Synthetic Biology Approaches for Engineering Diverse Bacterial Species

by

Jennifer Ann Noelani Brophy
B.S., University of California Berkeley (2009)

Submitted to the Department of Biological Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2016

[June 2016]

Signature redacted

Jennifer A.N. Brophy
Author

Signature redacted

Christopher A. Voigt
Professor of Biological Engineering
Thesis Advisor

Signature redacted

Alan D. Grossman
Praecis Professor of Biology
Head, Department of Biology
Thesis Advisor

Signature redacted

Forest M. White
Professor of Biological Engineering
Graduate Program Chair

This doctoral thesis has been examined by a committee of the Department of Biological Engineering as follows:

Douglas A. Lauffenburger
Ford Professor of Biological Engineering, Chemical Engineering and Biology
Head, Department of Biological Engineering
Thesis Committee Chair

Christopher A. Voigt
Professor of Biological Engineering
Thesis Advisor

Alan D. Grossman
Praecis Professor of Biology
Head, Department of Biology
Thesis Advisor

Michael T. Laub
Associate Professor of Biology
Committee Member

Synthetic Biology Approaches for Engineering Diverse Bacterial Species

by

Jennifer Ann Noelani Brophy

Submitted to the Department of Biological Engineering
in May 2016 in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Biological Engineering

Abstract

When engineers control gene expression, cells can be re-programmed to create living therapeutics or materials by initiating expression of biosynthetic pathways in response to specific signals. In this thesis, two new genetic tools were developed to aid the construction of genetic circuits and facilitate their delivery to bacteria isolated from diverse environments. First, antisense transcription was explored as a new tool for tuning gene expression in *Escherichia coli*. Antisense transcription was found to reliably repress gene expression and was applied tune simple genetic circuits. Second, an integrative conjugative element from *Bacillus subtilis*, ICE*Bs1*, was engineered to deliver exogenous DNA to diverse strains of undomesticated Gram-positive bacteria. Engineered ICE*Bs1* conjugation was demonstrated in twenty different bacterial strains, spanning sixteen species and five genera. To demonstrate ICE's utility in creating new probiotics, the element was used to deliver functional nitrogen fixation pathways (*nif* clusters) to bacteria isolated from agricultural soils. Collectively, the tools presented here in provide a platform for programing bacteria from diverse environments for advanced applications.

Supervisors: Christopher A. Voigt & Alan D. Grossman

Acknowledgements

I would first like to thank my advisors, Professors Christopher Voigt and Alan Grossman for their guidance and support throughout my time at MIT. Thank you both giving me the freedom to pursue the research presented in this thesis and for providing me with so many opportunities to learn about different aspects of academic science. Thank you for continually challenging me to learn new skills and for supporting every scientific endeavor I have attempted, even the unsuccessful ones. I have been incredibly fortunate to work in both of your labs and appreciate all that you have done to help me grow as a scientist.

I would next like to thank my thesis committee members, Professors Mike Laub and Doug Lauffenburger, who provided helpful guidance and thoughtful contributions to my research throughout my time at MIT. I am especially grateful to Doug for your unwavering encouragement and support. Your support of me and your maintenance of an incredibly positive environment within MIT's Biological Engineering department have made me feel welcome and supported throughout my PhD.

A very special thank you to fellow residents at Simmons Hall. I would never have made it as a GRT without the help of the other Simmons GRTs, housemasters Professor John Essigmann, Dr. Ellen Essgimann, and Professor Steve Hall, and Area Director Joshua Gonzolez. Your unwavering positivity and strength has been invaluable over the years, so thank you for your kindness and friendship. I'd also like to thank Professor Katharina Ribbeck for initiating Zen Café at Simmons Hall and infusing some peace into the stressful whirlwind that sometimes accompanied graduate school. To all of the fabulous undergraduates who lived on the 7th floor of Simmons during my tenure, thank you for not burning down the building on my watch. It has been an honor and a pleasure to live with you all for the last four years.

Many many thanks to all the members of the Voigt lab, past and present, for making everything about graduate school a little bit better. Thank you for all of the technical help. Your scientific advice was truly indispensible for the success of the projects presented in this thesis and thank you for enabling so much goofing off over the years. The memes, gifs, coffee dates, Tatte breaks, and happy hours made the long days spent in lab much more fun. Special scientific thanks to Brian Caliando, MinHyung Ryu, Emerson Glassey, and Adam Meyer for patiently

answering all of my questions about basic science (Brian), *nif* clusters (MinHyung), and violacein production (Adam/Emerson).

To my friends in the BE department, thank you for getting me through the tough classes and long winters. Special thanks to Jen Wilson, Marcus Parrish, Erin Turowski, and Byron Kwan for helping me re-enact scenes from the Big Bang Theory on a weekly basis by getting together to eat Thai take-out and play increasingly complicated board games. You are like an East-coast family to me. Additional thanks to the California girls in Boston, Nitzan Koppel, Joy Makdisi, Anna Chen, and Nina Soltero (yes, I'm lumping you into this category now), for your friendship and love throughout my PhD.

Lastly, immense thanks and love to my parents, John and Bobbie, and brother, Jonathan, for your unconditional love and support. Thank you for travelling out to Boston so many times over the years, hosting thanksgiving dinners in my tiny apartment, braving the winters, and for always encouraging/supporting me. I am so incredibly blessed that you are my family. Thank you, for everything.

I gratefully acknowledge the funding sources that made this work possible: a fellowship from ENI through the MIT Energy Initiative, a National Science Foundation (NSF) Graduate Research Fellowship, the NSF-funded Synthetic Biology Engineering Research Center (SynBERC), the MIT Biological Engineering Department, and the Defense Advanced Research Projects Agency (DARPA).

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1-1: The history of interactions between humans and bacteria

Humans and bacteria have a complicated relationship. For the bulk of human history, bacteria were credited almost exclusively with causing disease. Plagues and pandemics caused by bacteria, such as the 5th Plague of Egypt (1250 B.C.E.)[1], the Plague of Athens (430 B.C.E.)[2], the Black Death (1346 to 1361 C.E.)[3], and the Paris Whooping Cough Epidemic (1578 C.E.)[4], wiped out significant portions of human populations in the past and often lead to a weakening of society in cities and nations (Fig 1-1). At first, doctors believed that these diseases were spread by *miasma*, a noxious form of "bad air" emanating from rotting organic matter. In the miasma theory of disease, illnesses were caused by environmental factors (miasma) present at specific locations and not by transmission between individuals[5]. However in the 19th century, humans began to realize that bacteria were responsible for causing some diseases. European physicians developed the germ theory of disease, which stated that microorganisms cause diseases when they grow and reproduce within a human host[6]. Acceptance of the germ theory of disease led to an increase in sanitation practices and a decline in devastating human pandemics. Since the mid-19th century researchers have identified several of the pathogenic bacteria that caused the major pandemics, *e.g Bacillus anthracis* (5th Plague of Egypt)[1], *Salmonella enterica serovar typhi* (Plague of Anthens)[2], *Yersinia pestis* (Black Death)[3], and *Bordetella pertussis* (Whooping cough)[4], using DNA recovered from teeth in mass graves and historical descriptions of illness.

Although the germ theory of disease correctly identified pathogenic bacteria as causative agents of illness, it missed the beneficial effects of other bacteria. In the late 19th century, researchers began to identify bacteria capable of positively influencing their hosts. One of the first symbiotic relationships between bacteria and a eukaryotic host was reported in 1889 by a Dutch microbiologist and botonist named Martinus Willem Beijerinck (1851-1931)[7]. Beijerinck demonstrated that legume roots contained bacteria that turned atmospheric nitrogen (N_2) into ammonium (NH_3)[8]. He named the bacteria *Rhizobia* and demonstrated that they provide the nitrogen needed for legume plants to grow. Around the same time, another bacterium *Bacillus thuringiensis* was isolated from silkworms by a Japanese biologist named Ishiwata Shigetane[9].

Strains of this bacterium kill insects and can protect plants from predators by living on its surfaces. More recently, researchers have identified beneficial bacteria in humans. Human symbiotic bacteria are typically studied en masse as a microbiota, *i.e.* a community of microorganisms living in a specific environment. Research into the human gut microbiota has demonstrated the importance of bacteria in defining human susceptibility to disease, nutrient absorption, immune function, and neural function[10]. In fact, the human gut microbiotia is often characterized as a "forgotten organ" because it plays so many important roles in human health[11]. Studies of the human skin[12], vaginal[13], and mouth[14] microbiota have revealed other important roles for bacteria in maintaining human health.

