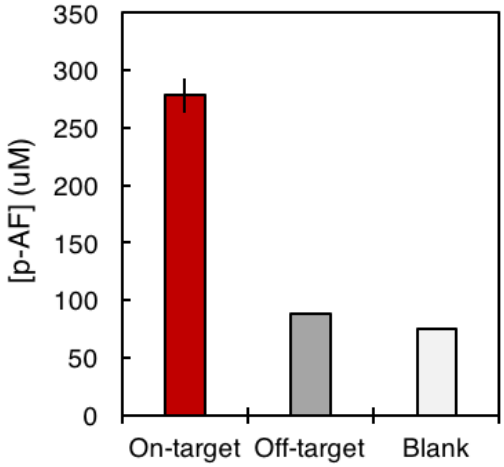


Figure 3.3. CRISPRa-controlled, plasmid-based *p*-AF production in *P. putida*. While the off-target control is only slightly elevated above the blank sample, CRISPR activation results in much higher amounts of pathway intermediate *p*-AF. The downstream enzyme PAL is not expressed in this system. The plasmid is on a medium-copy pBBR1 origin, with *aroGL* and *papABC* controlled by the synthetic activatable promoters J3 and J5, respectively. Error bars represent standard deviation of *n*=3 biological replicates.

From this baseline of *p*-AF production, we worked toward producing *p*-ACA, first with the *A. thaliana* PAL2, and later replacing it with *R. glutinis* PAL[75]. The difficulty of this process was the balance of burden contributed by plasmid-based genetics versus the need to express enough enzyme to produce detectable amounts of metabolite. The pathway including PAL was large enough to present difficulties fitting onto one stable

plasmid, *P. putida* is severely burdened by a second plasmid, and strengthening the base CRISPRa promoters in preparation for genomic integration (and its reduction in copy number relative to plasmid) proved difficult. Even suboptimal expression, however, was enough to produce small amounts of *p*-ACA, a first from a bacterial host.

3.3: Output copy number, burden, and integration

While the plasmid-based, CRISPRa-controlled *p*-AF pathway was producing up to 1.3 mM *p*-AF extracellularly (Figure 3.5), we sought to expand the circuit, through both additional enzymes (namely, PAL) and additional scRNAs and gRNAs with endogenous targets. Worried about the eventual size of the already-large pBBR1 plasmid, we therefore worked toward integrating the enzyme genes, driven by their synthetic CRISPRa promoters, while keeping the arbitrarily large scRNA/gRNA array on the plasmid for ease of adjustment. Given the eventual goal of several DBTL cycles optimizing the effects of these adjusted CRISPRa/i interventions, we considered this ease of adjustment to be an important design factor balancing the substantial burden of carrying a plasmid in *P. putida*. This burden is mitigated somewhat by limiting the size of the plasmid (Figure 3.4), so we aimed to integrate as much of the heterologous genetics as possible, aided by the *trans*-acting nature of scRNAs.

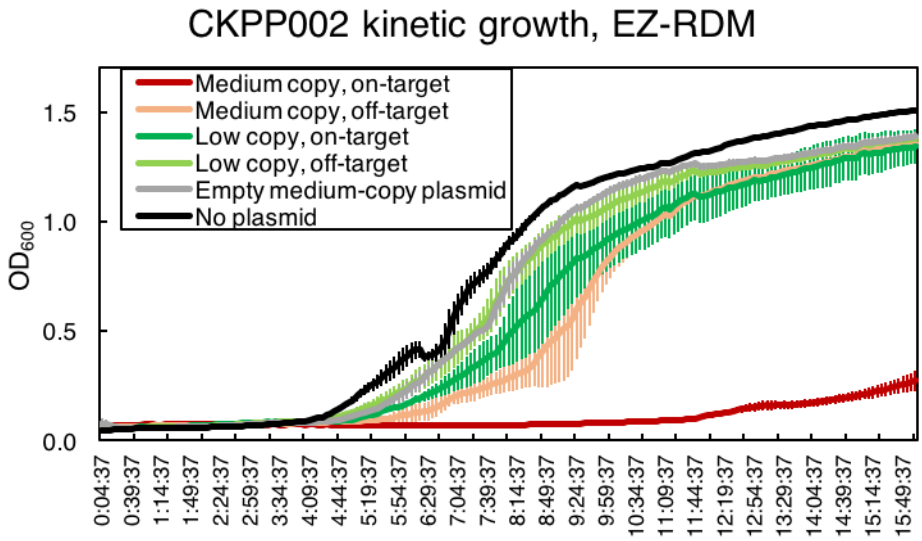


Figure 3.4. Effect of plasmid burden on kinetic growth of *P. putida*. Relative to an empty medium-copy plasmid (grey), the activation of the *p*-AF pathway genes borne on a medium-

copy plasmid (red) is burdensome enough to drastically impact growth. When unactivated, however, the impact is diminished (orange). The eventual goal of a scRNA-only plasmid will likely be even less burdensome than the medium-copy, off-target plasmid. We abandoned the low-copy plasmid (a mutation of pBBR1)[79] due to concerns about its stability arising from high production variability between replicates. Error bars represent standard deviation of *n*=3 biological replicates.

The problem with integration is the reduction in DNA copy number from the medium-copy plasmid to the single-copy genome: this reduction in gene dosage reduces overall expression levels, even when activated by CRISPRa. Upon integration of *papABC* and *aroGL*, we found that the lower expression level was not producing enough enzyme to produce measureable *p*-AF in the culture supernatant (Figure 3.5), confirming our suspicion that integration would require stronger base promoters within the synthetic CRISPRa promoters. Guided by a small promoter strength library controlling RFP expression by CRISPRa (Figure 3.S4), we attempted to clone versions of the pathway using J23110, J23106, and J23105 base promoters, combined with low-leak upstream sequences (between the spacer target and the base promoter). This presented another problem in turn: the integration process[80] required an initial step of plasmid cloning in *E. coli*, in which the copy numbers of even low-copy pSC101**, pBBR1, and pGNW, combined with the stronger base promoters and the substantial size of the output genes, was not tolerated by even cloning-specific *E. coli* strains.

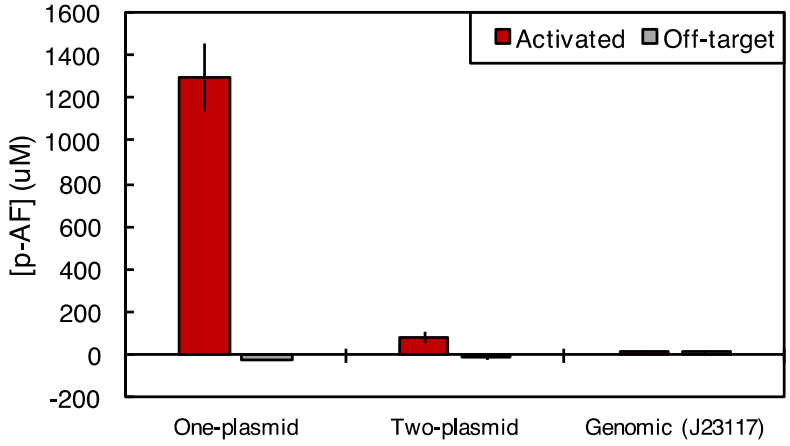


Figure 3.5. Effects of second-plasmid burden, and genomic copy number, on *p*-AF production. The burden of carrying the second plasmid is substantial in *P. putida*, especially using a

kanamycin resistance marker, reducing HPLC-measured *p*-AF production 16-fold in this experiment. Likely, this reduction is a result of lower heterologous enzyme expression due to lower global expression capacity. Also probably due to low enzyme levels, the same promoters used in plasmid-based production fail to produce any *p*-AF when integrated in the genome. In all cases, only *papABC* and *aroGL* are activated. Error bars represent standard deviation of *n*=3 biological replicates.

To address this problem, we devised a cloning workflow in which In-Fusion reactions were co-transformed with a "helper" plasmid consisting of dCas9 and a gRNA targeted to repress any output of a J23110 promoter (Figure 3.S3). Due to sequence similarities between J23110, J23106, and J23105 in particular, we reasoned that the same helper plasmid would sufficiently repress any of these promoters. This helper plasmid was maintained throughout the plasmid-cloning phase of integration, until eventual transformation into *P. putida*, before which it was restriction digested into non-replicable linear fragments. Combined with a highly competent *pir*+ cloning strain, this strategy resulted in successful cloning of the plasmid-based phase of the integration workflow, and successful transformation of the integration plasmid into the *P. putida* recipient strain. Production of *p*-AF and *p*-ACA by strains with J23110 and J23105 base promoters driving *papABC* and *aroGL* expression is shown in Figures 3.8 and 3.S11, though even the integration of these promoter strengths seems to be accompanied by genetic instability in some cases.

Worried about this copy number problem from the start, we simultaneously pursued an alternative strategy that used a second (pRK2 origin) plasmid, into which we cloned *A. thaliana* PAL2 under control of the J6 synthetic promoter, with an optional inclusion of *E. coli* TyrB in the same operon. We found that the burden of the second plasmid was not well-tolerated by *P. putida*, resulting in diminished growth rate and greatly reduced *p*-AF production (Figure 3.5). Despite the low concentration of its substrate, activation of PAL2 in this system resulted in a miniscule amount of *p*-ACA production (Figure 3.6), demonstrating the viability of even this suboptimal strategy. Because the extracellular *p*-ACA titers were so low, however, we continued optimization of the base pathway before implementing endogenous CRISPRa/i or iterating DBTL

cycles, aiming to have more certainty in the quantification of *p*-ACA production differences arising from these interventions.

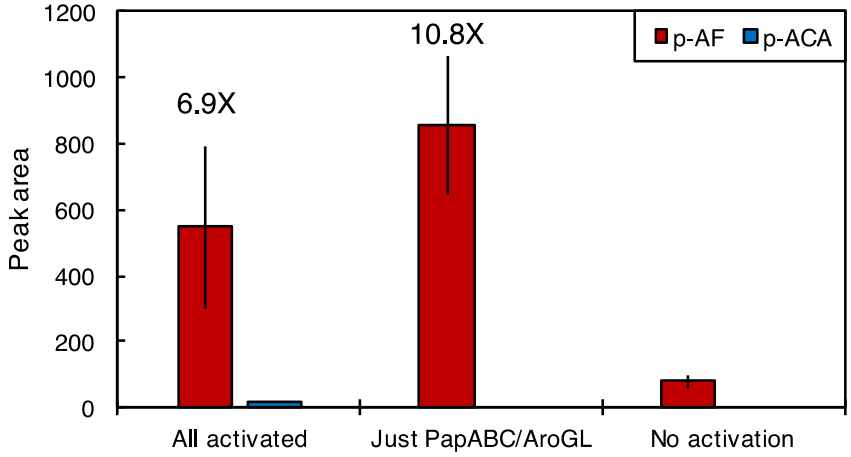
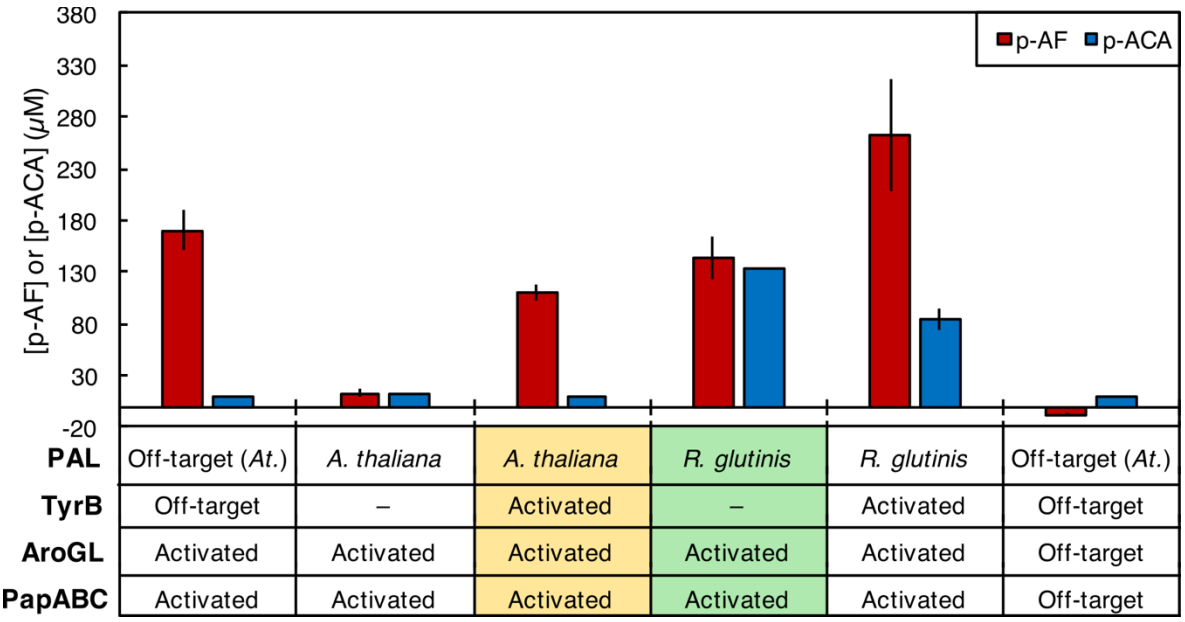


Figure 3.6. Production of *p*-ACA from the two-plasmid system in *P. putida*. Despite the small amount of *p*-AF substrate (Figure 3.5), activating *At*-PAL2 on the second plasmid results in a very small, but detectable, amount of *p*-ACA. Detection of *p*-ACA by this HPLC method is very sensitive due to the lack of interfering peaks in this area of the chromatogram; detecting even this amount of *p*-ACA is highly reproducible. The reduction in *p*-AF when PAL2 is activated could be in part due to substrate consumption, but is probably mostly due to increased burden from the additional enzyme expression. This graph is presented with peak area on the *y*-axis, to better distinguish the presence or absence of *p*-ACA. Figure 3.S5 shows the same data in molarity, but the produced *p*-ACA is hard to distinguish from the standard curve's zero point. Noted fold-changes are relative to the no-activation *p*-AF value. Error bars represent standard deviation of *n*=3 biological replicates.

3.4: Pathway engineering for *p*-ACA production improvement

Within the two-plasmid system, we replaced the *A. thaliana* PAL2 with a PAL enzyme from the yeast *Rhodotorula glutinis*, despite the system's low production of *p*-AF. It has been reported that *Rg*-PAL shows more promiscuous activity than *A. thaliana*'s similar enzyme PAL4, which is more specific to the native substrate phenylalanine[75]. We reasoned, therefore, that *Rg*-PAL might have more activity on our heterologous substrate *p*-AF than *At*-PAL2. Even with low amounts of that substrate, *Rg*-PAL indeed produced a substantial increase in *p*-ACA (Figure 3.7), resulting in a titer that inspires confidence for later quantification during DBTL cycles.



PAL	Off-target (*At.*)	*A. thaliana*	*A. thaliana*	*R. glutinis*	*R. glutinis*	Off-target (*At.*)
TyrB	Off-target	–	Activated	–	Activated	Off-target
AroGL	Activated	Activated	Activated	Activated	Activated	Off-target
PapABC	Activated	Activated	Activated	Activated	Activated	Off-target

Figure 3.7. *p*-ACA production using *R. glutinis* PAL. Changing the *At*-PAL2 ORF to that of *Rg*-PAL in the two-plasmid system results in dramatically increased *p*-ACA production. TyrB is only sometimes included in heterologous expression because there is also an endogenous copy. Its overexpression does seem to boost *p*-AF production, but its net effect on *p*-ACA production is less clear. The burden of another heterologous gene could be avoided if CRISPRa targeted to endogenous *tyrB* proves fruitful. Otherwise, a heterologous TyrB might be best expressed by a separate promoter than J6, as its inclusion in that operon could reduce PAL expression. The original *p*-ACA production strain (from Figure 3.6) is shaded in yellow, while the best two-plasmid producer is shaded in green. Error bars represent standard deviation of *n*=3 biological replicates.

If we may assume a proportional increase in *p*-ACA production as in *p*-AF production (the 16-fold increase seen in Figure 3.5), this new, *Rg*-PAL-containing pathway worked into a one-plasmid μM system would theoretically predict *p*-ACA titers reaching into the millimolar range, even before flux optimization by endogenous CRISPRa/i. Upon building multiple versions of one-plasmid *p*-ACA production strains, we found *p*-ACA titers to be not quite so dramatic, but still an improvement over the two-plasmid system. The strategies for constructing this system were either a large plasmid containing scRNAs, *papABC, aroGL*, and *Rg*-PAL, but excluding *tyrB*; or a smaller plasmid containing only scRNAs and *Rg*-PAL, used in the one of the strains with *papABC* and *aroGL* integrated, driven by either the J23110 and J23105 base promoters (Section 3.3). Avoiding the burden of maintaining and replicating the second plasmid

resulted in up to four-fold improvement of *p*-ACA titer, so this one-plasmid strain will be the basis for production improvement through endogenous CRISPRa/i during DBTL cycles. Early indications of the stability of the large plasmid containing the entire *p*-ACA pathway are encouraging, though it remains to be seen how well additional scRNAs/gRNAs are tolerated, especially under the potential metabolic stress provided by endogenous manipulation.

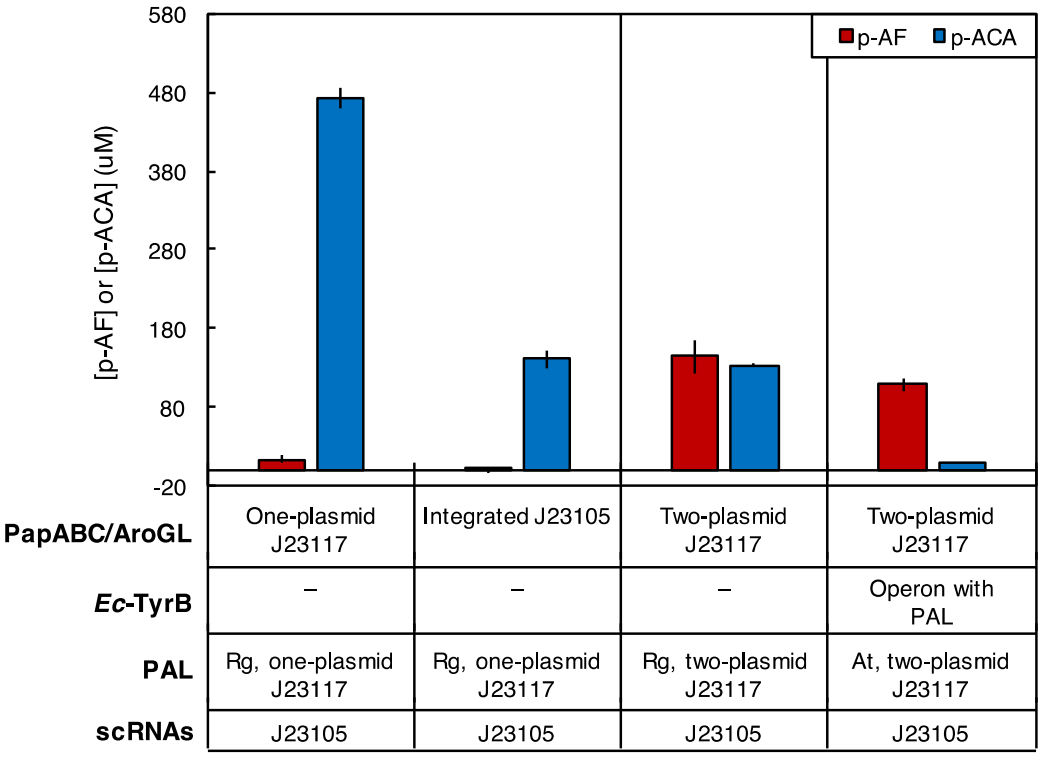
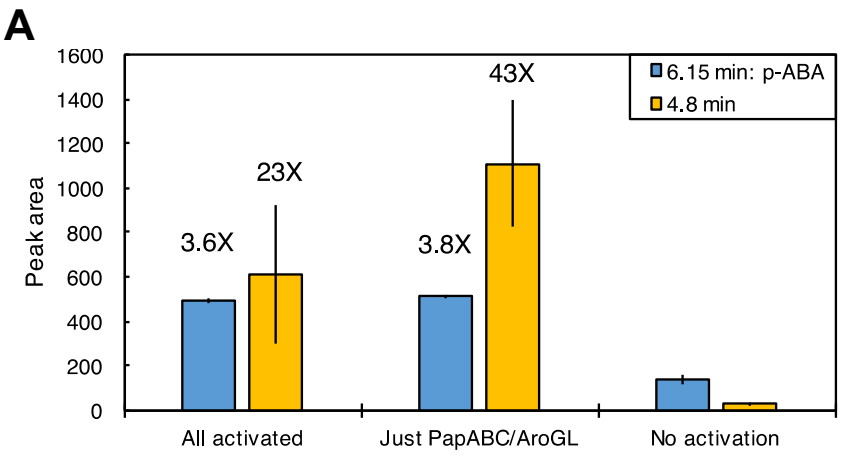


Figure 3.8. Summary of one-plasmid *p*-ACA production using *Rg*-PAL. Compared to the minimal *p*-ACA production by the two-plasmid *At*-PAL system (right) and the modest production by the two-plasmid *Rg*-PAL system (middle right), the one-plasmid *Rg*-PAL systems can produce one- to four-fold higher *p*-ACA titer if they can remain genetically stable (see Figure 3.S11). At far left, the entire pathway on one plasmid produces relatively high *p*-ACA titer—and leaves little *p*-AF unconverted, suggesting that enzyme stoichiometry is well-balanced at these expression levels. At middle left, *papABC* and *aroGL* are integrated and driven by CRISPR-activated J23105 promoters, resulting in *p*-ACA production on par with the two-plasmid system. See Figure 3.S4 for relative promoter strengths in *P. putida*. Error bars represent standard deviation of *n*=3 biological replicates.

To be sure we had indeed identified *p*-ACA as our product, and that its accumulation in culture supernatant was stable, we performed numerous follow-up

experiments verifying the peak location within the chromatogram (Figure 3.S6), the lack of consumption of extracellular *p*-ACA by growing *P. putida* cultures (Figure 3.S7), and the lack of *p*-ACA toxicity even in a strain with the putative catabolic pathway knocked out (Figure 3.S8).

Not only does HPLC indicate *p*-ACA production, it reveals several side products whose titers are increased by CRISPR activation of *papABC/aroGL* (Figure 3.9A) or of either PAL (Figure 3.9B). These metabolites are produced by heterologous enzymes acting on endogenous substrates, or by endogenous enzymes acting on heterologous substrates, especially the pathway intermediates. We can make use of the independent activation of each promoter to determine which side products are associated with each heterologous gene expression, finding that metabolites with peaks occurring at 4.8 minutes and 6.1 minutes in the HPLC chromatogram are produced by PapABC and/or AroGL, while a metabolite with a 17.7-minute peak is produced by PAL. Interestingly, they all undergo significant CRISPR activation even in the weak (J23117) integrated strain (Figure 3.S9), leading us to believe that even in that system there is enough enzyme to affect overall metabolism, and suggesting that reducing side pathway activity through endogenous CRISPRa/i could result in *p*-ACA production even from that non-producing strain (Figure 3.5). Regardless of whether we can achieve production from the J23117 integrated strain, we aim to use endogenous CRISPRa/i to improve a strain that has already demonstrated production: either the two-plasmid system or one of the one-plasmid alternatives.



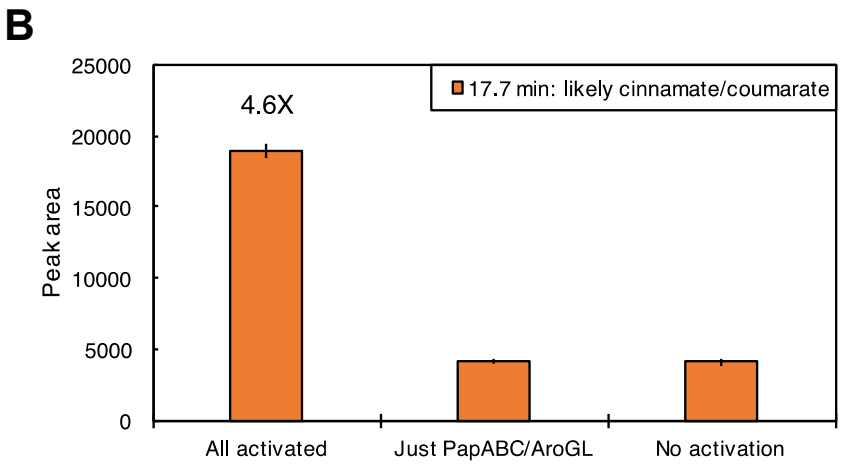


Figure 3.9. Side products generated by (A) *papABC/aroGL* activation or (B) PAL activation. Observation of HPLC chromatograms revealed several non-*p*-AF, non-*p*-ACA peaks that were nonetheless increased by CRISPR activation. These side products are clearly associated with either PapABC/AroGL (4.8-minute and 6.1-minute peaks) or PAL (17.7-minute peak) activity, and could represent lost pathway flux—and therefore reveal endogenous CRISPRi targets—if they are the result of endogenous enzymes acting on heterologous intermediates. Error bars represent standard deviation of *n*=3 biological replicates.