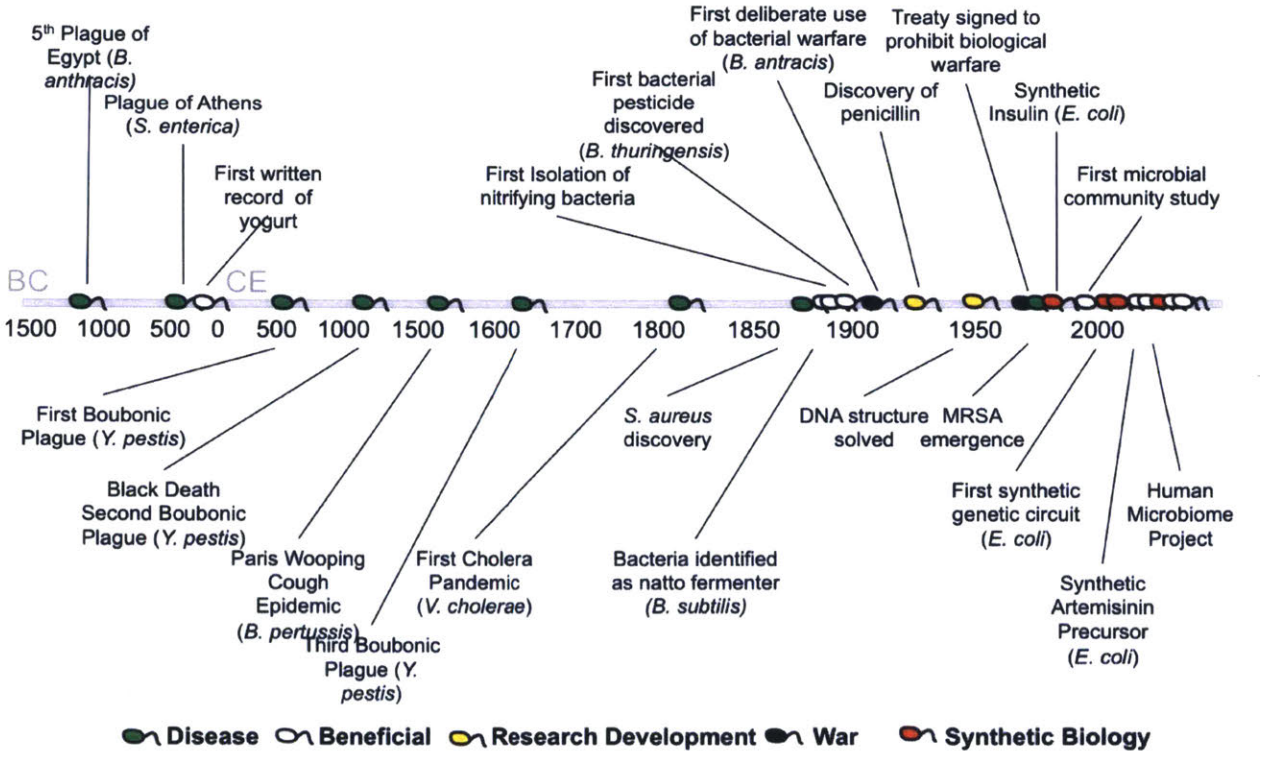

Figure 1-1: Timeline of human/bacteria interactions.

Although pathogenic and beneficial interactions between humans and bacteria typically arise without conscious human action, humans can deliberately use bacteria to produce goods. Bacterially fermented foods, such as yogurt (fermented milk), beer (fermented grain), and natto (fermented soy bean), have been produced by humans using bacteria since at least the 18th century B.C.E.[15]. Production of these goods relies on naturally occurring bacteria to ferment

milk, grain, and beans. However relying on naturally occurring bacteria limits the types of products that can be made using microorganisms. Fortunately, researchers developed cloning techniques that allow biologists to engineer bacteria and alter their behavior. Since the late 1970s, bacterial engineering has enabled fermentation-based production of drugs to treat diabetes (insulin)[16], malaria (artemisinin)[17], and cancer (taxol)[18]. Bacteria have also been engineered to detect toxins (arsenic)[19] and human pathogens (*Pseudomonas aeruginosa*)[20] to prevent illness. Though naturally occurring bacteria can positively impact our world, engineering opens up new ways for bacteria and humans to interact and may create new solutions to today's problems.

1-2: THE PROCESS OF ENGINEERING BACTERIA

The process of engineering bacteria can be crudely broken down into three steps (Fig 1-2). First, a researcher must select a strain of bacteria to modify. There is no shortage of bacteria with unique capabilities that can be used as starting strains for engineering projects. There are approximately $5x10^{30}$ bacteria on earth[21], many of which have interesting innate abilities, such as tolerance of extreme environments, utilization of unusual carbon sources, sensing/production of interesting molecules, etc. Ideally a researcher would engineer bacteria with the most advantageous characteristics for their purpose. These characteristics may include the ability to withstand a specific harsh condition, thrive in a competitive environment, secrete a massive amount of protein, or interact with other cell types. When selecting strains, researchers must also consider a bacterium's capacity to uptake and express genes from foreign DNA. Since most genetic engineering techniques/tools, *e.g.*, transformation methods, DNA vectors, antibiotic resistance cassettes, were developed for laboratory strains of model microorganisms, *e.g.*, *Escherichia coli* K12, *Saccharomyces cereviseae*, these strains are often used by engineers to create complex genetic programs, *e.g.*, multi-input logic gates[22], synthetic oscillators[23]. Unfortunately, genetic engineering techniques/tools can be difficult to use with non-model organisms. Thus, although unique strains of bacteria may be more desirable for specific applications, they may not be selected for engineering projects because they can be difficult to work with. The pros and cons of working with well-characterized model organisms need to be weighed when selecting strains to engineer.

The second step in engineering bacteria is to design and build DNA that alters the bacteria's behavior. Genes that produce the desired effect can be sourced from any organism. For example, if researchers want to produce the neuroactive compound phenylethylamine (PEA), an aromatic amino acid decarboxylase can be used to convert phenylalanine to the desired product[24]. This decarboxylase can be sourced from any organism, including complex eukaryotes, as long as it can catalyze the desired reaction. DNA sequences that encode the enzyme just need to be introduced to the bacteria with the correct control elements, *i.e.*, promoters, ribosome binding sites (RBSs), terminators, etc. These parts control the amount of gene expression, *i.e.*, mRNA levels and protein concentration, and are usually carefully selected to express the correct amount of mRNA and protein in the target bacteria. Over expression of the target molecules can overburden cells and slow growth[25], however under expression can result in weak phenotypes or poor yields of the desired product. Several large promoter, RBS, and terminator libraries have been built and tested in model organisms to facilitate the design of DNA that expresses the right amount of mRNA and protein[26,27]. These parts can be used with additional biochemical tuning knobs (discussed in Chapter 2) to achieve the desired expression levels.

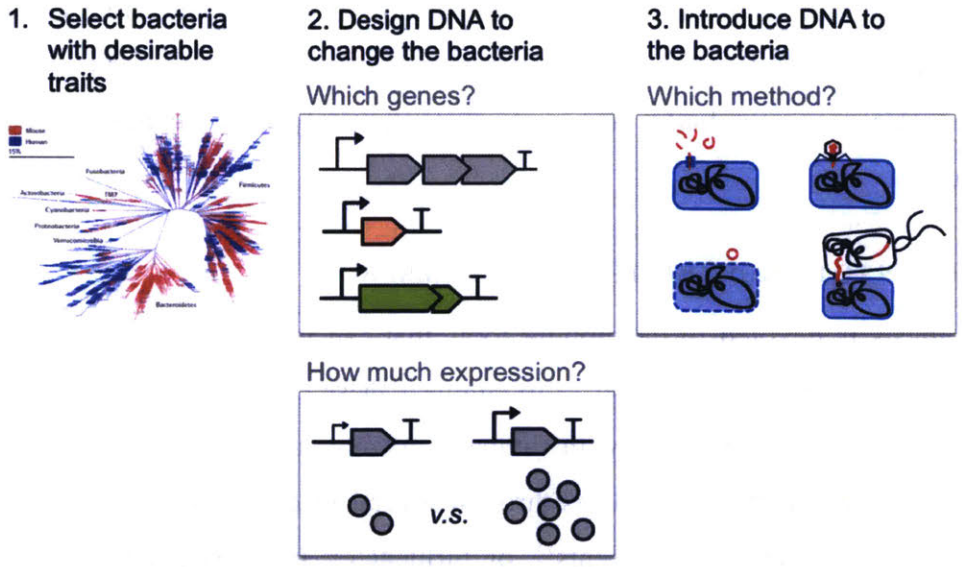

Figure 1-2: Steps to engineer bacteria.

Once the DNA is designed and built, the last step in engineering is to introduce it to bacteria. There are four commonly used strategies for introducing exogenous DNA to bacteria

(Fig 1-2). The first is to use a bacteria's natural ability to uptake extracellular DNA from their environment. This is a genetically encoded ability that few bacteria possess[28]. Bacteria that are not capable of natural competence can be made artificially competent with chemical or electrical shocks. Chemical and electrical shocks make bacterial membranes transiently permeable to DNA. Unfortunately, these methods can also kill cells and are not suitable for all cell types[29,30]. A third strategy is to use engineered bacteriophages to deliver DNA[31]. Delivering DNA with phage engineering requires isolating phage that can infect the target bacteria without causing lysis. Phage genomes can be modified to include exogenous segments of DNA. Unfortunately, the amount of DNA that can be delivered using engineered phage is limited by the phage capsid, which must package the DNA for delivery into target bacteria[32]. Finally, conjugation can be used to transfer genetic material between cells. Conjugation is a form of horizontal gene transfer that occurs frequently in nature[33]. Using conjugation to engineer bacteria in the lab requires two steps. First the foreign DNA must be introduced to a genetically tractable donor strain, then it can be conjugated into the bacteria of interest. Fortunately, it is easy to introduce foreign DNA to lab strains of *E. coli* and *B. subtilis*, which can be used as donors for conjugation.

After the new DNA is introduced to the bacteria of interest, the strain can be tested for the desired phenotype. If the phenotype is suboptimal, the second two steps can be iteratively repeated to create a design, build, test cycle that ultimately leads to the desired engineered bacteria product.