3.5: Endogenous CRISPRa/i for *p*-ACA production improvement

Once a base strain is optimized (to the point where it's stable, produces a reasonably-quantifiable amount of *p*-ACA, and easily accepts changes to the scRNA/gRNA program), the overall aim of this project is to improve production through iterative improvement of the CRISPR program responsible for modulating endogenous metabolism. The detection of so much side production in the heterologous-only pathway suggests the potential for substantial improvement, because there is so much metabolic flux adjacent to the desired pathway. Whether the side products we have detected arise from endogenous enzymes or endogenous substrates, they can provide clues for rational selection of endogenous CRISPRa/i targets. For example, we have determined through spike-in experiments that the 6.1-minute peak corresponds to *p*-aminobenzoic acid, likely resulting from endogenous PabC acting on a heterologous intermediate and competing with PapB for that substrate. It is reasonable to expect, then, that *pabC* is a high-priority target for CRISPRi. Such a knockdown would relieve some of the competition with PapB and redirect metabolic flux into the heterologous

pathway. Efforts are currently ongoing to identify the other side products, and beyond even those targets, we envision a wide array (Figure 3.1, blue names) of endogenous CRISPRa/i that will work in combination to improve product titer.

3.6: Circuit size considerations

It is difficult to predict the size of an array of endogenous CRISPRa/i scRNAs/gRNAs that will optimize *p*-ACA production, but Chapters 1 and 2 give us a measure of confidence that it will not be limited to a small size due to the metabolic burden of expressing so many RNA regulators. Other factors that are likely to limit its size include competition for dCas9 binding and the metabolic effects of the interventions themselves. To investigate the former's effect, and perhaps to quantify the burden-limited size of a CRISPRa/i circuit in this real-world implementation, we propose an experiment observing the effects of an arbitrary number of off-target scRNAs/gRNAs on both a CRISPR-activated reporter gene and a different constitutive reporter gene. The former reporter will determine the circuit size's effect on CRISPR functionality, while the latter will determine the circuit size's effect on overall expression capacity (and normalize this effect out of the CRISPR-specific effect).

3.7: Conclusions

Turning to a more real-world control circuit than in previous chapters, with pathway enzymes as outputs instead of reporter proteins, we see the utility and portability of CRISPR-based control. Not only is *p*-AF production on par with tet-inducible control, we demonstrate *p*-ACA production in a bacterial host, made possible by porting the circuit to a host better-suited to the pathway chemistry. That would have required a different transcription factor in the new host, and, compared to TF control, orthogonal CRISPR-activatable promoters allow for more independent control of individual operons and endogenous targets, while still retaining the ability to use dCas9 or MCP-SoxS expression as a master regulator. We can make use of this independence to tune enzyme stoichiometry within our heterologous pathway, and to rationally prioritize endogenous CRISPRa/i targets based on observed side products.

And our sense of the expression cost of additional scRNAs/gRNAs suggests that we should be able to build quite a wide circuit encompassing heterologous and endogenous targets, even in a non-model host, though we have yet to quantify the permissible size in this implementation.

Another potential pitfall of large CRISPR-controlled circuits is competition between RNAs for binding to dCas9, with a recent report noting a ten-fold reduction in CRISPRi efficacy when coexpressing 5-10 gRNAs due to this effect[13], though CRISPRa efficacy may be more resistant[81]. To try to boost this circuit size, autoregulation of CRISPR activity is an option[49], and importantly can be controlled by CRISPR itself, or we could attempt to equalize dCas9 (and MCP) affinity between different scRNAs/gRNAs—which might differ in their ensemble fractions that properly fold, due to their different spacer sequences—by incorporating different sequences in the dCas9 handle that modulate affinity[9,82]. Observations of very low CRISPRa fold-activation at high base promoter strengths could form the basis of a small autoregulatory boost to unbound shared component levels when reduced by binding competition.

This system, then, provides a useful testbed for the idea that CRISPR-based control circuits are well-suited to a large size with many nodes—as does the circuit controlling LNT production in Chapter 4. Because that circuit is implemented in an *E. coli* host, comparison of the two may reveal organism-dependent differences of expression capacity and therefore circuit size limitation.

3.8: Methods

Plasmid and strain construction

All plasmids were constructed using either circular polymerase extension cloning (CPEC), Infusion (Takara Bio), or Golden Gate (NEB). All cloning-related transformations were performed with *E. coli* DH10b cells, and DNA oligonucleotides were purchased from Integrated DNA technologies or Genewiz. Sanger sequencing to verify constructs was performed by Eurofins Genomics or Genewiz. DNA encoding the *papABC* operon responsible for *p*-AF production was taken from *P. fluorescens* SBW25[78]. The *aroGL* operon responsible for elevated chorismate levels, encoding AroL and a feedback-resistant mutant of AroG, were obtained from a previous report[83]. The *A. thaliana* PAL2 was obtained from a previous report[42],

while the *R. glutinis* PAL was synthesized by Twist Bioscience. Promoter sequences placing these outputs under CRISPRa control were obtained from Jason Fontana[9]. Plasmid backbones containing p15A or ColE1 origins were taken from BglBrick plasmids[39], gifts from Jay Keasling (Addgene plasmids 35308, 35280, 35316, 35322, and 35279). RBS sequences were generated using the RBS Calculator v2.0[84]. Integrations were achieved using an R6K delivery plasmid and an I-*Sce*I-mediated protocol adapted from a previous report[80], with *sacB* included as a counterselection marker.

Production assay in *E. coli*

E. coli DH10b chemically competent cells were heat-shock transformed with pathway and CRISPR plasmids, while a control strain was transformed with similar plasmids not containing pathway genes and used for standard curve diluent. Single colonies were picked in triplicate and used to inoculate 2 mL of MOPS EZ-Rich defined media (Teknova), supplemented with 10% v/v LB (Miller) (Teknova) and appropriate antibiotics, in 14 mL polypropylene culture tubes. Cultures were grown at 30˚C and shaken at 200 rpm for 24 hours.

Production assay in *P. putida*

Constitutive dCas9 and MCP-SoxS were previously integrated into *P. putida* KT2440 to make CKPP002. This strain, or its derivative IFPP006 with integrated *papABC* and *aroGL*, was transformed by electroporation with a pBBR1 plasmid containing either scRNAs only or pathway genes and scRNAs. Two-plasmid production strains including the additional pRK2 plasmid were doubly transformed in series, using competent cells containing the *papABC*/*aroGL* plasmid. A control strain for standard curve diluent was transformed with similar plasmids not containing pathway genes. Single colonies were picked in triplicate and used to inoculate 2 mL of MOPS EZ-Rich defined media (Teknova), supplemented with appropriate antibiotics, in 14 mL polypropylene culture tubes. Cultures were grown at 30˚C and shaken at 200 rpm for 24 hours.

HPLC for production analysis

Culture supernatants were filtered by centrifuging at 14000 g for 20 minutes using an Amicon Ultracel 10K centrifuge filter (Millipore). Filtered supernatants were supplemented with 0.2% trifluoroacetic acid (TFA) and assessed using an Agilent HPLC with a diode array detector set at 210 nm. *p-AF*, *p*-ACA, and other components were separated using a ZORBAX Eclipse Plus phenyl-hexyl column (Agilent) with water plus 0.2% TFA as solvent A and methanol plus 0.2% TFA as solvent B. The mobile phase gradient was as follows: 100% solvent A at 1 mL/min from 0 to 4 min, ratio increased to 95% solvent B and 5% solvent A at 1 mL/min from 4 to 18 min,

100% solvent A at 1 mL/min from 18 to 22 min. Sample concentrations were determined by interpolation from standard curves ranging from about 10 µM to 1 mM. *p*-AF, *p*-ACA, and *p*-ABA for standard curves were obtained from Santa Cruz Biotechnology.

***p*-ACA toxicity and kinetic growth experiments**

Overnight LB (Teknova) cultures of *E. coli* MG1655 and *P. putida* KT2440 were inoculated 1:100 into fresh MOPS EZ-Rich defined media (Teknova) and allowed to reach an OD_{600} of 0.05 – 0.1, as measured in a cuvette with 1 cm pathlength on a Nanodrop spectrophotometer. 196 uL of culture was then placed into wells of a black 96-well plate with a clear bottom (Corning) and 4 uL of DMSO containing varying amounts of *p*-ACA were added to each culture. Noted concentrations are final *p*-ACA concentrations in the cultures. Final DMSO concentration in media (vol/vol) was kept at 2% to equalize DMSO toxicity. The plate was then placed into a Biotek Synergy HT plate reader with orbital shaking set to medium and temperature set to 30˚C for 16 hours. OD_{620} or OD_{600} measurements were taken every 5 minutes. Error bars are standard deviation across three platewell replicates from the same overnight culture. Percent reduction was calculated for each timepoint as (OD_0mM - OD_20mM) / OD_0mM within each replicate and then averaged.

3.9: Supplementary Figures

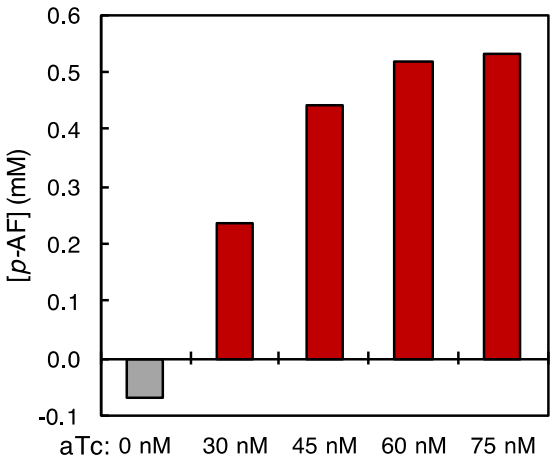


Figure 3.S1. Tet-inducible *p*-AF production in *E. coli*. A two-plasmid, medium-copy system bears the *aroGL* and *papABC* operons, both expressed by tet promoters. An induction delay allows cultures to enter log phase before expressing the pathway enzymes. Induction response, as measured by *p*-AF-detecting HPLC, reaches maximal levels at about 60 nM aTc. This experiment was performed in DH10B cells to maximize production.

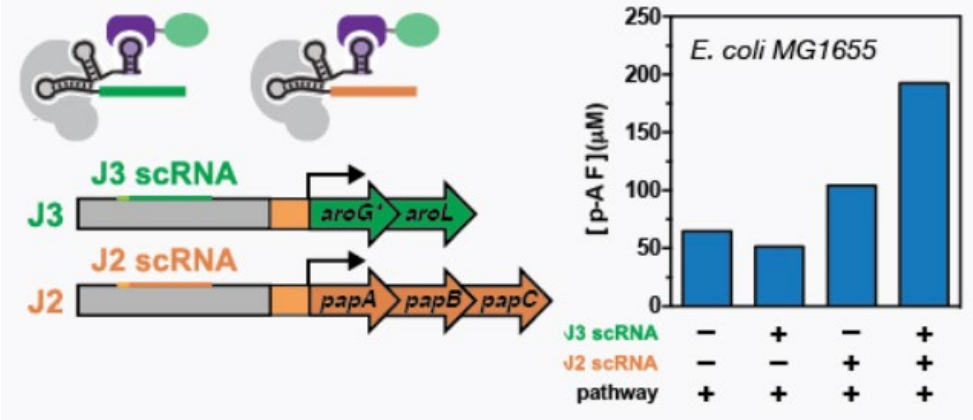


Figure 3.S2. CRISPRa-controlled *p*-AF production in *E. coli*. The same operons as in Figure 3.S1 are arranged under the control of J3 and J2 synthetic promoters, and are controlled by constitutive CRISPRa (removing the induction delay). HPLC-detected *p*-AF production increases when *papABC* is activated, especially when *aroGL* is activated as well. This experiment was performed in MG1655 cells to be sure of CRISPRa function.

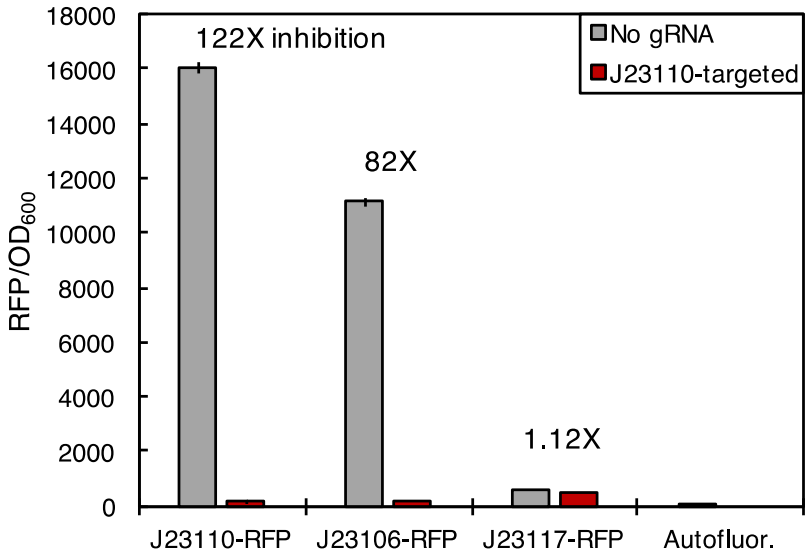


Figure 3.S3. A self-contained CRISPRi plasmid to suppress strong-promoter output. This is a separate plasmid intended to aid cloning reactions boosting the base promoter strength of our synthetic CRISPRa promoters, for use as a co-transformation to reduce output burden from the plasmid being cloned. The gRNA spacer sequence targets the base promoter itself, achieving repression regardless of output gene. Here, we test its effect on unactivated levels of such promoters expressing RFP, observing almost complete knockdown to autofluorescence, except

in the more-spacer-mismatched case of J23117. Error bars represent standard deviation of *n*=3 biological replicates.

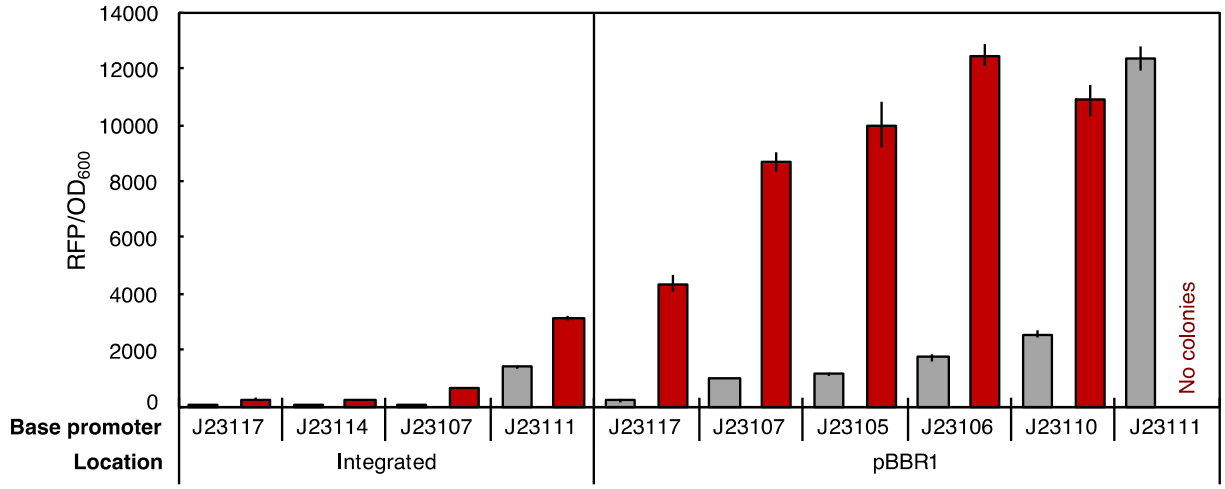


Figure 3.S4. RFP activation from integrated synthetic CRISPRa promoters with varying strength. Activated output levels were used as a guide for selecting stronger base promoters than J23117 for integrated *papABC* and *aroGL* expression, to counteract the lower genomic copy number than pBBR1. Something between integrated J23107 and J23111 (left side) would be reasonably close to the pBBR1–J23117 combination that produces about 1 mM *p*-AF in a one-plasmid system. Plasmid-based output (right side) suggests that J23105, J23106, and J23110 would be good candidates for integration, because they are between J23107 and J23111 and keep unactivated expression relatively low. Error bars represent standard deviation of *n*=3 biological replicates.

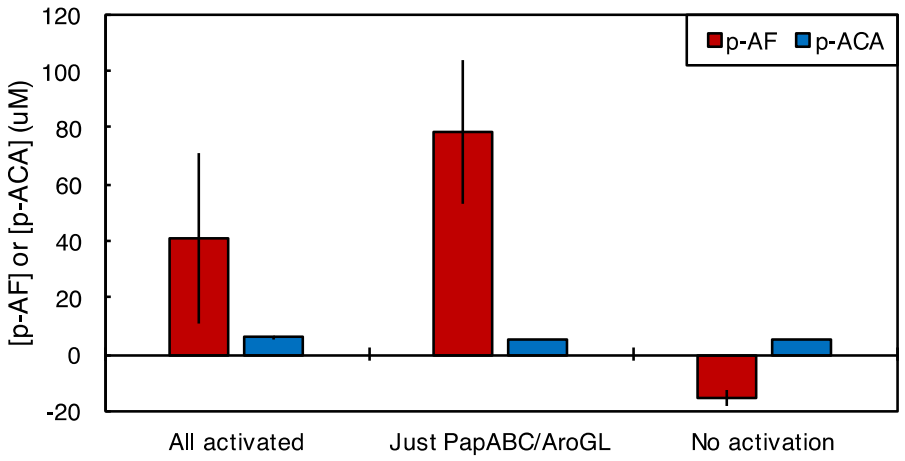


Figure 3.S5. Two-plasmid *p*-ACA production using *At*-PAL2. Shown with molarity on the *y*-axis, it is difficult to discern *p*-ACA production when PAL2 is activated (left). The standard curve fit results in non-zero molarity values predicted from samples with zero peak area. *p*-AF molarity more clearly shows a decrease upon PAL2 activation, probably due to a combination of added expression burden and substrate consumption. Error bars represent standard deviation of *n*=3 biological replicates.

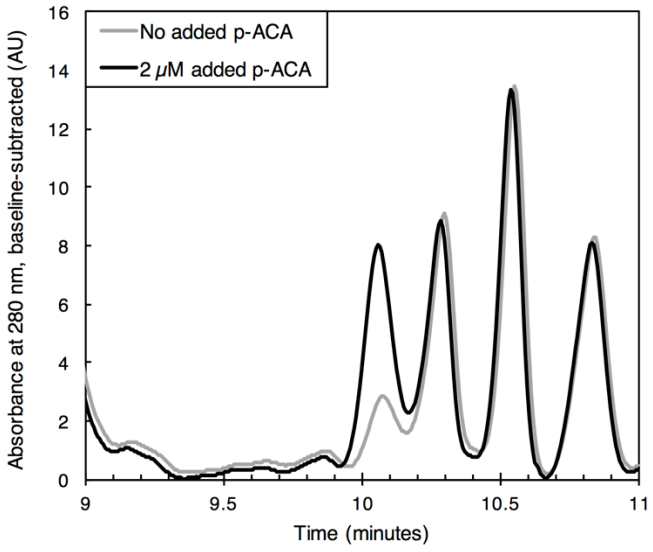


Figure 3.S6. Overlaid HPLC chromatograms of *p*-ACA production and low-concentration spike-in. Initial, low-concentration production of *p*-ACA (grey) was verified by superimposition of a larger peak (black) after spiking a small amount of known *p*-ACA (2 μM) into the same supernatant. *p*-ACA is observed at 10.5 minutes.

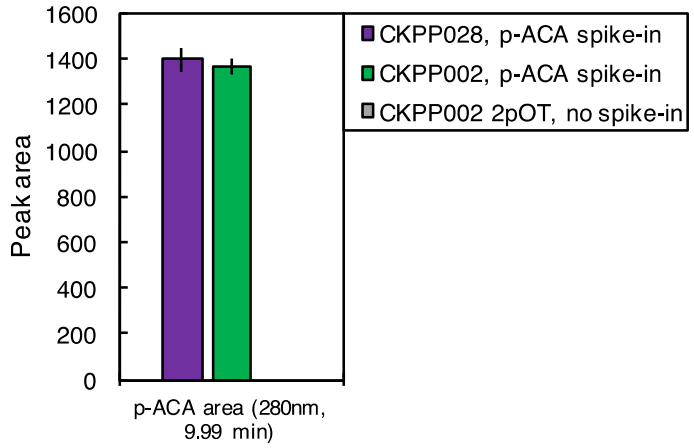


Figure 3.S7. Lack of *p*-ACA consumption by either knockout or intact strain. We hypothesize that *p*-ACA, if catabolized by *P. putida*, would occur through the *p*-coumarate pathway. The knockout strain, then, lacks *fcs, ehs,* and *vdh*. Here, we find that extracellular *p*-ACA is not consumed by either the knockout CKPP028 or the intact CKPP002 when supplied to the media at the time of inoculation, and measured by HPLC after a day of growth. Barring issues of transport, this indicates that our production strain is not likely catabolizing produced *p*-ACA. The control strain (grey) is the intact strain bearing the two *p*-ACA pathway plasmids, but unactivated, and with no extracellular *p*-ACA. Error bars represent standard deviation of *n*=3 biological replicates.

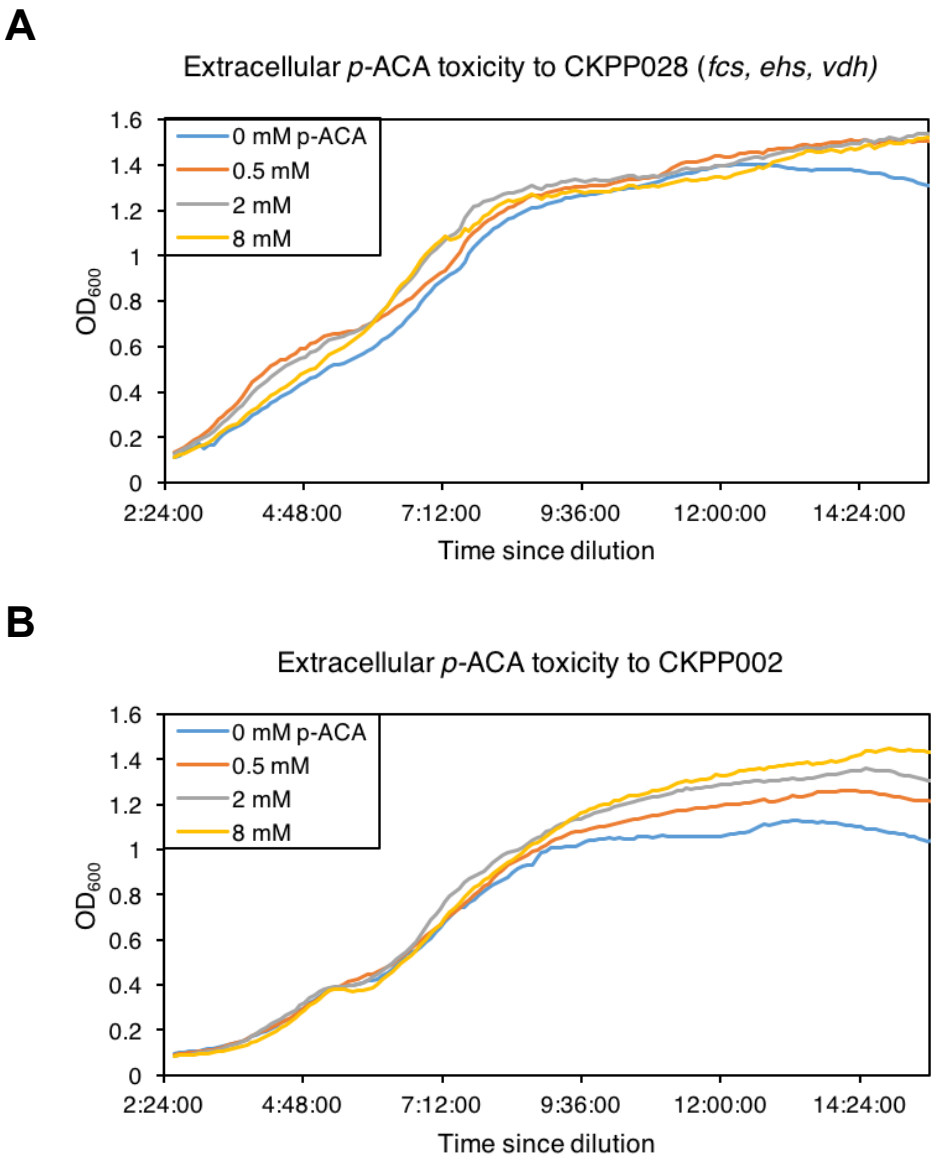


Figure 3.S8. Lack of *p*-ACA toxicity in the knockout strain. Using the same strain described in Figure 3.S7, we verify that catabolism is not the mechanism of *P. putida*'s lack of *p*-ACA toxicity. As in Figure 3.2, extracellular *p*-ACA is supplied in the media at the indicated concentrations, and cultures are grown in a kinetic growth assay in a plate reader. **A)** Generally similar growth curves indicate no toxic effect of *p*-ACA on the knockout strain CKPP028. **B)** In the non-knockout strain CKPP002, endpoint OD600 increases slightly with *p*-ACA concentration; however, we have seen that *P. putida* cannot grow with *p*-ACA as a sole carbon source (data not shown). The observed difference here is likely due to plate effects: namely, excess evaporation from the wells on the door-side edge of the plate. Error bars are omitted here for clarity.

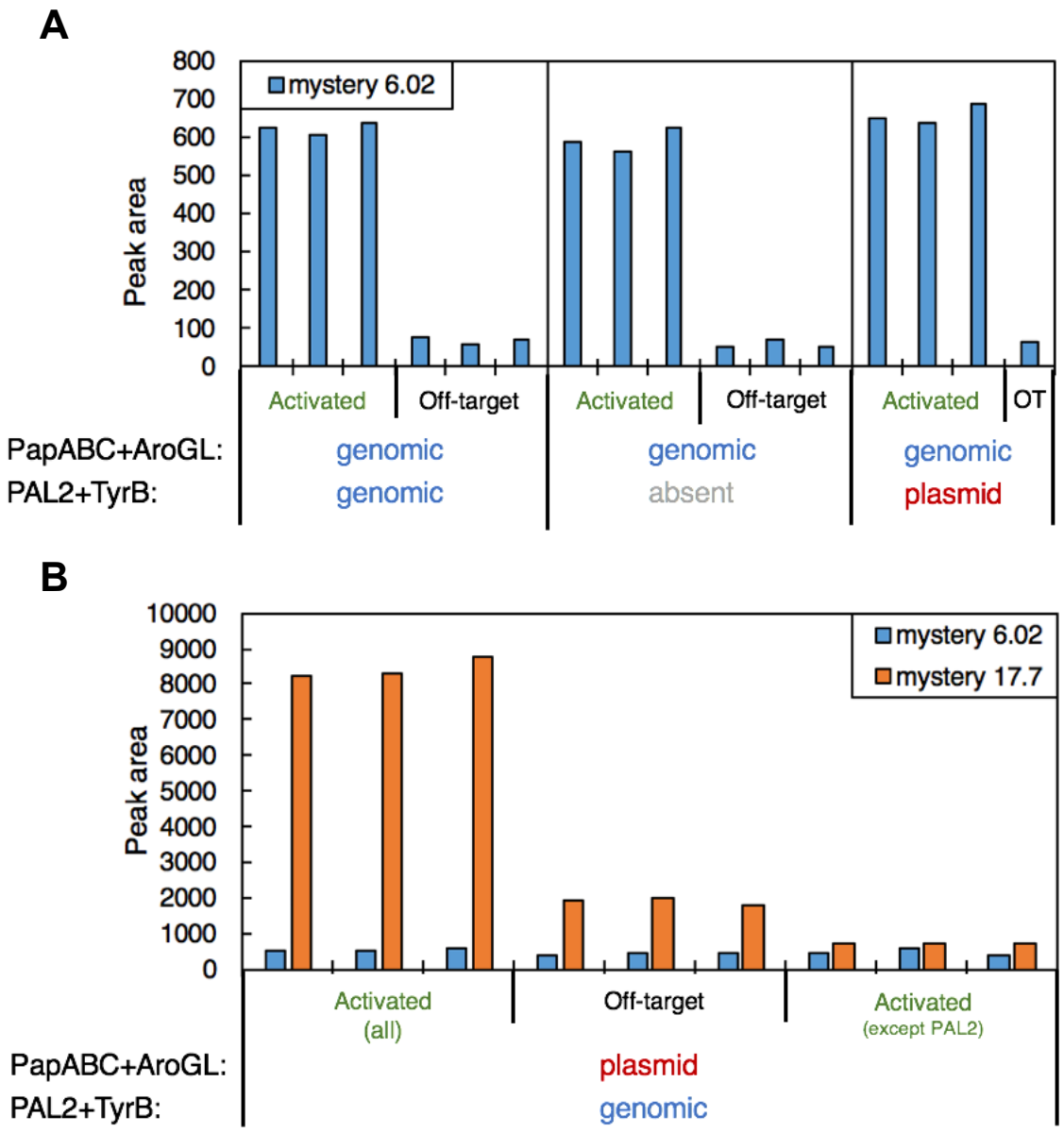
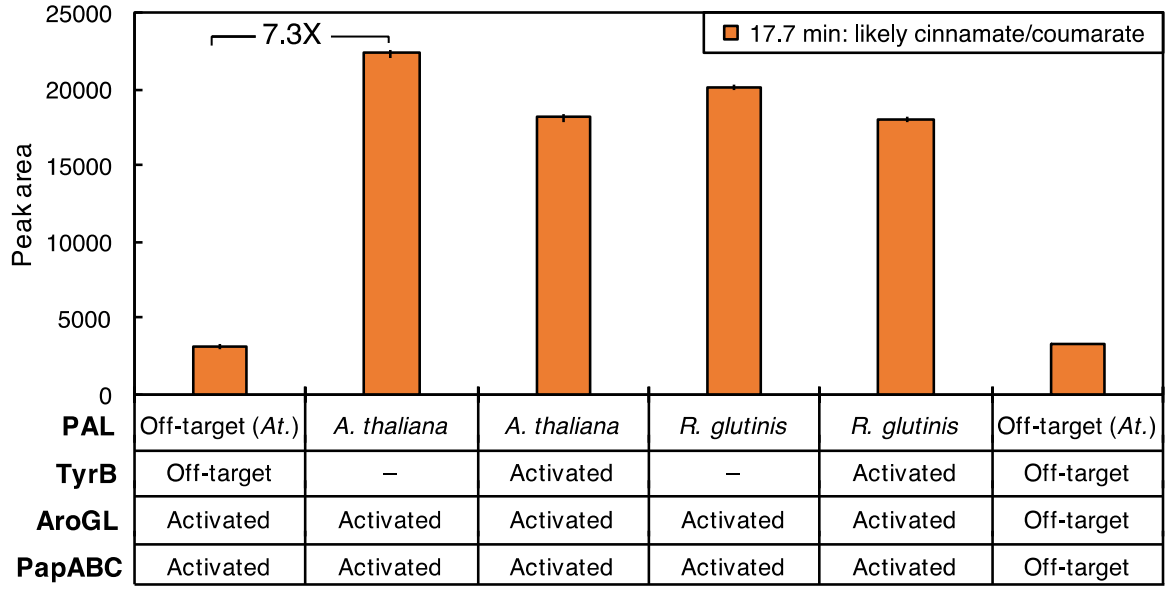
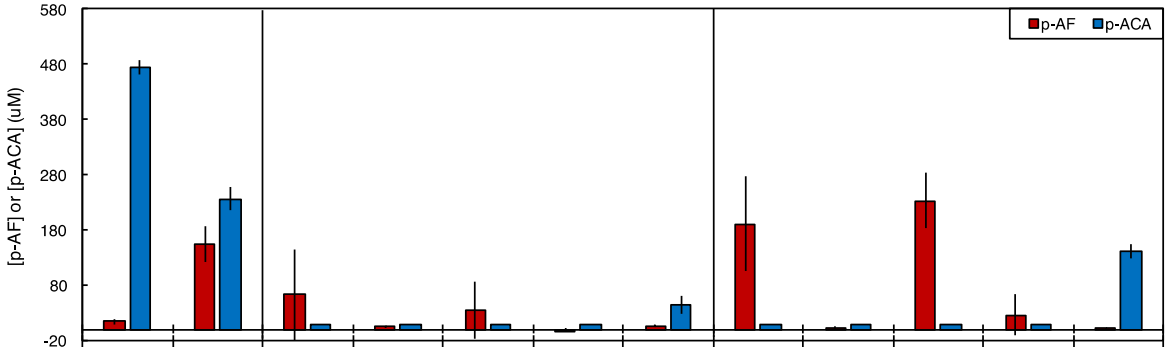


Figure 3.S9. Weak integrated strain nonetheless activates side products. Despite the low gene copy number, which prevents production of desired product (Figure 3.5), significant amounts of side products accumulate as a result of heterologous gene activation. Here we see activation of *p*-aminobenzoic acid, the 6-minute peak (A), and the 17-minute peak (B), again associated with PapABC/AroGL and *At*-PAL2, respectively. This suggests that reducing endogenous substrates or enzymes through endogenous CRISPRa/i could result in more heterologous product, if it frees the heterologous enzymes to act upon the main pathway substrates. Three biological replicates are shown separately.



PAL	Off-target (*At.*)	*A. thaliana*	*A. thaliana*	*R. glutinis*	*R. glutinis*	Off-target (*At.*)
TyrB	Off-target	–	Activated	–	Activated	Off-target
AroGL	Activated	Activated	Activated	Activated	Activated	Off-target
PapABC	Activated	Activated	Activated	Activated	Activated	Off-target

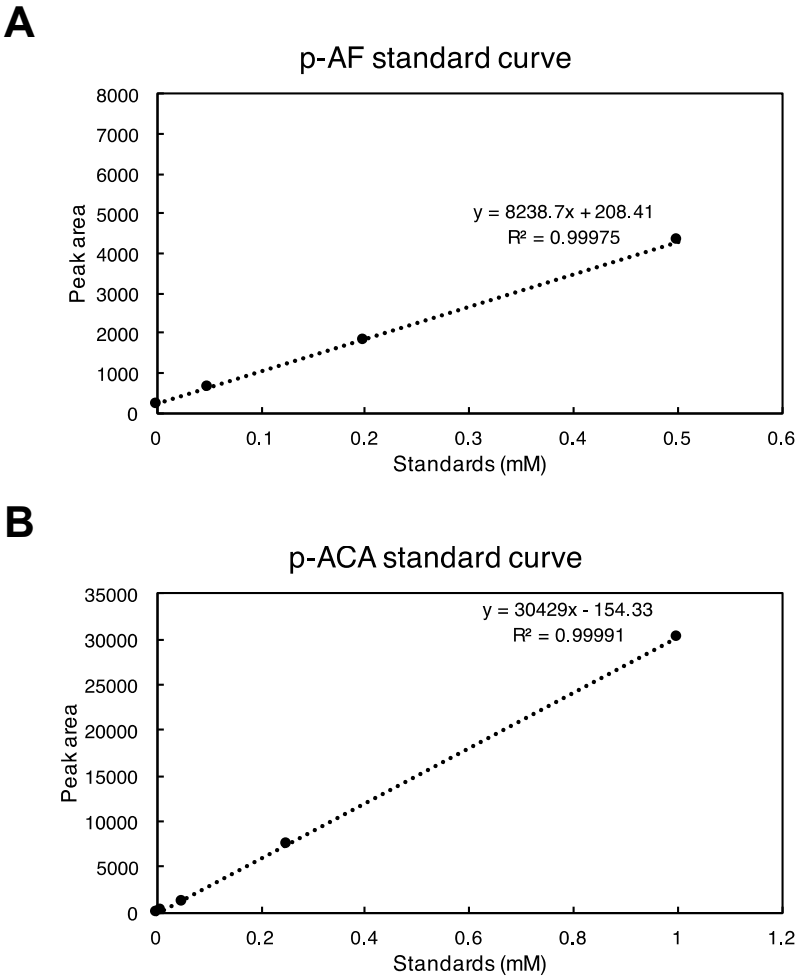
Figure 3.S10. Side product generated by different PAL enzymes. Despite its greater activity on *p*-AF substrate, *Rg*-PAL does not produce any less of its main side product (from some endogenous substrate, likely phenylalanine) than *At*-PAL does in a two-plasmid system. This suggests that promiscuity of substrate activity is potentially the most important determinant of *p*-ACA production at this enzyme expression level—that is, the PAL is not saturated to its V_{max}. Error bars represent standard deviation of n=3 biological replicates.



PapABC/AroGL	Plasmid J23117	Plasmid J23117	Integrated J23110	Integrated J23110	Integrated J23110	Integrated J23110	Integrated J23110	Integrated J23105	Integrated J23105	Integrated J23105	Integrated J23105	Integrated J23105
***Ec*-TyrB**	–	Operon with PAL	–	–	–	–	–	–	–	–	–	–
***Rg*-PAL**	Plasmid J23117	Plasmid J23117	–	–	–	–	Plasmid J23117	–	–	–	–	Plasmid J23117
scRNAs	J23105	J23105	J23105	Off-target	J23110 (x3)	J23110 (x5)	J23105	J23105	Off-target	J23110 (x3)	J23110 (x5)	J23105

Figure 3.S11. One-plasmid *p*-ACA production using *Rg*-PAL. With the entire pathway contained on one large plasmid driven by CRISPR-activated J23117 promoters (left), *p*-ACA

production is maximal at 480 μM, but drops off considerably when TyrB is included in an operon with PAL. This is presumably because of effects on PAL enzyme levels, because it seems that conversion of *p*-AF to *p*-ACA is the most diminished factor. When *papABC* and *aroGL* are integrated and driven by CRISPR-activated J23105 promoters (right), *p*-AF production exceeds that of a two-plasmid system by about two-fold, and *p*-ACA production is on par with the two-plasmid system when *Rg*-PAL is added. When *papABC* and *aroGL* are integrated and driven by CRISPR-activated J23110 promoters (middle), genetic instability is observed, resulting in minimal and variable production across replicates. Seemingly the extra burden of the J23110 promoter strength overwhelms the stability of the strain's production, but it is unclear whether this instability manifests at the scRNA plasmid or at the integration site. See Figure 3.S4 for relative promoter strengths in *P. putida*. Error bars represent standard deviation of *n*=3 biological replicates.



C

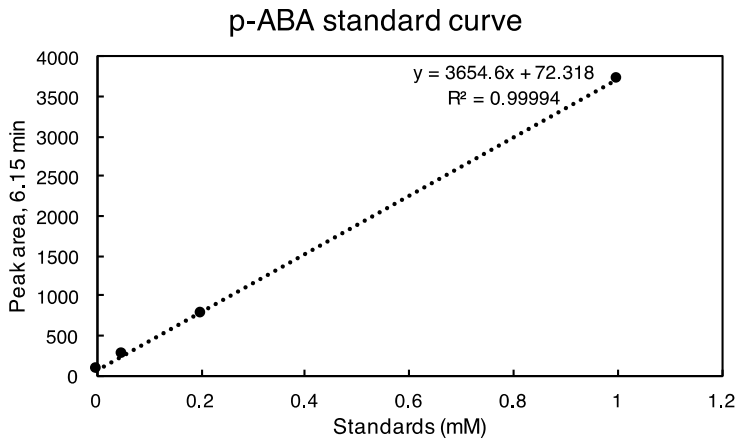


Figure 3.S12. Typical standard curves for HPLC quantification of (A) *p*-AF, (B) *p*-ACA, (C) *p*-ABA. While *p*-AF and *p*-ACA are heterologous metabolites, *p*-aminobenzoic acid is a side product seemingly resulting from endogenous PabC acting on heterologous intermediate 4-amino-4-deoxychorismate (Figures 3.8, 3.S9, and 3.S10). We used the standard curve in (C) as a spike-in experiment to confirm its identity in the chromatogram in which CRISPR activation was apparent.

Chapter 4: Production and detection of LNT by a ligand-responsive scRNA

Abstract

Human milk oligosaccharides (HMOs) are an abundant component of human milk, a feature unique among mammals. Their positive effects on infant development, both confirmed and putative, combined with their difficulty of production or extraction by traditional means, lead to a demand for microbially-produced HMOs. Here we describe CRISPRa control of a pathway capable of producing one of the most abundant core structures within the HMO family, lacto-*N*-tetraose (LNT), and take advantage of CRISPRa features to screen variants of pathway enzyme expression and stoichiometry to optimize production. Further, we describe the development of an *in vivo* biosensor, combining an RNA aptamer with a CRISPRa scRNA, capable of reporting intracellular production of LNT. This sensor represents another direction for CRISPR control circuit expansion, one that opens possibilities of high-throughput strain screening, dynamically-controlled pathway expression, dynamic multi-layered control circuits, or production improvement through auxotrophic selection.

4.1: Introduction

Oligosaccharides that are more complex than lactose are present in great variety in human milk[85,86], and are formed by elongation of lactose by glycosyltransferases into diverse linear or branched oligosaccharide structures[87], totaling more than 200 different molecules[88]. These human milk oligosaccharides (HMOs) are the third most abundant solid component of human milk[85] at up to 15 g/L, and are thought to have important effects on the health of the developing infant. These effects have little to do with direct nourishment of the infant[89] and are thought to relate more to immune development[90,91], microbiome health[92,93], and pathogen inhibition[94].

Milk from common farm animals, which—especially bovine milk—is the basis of most infant formula, contains drastically lower levels of oligosaccharides than human milk, and lacks the diversity and complexity of HMO structure[95–97]. Similarly, chemical synthesis suffers from complex protocols often involving reactive group protection and deprotection, resulting in high expense and low yield[88]. The lack of natural sources of HMOs, then, creates a demand for microbially-produced HMOs, not only for research purposes, but to be included in infant formula that is better-matched to breast milk[98].

Fortunately, oligosaccharide structures are often similar between HMOs and bacterial surface polysaccharides[99,100], so there are many bacterial glycosyltransferases that can be expressed in bacterial hosts to produce HMOs, usually from a lactose feedstock[101] easily derived from bovine or other sources. Specifically, certain glycosyltransferases can make use of an *E. coli* host's endogenous substrates UDP-galactose (UDP-Gal) and UDP-*N*-acetylglucosamine (UDP-GlcNAc) to synthesize the common HMO lacto-*N*-tetraose[102], a valuable formula additive and often a core structure for synthesis of more complex HMOs.

In this work, we construct an LNT synthesis pathway in *E. coli* under the control of CRISPR activation and explore the effects of relative pathway enzyme expression levels. Additionally, because product titer depends in part upon the endogenous substrates UDP-Gal and UDP-GlcNAc[103], we aim to investigate an array of endogenous CRISPRi interventions aimed at boosting the availability of those substrates[104,105].

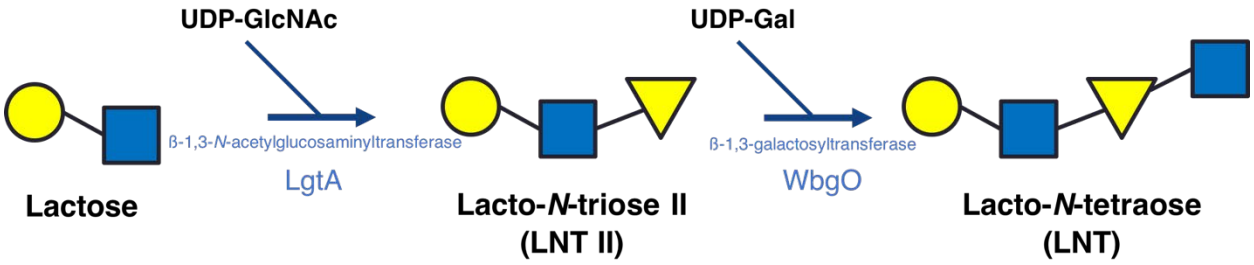


Figure 4.1. The enzymatic portion of the LNT production pathway. The β-1,3-galactosyltransferase, WbgO, comes from *E. coli* O55:H7, while the β-1,3-*N*-acetylglucosaminyltransferase, LgtA, comes from *Neisseria meningitidis*. LgtA activity is preceded by LacY import of lactose into the cytoplasm. Importantly, the substrates UDP-GlcNAc and UDP-Gal are derived from endogenous metabolism.

To better facilitate LNT production strain development and screening, we also develop a genetic biosensor capable of reporting on LNT production levels. Whether it controls a reporter gene for high-throughput screening or a more selective gene enabling, for example, auxotrophic selection, such a sensor could result in dramatic improvement of production titer, potentially leading, for example, to economically-viable inclusion of microbially-derived LNT in infant formula. The sensor is an RNA composed of an aptamer domain linked to a transcriptional output domain compatible with CRISPRa.

Nucleic acid aptamers have long been selected *in vitro* for binding of target molecules[106,107], and we make use of a preconjugated LNT column matrix[108] to select for binding[109]. Though many modes of RNA switch behavior are possible, we develop this sensor as a kinetic scRNA switch[110] to preserve compatibility with the rest of the control circuit, and to make use of its OFF-to-ON response to ligand production for easier interpretation and easier potential connection with a dynamic regulator of cellular fitness. This type of switch is designed to use differential aptamer folding to affect the structure of a scRNA's dCas9-binding handle, aiming to disrupt that structure in the absence of ligand and thereby prevent CRISPR activation of the output gene. Aptamer–ligand binding, on the other hand, allows proper handle formation. Due to the *trans*-acting nature of scRNAs within CRISPRa, another advantage of this switch type is modularity of output gene. Eventual adoption of the sensor into an auxotrophic

selection scheme, or ligand-responsive dynamic pathway regulation, for example, would only require a change of the output gene, not a reengineering of the sensor.

4.2: Heterologous genetics

To build a strain capable of producing LNT, we placed a codon-optimized version of *lgtA*, a *Neisseria meningitidis* gene encoding a β-1,3-*N*-acetylglucosaminyltransferase[111], under the control of the J5 synthetic CRISPR-activatable promoter; and *wbgO*, an *E. coli* O55:H7 gene encoding a β-1,3-galactosyltransferase[112], under the control of the J6 promoter. Additionally, to ensure the import of the media-supplied feedstock lactose, we placed an additional copy of the *E. coli lacY* gene, encoding the lactose permease, under the control of the J3 promoter. These genes were placed into one plasmid, while the other plasmid in the system constitutively expressed dCas9, MCP-SoxS, and an array of three scRNAs targeted to the synthetic promoters controlling the pathway genes (Figure 4.2). We chose JM109 as the *E. coli* strain in which to base this two-plasmid system, due to its Δ*lacZ* genotype limiting feedstock loss.

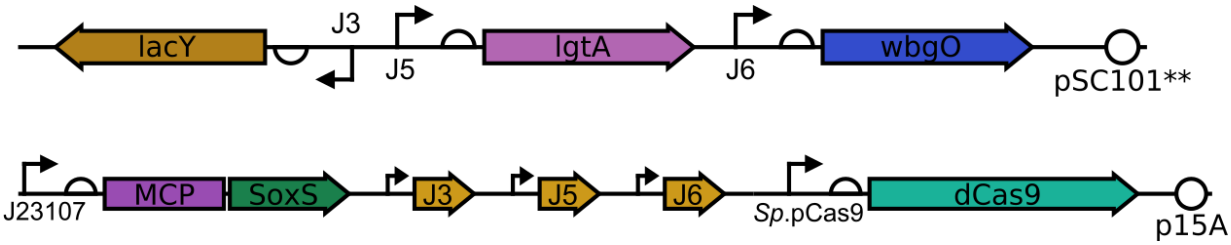


Figure 4.2. Two-plasmid system for expressing the LNT production pathway. A low-copy plasmid containing pathway genes *lacY, lgtA,* and *wbgO* is separate from a medium-copy plasmid expressing the CRISPR components, including a scRNA for each of the outputs. Either plasmid in isolation should lead to effectively no LNT production.

Because these three channels of CRISPRa are independent and orthogonal, we were able to tune expression by introducing spacer sequence truncations—which diminish activation (Figure 4.S1)—into each of the scRNAs independently, resulting in a combinatorial library of high, medium, low, and base expression levels for each of the three pathway genes (Figure 4.3). Combining these 64 library members with the

pathway plasmid yielded a library of strains with variable LNT production, which we used to determine productive combinations of enzyme expression levels and stoichiometry.

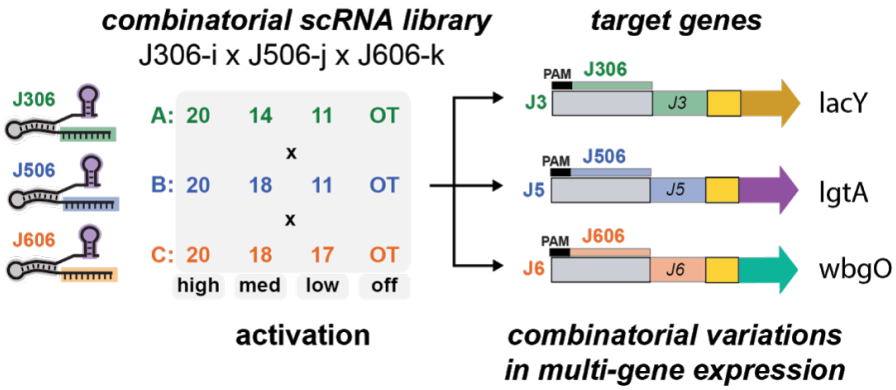


Figure 4.3. Orthogonal, tunable scRNA-promoter pairs constituting a combinatorial library of pathway gene expression. Selection of four discrete activation levels, guided by the response functions in Figure 4.S1 and including no-activation, is achieved by truncating the scRNAs' spacer sequences. Across three promoters and four activation levels, 64 possible combinations of gene expression can be applied to the three outputs. In this work, the outputs are the lactose permease LacY and the LNT pathway enzymes LgtA and WbgO.