1-3: Techniques to improve bacterial engineering

This thesis is devoted to developing tools that can be used to engineer environmental isolates of non-model bacteria. The desire to engineer these organisms was inspired by bacteria that form symbiotic relationships with humans, plants, and animals. Symbiotic bacteria positively impact the health of their hosts through several mechanisms, such as protecting hosts from pathogens[12,13,34], providing nutrients[35], or influencing host behavior[36–38]. However, these bacteria could be engineered further improve our world by performing new functions. Some examples of new advantageous functions include: producing therapeutic drugs in specific locations in the human body to fight cancer and providing nutrients to specific food crops to reduce fertilizer input and weed growth. Engineered symbiotic bacteria can be thought of as

smart probiotics, which are expected to perform a beneficial function after being delivered to a target environment. One key challenge in developing smart probiotics is ensuring survival of the engineered bacterium in the target environment. The most well studied strains of bacteria, *e.g.*, *Escherichia coli* K-12, are poorly adapted for survival outside lab conditions. Therefore, smart probiotics would ideally be engineered using bacteria that have evolved to survive in the target environment. The tools developed herein can be used to deliver DNA to environmental isolates of bacteria and tune expression of the delivered genes. In addition to creating smart probiotics, the tools can be applied to engineer bacteria that possess unique properties for other purposes, such as solvent tolerant bacteria for the industrial production of enzymes or chemicals.

The first chapter of this thesis is a distillation of what is currently known about how to construct complex genetic programs in *Escherichia coli*. The chapter is intended to serve as a benchmark for current state of the art of constructing genetic circuits in model organisms. There is an emphasis on the tools that are available for constructing genetic circuits and methods that can be used to tune gene expression. Studies that characterize gene expression control elements, *e.g.*, sRNAs, ribosome binding sites (RBSs), UP-elements, for the specific purpose of using them as 'parts' to tune gene expression in synthetic circuits are summarized using mathematical models that demonstrate the impact of modifying/including these 'parts' in synthetic circuits. Commonly encountered failure modes, and techniques to circumvent them, are also described to emphasize the practical challenges associated with constructing complex genetic programs in bacteria.

After surveying the tools available for constructing genetic circuits in *E. coli*, we decided to characterize a method for tuning gene expression with potentially broad applicability across organisms. In the second chapter of this thesis, antisense transcription is rigorously characterized using a fluorescent reporter system and mathematical modeling. Antisense transcription is a pervasive biological phenomenon that has been detected in all three domains on the tree of life. We sought to rigorously quantify its impact on gene expression so that it could be used as another tool for precisely controlling gene expression. We envision engineers using antisense transcription as a secondary tool that can be layered on top of other methods of regulation to make small adjustments in expression. Since convergent promoters on the same segment of DNA generate antisense transcription, operon arrangement can now be used as a tuning knob. Using

the results of our study, engineers can quantitatively predict the change in gene expression produced by designing convergent transcriptional units in synthetic gene clusters.

Two aspects of the antisense transcription project are especially important. First, antisense transcription may be a method of repressing gene expression that is applicable across all cell types. Although we characterize the phenomena in *E. coli*, the basic tenants of antisense transcription should hold true in other organisms, including eukaryotes. The qualitative effect that antisense transcription has on gene expression (repression) should be universal, even if the quantitative predictions made by our model are not. Second, this project is an example of a synthetic biology study that makes mechanistic predictions about a biological process. Specifically, the mathematical model of RNA polymerase collision that we built and parameterized using experimental data predicts that RNA polymerase collision rarely results in the release of the polymerase(s) from the DNA ($\leq$ 50%). We believe that mechanistic insight into biological processes could be obtained by analyzing more data from synthetic biology studies, especially because synthetic biology projects often produce large datasets that could be mined for more information.

The last chapter of this thesis is aimed at generating a tool for delivering DNA to undomesticated bacterial species. We recognize that DNA delivery is the first challenge, after isolation and culturing, associated with engineering new bacterial species. In order to begin building sophisticated genetic programs in new strains/isolates of bacteria, we must first be able to reliably introduce foreign DNA to the organism. We adapt an integrative conjugative element from the soil bacterium *Bacillus subtilis* as a tool for introducing DNA to undomesticated isolates of Gram-positive bacteria. This chapter fits into the broader goal of being able to engineer the *best* bacteria for a specific application by developing a tool that can be used to naively engineer new bacterial isolates.

Chapter 2: Practical applications of genetic circuits

This chapter is an introduction to genetic circuit design. It is focused on the tools that are currently available for constructing genetic circuits and the potential applications of genetic circuits in medicine, industrial chemical production, and agriculture. First, the parts that are currently available for constructing genetic circuits are presented and ordinary differential equation models are used to show how regulator choices and "tuning knobs" can influence circuit dynamics. Next, commonly encountered failure modes and the constraints that arise from operating within a living cell are discussed. Ultimately, this introduction attempts to demonstrate that better tools, well-characterized parts, and a comprehensive understanding of how to compose circuits are leading to a breakthrough in the ability to program living cells for advanced applications.

2-1. Introduction

The ability to perform computation in a living cell will revolutionize biotechnology by improving existing products and enabling new applications. In the short term, genetic circuits could be used to improve bio-based chemical production by inducing gene expression at different stages of fermentation or limiting expression of an enzyme to specific conditions (*e.g.*, low oxygen)[39–44]. As circuits become more advanced, entire algorithms from control theory could be applied to further improve biochemical production (Fig 2-1a)[45–54]. Synthetic regulation could also be used to discover new natural products, including pharmaceuticals and insecticides, by stimulating expression of biosynthetic gene clusters that are not expressed in laboratory conditions[55–60]. Outside the fermenter, living cells could be programmed to serve as therapeutic agents that correct genetic disease (Fig 2-1b) or colonize niches in the human microbiome to perform a therapeutic function (Fig 2-1c)[61–73]. Longer-term capabilities include "smart plants" that sense and adapt to environmental challenges (Fig 2-1d) and bacteria that organize to weave functional materials with nanoscale features[74–80].

Despite its potential, genetic circuit design remains one of the most challenging aspects of genetic engineering[81]. The earlier fields of protein and metabolic engineering have yielded

tools to optimize enzymes and flux through a metabolic network. These tools include computational methods that can predict the impact of an amino acid substitution on protein thermostability[82] or the distribution of flux through modified metabolic networks[83]. Biotech companies often have research groups dedicated to protein and metabolic engineering that have specialized training in these tools. However, industrial groups dedicated to building synthetic regulation are rare and even simple tasks, like building a switch or inducible system, tend to be one-off projects performed by a non-specialist.

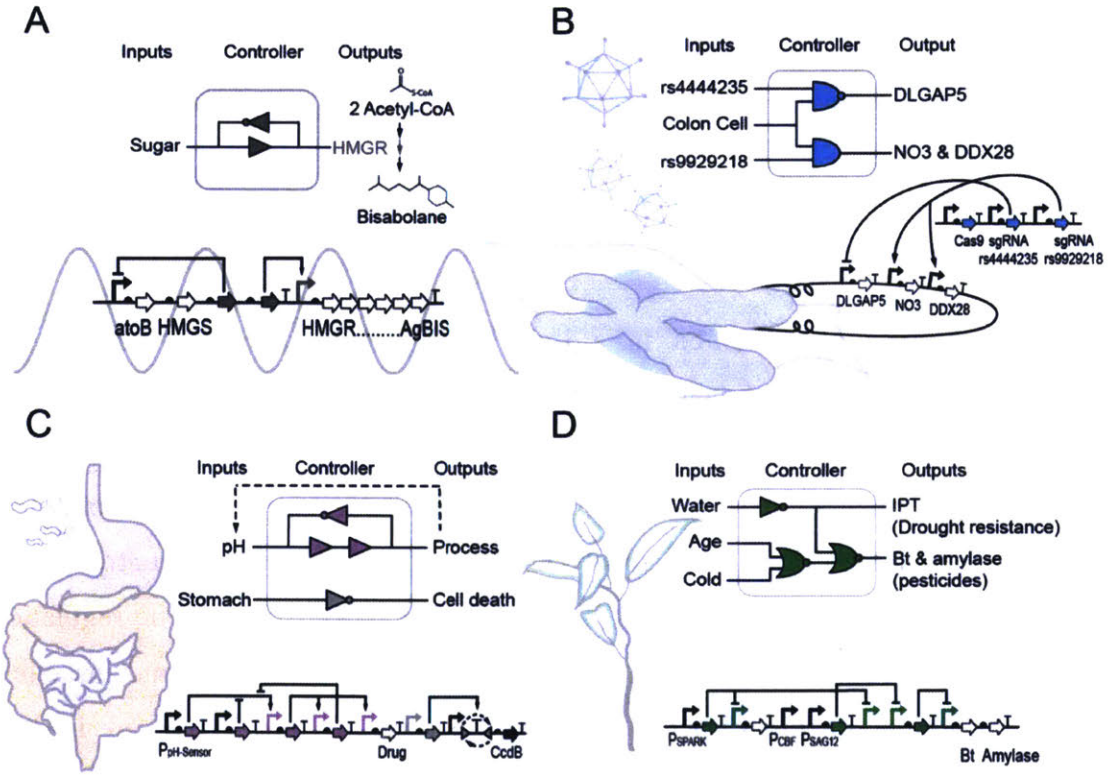

Figure 2-1: Potential uses of synthetic genetic circuits.
(a) In industrial applications, most synthetic metabolic pathways are overexpressed at all times or are under simple inducible control. This could be improved by incorporating timing, feedback of metabolic intermediates, or dynamic control. Here, we show a circuit that is controlling the production of a diesel fuel alternative (bisobolane[84]) by regulating the accumulation of a toxic intermediate (HMG-CoA) by sensing sugar, which induces oscillations in the production of HMGR. This type of oscillatory control occurs in natural metabolic networks[85]. (b) Gene therapy circuits could be built based on CRISPRi technology by detecting SNPs and integrating this information with tissue-specific sensors. As a hypothetical example, we show a circuit that could detect two SNPs associated with colon cancer susceptibility (rs4444235 and rs9929218)[86] and this is integrated with a promoter that is specific to colon cells (pAMUC2) to control the expression of misregulated genes (DLGAP5, NO3, and DDX28). (c) Bacteria could be programmed to colonize human gut and implement a therapeutic response. An example is

envisioned where a commensal bacterium is used to stabilize pH and treat gastoesophageal acid reflux (GERD). A bacterium that naturally resides in the stomach could be programmed to maintain a specific pH using a set point control circuit[47,87] whose output is a proton pump inhibitor (PPI). The imagined circuit would restrict acid regulation to the stomach by terminating the bacterium via an irreversible switch if it leaves this organ[88]. (d) Genetic circuits could also be used to build "smart plants" that are able to sense environmental stimuli and implement a response. Currently, traits are produced all the time whether or not they are needed by the plant. Here, we envision a circuit that would operate in the chloroplast integrate sensors for drought (pSpark), temperature (pCBF), and plant maturity (pSAG12) to control multiple traits. This could reduce the amount of recombinant protein that is produced and enters the food supply without reducing the effectiveness of the trait.