4.3: Library screen and production

The combinatorial expression library allows us to pick optimal variants of the lumped effects of enzyme activity, enzyme stoichiometry, burden from expression and from enzyme activity, and others; and each enzyme contributes to these effects in a manner specific to each of the enzymes and their interactions. Some combinations might work well for different reasons than others, and understanding the specific factors behind those reasons can sometimes be difficult. Nonetheless, we assessed LNT production in each of the library members by HPLC of culture supernatants, and observed a few patterns in the ~1/3 of strains that demonstrated significant LNT expression (Figure 4.4).

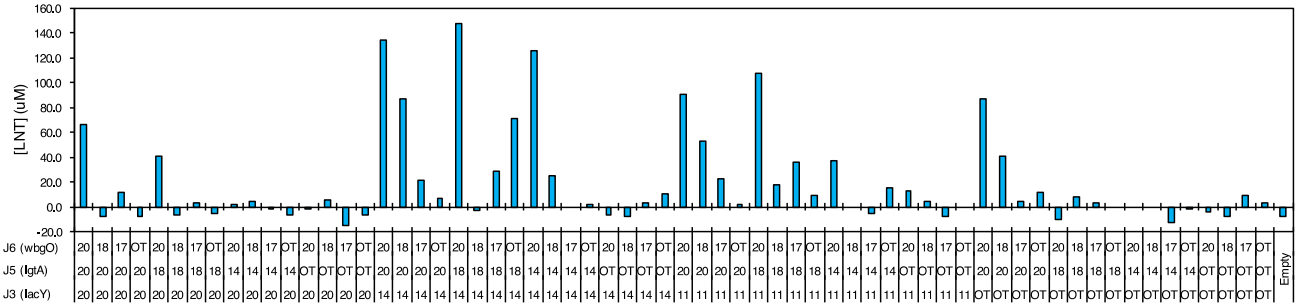


Figure 4.4. Combinatorial expression library highlights LNT-producing strains. The number of low- or non-producing strains suggests that this system is quite sensitive to these variations in gene expression. High levels of WbgO and low or medium levels of LacY seem to be important factors for production. For comparison, levels of the intermediate LNT II are shown in Figure 4.S2.

We found that LNT production was generally highest with a low or medium level of *lacY* overexpression, probably due to a combination of activity-associated stress from LacY[113] and the entropic stress and aggregation arising from overexpression of a membrane protein[114]. Because our HPLC detection method is capable of assessing production of both LNT and LNT II, we were able to observe the effect of varying pathway enzyme expression on both the final product and its intermediate. We found high levels of the LNT II intermediate resulting from many expression levels of *lgtA*, whereas high expression of *wbgO* was generally required for efficient conversion of LNT II to LNT. This suggests a *wbgO* bottleneck in this system that, because the HPLC method assesses accumulation in culture supernatant, may be related to competition for LNT II between WbgO activity and export of LNT II. Once LNT II is exported, it is no longer accessible to WbgO, underscoring the importance of tuning the relative rates of these processes to maximize yield. Our combinatorial library approach suggests—in addition to minimizing *lacY* overexpression—maximizing *wbgO* expression to improve yield, or choosing a more active isoform[115], though knocking down LNT II export would be a viable alternative strategy.

4.4: Endogenous CRISPRa/i for LNT production improvement

Because some substrates of the LNT pathway are derived from endogenous metabolism (Figure 4.1), we aim to include an array of gRNAs targeting competing

pathways, as in Chapter 3. We have focused these efforts on endogenous enzymes pulling directly or indirectly on UDP-GlcNAc and UDP-Gal, the knockdown of which could boost metabolic flux through the heterologous pathway, depending on enzyme activities and levels of heterologous intermediates. This work is ongoing, and several targets have been tested individually, as a fourth position in the array of scRNAs/gRNAs alongside the pathway scRNAs. The targets assessed so far have mostly focused on UDP-GlcNAc levels, guided by previous reports engineering a similar pathway[103–105], but we also included a *lacZ*-targeted gRNA, which should allow LNT production from a strain like MG1655 that would normally cleave the feedstock, but should not directly affect production levels of JM109 strains except through burden or binding competition.

We found that knockdown of the UDP-GlcNAc-associated targets did not increase production and in many cases reduced it (Figure 4.S3), though this reduction was more pronounced for LNT than for LNT II. Because the gRNA in each case is expressed by a strong J23119 promoter, but the scRNAs are expressed by weaker J23105 promoters, it is possible that activation of LNT production suffers from competition between RNAs for dCas9 binding, or even from broader metabolic burden effects, resulting in a decrease in production that counteracts any effect of increased UDP-GlcNAc. Indeed, the *lacZ*-targeted gRNA decreases LNT production (but not LNT II production) substantially in JM109, despite the strain's lack of LacZ activity to knock down (Figure 4.S3A). We have seen from Figures 4.4 and 4.S2 that WbgO activity seems especially sensitive to changes in activation level, so it is perhaps unsurprising to see this LNT production sensitivity to the added effect of CRISPRi. Weaker gRNA expression, perhaps paired with weaker repression strength through spacer sequence truncations, could preserve the desired metabolic effects while diminishing deleterious effects on pathway CRISPRa. It is also possible that UDP-Gal level is a more important factor limiting LNT production than UDP-GlcNAc level; assessment of UDP-Gal-associated CRISPRi targets is still ongoing.

Interestingly, in the MG1655 strain that is normally a poor pairing with a lactose feedstock, due to its LacZ activity, CRISPRi targeting that activity boosts LNT (Figure 4.S3A, left) and LNT II (4.S3B, left) production to levels comparable to the JM109 strain

(Δ*lacZ*). Practically, this finding represents a useful and easily programmable expansion of usable host strains for this and similar pathways, and fundamentally, it represents an important proof of concept for unified CRISPR control of heterologous and endogenous genetics in a broad, single-layered circuit. Even with the above limitations affecting production, it is encouraging to see any amount of production still possible when circuit nodes encompassing heterologous CRISPRa and endogenous CRISPRi—and sharing dCas9 level accordingly—are present.

4.5: Circuit size considerations and intracellular sensor for screening

Similarly to the *p*-ACA production control circuit in Chapter 3, the inclusion of endogenously-targeted gRNAs represents an expansion of nodes in the LNT production control circuit, with little positive effect needed to justify its small metabolic cost. Our initial implementation adds one gRNA at a time, rather than a whole array, but an eventual goal is to assess combinations of endogenously-targeted gRNAs along with the heterologous expression library from Section 4.3. With every additional node, however, including spacer truncations, the size of the library to be screened increases exponentially, rendering low-throughput screening methods such as HPLC unworkable.

As an alternative—and higher-throughput—screening method, we developed an RNA biosensor capable of reporting on intracellular production of LNT. Importantly, this sensor was built upon ligand-sensitive folding of a scRNA (Figure 4.5), representing yet another expansion of the CRISPR control circuit, albeit one with the added burden of expressing a reporter gene. The principle of the scRNA switch is that the binding of the ligand, if present, to the aptamer domain of the RNA directs the secondary structure toward a stable dCas9 handle[116,117], allowing successful formation of the CRISPRa complex. If the ligand is absent, however, the aptamer domain directs folding in such a way that part of the dCas9 handle sequence is occluded by a complementary aptamer sequence, preventing dCas9 binding to the RNA and therefore preventing CRISPR activation.

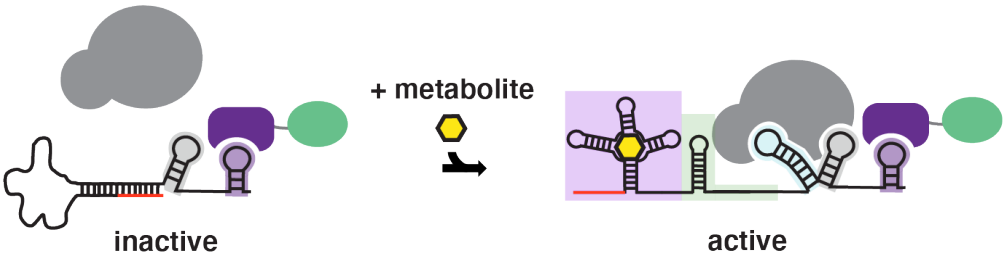


Figure 4.5. Function of a ligand-responsive scRNA switch. In the absence of ligand, complementarity between an aptamer-associated sequence (red) and the dCas9 handle sequence prevents that handle from forming, in turn preventing dCas9 binding and resulting in an inactive scRNA. If ligand is present during sensor folding, however, the altered conformation of the aptamer domain that is stabilized by ligand binding allows typical dCas9 handle folding, resulting in CRISPR activation. Aptamer domain is in purple, timer in green, dCas9 handle in blue.

To implement such a sensor in our LNT production strain, then, we needed to develop several new tools: an RNA aptamer capable of binding to LNT; a scRNA with a spacer and dCas9 handle capable of interacting with that aptamer; and a fourth, orthogonal, synthetic promoter capable of CRISPR-activating a reporter gene in response to proper folding of that scRNA.

4.6: Sensor development: generating an LNT-binding aptamer

Selecting an aptamer to bind a novel target is an *in vitro* evolution process starting from a diverse pool of sequences and enriching one or more binding-capable sequences through selective binding to a column. We started from a synthesized pool of roughly 1.9×10^{15} random 100-mer DNA sequences, and from there followed the iterative process of *in vitro* transcription, column binding and washing, elution, reverse transcription, and PCR amplification known as SELEX (Figure 4.6). Our column matrix was a methacrylic polymer functionalized with LNT by an external vendor, and we quantified the RNA in the purified column eluate by A_{260}. Though we did perform a blocking step during each cycle, we did not perform a counterselection against similar targets like LNT II or lactose.

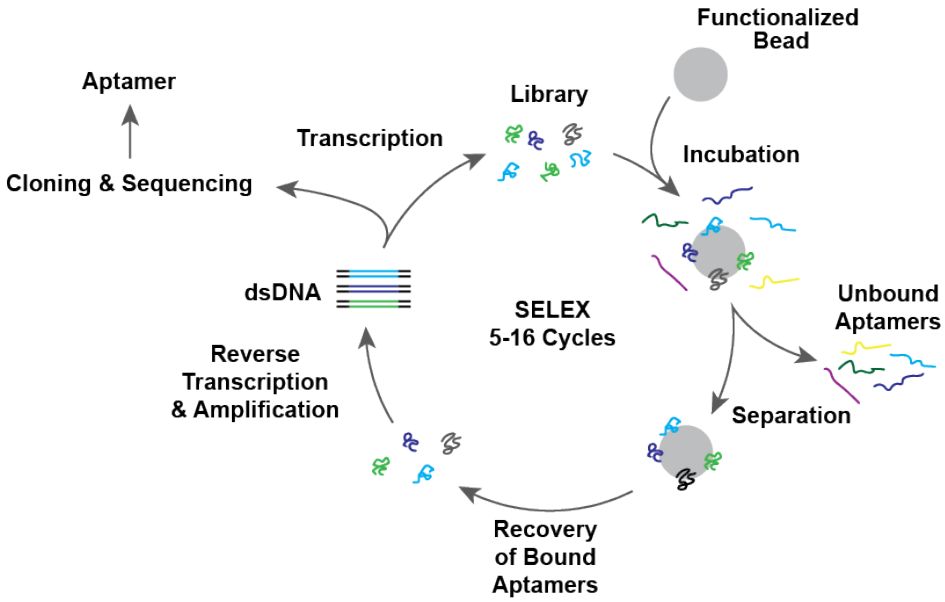


Figure 4.6. Selection of RNA aptamers by Systematic Evolution of Ligands by Exponential Enrichment. A diverse pool (up to 10^{15} sequences) of DNA sequences is transcribed, and the RNA applied to an affinity column. Unbound RNA is washed away and discarded, while bound RNA is eluted using free ligand, reverse transcribed, and PCR-amplified to make the pool for the next round of SELEX. Aptamer enrichment is usually indicated by high RNA concentration in the column eluate.

Usually, a selection is marked as finished after 5 or 10 rounds by the appearance of large amounts of RNA in the purified column eluate, indicating enrichment to a high population percentage of a particular, binding-capable sequence within that round's pool of sequences. Though we continually assessed eluate RNA concentration through round 12 of our first attempt at a selection, we never saw readings indicating enrichment. Thinking we needed column conditions more permissive to binding, we altered several aspects of the binding protocol (Table 4.S1) and performed a second selection. Again, the selection failed to indicate any enrichment by RNA concentration, through 12 rounds. However, when we cloned and sequenced some of the later rounds' DNA pools, several clones bore the same sequence, indicating some level of enrichment. Encouraged, we did the same for several earlier rounds and found a total of seven enriched sequences, some enriching as early as round five (Figure 4.7). Similar sequencing from the first selection yielded one more enriched sequence. The SELEX process usually expected to yield perhaps one aptamer instead yielded eight

aptamer candidates that could putatively bind LNT. Predicted minimum free energy structures for each candidate are shown in Figure 4.8.

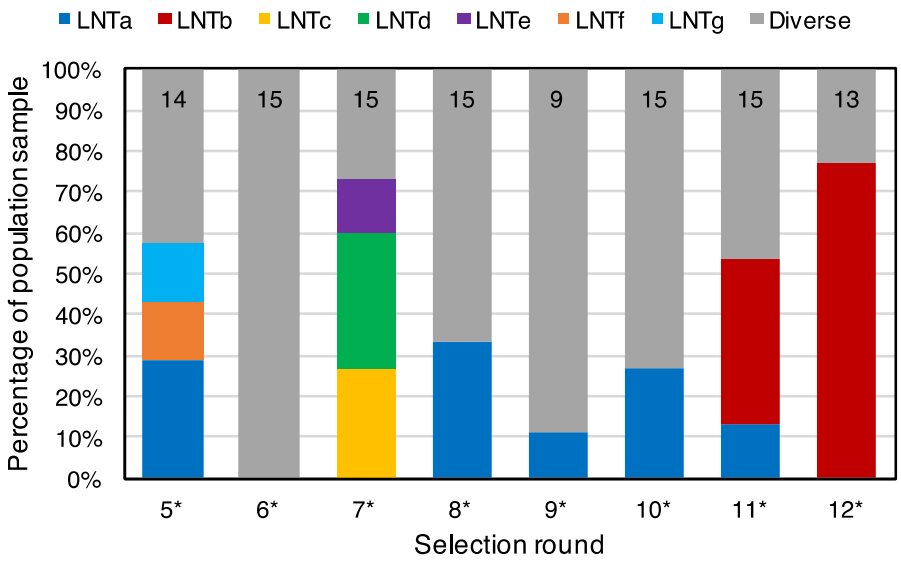


Figure 4.7. Sequencing-based investigation of aptamer candidate enrichment during SELEX. DNA pools from the indicated rounds were blunt-end-cloned into a blue-white screening vector, and a number of individual clones per round (indicated at the top of the bar) were prepped and sequenced. Non-grey colors indicate sequences seen in two or more clones, while grey indicates unique sequences (diversity, or rather, lack of enrichment). Individual aptamer candidates seem to arise, then disappear, or oscillate their frequency across rounds, explaining the lack of high-RNA-concentration column eluates. The reasons are unclear, but could relate to concentration-dependent effects like dimerization. Regardless, across all of these rounds, cloning and sequencing identified seven enriched sequences (LNTa-g) from the second selection (revised conditions for which are in Table 4.S1), while LNTh (not shown) was found to be enriched in the first selection.

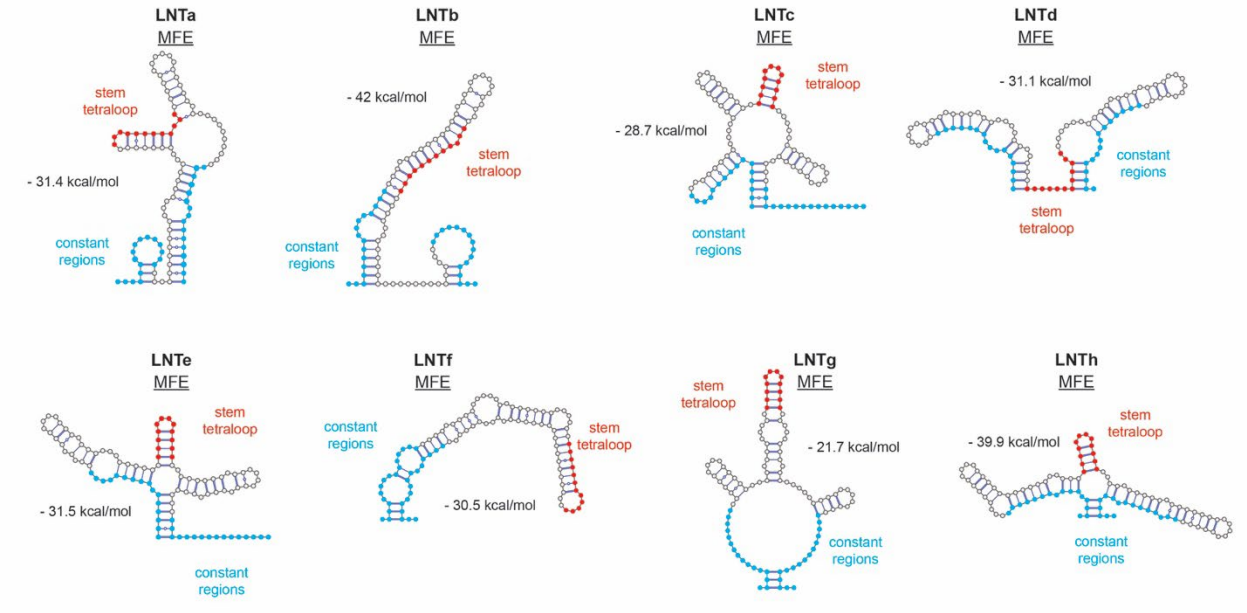


Figure 4.8. MFE structures of aptamer candidates. Folding predictions indicate potential of enriched sequences to function as aptamers. Blue and red bases were constant in the initial pool; the rest were diverse. LNTa, b, c, d, g, and h were assessed for ability to bind LNT using an in-line probing assay (Figure 4.S4), all of which provided evidence of binding. LNTc and g were chosen as the basis of sensor design.

We still lack full understanding of why so many enrichment events never resulted in high-RNA eluates, or why enriched sequences seem to die out in later rounds, but it is likely related to some counterbalancing selective pressure that is hopefully peculiar to this case, perhaps relating to concentration-dependent effects like dimerization. Regardless, the sequences clearly enriched as a result of some positive selective force, and we attempted to show that that force was binding capability by assessing the aptamer candidates' performance in *in vitro* binding assays. A theme of these assays is differentiating some signal due to differential RNA folding in binding reactions with and without the target ligand. An early attempt to use RT-qPCR to observe LNT-dependent differences in reverse transcription efficiency failed, but an in-line probing assay—which observes LNT-dependent differences in room-temperature RNA degradation over a timescale of days—resulted in different gel-banding patterns in the degraded reactions of all six tested aptamer candidates (Figure 4.S4). We found this evidence of LNT binding convincing enough to pick two of these aptamers for further development into LNT-responsive scRNA switches.

4.7: Sensor development: design of aptamers into ligand-responsive scRNAs

Six binding-validated aptamers was more than anticipated, so we used the predicted structures (Figure 4.8) to pick two of the most promising aptamers for further development into biosensors: LNTc and LNTg. As in Figure 4.5, this entails connection of part of the aptamer sequence with part of a scRNA sequence, including an intervening timer domain sequence. The transcriptional timer is intended to give the nascent aptamer more time to bind LNT if present, altering its still-folding structure into the ON-state or OFF-state, before that folding can be influenced by the nascent scRNA domain. Based on our understanding of several important design characteristics, we computationally screened a large set of candidate biosensor sequences, identifying four sequences likely to be functional LNT biosensors. Two of them used the LNTc aptamer and two used the LNTg aptamer, so the four sensor candidates were LNTc1, LNTc2, LNTg1, and LNTg2.

Because of the importance of the timer domain for sensor function and sensitivity, we strategized to screen through a pool of diverse timer sequences for each sensor candidate *in vivo*, aiming for a comfortable span of RFP output, or dynamic range, between the OFF- and ON-states of each sensor candidate at a given LNT concentration. The four sensor candidates, then, were synthesized with Type IIS restriction sites into which we would clone these timer pools just before the screen. Because this process would occur in *E. coli*, we needed to include an output construct to allow for easy visual screening.

4.8: Sensor development: a new CRISPRa promoter controlling RFP output

At this stage in sensor development, the most sensible output indicating *in vivo* sensor function would be a fluorescent protein, regardless of the sensor output in the final application. While a fluorescent plate reader allows for several color options in liquid culture, our screening process involves colony visualization on agar plates (Figure 4.11), so we opted to use RFP as our output protein to aid that visualization under natural light.

In order for the sensor to activate RFP expression, we devised a new, orthogonal, synthetic CRISPRa promoter intended specifically for sensor applications, known as J_{sen}. The new promoter is based on J23117, like the rest of the J-series promoters, but uses an N26 sequence (the region between the base promoter and the spacer target sequence) specifically designed to have low leak in the OFF-state. It also comes in different versions using a spacer target sequence that corresponds to the complementary spacer in the scRNA sensor, which is specific to each of the four sensor candidates (plus J5 as an aptamer-less control). ON-state output from this promoter, as determined by the aptamer-less J506 control scRNA, is quite low compared to the rest of the J-series promoters, but we considered this a desirable quality especially for a locus on a medium-copy plasmid, given the metabolic burden that high levels of RFP would contribute to an already-burdened system. The promoter's fold-change between OFF- and ON-states was nonetheless deemed acceptable for screening purposes, at 8.6-fold on a p15A plasmid (Figure 4.9).

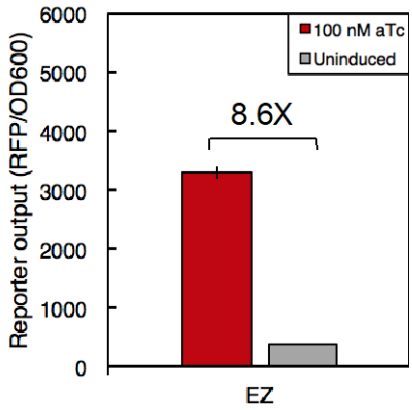


Figure 4.9. Control scRNA with no aptamer region activating J_{sen} promoter. The new synthetic promoter shows lower activated output than other J-series promoters, arguably making it well-suited to sensor output. Unactivated output does depend somewhat on spacer target sequence, with the LNTc2-compatible promoter in particular expressing high background output. Shown here is promoter behavior using a J5 spacer sequence. The gene is on a p15A plasmid, and the host in this case was DH10B, requiring more aTc than JM109 to reach maximal activation (Figure 4.S5). Error bars represent standard deviation of *n*=3 biological replicates.

As part of the screening process, we initially desired inducible control of the pathway, so we picked out a subset of the full scRNA library from Section 4.3 to convert to tet control of the MCP-SoxS, which would act as a master regulator. HPLC testing indicated that one of the variants (#33, low-high-high) showed inducible LNT production (Figure 4.10), and we used that strain in subsequent screening and testing of the biosensor candidates. Finally, before the screen, we further cloned the CRISPR plasmid to insert the reporter gene from Figure 4.9 and each of the four sensor candidates under constitutive control of the J23105 promoter.

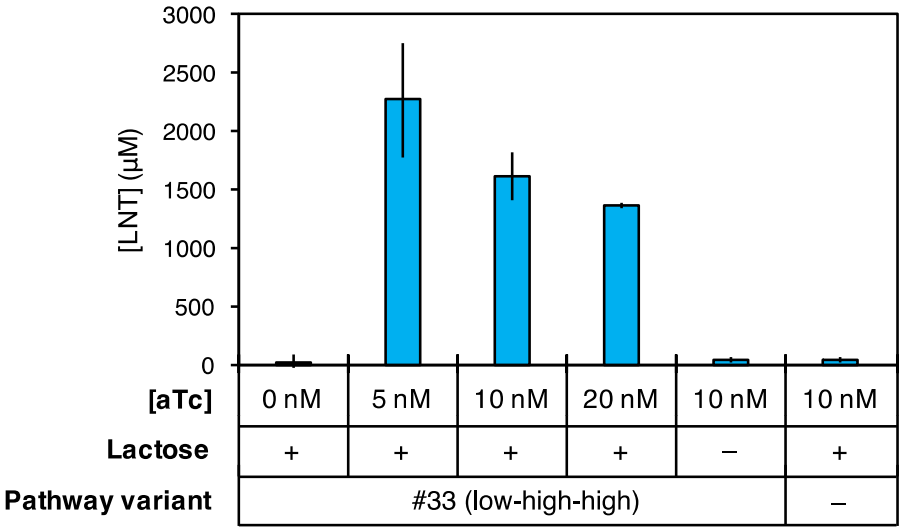


Figure 4.10. Inducible CRISPRa control of LNT production. The tet promoter controls expression of MCP-SoxS, so that protein acts as a master regulator for pathway function and sensor function. Here, we use HPLC to verify that strain 33 produces LNT at these levels of aTc, just as it did in Figure 4.4 when constitutively controlled. This strain was used for the sequential screen and for the final sensor testing. LNT production is abolished if lactose is absent from the media, or if the pathway genes are absent (replaced with mTag-BFP). Error bars represent standard deviation of *n*=3 biological replicates.