Several features of genetic circuits make them challenging to work with. First, circuits require the precise balancing of their component regulators to generate the proper response[89,90]. Computational tools and part libraries that enable the tuning of expression levels have only been developed recently[91–94]. Before this, only course-grained control was achievable with small sets of parts[89,90,95]. Second, many circuits are difficult to screen in directed evolution experiments. Digital logic has clear on and off states that can form the basis for a screen[50,96–102]. However, screening for functional dynamic circuits, such as oscillators, is significantly more complex[103] and it is hard to imagine how screens would be established for more sophisticated functions, like a PID controller with proscribed response properties. Third, there are few tools to measure circuit performance. Typically, a fluorescent reporter is used to measure the output, but fluorescence detection requires artificially high expression levels and fluorescent protein degradation rates can limit the ability to measure dynamics. Fourth, synthetic circuits are very sensitive to environment, growth conditions, and genetic context in ways that are poorly understood[104]. Finally, the process of building a large genetic circuit requires the assembly of many DNA parts and this process has been both technically challenging (until recently) and fraught with its own sources of errors[101,105–110].

This chapter can serve as a guide for designing prokaryotic transcriptional circuits, where both the inputs and outputs are promoters[96,98,111–114]. Transcriptional circuits maintain a common signal carrier, which simplifies the connection of circuits to build up sophisticated operations[115]. Post-transcriptional circuits, including those based on protein and RNA interactions, are covered in excellent reviews[116–118]. Although the majority of this guide is dedicated to bacterial circuits, many of the principles, albeit not the details, are relevant for eukaryotes, including human cells and plants[119,120].

2-2. GENETIC CIRCUIT DESIGN BASED ON DIFFERENT REGULATOR CLASSES

Transcriptional circuits operate by affecting the flow of RNA polymerase (RNAP) on DNA. There are a number of molecules that impact transcription that have been used as the basis for building synthetic circuits (Fig 2-2). For example, DNA-binding proteins can recruit or block RNAP to increase or decrease the flux, respectively. Analogously, the new CRISPRi system can use the Cas9 protein to bind DNA and alter transcription[121,122]. RNAP flux can also be altered with invertases that change the orientation of promoters, terminators, or gene sequences. Finally, RNA translational repressors, such as RNA-IN/OUT, can be converted to transcriptional regulators to control RNAP flux[123,124]. In this section, we describe recent advances in these methods and analyze the impact that each regulator has on circuit response.

2-2-1. DNA-binding Proteins

Many families of proteins can bind to specific DNA sequences (operators). The simplest way to use these proteins as regulators is to use them as repressors, which block the binding of RNAP to promoters or inhibit elongation. Repressors have been built out of zinc finger proteins (ZFPs)[125], transcription activator-like effectors (TALEs)[126,127], TetR homologues[114], phage repressors[128,129], and LacI homologues[130]. A core set of ~3 repressors were re-used in many of the first synthetic circuits (CI, TetR, LacI)[90,96,131–134]. However, recently there have been efforts to expand the number of DNA-binding proteins that are available for circuit design[97,135–142]. Expanding protein libraries can be challenging because each repressor has to be orthogonal; *i.e.*, only interact with their operator and not the others in the set. Because of their simple function, repressors are relatively easy to move between species, including to eukaryotes[135–140]. DNA-binding proteins can also function as activators that increase the flux of RNAP on DNA. Activators either recruit host RNAP to a promoter or are alternative RNAPs that transcribe genes directly. Recent efforts have increased the number of such proteins that are available for constructing circuits[97,141–143].

Many logic gates have been constructed with DNA binding proteins that recruit or block RNAP[114,144–152]. For example, NOT and NOR gates have been built with inducible promoters that drive expression of repressors (Fig 2-2a)[114,15390,96,131]. Additional transcriptional logic has been achieved with regulators that bind directly to the proteins to either inhibit or enhance their

function. For example, AND gates have been built with activators that require chaperones (Fig 2-2b)[98,144] and with artificially split proteins[154]. Similarly, NAND gates can be built with proteins that block the activity of an activator, such as anti-σ factors, which inhibit σ factors[143]. Of the regulators described in this review, DNA-binding proteins are the only class (so far) that has been used to build dynamic circuits. This includes pulse generators[155], bistable switches[90,96,156], counters[112], feedback loops, and oscillators that have different periods and amplitudes[113,131,157–159]. Analog computing modules have also been built with DNA binding proteins, which highlights the diverse signal processing capabilities of these regulators[98,114,144,146,153].

2-2-2. Invertases

Invertases are site-specific recombinase proteins that facilitate the inversion of DNA segments between binding sites[160]. All invertases mediate "cut-and-paste" recombination, during which DNA is looped, cleaved and re-ligated[161]. Two types of invertases have been used to build genetic circuits. The first are tyrosine recombinases, such as Cre, Flp, and FimBE, which require host-specific factors[112,162–164]. These recombinases can be reversible and flip the DNA in both directions, or irreversible and only flip DNA in a single direction. The second class of invertases is serine integrases, which catalyze unidirectional reactions that rely on double stranded breaks to invert DNA. Serine integrases typically do not require host factors and often have cognate excisionases that can be expressed independently to return the DNA to its original orientation.

Invertases have been used to build switches[162], memory circuits[163,164], counters[112], and logic gates [165,166]. These proteins are ideal for memory storage, because they flip DNA permanently and do not require the continuous input of materials or energy to maintain their new orientation. In invertase logic gates, discrete physical states of the DNA can correspond to on and off states (0 and 1). However, using invertases can be challenging because their reactions are slow (requiring 2-6 hours) and can generate mixed populations when targeting a multicopy plasmid[164]. Reversible invertases can also generate mixed populations, however this limitation was overcome recently by using a serine integrase to flip DNA in one direction and an integrase/exisionase pair to return it to the original state[167].

All two-input gates, including AND and NOR logic, have been constructed using orthogonal serine integrases (Fig 2-2cd)[165,166]. The gates are organized such that two input

promoters express a pair of orthogonal recombinases, which change RNAP flux by inverting unidirectional terminators, promoters, or entire genes. These gates are based on unidirectional serine integrases without excisionases, therefore they operate as memory circuits that remember exposure to two input signals. Once flipped, the circuits cannot be returned to their original state, therefore the gates do not distinguish the order they were exposed to the inputs or even if they occurred at the same time.

2-2-3. CRISPR

CRISPR (Clustered Regularly Spaced Short Palindromic Repeat) arrays function as a bacterial "immune system" that targets specific DNA sequence motifs for degradation[168]. CRISPR systems utilize a Cas nuclease and guide RNA to introduce double strand breaks to specific DNA sequences[169]. Mutant Cas proteins (dCas9[122], $Cas9_N$-[170]) that do not have nuclease activity have been developed and used as transcription factors that knock down gene expression by forming a DNA bubble that interferes with RNAP activity[121,122]. CRISPR machinery can also be used to activate transcription by fusing an RNAP recruiting domain to catalytically inactive Cas9[121,170–174]. Considering the needs of synthetic circuits, a significant advantage of CRISPR is the ability to design guide RNAs to target specific DNA sequences. A large set of orthogonal guide sequences that target different promoters would enable the construction of large genetic circuits.

CRISPRi is still relatively new and NOT gates are the most complex circuits built to date[122]. The NOT gates induce sgRNA and dCas9 expression simultaneously to repress transcription at an output promoter. In theory, a NOR gate could be created by introducing a second sgRNA that targets the same output promoter (Fig 2-2e). In general, CRISPRi circuits will probably resemble DNA binding protein circuits. However, unlike the repressor based NOR gate (Fig 2-2a), the CRISPRi NOR gate will need to have the sgRNAs expressed from separate input promoters because 5' RNA extensions can reduce or eliminate activity (Fig 2-2e)[175]. Circuits based on CRISPRi are expected to operate on similar timescales to protein-based circuits because of the stability of the regulatory dCas9/sgRNA/DNA duplex[122].