4.9: Screening strategy and results: shared activator

With aTc now inducing MCP-SoxS expression, the sensor and the pathway are under the control of the same master regulator, making it difficult to assess the sensor's off-state during the *in vivo* timer pool screen. Fortunately, we placed the reporter gene and pathway genes on separate plasmids, so LNT should only be present in the full

two-plasmid system. This allowed for a sequential screening strategy in which timer domain variants on the CRISPR-sensor-reporter plasmid were screened for low OFF-state in the absence of the pathway genes. These variants were prepped pooled, and then cotransformed with the pathway plasmid before screening for high ON-state (Figure 4.11).

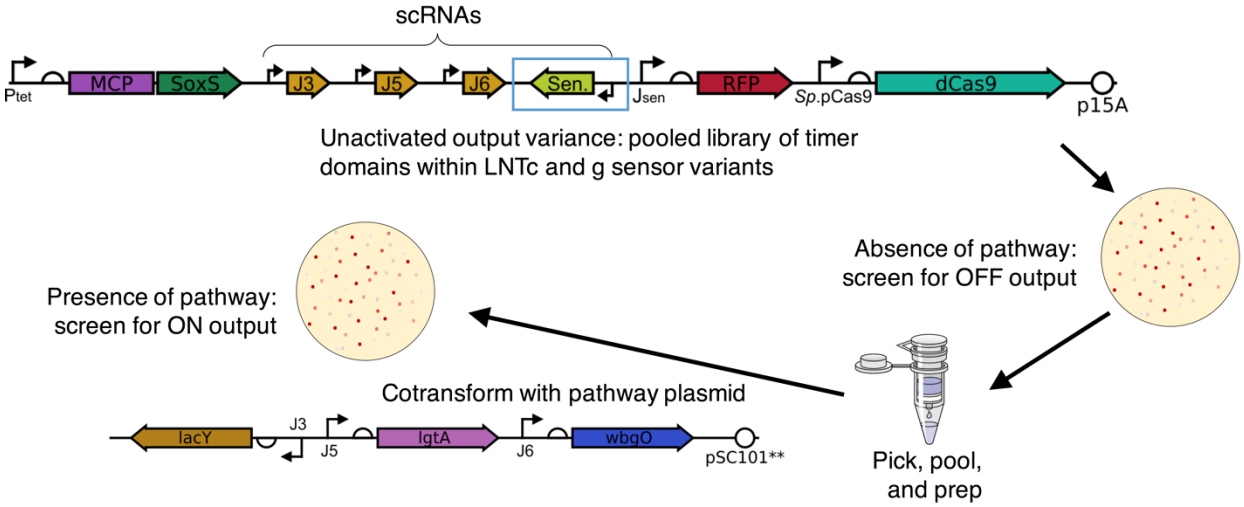


Figure 4.11. Scheme for sequential timer pool screen. The plasmid containing the sensor and reporter, as well as the necessary CRISPR proteins, is first screened alone for low RFP signal, indicating a prevalent OFF-state of the sensor when LNT is not possibly present. Within this population, the plasmid containing pathway genes is added, resulting in LNT production—according to our HPLC-based understanding of CRISPRa program productivity—and screened again for high RFP signal, indicating a prevalent ON-state of the sensor. aTc is always present.

We performed this screen on our four sensor candidates with timer pools inserted, but found it difficult to distinguish between different output levels in the colonies, mostly due to low output expression. As a result, the OFF-state screen was rather arbitrary, though signal was low in all picked colonies, as desired. The exception was the LNTc2 candidate, which showed high background RFP expression. Background was also high, though, in the –aTc control, indicating that the background actually arises from the J_{sen} promoter, but only the version with the LNTc2 target sequence, even in the absence of CRISPRa. While we expect slight differences in unactivated baseline output levels from different target sequence versions of J_{sen}, the LNTc2 version's level is surprisingly high.

Regardless of baseline promoter output levels, the plate-based screen provided two timer domain isolates per sensor candidate, as determined by sequencing colonies from the second plate, to take forward into liquid culture testing. This entailed synthesizing the identified sequences and cloning them into the sensors' restriction sites, resulting in clonal constructs of two timer variants per each of the four sensor candidates, or eight variants total.

4.10: Screening strategy: orthogonal activator

Initially, we had intended to use a simultaneous OFF- and ON-state screen with an orthogonal CRISPR activator protein (PCP-SoxS)[6,7] to address the problem of linked pathway and sensor induction. By effectively splitting the master regulator into two orthogonal channels—MCP-SoxS under tet control and PCP-SoxS constitutive—we would be able to easily assess sensor OFF-state in the same colony (with the same timer variant) as assessing ON-state by simply inoculating both –aTc and +aTc liquid cultures. However, we were unable to achieve PCP-dependent CRISPR activation of the J_{sen} promoter, even after attempting to optimize it for the PCP-SoxS system, a difficulty that precluded the simultaneous screen.

Later experiments, though, would demonstrate that a control condition with the lactose feedstock absent from the media completely abolished LNT production, according to HPLC results (Figure 4.10). This is entirely expected when assessed by HPLC, but Figure 4.12 indicates the same result according to LNTc2 output, suggesting that the absence or presence of lactose could be used as the basis of a simultaneous screen in the future (Figure 4.S7), just as the absence or presence of aTc was to be used in the initial plan.

4.11: Detection of intracellular LNT by a scRNA biosensor

Of the eight sensor candidates including timer variants, most had too little RFP signal for production differences to be detected (Figure 4.S9), but we found that the high background of the LNTc2 sensor and its reporter were actually activated to even higher levels by the presence of both lactose and the pathway genes—that is, in a strain that

produces LNT (Figure 4.12). The background from the –aTc control is subtracted in Figure 4.12 to better see the activation; unsubtracted background can be seen in Figure 4.S8B.

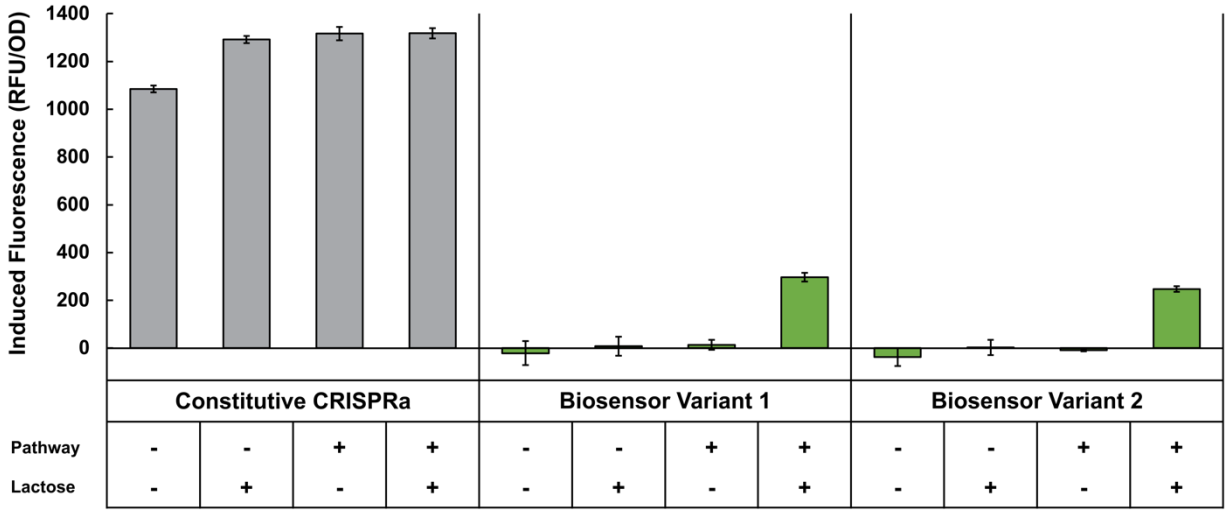


Figure 4.12. Detection of intracellular LNT by a scRNA biosensor. The sensors shown are both timer variants of the LNTc2 sensor. The *y*-axis indicates density-normalized fluorescence values in cultures induced with 5 nM aTc, when uninduced controls (including the high LNTc2 background) are subtracted. While the control sensor-less J5 scRNA (left) activates RFP regardless of pathway conditions, in the sensor strains both the pathway genes and the feedstock lactose must be present to achieve RFP activation. LNT production is verified by HPLC in Figure 4.S8. Error bars represent standard deviation of *n*=3 biological replicates.

While the activation indicates a functional biosensor, there are still many questions about the parameters of its function, including whether it its sensitivity is well-matched to the production level in this strain, whether the activation will scale with different levels of production, and its degree of specificity to LNT rather than LNT II (especially in the absence of an aptamer counterselection). To address these questions and further validate sensor functionality, we used the combinatorial library data from Figure 4.4 to select strong and medium LNT producers, a strong LNT II producer with off-target *wbgO* activation, and the all-off-target control; and tested sensor output in these strains using the same sensor and timer variant as in Figure 4.12, Variant 1. This experiment represents a return to the original intent of a production variant strain screen using a single sensor, rather than screening for functional sensors using a known

producing strain, and provided the interesting findings that LNT sensing signal seems to be saturated—at least in the range between about 30 and 100 μM—and that the LNTc2 sensor Variant 1 responds to LNT II production, probably indicating promiscuous binding of the aptamer to the two metabolites (Figure 4.13). If we wanted to adjust the sensitivity of LNT sensing to be relevant to these strains' level of production, then we can theoretically do so through adjustment of the timer domain length. The sensor's specificity to LNT could be improved by changing the timer domain, changing the aptamer domain to one of the other candidates from our selection (Figure 4.8), or, most drastically, by redoing the selection.

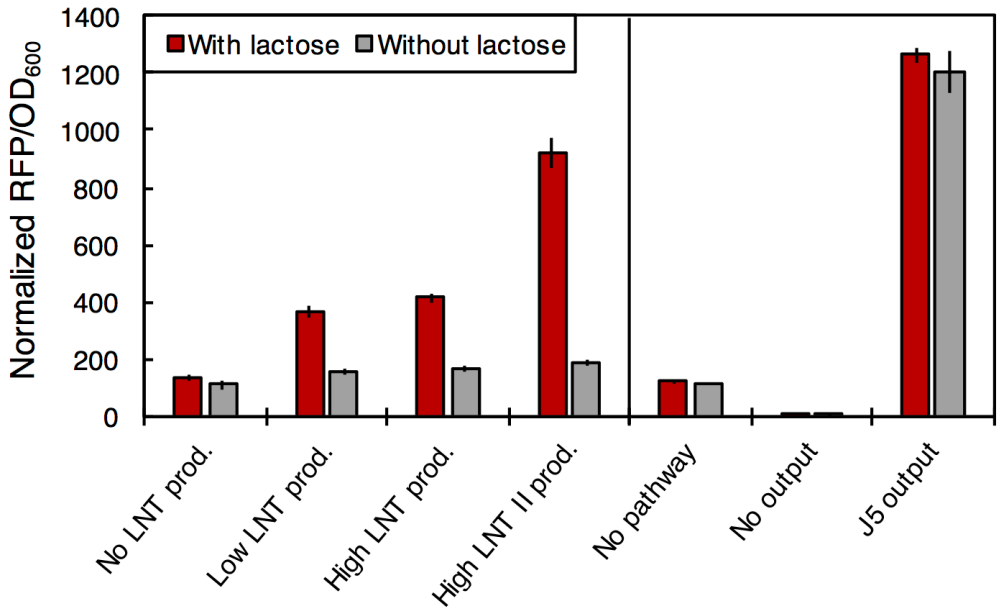


Figure 4.13. LNTc2 Variant 1 biosensor response to production variants. In enzyme expression variants (Figure 4.4) producing no LNT (variant #64), low LNT (#41), high LNT (#33), or no LNT with high LNT II (#4), the increase of sensor output when lactose and aTc are both present is minimal, low, low, and high, respectively. The small output difference between low LNT and high LNT indicates probable sensor saturation at these production levels, while the large increase in output in the high LNT II strain indicates promiscuous aptamer binding to LNT II. The *y*-axis indicates density-normalized fluorescence values in cultures induced with 5 nM aTc, when uninduced controls (including the high LNTc2 background) are subtracted. LNT and LNT II production levels are assessed by HPLC in Figure 4.S10B. Error bars represent standard deviation of *n*=3 biological replicates.

As an easy first step to address the specificity problem, we will reattempt the experiment in Figure 4.13 with the LNTc2 Variant 2 biosensor: the alternative timer domain might affect specificity as assessed by RFP signal if the issue is due to poor signal propagation between aptamer binding and scRNA folding. For example, we have noticed markedly different sensor output in two strains (Figure 4.13, “High LNT prod.” and “High LNT II prod.”) producing similar amounts of LNT II (as assessed by HPLC in Figure 4.S10B), even though the lower-output strain is additionally producing LNT. If sensor Variant 1 is indeed poorly propagating the signal from the LNT-bound aptamer domain to the scRNA domain, but is adequately propagating that signal when LNT II is bound, *and* if LNT–aptamer binding is sequestering most of the sensor population in the “High LNT prod.” strain without effectively propagating that signal to RFP production, then this odd finding can be explained, and can possibly be corrected by a different timer domain that more reliably and functionally connects the aptamer and scRNA domains. If, on the other hand, Variant 2 or other timer variants from the LNTc2 pool cannot improve specificity, the next step is to assess the remaining sensor candidates, especially those built upon a different aptamer sequence like LNTg. In Figure 4.S9, however, output generated by these other candidates was too low to determine differences of LNT production in liquid culture, or indeed differences between timer domain variants during the plate-based screen.

To address the remaining sensor candidates and their low output signals, we cloned stronger promoters (J23119) expressing the sensor scRNAs, following experimental indications of sensor concentration affecting output level, and anticipating that weak sensor expression could compound dCas9 competition effects, due to the time lag before sensor scRNAs can be expected to fold into the ON-state. Pathway-controlling scRNAs remain under the control of weaker J23105. Assessing the stronger promoter’s effects on output is still ongoing, but we expect that this change will boost output signal from the LNTc1, LNTg1, and LNTg2 sensors above levels seen under J23105 promoters (similar to OFF-states in Figure 4.S9), possibly resulting in a higher fold-change in response to LNT production. This is especially true if we include a ribozyme to cleave the scRNA from the rest of the sensor in the ON-state. A new

sequential screen is under way to test these constructs, and we remain hopeful that some of their responses will prove more specific to LNT than the LNTc2 sensor—especially those of the constructs based on the LNTg aptamer.

4.12: Conclusions

Similarly to the *p*-ACA pathway in Chapter 3, here we have explored another real-world implementation of a broad CRISPR control circuit. The expansion of this circuit, however, relates less to the endogenous CRISPRi program than to the inclusion of a ligand-responsive scRNA as a biosensor, especially when expressed by a strong promoter. We determined in Chapters 1 and 2 that such RNA expression is not particularly taxing on expression capacity, but it is worth considering the added burden from the sensor's output, in this case RFP. Larger, more costly, or potentially toxic outputs such as pathway enzymes would probably benefit from an autoregulatory element of the control circuit.

Further, in this example, we have made use of the independence and orthogonality of our synthetic promoters to examine a library of pathway gene dosage and stoichiometry, separating producing strains from non-producing strains. The difference could be due to pathway expression, pathway burden, substrate availability, or other factors, but it is the advantage of a screen that all of these factors are included. Moving such a screen toward higher throughput, we have successfully selected an RNA aptamer for LNT and successfully integrated it into a ligand-responsive scRNA switch platform, marking a rare complete journey from a novel target molecule to a functional *in vivo* biosensor.

Given the findings from Chapters 1 and 2 about burden and expression capacity, we should use caution when interpreting biosensor output to determine production capacity, especially using a low-output reporter like Jsen–RFP. Namely, output level is some function of sensor activity and global expression capacity, which itself can be differentially influenced by the enzyme expression levels of the genotypes being screened. Quality sensor design resulting in high dynamic range can help shift that balance so that output level is mostly determined by product sensing, but to specifically

eliminate burden effects from output interpretation, users of *in vivo* genetic biosensors should consider including a constitutive capacity monitor for normalization or some burden-associated closed-loop circuit to smooth out burden effects on capacity[52].

In this project, we have treated this technology as a strain screening platform, and such a use would be extremely helpful for rapid development of microbial production strains—in which screening is often the throughput bottleneck that prevents development of economically viable strains[118]—but the uses of such a programmable biosensor extend beyond activation of a fluorescent reporter. By changing the sensor's output, applications could vary in many exciting ways: a regulatory output like another scRNA could bring about dynamic control circuit induction, including helpful temporal delays in expression; a pathway enzyme output would result in dynamic pathway control; or an antibiotic resistance gene as output could force selection pressure toward enhanced production, or simply homogenous production cultures[14]. The ability to easily change output ORF to bring about these various applications is a fundamental advantage of a *trans*-acting scRNA biosensor.

4.13: Methods

Plasmid construction

Plasmids were cloned using either Gibson or In-Fusion protocols and transformed in DH10B *E. coli* cells. Plasmids expressing the CRISPRa components were constructed using a p15A vector. *S. pyogenes* dCas9 (*Sp*-dCas9) was expressed using the endogenous *Sp*.pCas9 promoter. The MCP-SoxS activation domain containing mutant SoxS was expressed using the BBa_J23107 promoter (http://parts.igem.org). The scRNAs, including the LR-scRNA, were expressed using the BBa_J23105 promoter, or BBa_J23119 promoter as noted. scRNAs used the b2 design, in which the endogenous tracr terminator hairpin upstream of MS2 is removed[7]. The pathway-expressing plasmid constructed using a low-copy pSC101** vector. mRFP1 and metabolic pathway genes were placed expressed from the weak BBa_J23117 minimal promoter (http://parts.igem.org) preceded by synthetic DNA sequences containing the CRISPRa target sites. Codon-optimized ORFs of *E. coli* LacY, *N. meningitidis* LgtA, and *E. coli* O55:H7 WbgO were synthesized by Twist, and assembled into the pathway plasmid by Gibson. RBS sequences were generated using the RBS Calculator v2.0[84].

***in vitro* transcription and preparation of RNA**

DNA pools were ordered as two separate oligonucleotides from IDT and were PAGE purified to ensure that only the correct length DNA was present. One oligo contained the variable positions of the pool, and was synthesized using equimolar base-ratios produce through hand-mixing. The second primer contained the sequence for the T7 promoter. Primer extension was performed using DreamTaq PCR Master Mix (2X) (Thermo) to produce the full-length transcription template. The resulting PCR product was treated with Exonuclease I (Thermo), and subsequently purified using a QIAquick PCR clean up kit (Qiagen).

Transcriptions were run overnight (10-12 hours) at 37˚C using T7 RNA polymerase (NEB), at 200 µL scale for all rounds except round 1, in which the extra diversity of the pool dictated a larger scale (2 mL). Transcriptions were terminated by adding 3 µL DNase I (Thermo) per 200 µL transcription and incubating for 30 minutes at 37˚C. The resulting RNA was run on a 1.5 mm PAGE gel at 20 W for 60-90 minutes, and the correct bands were identified using UV shadowing by comparison with xylene cyanol and bromophenol blue dye bands and then cut from the gel. RNA was recovered from the cut gel slices with 0.5X TBE using a Whatman Elutrap; for recovery, the Elutrap was operated at 200 volts for 2 hours. Eluted RNA was extracted from the Elutrap, ethanol precipitated with KCl, and then resuspended in water. Alternatively, RNA was extracted from the gel slices using a Small-RNA PAGE Recovery Kit (Zymo Research) used with IICR columns from an RNA Clean & Concentrator Kit (Zymo). Resuspended RNA totaled between 125-350 µg of RNA.

***in vitro* selection**

Purified RNA was resuspended or eluted in 200 µL water, melted for 1 min at 80˚C and reannealed at room temperature, and combined with selection buffer containing 2.5 mM magnesium. The affinity column was packed with 200 µL gravity-settled LNT-gel beads (IsoSep, cat. 84/01-0010) and equilibrated with 20 column volumes of water, followed by 20 CVs of selection buffer, followed by a final wash with 100 µg/mL BSA in selection buffer. The purified RNA in selection buffer was then added to the column and incubated for 30 minutes before flowing through. After incubation, the column was washed with 20 CVs of selection buffer. Finally, the RNA was eluted by adding 2 CVs at a time of 30 mM LNT in selection buffer, four times (8 total CVs), incubating the last for 1 hour, then collecting the flowthrough. All column work was performed at room temperature.

After elution from the column, the dilute RNA samples were concentrated using an RNA Clean & Concentrator Kit (Zymo). After elution into a small volume of elution buffer, RNA recovery

was quantified using a NanoDrop spectrophotometer (Thermo). RNA samples were then reverse-transcribed using a standard reverse primer and RevertAid reverse transcriptase (Thermo). A maximum of 5 ng or RNA was used per µL of RT reaction. RT reactions were run using the manufacturer's suggested protocol. After the RT reaction, the resulting cDNA was PCR purified using a QIAquick PCR clean up kit (Qiagen). The cDNA was then amplified to return the template concentration to a constant level between selection rounds. In order to prevent a loss of diversity due to exhaustion of amplification primers and subsequent self-priming, the correct number of PCR cycles was estimated using a predicted doubling-efficiency of 1.4, and checked by running on agarose gels as the predicted cycle number was approached. The resulting PCR product was treated with Exonuclease I (Thermo), and subsequently purified using a QIAquick PCR clean up kit (Qiagen).

Cloning and sequencing

DNA pools from the indicated selection rounds were cloned into plasmid vectors using the TOPO TA Cloning Kit (Invitrogen) and chemically transformed into competent NEB Turbo *E. coli* cells. Transformants were plated on LB agar plates with 50 µg/mL carbenicillin, 40 µg/mL X-gal, 100 µM IPTG, and 15 white colonies per round were grown overnight in LB, miniprepped with a Qiagen vacuum kit, and sent for Sanger sequencing provided by Genewiz.

In-line probing of aptamer candidates

Prior to analysis, aptamer sequences were structurally optimized to be suitable for subsequent use. Single-stranded tails were truncated, and the closing stem of the suggested structure were lengthened where necessary. Aptamer candidates were transcribed using the standard procedure above. RNA samples were purified as above. RNA samples were then incubated at room temperature for 2 days in 1X SB at either 0 mM or 10 mM of LNT. After incubation, reactions were quenched by mixing the samples with a formamide-EDTA loading buffer. 0.5 µL of each sample was run on an extra-long 0.75 mm denaturing (7.5 M) Urea-PAGE gel (10% monomer). Gel was stained for 15 minutes using SYBR Gold RNA stain, and imaged using a PharosFX Molecular Imager (Bio-Rad) with a SYBR Gold emission filter and high sample intensity.

Production cultures and fluorescence measurement

Plasmids containing the HMO-responsive scRNA candidates, along with Cas9 and MCP-SoxS, were co-transformed into E. coli strain JM109 with either the plasmid containing the 3 pathway genes, or a plasmid with the same origin of replication and resistance marker but expressing mTag-BFP instead. 3 individual colonies were picked for each co-transformation, and were

grown in 400 µL of MOPS EZ-Rich defined medium (Teknova) supplemented with 8 g/L glucose and 2 g/L of lactose, as well as the appropriate antibiotics. Cultures were grown in 96 deepwell plates with rapid shaking at 37C. After 24 hours of growth, the cultures were diluted 1:1000 into fresh media. The media contained all combinations of (0 or 2 g/L lactose) and (0, 5, 10, or 20 nM aTc). After 24 hours of growth, 150 µL of each culture was measured in a 96-well plate format in a Synergy HTX plate reader (BioTek) with gain 35.

Production analysis by HPLC

Single colonies from LB-agar plates were inoculated in 2 mL EZ-RDM (Teknova) with 10 g/L glucose, 2 g/L lactose and supplemented with appropriate antibiotics. Cultures were grown in 14 mL plastic culture tubes at 37 °C and shaking for 48 h. Roughly 200 uL of supernatant from each culture were loaded onto 10 kDa microcentrifuge filters (Millipore) and spun for 20 min at 14000 rcf. 1 uL of filtered supernatants were assayed with a Shimadzu HPLC using UV-vis detection at 210 nm. Lacto-*N*-tetraose (LNT) was separated using a Rezex ROA-Organic Acid H+ column (Phenomenex) and a 20 mM H_2SO_4 isocratic mobile phase. A standard curve was prepared by spiking known amounts of LNT into supernatants derived from cultures of JM109 *E. coli* transformed with empty vectors. Product LNT was observed at 9.5 minutes, and intermediate LNT II, a triose, was observed at 11 minutes.