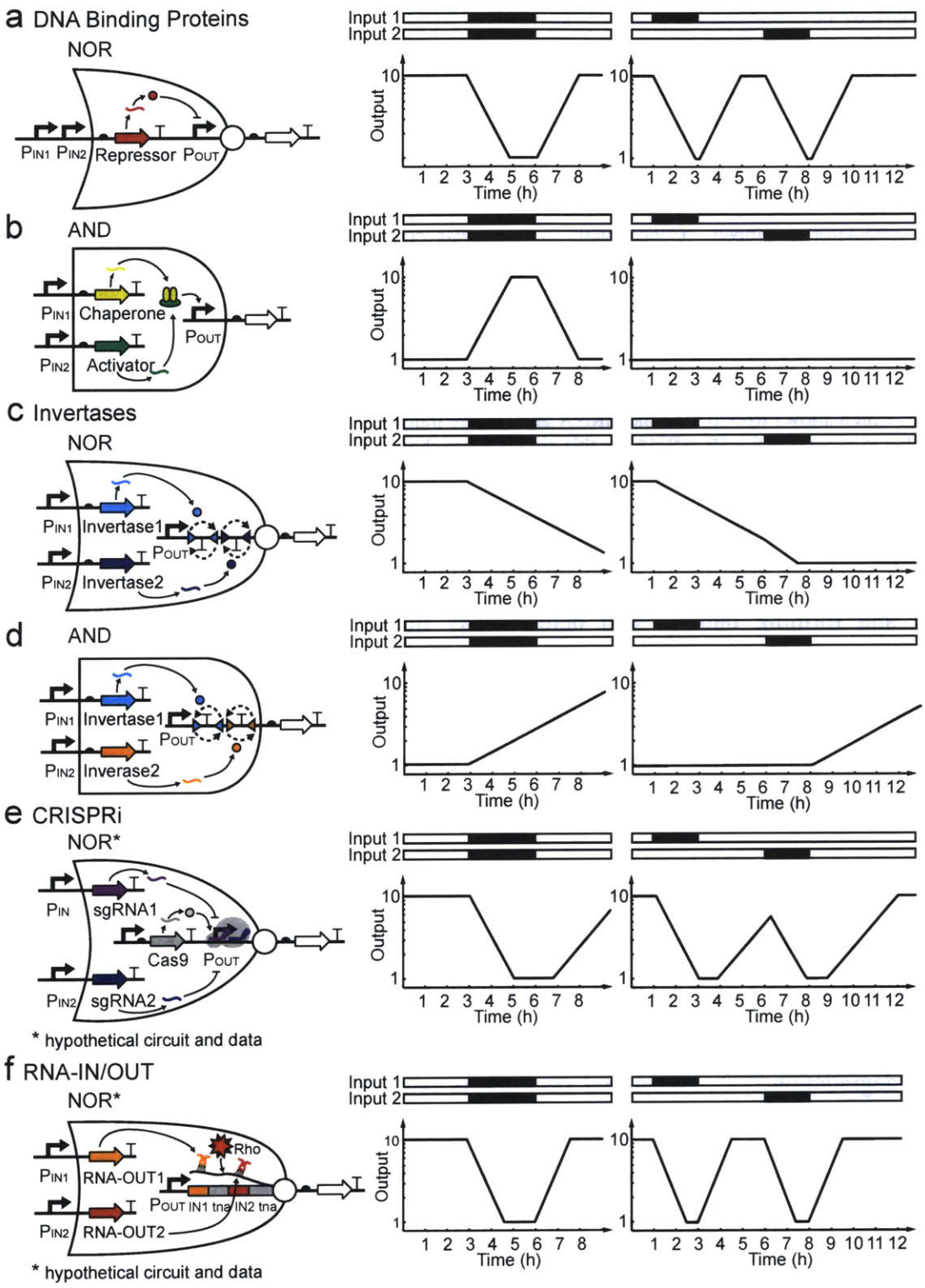

Figure 2-2: Logic gates built based on different regulator types.
All of the gates are transcriptional, where there are two input promoters (P_{IN1} and P_{IN2}) and one output promoter (P_{OUT}). Two-input transcriptional logic gates have not yet been built for CRISPRi and RNA-

IN/OUT so we hypothesize how these biochemistries could be used. The graphs at the right show how the gates will respond to inputs introduced at the same time (graphs at left) or sequentially (right). In all panels, the on state is assumed to generate ten fold higher than the off state. (a) A NOR gate is shown based on a repressor that binds DNA[153]. The lines are based on measured induction ($\tau_{1/2}$ ~36 min) and relaxation ($\tau_{1/2}$ ~35 min) half-lives[176]. (b) An AND gate based on an activator that binds DNA that requires a second protein to be active[98]. The lines are based on a measured induction ($\tau_{1/2}$ ~36 min)[98] and approximate relaxation ($\tau_{1/2}$ ~35 min) half-life. (c) A NOR gate based on integrases that flip two terminators to turn off the output[165,166]. We assume a small readthrough probability, which leads to a change in the rate when only one terminator is flipped. Conceptually, a NOR gate could also be constructed by having two input promoters in series drive the expression of a single integrase. The lines are based on an on-rate of 1.8 hours [162,164,165]. (d) An AND gate based on integrases[165]. The same on- and off- rates are used as in part c. (e) A NOR gate could be built based on CRISPRi by setting a constitutive level of Cas9 expression and then having the two input promoters drive the expression of two guide RNAs. The lines are based on measured induction ($\tau_{1/2}$ = 35 min) and relaxation ($\tau_{1/2}$ = 47 min) half-lives[122]. (f) A NOR gate could be built based on the RNA-IN/RNA-OUT system developed by Arkin and co-workers[123]. RNA-OUT represses translation of tnaC, which allows Rho to bind the mRNA and repress transcription of the output. The CRISPR machinery needed to process RNA-IN mRNA for this circuit is not shown. The lines are based on theoretical induction ($\tau_{1/2}$ ~30 min) and relaxation ($\tau_{1/2}$ ~35 min) half-lives[176,177].

A current challenge in implementing CRISPRi circuits is toxicity, which is difficult to control. Toxicity is most likely the result of Cas9 binding to the host genome at PAM sequences (NGG) and forming bubbles that deleteriously impact host gene expression[178,179]. Another consideration for building CRISPRi circuits is retroactivity[180], which could arise from using Cas9 as a shared resource (Section III). One way to circumvent retroactivity would be to express multiple orthogonal Cas9 homologues[181,182]. Finally, each guide RNA will need to be experimentally screened because predicting guide RNA orthogonality is complicated[175,181,183,184].

2-2-4. Adapted RNA-IN / RNA-OUT

The RNA-IN/OUT system from *E. coli* represses translation of a target protein when a short noncoding RNA (RNA-OUT) is expressed. In the natural system, RNA-OUT binds to a specific sequence at the 5'end of an mRNA (RNA-IN) to occlude ribosome binding and increase mRNA degradation[185–187]. Arkin and co-workers retooled this system to repress transcription, instead of translation, using a transcriptional adaptor from the *tna* operon[123]. The *tna* regulatory element is composed of a ribosome binding site (RBS), the coding sequence for a short peptide called *tnaC*, a Rho binding site and an RNAP pause site that facilitates Rho-mediated transcription termination. Translation of *tnaC* causes ribosomal stalling, which blocks Rho-factor binding and allows RNAP to transcribe genes downstream of *tnaC*. However, when translation of *tnaC* is prohibited, Rho binds the growing mRNA and knocks off RNAP thereby inhibiting

transcription elongation. RNA-IN/OUT RNAs regulate transcription elongation by altering translation of *tnaC*. Like CRISPRi, the adapted RNA-IN/OUT system could be used to generate a large set of orthogonal regulators because it is based on designable RNA-RNA interactions. To date, more than 150 different families of at least seven orthogonal RNA-IN/OUT mutants have been designed using an RNA-IN/OUT model and all of the mutants tested experimentally have been functional and orthogonal[124].

Adapted RNA-IN/OUT has been used to build two-, three-, and four-input NOR gates (Fig 2-2f)[123]. In these systems, orthogonal RNA-IN variants were connected such that expression of any cognate RNA-OUT represses transcription of the output gene. Additional layers of regulation could be engineered into the adapted RNA-IN/OUT system with ligand-responsive aptamers that regulate RNA-OUT activity[188] or tRNAs that control ribosomal pausing in *tnaC*[189]. A challenge in building larger RNA-IN/OUT circuits is that each transcriptional regulator requires the same *tna* regulatory element (~290bp). The re-use of this part in multiple circuits could lead to homologous recombination (Section III). Engineering TnaC to reduce the length of the repeated sequence[123] or using homologs from other organisms and alternative Rho binding sites could potentially attenuate recombination.

2-3. Selecting Parts to Tune the Circuit Response

Genetic circuits need to be tuned to meet the specifications required for a particular application. For example, a large dynamic range may be required to strongly activate a pathway. Similarly, low off states are desirable when expressing toxic proteins[190]. When the first synthetic circuits were built, there were few options available for tuning circuits and only course-grained changes were possible[89,90]. New libraries of well-characterized parts and computational tools have made it easier to design and tune genetic circuits. Moreover, new classes of insulators improve the reliability of these parts when they are placed in the local genetic context of a circuit. Additional biochemical interactions, such as small RNA (sRNA), have been incorporated into circuits in order to provide additional tuning knobs. In a review, graduate students from the Voigt lab detailed advances in part design and tools to obtain reliable expression levels[191]. Here, we show how the selection or modification of different parts impacts the response of a circuit.

Two circuits are used as model systems to demonstrate the effects of various tuning knobs. The first, a NOT gate, represents a simple logic operation (Fig 2-3a)[89,96]. Logic gates are often characterized by their response function, which captures how the steady-state output changes as a function of input. The shape of this function is defined by: 1. the ON and OFF states, which define the circuit's dynamic range, 2. the amount of input required to reach the half-maximum output (also referred to as the threshold), and 3. cooperativity of the switch[192,193]. An oscillator was selected as an example of a dynamic circuit (Fig 2-4a). These types of circuits can be very difficult to tune because they need to be balanced in a narrow region of parameter space in order to function properly[133,194,195]. For an oscillator, tuning will affect the period, amplitude, and shape of the oscillations. Tuning can also force the system out of the oscillating parameter space and cause the circuit to fail[133].