4.14: Supplementary Figures

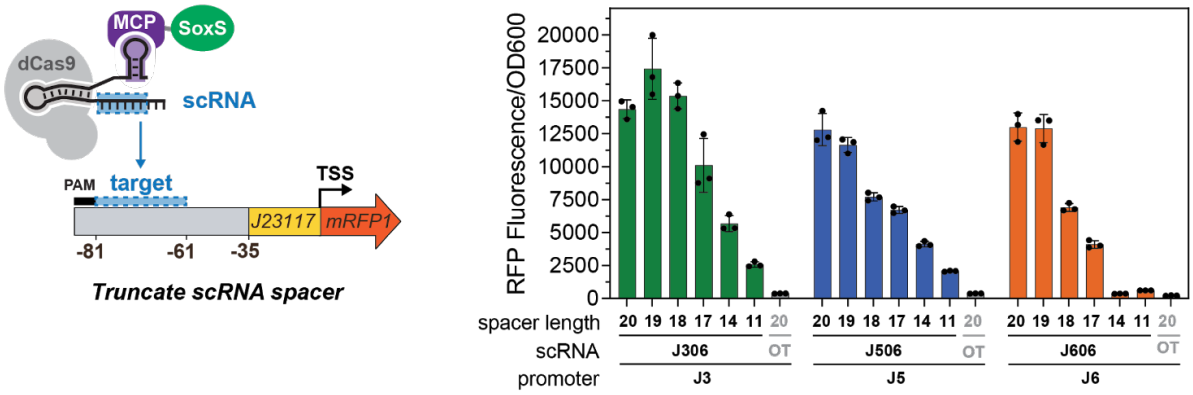


Figure 4.S1. scRNA spacer sequence truncations tune activation of output. For each of the three synthetic promoters, here we assess CRISPR activation levels as a function of scRNA spacer sequence truncation. While responses differ slightly in their steepness, we can pick out a set of spacer lengths that result in discrete grades of activation: high, medium, and low. The

lengths corresponding to those grades differ for each promoter (20, 14, 11 for J3; 20, 18, 14 for J5; 20, 18, 17 for J6), but the level of activation in each grade is similar across promoters, and their combination results in the library in Figure 4.3. Error bars represent standard deviation of *n*=3 biological replicates, and the replicates are also plotted individually.

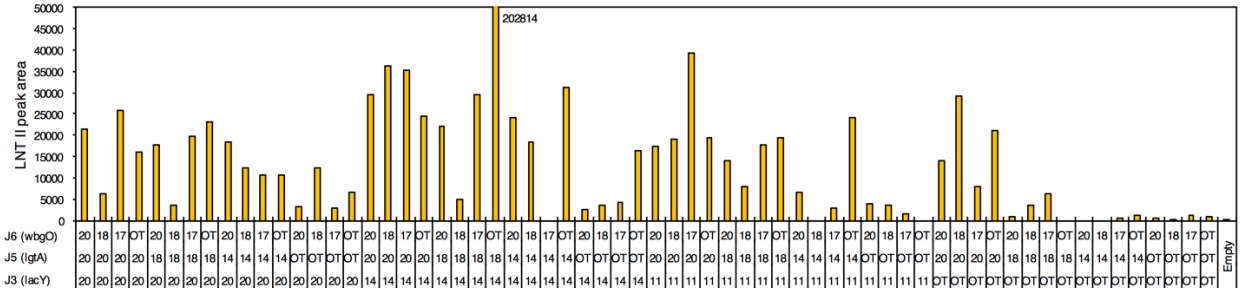
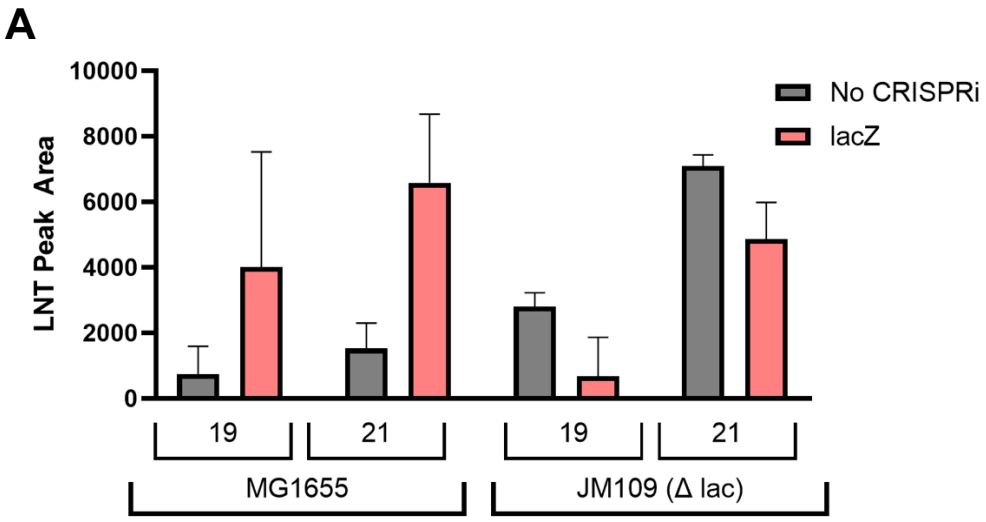


Figure 4.S2. Combinatorial expression library highlights LNT II-producing strains, assessed by HPLC as in Figure 4.4. More strains are producing medium or high levels of LNT II than LNT, suggesting that optimizing WbgO activity is worth devoting attention to. Unactivated levels of LacY overexpression seem to be sufficient for some lactose import. Though LNT II production levels as indicated here are related to the LNT production in Figure 4.4, it is important to remember that they represent the LNT II that is exported into the supernatant—that is, not exactly the LNT II levels that are directly converted into LNT.



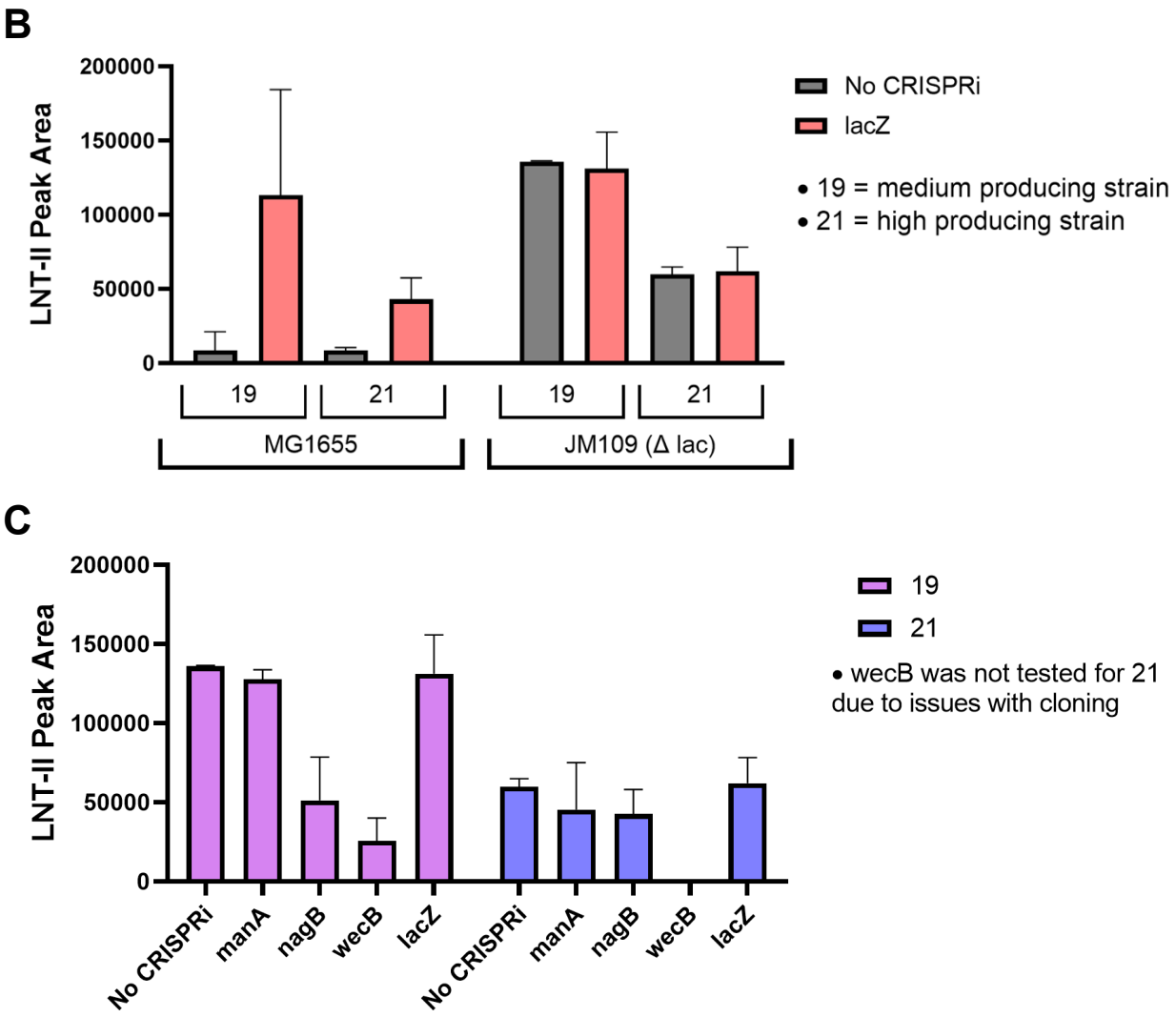


Figure 4.S3. CRISPRi of endogenous targets in MG1655 and JM109. **A)** Knockdown of *lacZ* in MG1655 boosts LNT production to near-JM109 levels in medium- and high-producing (determined by CRISPRa program) strains, presumably by limiting cleavage of the feedstock lactose. In JM109 (right), the gRNA is effectively off-target, but nonetheless reduces LNT production, either through metabolic burden or competition for dCas9 binding. **B)** The same *lacZ* knockdown similarly boosts LNT II production in MG1655, but has very little burden/competition effect on LNT II level when off-target in JM109. **C)** In JM109, CRISPRi targets vary in their unexpectedly negative effects on production (LNT II is shown here, due to these targets being associated with UDP-GlcNAc), ranging from the small reduction in the *manA* case to almost no production in the *wecB* case. Cloning difficulties with the *wecB* gRNA suggest an effect on cell health. No target in this set successfully boosted production, except for *lacZ* in MG1655 above. CRISPRa program #19 (medium-high-low) was selected as a

medium-production strain, and #21 (medium-medium-high) was selected as a high-production strain, even though its LNT II levels are lower than those of #19. It is important to remember that metabolite levels assessed here by HPLC have been exported to the culture supernatant. Error bars represent standard deviation of *n*=3 biological replicates.

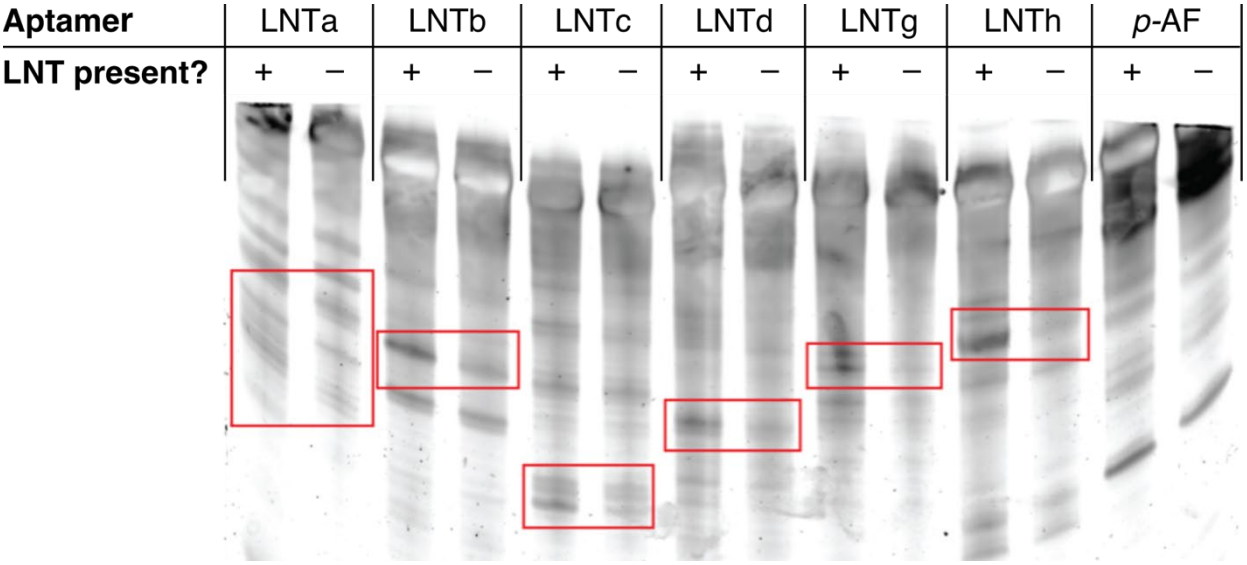


Figure 4.S4. In-line probing assay indicates aptamer-LNT binding by differential degradation. Binding biases the + and – RNA populations into different structural conformations, exposing different sites to degradation during room-temperature incubation. Red boxes highlight points of difference in banding pattern, indicating differential degradation of the original sequence into smaller fragments. The control aptamer, designed to bind the unrelated molecule *p*-AF, shows no such banding pattern difference.

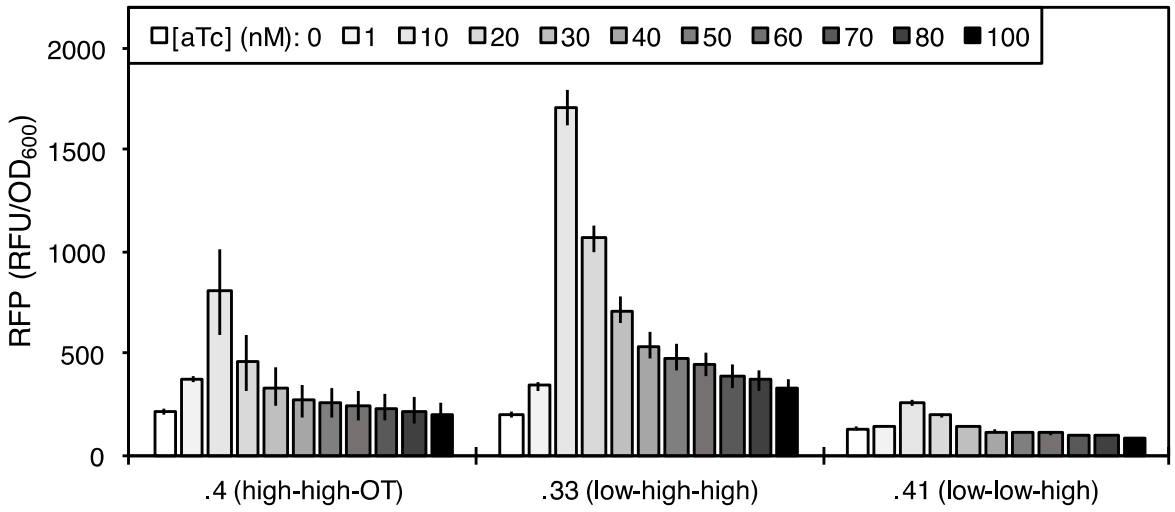


Figure 4.S5. Induction response to aTc in two-plasmid system in JM109. Induction response strongly diminishes above 10 nM aTc, unlike the response in other strains like DH10B or MG1655, in which output saturates and remains steady above roughly 100 nM. Though the low output is at first reminiscent of genetic instability, the lack of unusual variation between replicates instead suggests stability, and a high-induced culture subcultured into –aTc media retains normal inducibility (not shown). The tet promoter in this system (from Section 4.4 onward) controls MCP-SoxS expression. The maximum output level is low in strain .41 (right) because its scRNA is truncated to the "low" activation level (11 nt for J506). The maximum output level is low in strain .4 (left) because the pathway genes are being simultaneously activated in this experiment, and LacY is maximally activated in strain .4, additionally burdening the cells. OT: off-target. Error bars represent standard deviation of *n*=3 biological replicates.

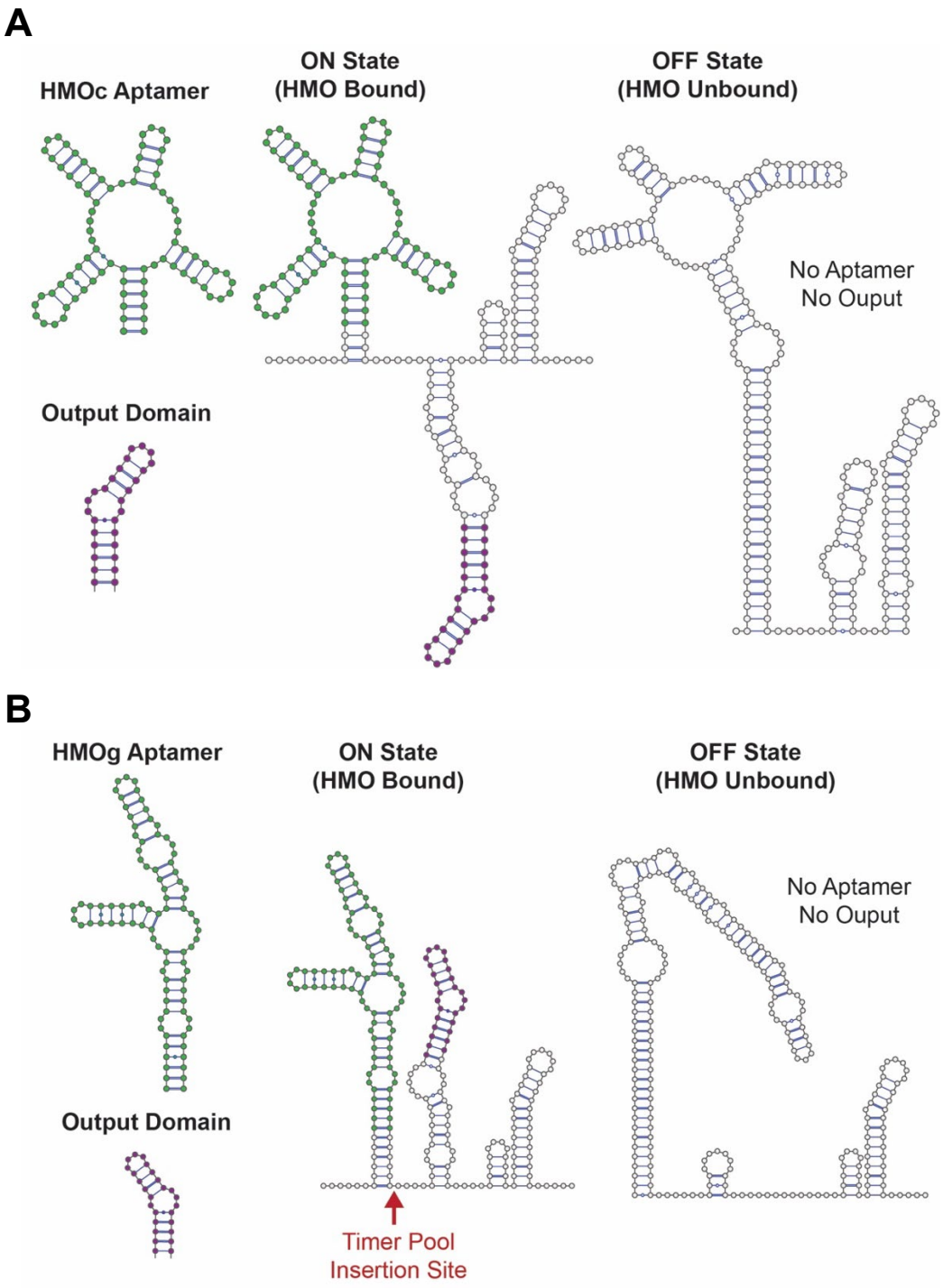


Figure 4.S6. LNT-responsive scRNA candidates: (A) LNTc and (B) LNTg. Output domain indicates the dCas9 binding handle, which is designed to fold into a functional handle only when LNT is present and influencing aptamer domain folding. In the absence of LNT, the dCas9 handle should not fold into a functional handle, resulting in no dCas9 binding and no output activation.

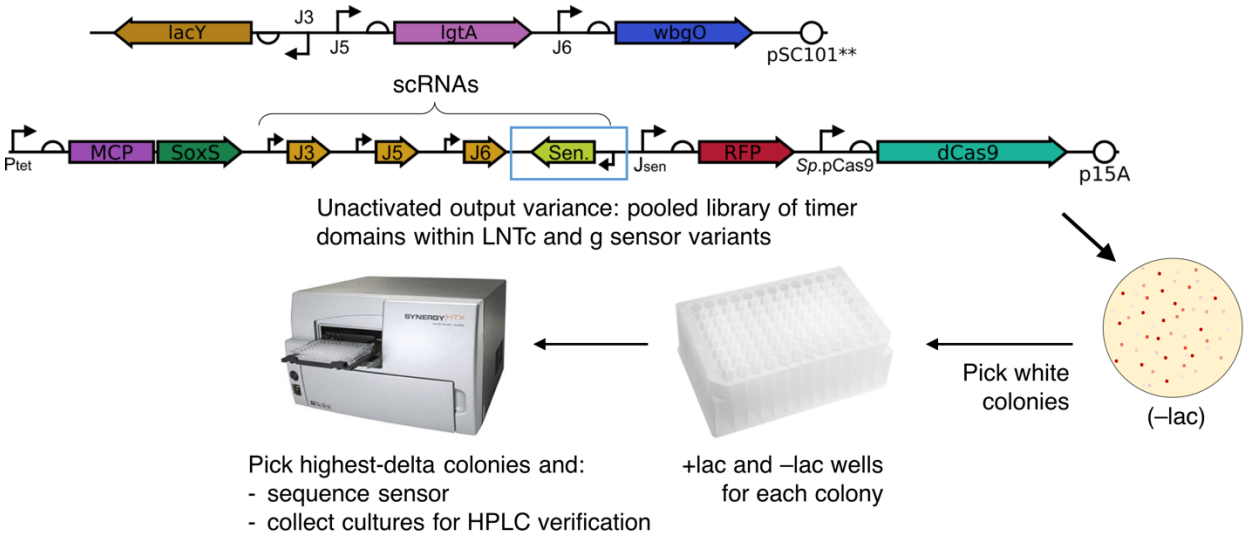
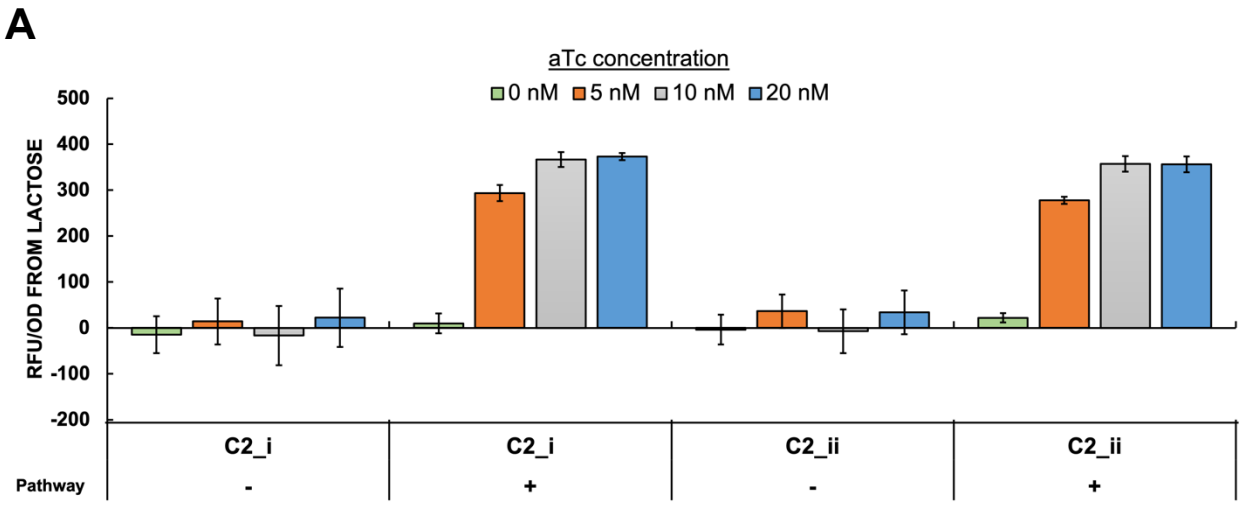


Figure 4.S7. Scheme for simultaneous timer pool screen. The plasmid containing the sensor and reporter and the plasmid containing pathway genes are co-transformed and screened for low RFP signal, indicating a prevalent OFF-state of the sensor, on an agar plate without lactose. These colonies are grown in liquid culture with or without lactose, resulting in differential LNT production—see Figures 4.10 and 4.S8C for production levels of –lac controls—and screened for high RFP signal in the +lac condition, indicating a prevalent ON-state of the sensor; and/or large fold-change between the two conditions, indicating high sensor discrimination for LNT presence. aTc is always present.