Two ordinary differential equation (ODE) models were used to show how the selection or modification of different parts impacts the performance of the logic gate and dynamic circuit. The following equations, 2-1 – 2-9, comprise the NOT gate ODE model, with the base parameters in Table 2-1.

$$\frac{dL}{dt} = 0 \tag{2-1}$$

$$\frac{dP_T}{dt} = P_{T+L}\delta_L - P_T L\alpha_L + P_{T+L}\gamma_{T+L} \tag{2-2}$$

$$\frac{dP_{T+L}}{dt} = P_T L\alpha_L - P_{T+L}\delta_L - P_{T+L}\gamma_{T+L} \tag{2-3}$$

$$\frac{dm_R}{dt} = P_T\eta_T - m_R\gamma_{mR} \tag{2-4}$$

$$\frac{dR}{dt} = m_R\tau_R - R\gamma_R - RP_R\alpha_R + P_{R+R}\delta_R \tag{2-5}$$

$$\frac{dP_R}{dt} = P_{R+R}\delta_R - RP_R\alpha_R + P_{R+R}\gamma_{R+R} \tag{2-6}$$

$$\frac{dP_{R+R}}{dt} = RP_R\alpha_R - P_{R+R}\delta_R - P_{R+R}\gamma_{R+R} \tag{2-7}$$

$$\frac{dm_Y}{dt} = P_R\eta_R - m_Y\gamma_{mY} \tag{2-8}$$

$$\frac{dY}{dt} = m_Y\tau_Y - Y\gamma_y \tag{2-9}$$

Here P_T and P_R are the input and output promoters, m_R and m_y are the mRNA species for the repressor and output reporter protein (YFP), L, R and Y are LacI, the repressor protein, and YFP

and P_{T+L} and P_{R+R} are the input promoter bound to LacI and the output promoter bound to the repressor, respectively. The model assumes that the LacI repressor concentration changes only as a function of added inducer. The concentration of LacI was determined using the following model:

$$[LacI] = \frac{(1-IPTG^2)}{0.01(IPTG^2+0.01^2)} \quad (2\text{-}10)$$

Parameter	Parameter value	Species	Species description	Initial value
LacI dissociation (δ_L)	0.1	L	Lac repressor (LacI)	Eq 1
LacI association (α_L)	0.5	P_T	P_{tac} promoter	25
LacI degradation (γ_{T+L})	0.000005	P_{T+L}	P_{tac} bound by LacI	0
m_R transcription (η_T)	0.005	m_R	Repressor mRNA	0
m_R degradation (γ_{mR})	0.004	R	Repressor	0
R translation (τ_R)	0.05	P_R	Repressible promoter	25
R degradation (γ_R)	0.0008	P_{R+R}	P_R bound by repressor	0
R dissociation (δ_R)	0.00036	m_y	YFP mRNA	0
R association (α_R)	0.0001	Y	YFP protein	0
R degradation on promoter (γ_{R+R})	0.000005	P_C	Constitutive promoter	25
m_y transcription (η_y)	0.01	s_R	sRNA	0
m_y degradation (γ_{mY})	0.004	d_{SR}	sRNA mRNA duplex	0
Y translation (τ_y)	0.03	O	Operator	1000
Y degradation (γ_y)	0.00028	O_R	O bound by repressor	0
sRNA transcription (η_S)	0.01			
sRNA/mRNA hybridization (σ_+)	0.001			
sRNA/mRNA dissociation (σ_-)	0.0001			
sRNA degradation (γ_S)	0.004			
RNA duplex degradation (γ_d)	0.04			

Table 2-1: Base parameters for the NOT gate ODE model.
Parameters were altered to simulate tuning knobs as described in the main text.

The response function of a digital logic gate can be shifted up or down by changing promoter strengths (Fig 2-3b)[196], ribosome binding sites (RBS), or the proteins' degradation rates (Fig 2-3c)[197]. Promoter strength can be altered with mutations in the promoter sequence[198] or by selecting new promoters from a characterized library[92,199]. Increased degradation can be achieved with protease tags or N-terminal degrons[197]. Circuit components are often distributed between multiple plasmids at different copy numbers in order to synthesize each component at the necessary level. However, when entire circuits are expressed on one plasmid, copy number can

be shifted to simultaneously alter the circuit's dynamic range and threshold (Fig 2-3d). Different origins of replication can generate complex and poorly understood effects on expression, for example, by changing localization and supercoiling. This can be minimized by using plasmid systems where the copy number can be controlled without changing the origin[200].

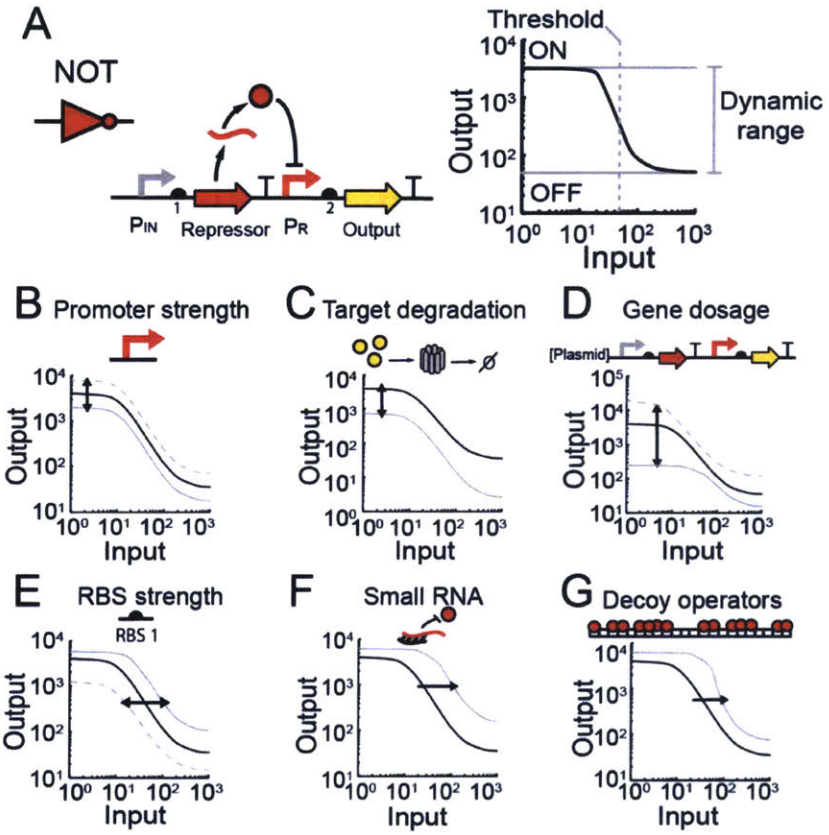

Figure 2-3: Methods of Modifying NOT Gate Behavior.
Every panel displays circuit outputs with original parameter values (black) or tuning knob variations (grey). Inputs in (a-f) are IPTG. (a) Architecture and ideal response functions for the NOT gate. (b) Promoter strength is increased (dashed grey line) or decreased (solid grey line) by a factor of two. (c) Enzymatic degradation of the reporter protein was modeled as a five fold increase in the protein degradation rate[201]. (d) Gene dosage. The NOT gate is moved between a high copy plasmid (dashed grey line) and the genome (solid grey line) to tune expression. The high copy plasmid is assumed to be ten times more abundant than the original circuit. (e) Ribosome binding site strength. Repressor RBSs (RBS1) are increased (dashed grey line) or decreased (solid grey line) by a factor of five. Altering the reporter RBS would shift the output of both circuits vertically (not pictured). (f) Small RNA designed to bind repressor mRNA are modeled with the introduction of a new species that binds repressor mRNA with the same affinity as a ribosome (this value was chosen arbitrarily and can be modulated to change circuit dynamics). In this model, small RNAs are produced constitutively and sRNA/mRNA duplexes are degraded faster than either RNA alone. (g) Decoy operators that bind repressor proteins. Decoy operators were modeled by introducing a new species that binds repressor protein with the same K_d as the repressible promoter. The model has 25 decoy operator sites, however circuits can be tuned with more or less as needed.

The threshold of the gate can be changed via several methods. Selecting a stronger or weaker RBS, adding multiple operators, or changing operator positions within the repressible promoter can change the threshold (Fig 2-3e)[102,114,202,203]. The threshold of a gate becomes steeper and more switch-like when small changes in the input have a large effect on the output[204]. This phenomenon, known as ultrasensitivity, can be important for controlling actuators where intermediate levels of expression are undesirable[204]. It can also make connecting gates easier by decreasing the range of input needed from an upstream circuit to span the induction threshold. One way to make a gate ultrasensitive is to change the cooperativity of repressor binding to the promoter or to introduce DNA looping[205,206]. Another approach is to express a sequestering molecule that binds a circuit component and prevents it from functioning. Sequestration has been achieved using sRNAs that bind to mRNA[177,207] (Fig 2-3f), proteins that bind to transcription factors[156,208,209], and decoy DNA operators that titrate the transcription factor away from the output promoter[210] (Fig 2-3g).

sRNA-based mRNA sequestration was modeled by adding three equations to the original NOT gate model (Eq 2-1 – 2-9) and modifying Eq 2-4. These equations (Eq 2-10 – 2-12, with 2-4M) describe a constitutively expressed sRNA that irreversibly binds mRNA encoding the repressor, leading to increased degradation (Fig 2-4).