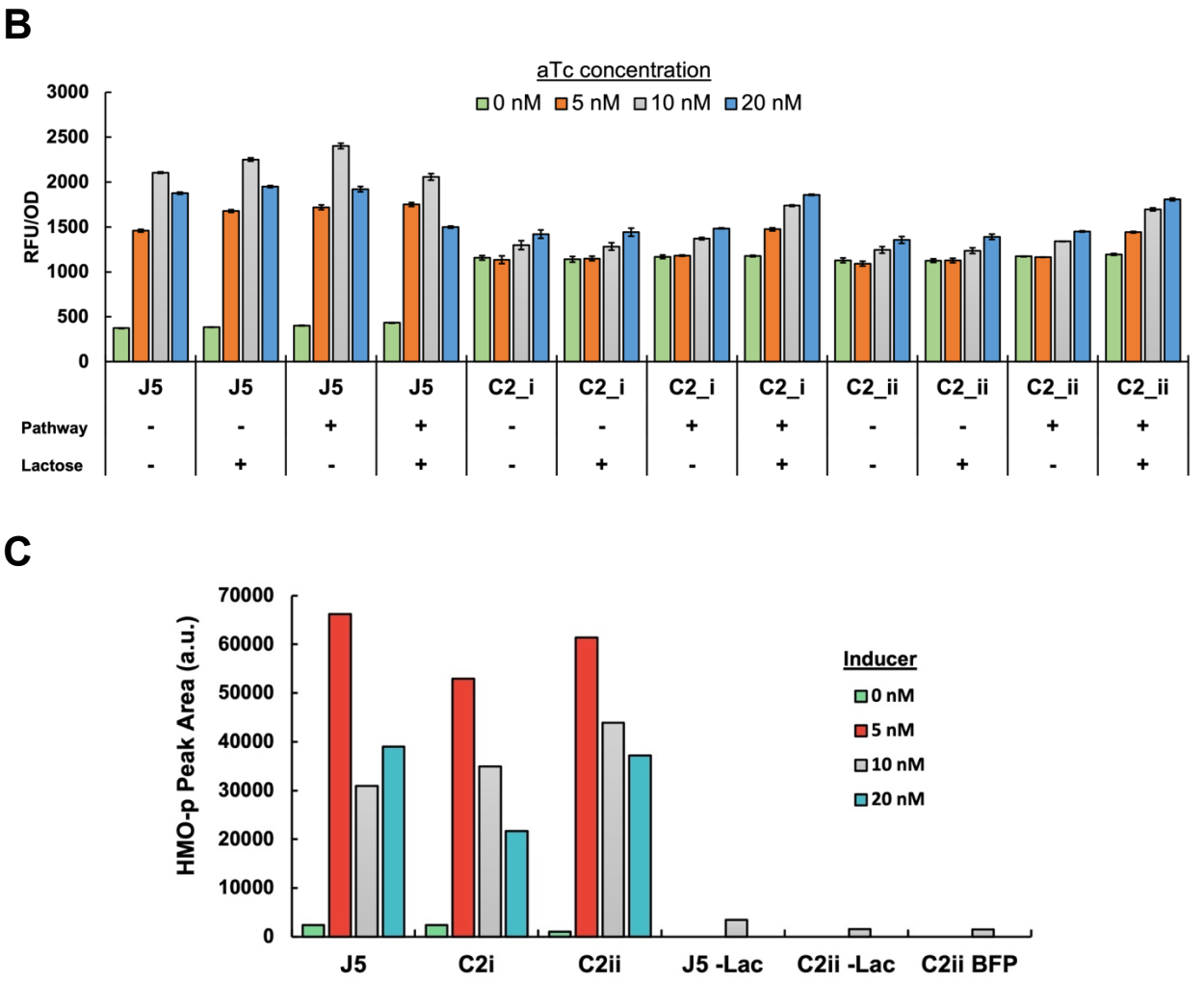


Figure 4.S8. Verification of LNT production in LNTc2 sensor strains. The same strains indicating production through sensor output were assessed by HPLC, indicating again that LNT production is inducible and dependent on lactose in the media and the presence of the pathway genes (the BFP control strain does not carry them). **A)** Full aTc response of the strains in Figure 4.12. There is no increase in signal in the absence of pathway or when MCP-SoxS is induced. RFP levels of the no-lactose controls are subtracted from each value. **B)** Unsubtracted RFP/OD_{600} values of the J5 no-sensor control and the two LNTc2 variants. Even in nonproducing cultures, the RFP signal increases slightly when aTc is added beyond 5 nM, but this effect is superseded by a larger increase in signal when LNT is being produced. **C)** Verification by HPLC of LNT production only in induced, pathway-containing strains with lactose added to the media. LNT is maximal at 5 nM aTc. We verified separately the lack of LNT production in a strain (DH10B) expressing LacZ (not shown). Error bars represent standard deviation of *n*=3 biological replicates.

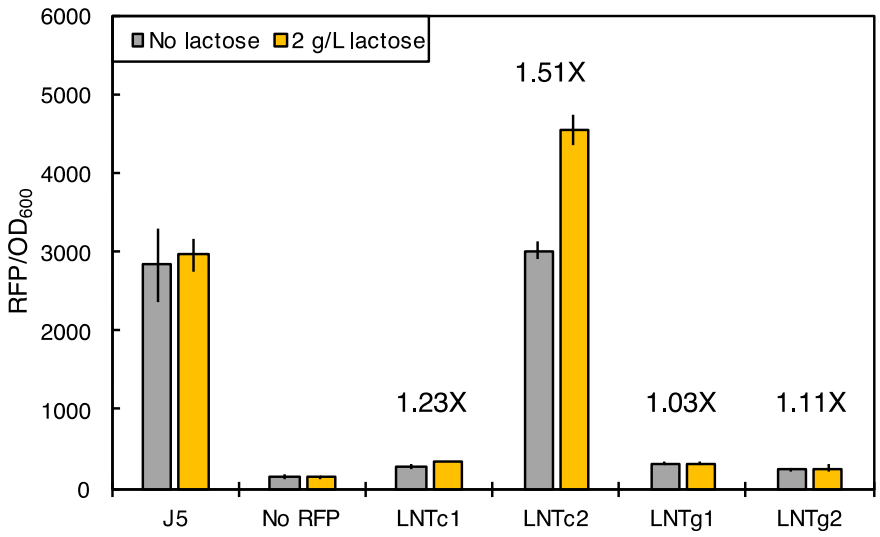
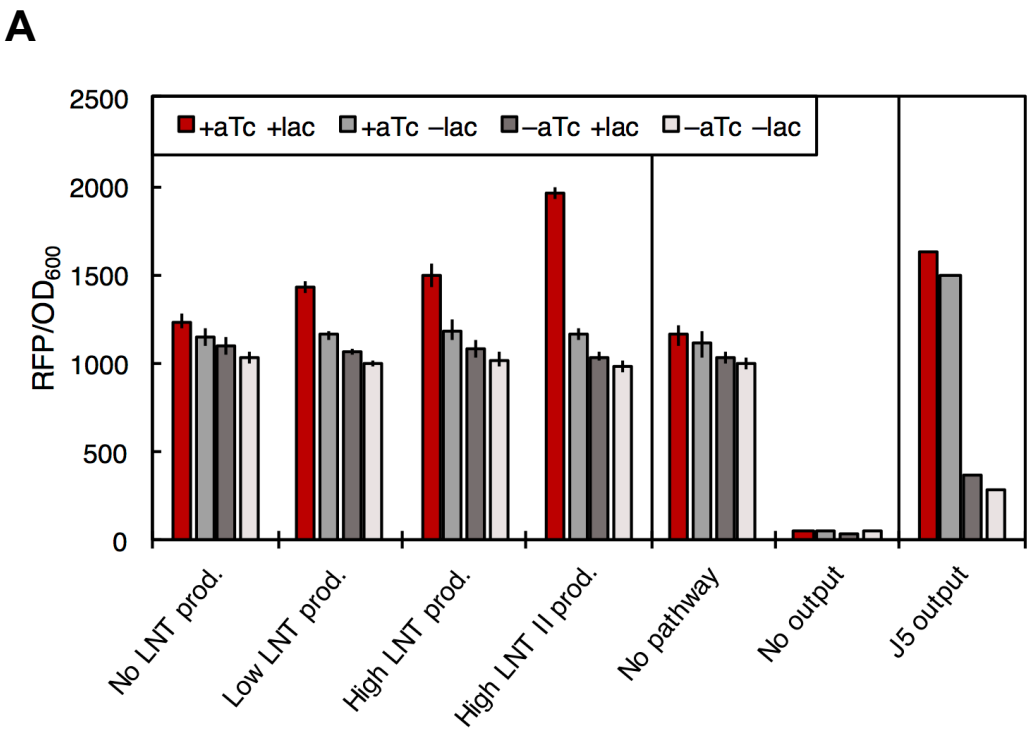


Figure 4.S9. RFP output levels of biosensor candidates. While LNTc2 starts at high background in the absence of production, addition of lactose increases output further. No such increase is observed in the other candidates, perhaps because the OFF-state output is so near the autofluorescence level, or because sensor folding into the ON-state is delayed by production until free dCas9 is rare. An increase in output scale or sensor expression, then, might increase dynamic range for LNTc1, g1, and g2. Sensors in this experiment had not yet received a timer domain. Error bars represent standard deviation of *n*=3 biological replicates.



B

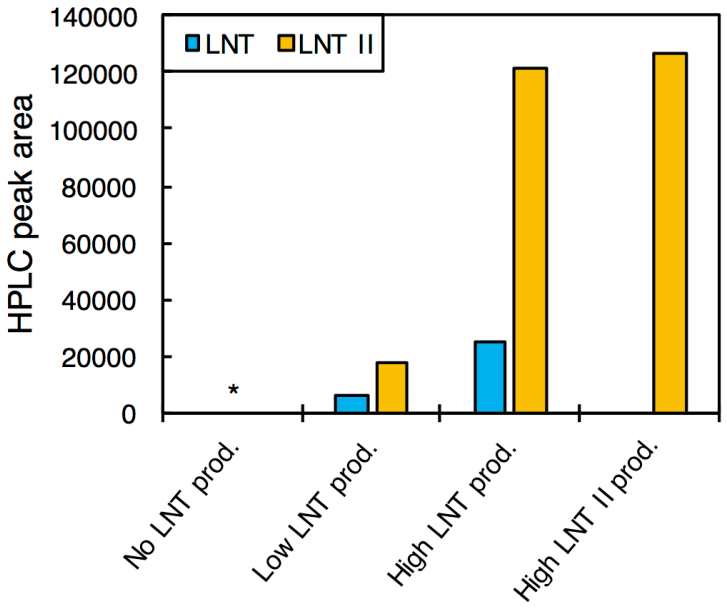
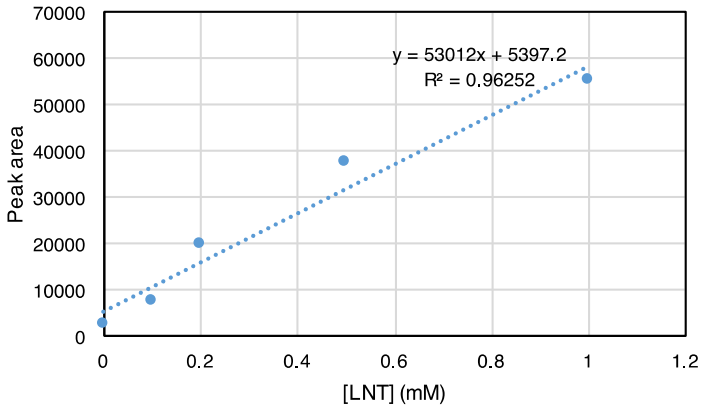


Figure 4.S10. Production variation in LNTc2 Variant 1 sensor strain. **A)** Unsubtracted RFP/OD_{600} values of the same strains assessing sensitivity and specificity in Figure 4.13. In all strains, the RFP signal increases slightly when aTc or lactose are added, but these effects are superseded in production strains by larger increases in signal when both aTc and lactose are added. **B)** The same production strains in Figure 4.13 were assessed by HPLC, verifying low and high LNT production and similarly high LNT II levels in two strains (right) with different RFP output levels. It is important to remember that HPLC reports here on extracellular metabolite levels in culture supernatants, while the biosensor reports on intracellular metabolite levels. Asterisk indicates a technically failed sample.

A



B

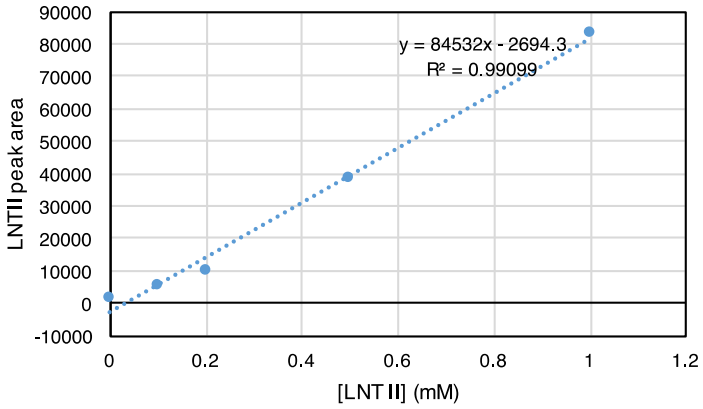


Figure 4.S11. Typical standard curves for HPLC quantification of (A) LNT and (B) LNT II. With this HPLC protocol, there is sometimes an interfering peak close to LNT, making peak area difficult to assess exactly. Quantification suffers, but qualitative presence or absence is well-indicated. LNT peak occurs at 9.5 minutes, LNT II peak occurs at 11 minutes.

	Selection 1	**Selection 2**
Days per round	2	1
[Mg^{2+}]	1 mM	2 mM
BSA on column	None	100 µg/mL prewash
RNA melting	None	80˚C
Column load incubation	2-5 minutes	30 minutes
Bead functionalization	25%	100%
[LNT] in elution	7.5 mM	30 mM
Elution incubation	Overnight	1 hour

Table 4.S1. Revised selection conditions. All of these changes after the first selection were aimed at optimizing binding of aptamers to the column, and making sure binders were successfully eluted. The elution incubation had to be shortened to allow for the compacted overall schedule (one round per day).

Chapter 5: Conclusions

In this work, we have seen how consideration of control circuit design is an important factor in the successful engineering of metabolic pathways for production of valuable chemicals. In many cases, the goal of achieving industrially-relevant product titers demands a large—preferably tunable and modular—control circuit, whether for variant screening or metabolic manipulation or both, and a particular attention to the metabolic demands of the heterologous system, relative to the capabilities of the host. Toward that end, we generated expression capacity data to show features of CRISPR-based expression control that are well-suited to these applications—namely, its inexpensive expandability into numerous orthogonal and tunable control nodes—and explored the effects of simple autoregulatory motifs on efficiency and stability. We then demonstrated the utility of CRISPR control of biosynthetic pathways not only for heterologous gene expression, but for adding in endogenous metabolism perturbations and metabolite sensing capabilities to the control circuit. Together, these efforts draw upon the versatility of CRISPR as a controller of heterologous expression and as a genetic biosensor platform, and indicate important future routes of metabolic engineering.

In Chapter 1, we developed an adapted method of measuring the host's expression capacity, as a marker of metabolic burden, and paired it with reference strains designed to isolate measurement of particular sources of burden. In doing so, we took advantage of the modularity of CRISPR expression control and orthogonal T7 RNAPs to separate controller burden from output burden, and the burden contributions of each of the CRISPR components. We found that the greatest contribution to overall burden was from the output expression rather than the controller, and that this output burden responded linearly to expression level, up to a certain point (sometimes reaching 90% reduction in expression capacity). Beyond this point, it is likely that energy or material resources are taxed to exhaustion, resulting in poorly-quantified (but harshly-observed) phenotypes like growth arrest, genetic instability, and population heterogeneity. Despite different controllers' propensities to reach this point more or less

easily, within the linear range the efficiency of output expression does not significantly differ for a given output. We take this to mean that controllers should be chosen to fit their purpose mainly in an effort to avoid the phenotypes of instability—for example, T7 RNAP is a poor choice for expression control within a complex circuit, because fluctuations in its expression level due to the circuit's retroactivity can easily push the T7's activity to levels that cause problematic phenotypes.

Because of the dramatic difference inside versus outside of the linear range of burden per output, in Chapter 2 we explored negative autoregulation as a means of forcing output expression to remain within that linear range. We found variable results from protein effectors of negative feedback, with LacI seemingly contributing amounts of burden that led to instability, and T7 lysozyme limiting output while leaving efficiency unchanged. We developed instead a method of connecting an RNA effector to output expression, therefore coupling its effectiveness to the main source of burden in the circuit. Again, output was reduced from two different controllers, while efficiency remained unchanged—though output was limited enough to prevent instability. A lack of tunability in the CRISPRi feedback led us to adopt an IFFL architecture, with an unconnected but simultaneously activatable gRNA, that did prove tunable in its output level, showed similar efficiency to other circuits, and again remained stable while an open-circuit control strain failed. Though they did not appear in the pathway control circuits in later chapters, these motifs show promise for enhancing the stability of CRISPR control circuits, as long as they can be tuned to provide adequate output.

In what was a demonstration of CRISPR control's portability, we constructed a pathway producing *p*-ACA in *P. putida* as part of a larger project implementing genome-scale modeling for iterative strain improvement through DBTL cycles, an effort detailed in Chapter 3. In the first reported bacterial production of *p*-ACA (from a feedstock within endogenous metabolism), we optimized expression levels of pathway enzymes from plasmids and the genome to achieve production that showed low unactivated titers, but reached hundreds of micromolar when activated. Not only are the heterologous genes independently tunable, but we developed a workflow for the eventual testing of arrays of

scRNAs and gRNAs aimed at perturbation of endogenous metabolism, for optimizing flux through the heterologous enzymes in upcoming DBTL cycles.

In Chapter 4, we made more systematic use of the independently tunable control channels, this time in a pathway producing LNT. We assessed every combination of four activation strengths of each pathway gene, highlighting the roughly one-third of strains that produced significant amounts of LNT. While metabolic perturbations using an array of gRNAs are also under way in this strain and show some promise, we primarily expanded the control circuit in this case to accommodate a novel ligand-responsive scRNA switch as an intracellular biosensor for LNT detection. This involved *in vitro* selection of an RNA aptamer to a novel target, computational screening of candidate connections of that aptamer to a scRNA effector domain, using previously determined design rules, and an *in vivo* screen of a pool of timer domains, culminating in two scRNAs responsive to LNT production. As currently implemented, their RFP output is useful for high-throughput screening of strain variants, potentially relieving a severe bottleneck in strain development—but the *trans*-acting nature of scRNAs allows easy adaptation of the sensor to different outputs involved in selective pressure or further strain control through dynamic circuits.

References

1. Nielsen, J. & Keasling, J. D. Engineering Cellular Metabolism. *Cell* **164**, 1185–1197 (2016).
2. Wu, G. *et al.* Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends in Biotechnology* **34**, 652–664 (2016).
3. Lee, S. Y. & Kim, H. U. Systems strategies for developing industrial microbial strains. *Nat Biotechnol* **33**, 1061–1072 (2015).
4. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).
5. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research* **41**, 7429–7437 (2013).
6. Zalatan, J. G. *et al.* Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell* **160**, 339–350 (2015).
7. Dong, C., Fontana, J., Patel, A., Carothers, J. M. & Zalatan, J. G. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nature Communications* **9**, (2018).
8. Fontana, J. *et al.* Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements. *Nature Communications* **11**, 1–11 (2020).
9. Fontana, J. *et al.* Rapid metabolic pathway profiling with rationally-designed multi-gene CRISPRa programs. *In preparation* (2021).
10. Jayanthi, S., Nilgiriwala, K. S. & Del Vecchio, D. Retroactivity Controls the Temporal Dynamics of Gene Transcription. *ACS Synthetic Biology* **2**, 431–441 (2013).
11. Brewster, R. C. *et al.* The Transcription Factor Titration Effect Dictates Level of Gene Expression. *Cell* **156**, 1312–1323 (2014).
12. Qian, Y., Huang, H.-H., Jiménez, J. I. & Del Vecchio, D. Resource Competition Shapes the Response of Genetic Circuits. *ACS Synthetic Biology* **6**, 1263–1272 (2017).
13. Zhang, S. & Voigt, C. A. Engineered dCas9 with reduced toxicity in bacteria: implications for genetic circuit design. *Nucleic Acids Research* (2018) doi:10.1093/nar/gky884.
14. Michener, J. K., Thodey, K., Liang, J. C. & Smolke, C. D. Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metabolic Engineering* **14**, 212–222 (2012).
15. Varman, A. M., Xiao, Y., Leonard, E. & Tang, Y. J. Statistics-based model for prediction of chemical biosynthesis yield from Saccharomyces cerevisiae. *Microb Cell Fact* **10**, 45 (2011).

16. Gorochowski, T. E., Avcilar-Kucukgoze, I., Bovenberg, R. A. L., Roubos, J. A. & Ignatova, Z. A Minimal Model of Ribosome Allocation Dynamics Captures Trade-offs in Expression between Endogenous and Synthetic Genes. *ACS Synthetic Biology* **5**, 710–720 (2016).

17. Klumpp, S. & Hwa, T. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proceedings of the National Academy of Sciences* **105**, 20245–20250 (2008).

18. Carrera, J., Rodrigo, G., Singh, V., Kirov, B. & Jaramillo, A. Empirical model and in vivo characterization of the bacterial response to synthetic gene expression show that ribosome allocation limits growth rate. *Biotechnology Journal* **6**, 773–783 (2011).

19. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science* **330**, 1099–1102 (2010).

20. Tan, C., Marguet, P. & You, L. Emergent bistability by a growth-modulating positive feedback circuit. *Nature Chemical Biology* **5**, 842–848 (2009).

21. Klumpp, S., Zhang, Z. & Hwa, T. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell* **139**, 1366–1375 (2009).

22. Dong, H., Nilsson, L. & Kurland, C. G. Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. *Journal of bacteriology* **177**, 1497–1504 (1995).

23. Friehs, K. Plasmid Copy Number and Plasmid Stability. in *New Trends and Developments in Biochemical Engineering* vol. 86 47–82 (Springer, Berlin, Heidelberg, 2004).

24. Albermann, C., Trachtmann, N. & Sprenger, G. A. A simple and reliable method to conduct and monitor expression cassette integration into the *Escherichia coli* chromosome. *Biotechnol. J.* **5**, 32–38 (2010).

25. 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.* **2**, 519–528 (2013).

26. Rugbjerg, P. & Sommer, M. O. A. Overcoming genetic heterogeneity in industrial fermentations. *Nat Biotechnol* **37**, 869–876 (2019).

27. Davanloo, P., Rosenberg, A. H., Dunn, J. J. & Studier, F. W. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proceedings of the National Academy of Sciences* **81**, 2035–2039 (1984).

28. Ceroni, F., Algar, R., Stan, G.-B. & Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nature Methods* **12**, 415–418 (2015).

29. Meyer, A. J., Ellefson, J. W. & Ellington, A. D. Directed Evolution of a Panel of Orthogonal T7 RNA Polymerase Variants for *in Vivo* or *in Vitro* Synthetic Circuitry. *ACS Synthetic Biology* **4**, 1070–1076 (2015).

30. Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F. & Voigt, C. A. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Research* **40**, 8773–8781 (2012).

31. Gyorgy, A. *et al.* Isocost Lines Describe the Cellular Economy of Genetic Circuits. *Biophysical Journal* **109**, 639–646 (2015).

32. 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. *Molecular Systems Biology* **10**, 742–742 (2014).

33. Jones, R. D. *et al.* An endoribonuclease-based feedforward controller for decoupling resource-limited genetic modules in mammalian cells. *Nat Commun* **11**, 5690 (2020).

34. Vind, J., Sorensen, M., Rasmussen, M. & Pedersen, S. Synthesis of Proteins in Escherichia coli is Limited by the Concentration of Free Ribosomes. *Journal of Molecular Biology* **231**, 678–688 (1993).

35. Iost, I., Guillerez, J. & Dreyfus, M. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. *Journal of bacteriology* **174**, 619–622 (1992).

36. Lopez, P. & Dreyfus, M. The lacZ mRNA can be stabilised by the T7 late mRNA leader in E coli. *Biochimie* **78**, 408–415 (1996).

37. Makarova, O. V., Makarov, E. M., Sousa, R. & Dreyfus, M. Transcribing of Escherichia coli genes with mutant T7 RNA polymerases: stability of lacZ mRNA inversely correlates with polymerase speed. *Proceedings of the National Academy of Sciences* **92**, 12250–12254 (1995).

38. Borkowski, O. *et al.* Cell-free prediction of protein expression costs for growing cells. *Nature communications* **9**, 1457 (2018).

39. Lee, T. S. *et al.* BglBrick vectors and datasheets: a synthetic biology platform for gene expression. *Journal of biological engineering* **5**, 1 (2011).

40. Fontana, J., Dong, C., Ham, J. Y., Zalatan, J. G. & Carothers, J. M. Regulated Expression of sgRNAs Tunes CRISPRi in *E. coli*. *Biotechnology Journal* **13**, 1800069 (2018).

41. Kosuri, S. *et al.* Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **110**, 14024–14029 (2013).

42. Cochrane, F. C., Davin, L. B. & Lewis, N. G. The Arabidopsis phenylalanine ammonia lyase gene family: kinetic characterization of the four PAL isoforms. *Phytochemistry* **65**, 1557–1564 (2004).

43. Gonçalves, G. A. L., Prazeres, D. M. F., Monteiro, G. A. & Prather, K. L. J. De novo creation of MG1655-derived E. coli strains specifically designed for plasmid DNA production. *Appl Microbiol Biotechnol* **97**, 611–620 (2013).

44. Gander, M. W., Vrana, J. D., Voje, W. E., Carothers, J. M. & Klavins, E. Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nature Communications* **8**, 15459 (2017).

45. Milo, R. & Alon, U. Network Motifs: Simple Building Blocks of Complex Networks. *Science* **298**, (2002).

46. Nielsen, A. A. & Voigt, C. A. Multi-input CRISPR / C as genetic circuits that interface host regulatory networks. *Mol Syst Biol* **10**, 763 (2014).

47. Nevozhay, D., Adams, R. M., Murphy, K. F., Josić, K. & Balázsi, G. Negative autoregulation linearizes the dose–response and suppresses the heterogeneity of gene expression. *Proceedings of the National Academy of Sciences* **106**, 5123–5128 (2009).

48. Lv, Y. *et al.* Coupling feedback genetic circuits with growth phenotype for dynamic population control and intelligent bioproduction. *Metabolic Engineering* **54**, 109–116 (2019).

49. Huang, H.-H. *et al.* dCas9 regulator to neutralize competition in CRISPRi circuits. *Nat Commun* **12**, 1692 (2021).

50. Segall-Shapiro, T. H., Sontag, E. D. & Voigt, C. A. Engineered promoters enable constant gene expression at any copy number in bacteria. *Nature Biotechnology* **36**, 352–358 (2018).

51. Strovas, T. J., Rosenberg, A. B., Kuypers, B. E., Muscat, R. A. & Seelig, G. MicroRNA-Based Single-Gene Circuits Buffer Protein Synthesis Rates against Perturbations. *ACS Synthetic Biology* **3**, 324–331 (2014).

52. Mishra, D., Rivera, P. M., Lin, A., Del Vecchio, D. & Weiss, R. A load driver device for engineering modularity in biological networks. *Nature Biotechnology* **32**, 1268–1275 (2014).

53. Cheng, X., Zhang, X., Pflugrath, J. W. & Studier, F. W. The structure of bacteriophage T7 lysozyme, a zinc amidase and an inhibitor of T7 RNA polymerase. *Proceedings of the National Academy of Sciences* **91**, 4034–4038 (1994).

54. Jeruzalmi, D. & Steitz, T. A. Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *The EMBO Journal* **17**, 4101–4113 (1998).

55. Moffatt, B. A. & Studier, F. W. T7 Lysozyme Inhibits Transcription by T7 RNA Polymerase. *Cell* **49**, 221–227 (1987).

56. Samuelson, J. C., Davis, T. B., Raleigh, E. A. & Southworth, M. W. *Expression of toxic genes in vivo in a non-natural host*. (Google Patents, 2013).

57. Kushwaha, M. & Salis, H. M. A portable expression resource for engineering cross-species genetic circuits and pathways. *Nature Communications* **6**, 7832 (2015).

58. Gao, Y. & Zhao, Y. Self-processing of ribozyme-flanked RNAs into guide RNAs *in vitro* and *in vivo* for CRISPR-mediated genome editing. *J. Integr. Plant Biol.* **56**, 343–349 (2014).

59. Tickman, B. I. *et al.* Multi-Layer CRISPRa/i Circuits Generate Rationally-Tunable Dynamic Expression Programs in Cell-Free and Bacterial Systems. *Cell Systems* In Press (2021).

60. Bellato, M. *et al. CRISPR interference as low burden logic inverters in synthetic circuits: characterization and tuning*. http://biorxiv.org/lookup/doi/10.1101/2020.08.03.234096 (2020) doi:10.1101/2020.08.03.234096.

61. Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249–253 (2012).

62. Brunk, E. *et al.* Characterizing Strain Variation in Engineered E. coli Using a Multi-Omics-Based Workflow. *Cell Systems* **2**, 335–346 (2016).

63. Carbonell, P. *et al.* An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals. *Commun Biol* **1**, 66 (2018).

64. Liu, R., Bassalo, M. C., Zeitoun, R. I. & Gill, R. T. Genome scale engineering techniques for metabolic engineering. *Metabolic Engineering* **32**, 143–154 (2015).

65. Kiattisewee, C. *et al.* Portable bacterial CRISPR transcriptional activation enables metabolic engineering in Pseudomonas putida. *Metabolic Engineering* **66**, 283–295 (2021).

66. Tan, S. Z., Reisch, C. R. & Prather, K. L. J. A Robust CRISPR Interference Gene Repression System in Pseudomonas. *J Bacteriol* **200**, (2018).

67. Tsuge, Y., Kawaguchi, H., Sasaki, K. & Kondo, A. Engineering cell factories for producing building block chemicals for bio-polymer synthesis. *Microb Cell Fact* **15**, 19 (2016).

68. Sariaslani, F. S. Development of a Combined Biological and Chemical Process for Production of Industrial Aromatics from Renewable Resources. *Annu. Rev. Microbiol.* **61**, 51–69 (2007).

69. Kawaguchi, H., Hasunuma, T., Ogino, C. & Kondo, A. Bioprocessing of bio-based chemicals produced from lignocellulosic feedstocks. *Current Opinion in Biotechnology* **42**, 30–39 (2016).

70. Stevens, J. T. & Carothers, J. M. Designing RNA-Based Genetic Control Systems for Efficient Production from Engineered Metabolic Pathways. *ACS Synthetic Biology* **4**, 107–115 (2015).

71. Qi, W. *et al.* Functional expression of prokaryotic and eukaryotic genes in Escherichia coli for conversion of glucose to pp-hydroxystyrene. *Metabolic Engineering* **9**, 268–276 (2007).

72. Suvannasara, P. *et al.* Biobased Polyimides from 4-Aminocinnamic Acid Photodimer. *Macromolecules* **47**, 1586–1593 (2014).

73. Molina-Santiago, C. *et al.* Pseudomonas putida as a platform for the synthesis of aromatic compounds. *Microbiology* **162**, 1535–1543 (2016).

74. Ger, Y.-M., Chen, S.-L., Chiang, H.-J. & Shiuan, D. A Single Ser-180 Mutation Desensitizes Feedback Inhibition of the Phenylalanine-Sensitive 3-Deoxy-D-Arabino-Heptulosonate 7-Phosphate (DAHP) Synthetase in Escherichia coli1. *The Journal of Biochemistry* **116**, 986–990 (1994).

75. Konishi, K., Takaya, N., Masuo, S. & Zhou, S. 4-amino Cinnamic Acid Production Method Using Enzyme. (2018).

76. Bentley, G. J. *et al.* Engineering glucose metabolism for enhanced muconic acid production in Pseudomonas putida KT2440. *Metabolic Engineering* **59**, 64–75 (2020).

77. Tian, T., Kang, J. W., Kang, A. & Lee, T. S. Redirecting Metabolic Flux *via* Combinatorial Multiplex CRISPRi-Mediated Repression for Isopentenol Production in *Escherichia coli*. *ACS Synth. Biol.* **8**, 391–402 (2019).

78. Masuo, S., Zhou, S., Kaneko, T. & Takaya, N. Bacterial fermentation platform for producing artificial aromatic amines. *Scientific Reports* **6**, (2016).

79. Mi, J. *et al.* Investigation of plasmid-induced growth defect in Pseudomonas putida. *Journal of Biotechnology* **231**, 167–173 (2016).

80. Wirth, N. T., Kozaeva, E. & Nikel, P. I. Accelerated genome engineering of *Pseudomonas putida* by I- *Sce* I—mediated recombination and CRISPR -Cas9 counterselection. *Microb. Biotechnol.* **13**, 233–249 (2020).

81. Clamons, S. & Murray, R. *Modeling predicts that CRISPR-based activators, unlike CRISPR-based repressors, scale well with increasing gRNA competition and dCas9 bottlenecking*. http://biorxiv.org/lookup/doi/10.1101/719278 (2019) doi:10.1101/719278.

82. Reis, A. C. *et al.* Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays. *Nat Biotechnol* **37**, 1294–1301 (2019).

83. Juminaga, D. *et al.* Modular Engineering of l-Tyrosine Production in Escherichia coli. *Appl. Environ. Microbiol.* **78**, 89–98 (2012).

84. Tian, T. & Salis, H. M. A predictive biophysical model of translational coupling to coordinate and control protein expression in bacterial operons. *Nucleic Acids Res* **43**, 7137–7151 (2015).

85. Bode, L. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* **22**, 1147–1162 (2012).

86. Deng, J. *et al.* Recent advances and challenges in microbial production of human milk oligosaccharides. *Syst Microbiol and Biomanuf* **1**, 1–14 (2021).

87. Kobata, A. Structures and application of oligosaccharides in human milk. *Proc. Jpn. Acad., Ser. B* **86**, 731–747 (2010).

88. Sprenger, G. A., Baumgärtner, F. & Albermann, C. Production of human milk oligosaccharides by enzymatic and whole-cell microbial biotransformations. *Journal of biotechnology* **258**, 79–91 (2017).

89. Engfer, M. B., Stahl, B., Finke, B., Sawatzki, G. & Daniel, H. Human milk oligosaccharides are resistant to enzymatic hydrolysis in the upper gastrointestinal tract. *The American Journal of Clinical Nutrition* **71**, 1589–1596 (2000).

90. Kulinich, A. & Liu, L. Human milk oligosaccharides: The role in the fine-tuning of innate immune responses. *Carbohydrate Research* **432**, 62–70 (2016).

91. Peterson, R., Cheah, W. Y., Grinyer, J. & Packer, N. Glycoconjugates in human milk: Protecting infants from disease. *Glycobiology* **23**, 1425–1438 (2013).

92. Newburg, D. S. & Morelli, L. Human milk and infant intestinal mucosal glycans guide succession of the neonatal intestinal microbiota. *Pediatr Res* **77**, 115–120 (2015).

93. Asakuma, S. *et al.* Physiology of Consumption of Human Milk Oligosaccharides by Infant Gut-associated Bifidobacteria. *Journal of Biological Chemistry* **286**, 34583–34592 (2011).

94. Hill, D. R. & Newburg, D. S. Clinical applications of bioactive milk components. *Nutrition Reviews* **73**, 463–476 (2015).

95. Urashima, T., Saito, T., Nakamura, T. & Messer, M. Oligosaccharides of milk and colostrum in non-human mammals. *Glycoconjugate Journal* **18**, 357–371 (2001).

96. Tao, N. *et al.* Bovine Milk Glycome. *Journal of Dairy Science* **91**, 3768–3778 (2008).

97. Tadasu Urashima, Tadao Saito, Kenzi Ohmisya, & Keiichi Shimazaki. Structural determination of three neutral oligosaccharides in bovine (Holstein-Friesian) colostrum, including the novel trisaccharide; GalNAcαl-3Galβ1-4Glc. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1073**, 225–229 (1991).

98. Faijes, M., Castejón-Vilatersana, M., Val-Cid, C. & Planas, A. Enzymatic and cell factory approaches to the production of human milk oligosaccharides. *Biotechnology Advances* **37**, 667–697 (2019).

99. Pritchard, D. G., Gray, B. M. & Egan, M. L. Murine Monoclonal Antibodies to Type Ib Polysaccharide of Group B Streptococci Bind to Human Milk Oligosaccharides. *INFECT. IMMUN.* 5.

100. Appelmelk, B. J. *et al.* Potential role of molecular mimicry between Helicobacter pylori lipopolysaccharide and host Lewis blood group antigens in autoimmunity. *Infect Immun* **64**, 2031–2040 (1996).

101. Priem, B., Gilbert, M., Wakarchuk, W. W., Heyraud, A. & Samain, E. A new fermentation process allows large-scale production of human milk oligosaccharides by metabolically engineered bacteria. *Glycobiology* **12**, 235–240 (2002).

102. Baumgärtner, F., Conrad, J., Sprenger, G. A. & Albermann, C. Synthesis of the human milk oligosaccharide lacto-N-tetraose in metabolically engineered, plasmid-free E. coli. *Chembiochem* **15**, 1896–1900 (2014).

103. Dong, X. *et al.* Modular pathway engineering of key precursor supply pathways for lacto-N-neotetraose production in Bacillus subtilis. *Biotechnology for biofuels* **12**, 1–11 (2019).

104. Dong, X. *et al.* CRISPRi-guided multiplexed fine-tuning of metabolic flux for enhanced Lacto-N-neotetraose production in Bacillus subtilis. *Journal of Agricultural and Food Chemistry* (2020).

105. Zhang, W. *et al.* Metabolic engineering of Escherichia coli for the production of Lacto-N-neotetraose (LNnT). *Syst Microbiol and Biomanuf* **1**, 291–301 (2021).

106. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).

107. Werstuck, G. Controlling Gene Expression in Living Cells Through Small Molecule-RNA Interactions. *Science* **282**, 296–298 (1998).

108. Blomberg, L., Wieslander, J. & Norberg, T. Immobilization of Reducing Oligosaccharides to Matrices by a Glycosylamide Linkage. *J Carb Chem* **12**, 265–276 (1993).

109. Hwang, C. & Carothers, J. M. Label-free selection of RNA aptamers for metabolic engineering. *Methods* **106**, 37–41 (2016).

110. Fontana, J., Sparkman-Yager, D., Zalatan, J. G. & Carothers, J. M. Challenges and opportunities with CRISPR activation in bacteria for data-driven metabolic engineering. *Current Opinion in Biotechnology* **64**, 190–198 (2020).

111. Blixt, O., van Die, I., Norberg, T. & van den Eijnden, D. H. High-level expression of the Neisseria meningitidis lgtA gene in Escherichia coli and characterization of the encoded N-acetylglucosaminyltransferase as a useful catalyst in the synthesis of GlcNAcβ1→ 3Gal and GalNAcβ1→ 3Gal linkages. *Glycobiology* **9**, 1061–1071 (1999).

112. Liu, X. *et al.* Characterization and synthetic application of a novel β1,3-galactosyltransferase from Escherichia coli O55:H7. *Bioorganic & Medicinal Chemistry* **17**, 4910–4915 (2009).

113. Eames, M. & Kortemme, T. Cost-Benefit Tradeoffs in Engineered lac Operons. *Science* **336**, 911–915 (2012).

114. Wagner, S. *et al.* Consequences of Membrane Protein Overexpression in Escherichia coli*. *Molecular & Cellular Proteomics* **6**, 1527–1550 (2007).

115. McArthur, J. B., Yu, H. & Chen, X. A Bacterial β1–3-Galactosyltransferase Enables Multigram-Scale Synthesis of Human Milk Lacto- *N* -tetraose (LNT) and Its Fucosides. *ACS Catal.* **9**, 10721–10726 (2019).

116. Kundert, K. *et al.* Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs. *Nat Commun* **10**, 2127 (2019).

117. Tang, W., Hu, J. H. & Liu, D. R. Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nat Commun* **8**, 15939 (2017).

118. Rogers, J. K., Taylor, N. D. & Church, G. M. Biosensor-based engineering of biosynthetic pathways. *Current Opinion in Biotechnology* **42**, 84–91 (2016).

119. Iost, I. & Dreyfus, M. The stability of Escherichia coli lacZ mRNA depends upon the simultaneity of its synthesis and translation. *The EMBO journal* **14**, 3252 (1995).

120. Avcilar-Kucukgoze, I. & Ignatova, Z. Rewiring host activities for synthetic circuit production: a translation view. *Biotechnology Letters* (2016) doi:10.1007/s10529-016-2229-6.

121. Rogers, J. K. & Church, G. M. Genetically encoded sensors enable real-time observation of metabolite production. *Proceedings of the National Academy of Sciences* **113**, 2388–2393 (2016).

122. Wu, S. G., He, L., Wang, Q. & Tang, Y. J. An ancient Chinese wisdom for metabolic engineering: Yin-Yang. *Microbial Cell Factories* **14**, (2015).

123. Kotte, O., Volkmer, B., Radzikowski, J. L. & Heinemann, M. Phenotypic bistability in Escherichia coli's central carbon metabolism. *Molecular Systems Biology* **10**, 736–736 (2014).

124. Emory, S. A., Bouvet, P. & Belasco, J. G. A 5'-terminal stem-loop structure can stabilize mRNA in Escherichia coli. *Genes & development* **6**, 135–148 (1992).

125. Der, B. S. *et al.* DNAplotlib: programmable visualization of genetic designs and associated data. *ACS Synthetic Biology* (2016) doi:10.1021/acssynbio.6b00252.

126. Lo, T.-M., Chng, S. H., Teo, W. S., Cho, H.-S. & Chang, M. W. A Two-Layer Gene Circuit for Decoupling Cell Growth from Metabolite Production. *Cell Systems* **3**, 133–143 (2016).

127. Pu, J., Zinkus-Boltz, J. & Dickinson, B. C. Evolution of a split RNA polymerase as a versatile biosensor platform. *Nature Chemical Biology* **13**, 432–438 (2017).

128. Bergmiller, T. *et al.* Biased partitioning of the multidrug efflux pump AcrAB-TolC underlies long-lived phenotypic heterogeneity. *Science* **356**, 311–315 (2017).

129. Ceroni, F. *et al.* Burden-driven feedback control of gene expression. *Nature Methods* **15**, 387–393 (2018).

130. Chen, Y.-J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nature methods* **10**, 659 (2013).

131. Matsumoto, I., Ito, Y. & Seno, N. Preparation of affinity adsorbents with Toyopearl gels. *Journal of Chromatography A* **239**, 747–754 (1982).

132. Gilbert, M. *et al.* Cloning of the lipooligosaccharide α-2, 3-sialyltransferase from the bacterial pathogens Neisseria meningitidis and Neisseria gonorrhoeae. *Journal of Biological Chemistry* **271**, 28271–28276 (1996).

133. Wakarchuk, W., Martin, A., Jennings, M. P., Moxon, E. R. & Richards, J. C. Functional relationships of the genetic locus encoding the glycosyltransferase enzymes involved in expression of the lacto-N-neotetraose terminal lipopolysaccharide structure in Neisseria meningitidis. *Journal of Biological Chemistry* **271**, 19166–19173 (1996).

134. Carothers, J. M., Goler, J. A., Kapoor, Y., Lara, L. & Keasling, J. D. Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity. *Nucleic acids research* **38**, 2736–2747 (2010).

135. Carrier, T. A. & Keasling, J. D. Library of Synthetic 5' Secondary Structures To Manipulate mRNA Stability in Escherichia coli. *Biotechnol. Prog.* **15**, 58–64 (1999).

136. Philp, J. C., Ritchie, R. J. & Allan, J. E. M. Biobased chemicals: the convergence of green chemistry with industrial biotechnology. *Trends in Biotechnology* **31**, 219–222 (2013).

137. Grubbe, W. S., Rasor, B. J., Krüger, A., Jewett, M. C. & Karim, A. S. Cell-free styrene biosynthesis at high titers. *Metabolic Engineering* **61**, 89–95 (2020).

138. Walton, S. J., Clamons, S. E. & Murray, R. M. *Analysis of Circuits for Dosage Control in Microbial Populations*. http://biorxiv.org/lookup/doi/10.1101/2020.12.18.423556 (2020) doi:10.1101/2020.12.18.423556.

139. Venayak, N., Anesiadis, N., Cluett, W. R. & Mahadevan, R. Engineering metabolism through dynamic control. *Current Opinion in Biotechnology* **34**, 142–152 (2015).

140. Frumkin, I. *et al.* Gene Architectures that Minimize Cost of Gene Expression. *Molecular Cell* **65**, 142–153 (2017).

141. González-Colell, M. & Macía, J. General Analyses of Gene Expression Dependencies on Genetic Burden. *Front. Bioeng. Biotechnol.* **8**, 1017 (2020).

142. Brockman, I. M. & Prather, K. L. J. Dynamic knockdown of E. coli central metabolism for redirecting fluxes of primary metabolites. *Metabolic Engineering* **28**, 104–113 (2015).

143. Minakawa, H., Masuo, S., Kaneko, T. & Takaya, N. Fermentation and purification of microbial monomer 4-amminocinnamic acid to produce ultra-high performance bioplastics. *Process Biochemistry* **77**, 100–105 (2019).

144. Zhao, D. *et al.* CRISPR-based metabolic pathway engineering. *Metabolic Engineering* **63**, 148–159 (2021).

145. Cuba Samaniego, C. & Franco, E. Ultrasensitive molecular controllers for quasi-integral feedback. *Cell Systems* **12**, 272-288.e3 (2021).

146. Kim, J., Darlington, A. P. S., Bates, D. G. & Jimenez, J. I. *The interplay between growth rate and nutrient quality defines gene expression capacity*. http://biorxiv.org/lookup/doi/10.1101/2021.04.02.438188 (2021) doi:10.1101/2021.04.02.438188.

147. Lim, H. G., Jang, S., Jang, S., Seo, S. W. & Jung, G. Y. Design and optimization of genetically encoded biosensors for high-throughput screening of chemicals. *Current Opinion in Biotechnology* **54**, 18–25 (2018).

148. Xu, L. L. & Townsend, S. D. Synthesis as an Expanding Resource in Human Milk Science. *J. Am. Chem. Soc.* **143**, 11277–11290 (2021).

149. Win, M. N. & Smolke, C. D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proceedings of the National Academy of Sciences* **104**, 14283–14288.

150. Chauvier, A. *et al.* Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. *Nat Commun* **8**, 13892 (2017).

151. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).

152. Lai, D., Proctor, J. R. & Meyer, I. M. On the importance of cotranscriptional RNA structure formation. *RNA* **19**, 1461–1473 (2013).

Acknowledgements

Modern science is no individual endeavor, and involves a degree of inherent collaboration that is almost impossible to fully acknowledge. From the shoulders of giants we explicitly cite some works forming a foundation to ours, but implicitly understand the partial works and overall body of knowledge that hold the foundation together. This dependence is as relevant to an individual journey toward a Ph.D. as it is to broader scientific progress. This work, and my ability to perform it, would not exist remotely in this form without the contributions, support, and adjacent work of all former and current members of the Carothers lab and collaborating labs, and they share some responsibility for any accomplishments presented in this document. Though all of these contributions are too numerous and extensive to list here, the following are many of the much-appreciated bricks of my scientific and personal foundations:

James Carothers, who as an advisor afforded the scientific freedom to follow my interests, however long they took to develop into interesting answers to interesting questions, who is dedicated to letting students develop on their own into experts in their field, and who, with an open mind, encourages students to consider realistic constraints on potential projects, all of which strongly develop self-directed learning and critical thinking skills,

The members of my committee, Jesse Zalatan, Cole DeForest, Maitreya Dunham, Georg Seelig, and Herbert Sauro, whose guidance has perpetually improved my skills as a scientist and as a communicator,

The funders who have made this work possible—the NSF for the metabolic burden work, the DOE for the p-ACA production work, and especially BASF for the LNT biosensor work—and who have shown remarkable understanding, patience, and flexibility during the disruptions of the COVID-19 pandemic.

David Sparkman-Yager, whose deep knowledge of RNA design borders on wizardry (but is much more rational, predictable, and tunable!) and was the driving force behind the LNT biosensor,

Aria Walls, the hands behind many of the most intricate experiments presented in this work, whose development as a scientist has been delightful to observe, shape, and benefit from,

Chuhern Hwang, whose efforts to understand and optimize aptamer selections were a laudable sacrifice for the greater good, and whose inexorable mountaineering is an inspiration,

Jason Fontana and Chen Dong, whose creativity and work toward standardization of expression has made so much of our lab's current work possible,

Jason Stevens, whose research, construction, and troubleshooting laid the groundwork for bacterial *p*-ACA production,

Cassie Burke, legend of the Carothers lab, whose drive to help others with tasks ranging from mundane to complex is almost automatic,

Willy Voje, a source of diverse wisdom and most entertaining of late-night grading partners, who has directly and indirectly affected the graduate school experience in many ways,

Cholpisit Kiattisewee, Ice, master of *P. putida* and other bacteria, whose dedication to the advancement of practical tools of biotechnology enables so much impactful work,

Ben Tickman, whose need for scientific exploration is matched only by his need for topographical exploration,

Pramod Chavali, purveyor of boundless enthusiasm for lab work and lab conversation,

Ryan Cardiff, who spearheaded endogenous CRISPRi for LNT production,

Sai Samineni, Maritza Zaldivar, and Aamy Woldesenbet, who contributed not only lab work but valuable conversation that helped frame and guide this work,

And friends and family, most of all Katie Schmidt, who is an extraordinary partner in life and in all our endeavors, joint or individual, familiar or novel, and who proved that true dedication of time to a singular purpose can be enriching, rewarding, even delightful,

My parents, Dave and Peggy, and my extended family, whose wonder acted like fuel for accomplishing interesting and impactful science,

Maggie, whose knowledge, passion, and occasional disruption are inspirational, and Garrett, whose musical selections powered the creation of this document,

My industry coworkers, especially Soma, Lauren, Shawn, Erin, Jon, Mack, Moses, Geoff, and Denitza, whose encouragement that graduate research was a good fit proved convincing and correct,

Andrew, the source of many stimulating skintrack discussions about this work, and the many friends with whom I've shared delightful and revitalizing experiences over these years, most often Alex, Michaela, Alaina, Carly, Chris, Avnie, Jimmy, Denise, Carolyn, Dan, Kirill, and Eugene,

And finally, the classmates from my cohort and beyond, especially Jay, Claire, Ben, Neal, Chris, Ryan, Kelly, Jon, Barry, Caitlyn, and Victor, who helped keep me focused on the important things.