$$\frac{dP_C}{dt} = 0 \tag{2-10}$$

$$\frac{ds_R}{dt} = P_C\eta_S - s_R m_R \sigma_+ + d_{Rs}\sigma_- - s_R\gamma_S \tag{2-11}$$

$$\frac{dd_{Rs}}{dt} = s_R m_R \sigma_+ - d_{Rs}\sigma_- - d_{Rs}\gamma_d \tag{2-12}$$

$$\frac{dm_R}{dt} = P_T\eta_T - m_R\gamma_{mR} - s_R m_R \sigma_+ + d_{Rs}\sigma_- \tag{2-4M}$$

Decoy operator-based sequestration was modeled by adding two equations to the original NOT gate model (Eqs 2-1 – 2-9) and modifying Eq 2-5. These equations (2-13 – 2-14 with 2-5M) simulate 1,000 decoy operator sequences that the repressor R binds to with the same affinity as the promoter P_R.

$$\frac{dO}{dt} = O_R\delta_R - OR\alpha_R + O_R\delta_R \tag{2-13}$$

$$\frac{dO_R}{dt} = OR\alpha_R - O_R\delta_R - O_R\gamma_{R+R} \tag{2-14}$$

$$\frac{dR}{dt} = m_R\tau_R - R\gamma_R - RP_R\alpha_R + P_{R+R}\delta_R - OR\alpha_R + O_R\delta_R \quad \text{(2-5M)}$$

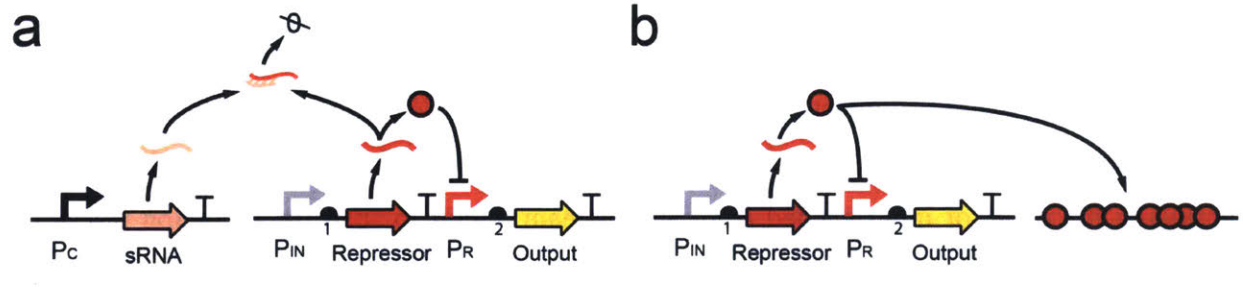

Figure 2-4: Architecture of the NOT gate with sRNA or decoy operator regulation.
Inputs and parameters/initial conditions are the same as in Fig 2-3. (A) NOT gate with sRNA regulation. P_C is a constitutive promoter driving expression of the sRNA. (B) NOT gate with decoy operators added to tune the response function.

The next set of equations, 2-15 – 2-31 describe the oscillator model, with parameters in Table 2-2.

$$\frac{dP_1}{dt} = P_{1R}\delta_R - P_1 d_R\alpha_R + P_{1A}\delta_A - P_1 d_A\alpha_A + P_{1R}\gamma_{PR} + P_{1A}\gamma_{PA} \quad \text{(2-15)}$$

$$\frac{dP_{1R}}{dt} = P_1 d_R\alpha_R - P_{1R}\delta_R - P_{1R}\gamma_{PR} \quad \text{(2-16)}$$

$$\frac{dP_{1A}}{dt} = P_1 d_A\alpha_A - P_{1A}\delta_A - P_{1A}\gamma_{PA} \quad \text{(2-17)}$$

$$\frac{dm_R}{dt} = P_1\eta_R - P_{1A}\eta_{R+} - m_R\gamma_{mR} \quad \text{(2-18)}$$

$$\frac{dR}{dt} = m_R\tau_R - RR\sigma_{R+} + d_R\sigma_{R-} + d_R\gamma_{dR} - R\gamma_R \quad \text{(2-19)}$$

$$\frac{dd_R}{dt} = RR\sigma_{R+} - d_R\sigma_{R-} - d_R\gamma_{dR} - P_1 d_R\alpha_R + P_{1R}\delta_R - P_2 d_R\alpha_R + P_{2R}\delta_R - P_3 d_R\alpha_R + P_{3R}\delta_R \quad \text{(2-20)}$$

$$\frac{dP_2}{dt} = P_{2R}\delta_R - P_2 d_R\alpha_R + P_{2A}\delta_A - P_2 d_A\alpha_A + P_{2R}\gamma_{PR} + P_{2A}\gamma_{PA} \quad \text{(2-21)}$$

$$\frac{dP_{2R}}{dt} = P_2 d_R\alpha_R - P_{2R}\delta_R - P_{2R}\gamma_{PR} \quad \text{(2-22)}$$

$$\frac{dP_{2A}}{dt} = P_2 d_A\alpha_A - P_{2A}\delta_A - P_{2A}\gamma_{PA} \quad \text{(2-23)}$$

$$\frac{dm_A}{dt} = P_2\eta_A - P_{2A}\eta_{A+} - m_A\gamma_{mA} \quad \text{(2-24)}$$

$$\frac{dA}{dt} = m_A\tau_A - AA\sigma_{A+} + d_A\sigma_{A-} + d_A\gamma_{dA} - A\gamma_A \quad \text{(2-25)}$$

$$\frac{dd_A}{dt} = AA\sigma_{A+} - d_A\sigma_{A-} - d_A\gamma_{dA} - P_1 d_A\alpha_A + P_{1A}\delta_A - P_2 d_A\alpha_A + P_{2A}\delta_A - P_3 d_A\alpha_A + P_{3A}\delta_A \quad (2\text{-}26)$$

$$\frac{dP_3}{dt} = P_{3R}\delta_R - P_3 d_R\alpha_R + P_{3A}\delta_A - P_3 d_A\alpha_A + P_{3R}\gamma_{PR} + P_{3A}\gamma_{PA} \quad (2\text{-}27)$$

$$\frac{dP_{3R}}{dt} = P_3 d_R\alpha_R - P_{3R}\delta_R - P_{3R}\gamma_{PR} \quad (2\text{-}28)$$

$$\frac{dP_{3A}}{dt} = P_3 d_A\alpha_A - P_{3A}\delta_A - P_{3A}\gamma_{PA} \quad (2\text{-}29)$$

$$\frac{dm_Y}{dt} = P_3\eta_Y - P_{3A}\eta_{Y+} - m_Y\gamma_{mY} \quad (2\text{-}30)$$

$$\frac{dY}{dt} = m_Y\tau_Y - Y\gamma_Y \quad (2\text{-}31)$$

Here P_1, P_2, and P_3 are identical promoters that can be activated or repressed by dimers of activator (A) or repressor (R) proteins, respectively. The model assumes monomers of the activator and repressor cannot bind the DNA and that a single promoter cannot be bound by both repressors and activators simultaneously. Activator binding to the promoters increases transcription rate 50-fold.

Parameter	Parameter value	Species	Species description	Initial value
R dissociation w/promoter (δ_R)	0.001	R	Repressor	0.804
R association w/promoter (α_R)	5000	d_R	Repressor dimer	0
R degradation on promoter (γ_{PR})	0.0005	m_R	Repressor mRNA	0.31
A dissociation w/promoter (δ_A)	0.01	P_1	Repressor promoter	0
A association w/promoter (α_A)	0.1	P_{1R}	P_1 bound by repressor	25
A degradation on promoter (γ_{PA})	0.0005	P_{1A}	P_1 bound by activator	0
m_R transcription (η_R)	0.1	A	Activator	0.32
m_R transcription + activator (η_{R+})	5	d_A	Activator dimer	0.05
m_R degradation (γ_{mR})	0.01	m_A	Activator mRNA	0.021
R translation (τ_R)	0.15	P_2	Activator promoter	0.0182
R degradation (γ_R)	0.008	P_{2R}	P_2 bound by repressor	25
m_A transcription (η_A)	0.1	P_{2A}	P_2 bound by activator	0
m_A transcription + activator (η_{A+})	5	Y	YFP	126
m_A degradation (γ_{mA})	0.01	m_Y	YFP mRNA	0.2
A translation (τ_A)	0.05	P_3	YFP promoter	0.0182
A degradation (γ_A)	0.008	P_{3R}	P_3 bound by repressor	25
m_Y transcription (η_Y)	0.1	P_{3A}	P_3 bound by repressor	0
m_Y transcription + activator (η_{Y+})	5	P_C	sRNA promoter	50
m_Y degradation (γ_{mY})	0.01		sRNA	0

Y translation (τ_Y)	2.5	sRNA mRNA duplex	0
Y degradation (γ_Y)	0.008	Operator	1000
R dimerization (σ_{R+})	0.1	O bound by repressor	0
R dimer dissociation (σ_{R-})	0.001		
A dimerization (σ_{A+})	0.01		
A dimer dissociation (σ_{A-})	0.01		
sRNA transcription (η_S)	0.001		
sRNA/mRNA hybridization (σ_+)	0.02		
sRNA/mRNA dissociation (σ_-)	0.001		
sRNA degradation (γ_S)	0.01		
RNA duplex degradation (γ_d)	0.05		

Table 2-2: Base parameters for the oscillator ODE model.
Parameters were altered to simulate tuning knobs as described in the main text.

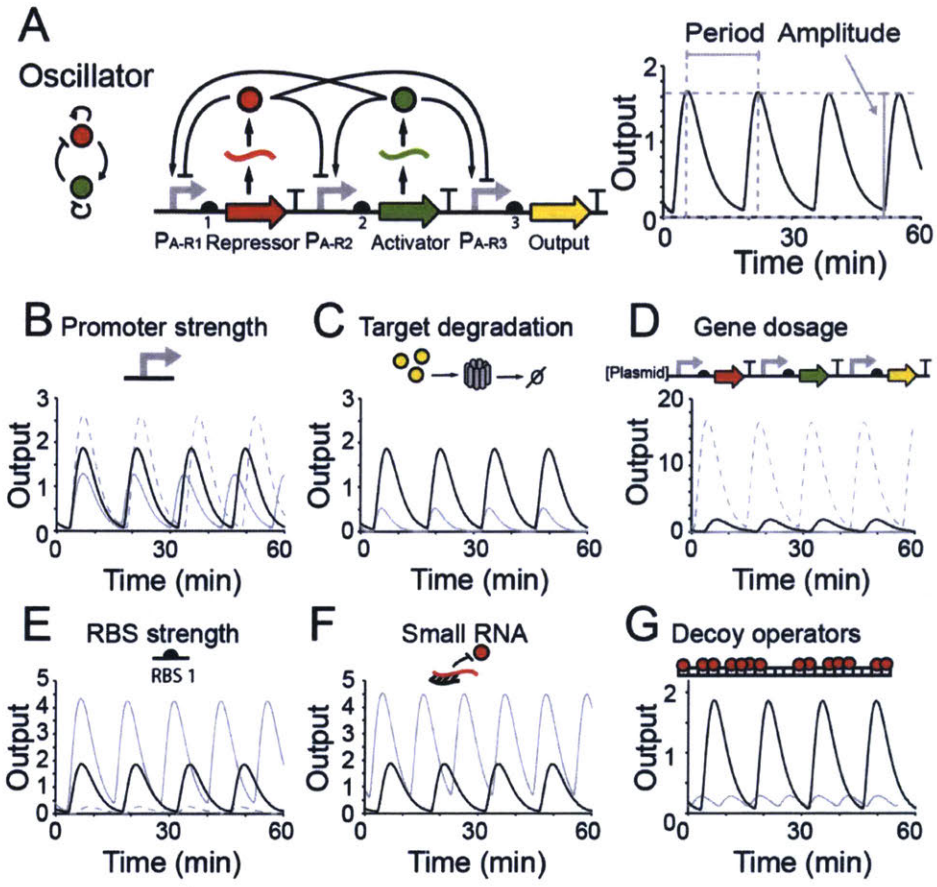

Figure 2-5: Methods of Modifying Oscillator Behavior.
Every panel displays circuit outputs with original parameter values (black) or tuning knob variations (grey). (a) Architecture and ideal response functions for the oscillator. (b) Promoter strength is increased (dashed grey line) or decreased (solid grey line) by a factor of two. (c) Enzymatic degradation of the reporter protein was modeled as a five fold increase in the protein degradation rate[201]. (d) Gene dosage. The circuit is moved between a high copy plasmid (dashed grey line) and the genome (solid grey line) to tune

expression. The high copy plasmid is assumed to be ten times more abundant than the original circuit. (e) Ribosome binding site strength. Repressor RBSs (RBS1) are increased (dashed grey line) or decreased (solid grey line) by a factor of five. Altering the reporter RBS would shift the output of the circuit vertically (not pictured). (f) Small RNA designed to bind repressor mRNA are modeled with the introduction of a new species that binds repressor mRNA with the same affinity as a ribosome (this value was chosen arbitrarily and can be modulated to change circuit dynamics). In this model, small RNAs are produced constitutively and sRNA/mRNA duplexes are degraded faster than either RNA alone. (g) Decoy operators that bind repressor proteins. Decoy operators were modeled by introducing a new species that binds repressor protein with the same K_d as the repressible promoter. The model has 25 decoy operator sites, however circuits can be tuned with more or less as needed.

In an oscillator, parts that impact the rate of gene expression change the amplitude of the response and can shift the period (Fig 2-5be). Rapid protein degradation is critical for dynamic circuits to function correctly. If proteins are slow to degrade, then the circuit may slow down or stop functioning altogether (Fig 2-5c)[211]. Protease tags can be used to decrease the degradation rate from several hours to ~20 minutes, which will increase the rate at which a gate switches[197113,132,155]. Cooperativity is critical for obtaining robust oscillators because it increases the region of phase space that produces oscillations[205]. Therefore, sequestration approaches (*e.g.*, sRNA or dummy operators) are predicted to have a large impact on the period and amplitude of oscillations (Fig 2-5fg)[212].

sRNA-based mRNA sequestration was modeled by adding three equations to the original oscillator model (Eq 2-15 – 2-31) and modifying Eq 2-18. These equations (Eq 2-32 – 2-34, with 2-18M) describe a constitutively expressed sRNA that irreversibly binds mRNA encoding the repressor, leading to increased degradation (Fig 2-6).

$$\frac{dP_C}{dt} = 0 \tag{2-32}$$

$$\frac{ds_R}{dt} = P_C\eta_s - s_R m_R \sigma_{d+} + d_{Rs}\sigma_- - s_{dR}\gamma_S \tag{2-33}$$

$$\frac{dd_{Rs}}{dt} = s_R m_R \sigma_{d+} - d_{Rs}\sigma_{d-} - d_{Rs}\gamma_d \tag{2-34}$$

$$\frac{dm_R}{dt} = P_1\eta_R - P_{1A}\eta_{R+} - m_R\gamma_{mR} - s_R m_R \sigma_+ + d_{Rs}\sigma_- \tag{2-18M}$$

Decoy operator-based sequestration was modeled by adding two equations to the original oscillator model (Eq 2-15 – 2-31) and modifying Eq 2-20. These equations (2-35 – 2-36 with 2-20M) simulate 1,000 decoy operator sequences that the repressor dimer d_R binds to with the same affinity as the promoters P_1, P_2, and P_3.

$$\frac{dO}{dt} = O_R\delta_R - Od_R\alpha_R + O_R\gamma_{PR} \tag{2-35}$$

$$\frac{dO_R}{dt} = Od_R\alpha_R - O_R\delta_R - O_R\gamma_{PR} \quad (2\text{-}36)$$

$$\frac{dd_R}{dt} = RR\sigma_{R+} - d_R\sigma_{R-} - d_R\gamma_{dR} - P_1d_R\alpha_R + P_{1R}\delta_R - P_2d_R\alpha_R + P_{2R}\delta_R - P_3d_R\alpha_R + P_{3R}\delta_R - OR\alpha_R + O_R\delta_R \quad (2\text{-}20M)$$

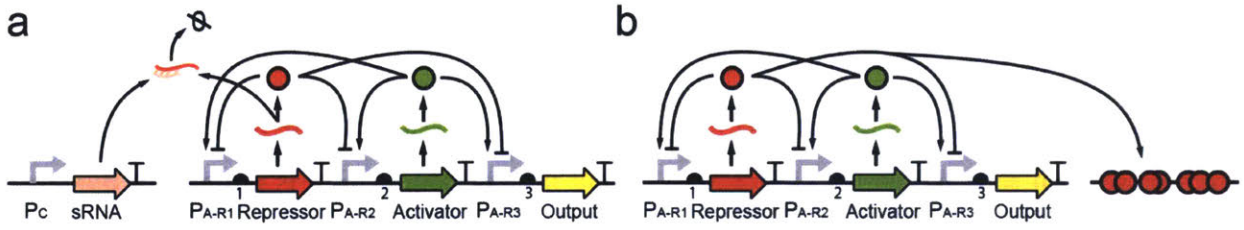

Figure 2-6: Architecture of the oscillator with sRNA or decoy operator regulation.
Inputs and parameters/initial conditions are the same as in Fig 2-5. (A) Oscillator with sRNA regulation. P_C is a constitutive promoter driving expression of the sRNA. (B) Oscillator with decoy operators added to tune the response function.

2-4. COMMON FAILURE MODES FROM CONNECTING CIRCUITS

Gates can be combined to build larger circuits that implement more sophisticated computational operations. To connect transcriptional gate, the output promoter of one circuit is used as the input promoter to the next. This method applies for all transcriptional circuits, including digital, analog and dynamic circuits or a combination of types. To connect circuits, they have to broken up into their component parts and then combined in a particular order (Fig 2-7a). Reorganizing the parts places them in new local contexts that are different from those where they were characterized. This can be problematic because circuit components can behave differently in new genetic contexts and small circuits may have identical component parts (*e.g.*, terminators) that interfere with each other in the larger circuit. In this section, we discuss failure modes that can arise when building larger circuits, show the impact that each failure has on circuit function, and discuss engineering approaches to mitigate these problems.

When connecting circuits, a common problem is that the upstream circuit's output does not span the dynamic range required to stimulate next circuit in series (Fig 2-7b). In digital logic, this 'mismatch' manifests as either a decrease in the dynamic range of the complete circuit or a loss of function. Connectivity mismatches can be corrected by selecting parts that shift the thresholds of individual gates. For example, RBSs can be mutated to force the threshold of a gate

to fall within the dynamic range produced from an upstream circuit[89,145]. 'Mismatches' in an oscillator can dampen oscillations or force the system outside the functional parameter space (Fig 2-7b). Mathematical models could be used to streamline circuit design by predicting the functional parameter space and selecting appropriate RBSs and promoters to achieve the required expression levels[91,111,203].

Genetic parts are often context dependent, meaning their functions change when the DNA sequences on either side of the part are altered[213,214]. Context dependencies complicate part substitutions because part characterizations are often carried out in isolation and their activity in a new context may not match the measured strength. For example, promoters are that are defined as DNA sequences <50 bp may behave differently in new contexts because the a-domain of *E. coli* RNAP can contact the DNA ~100 bp upstream of the transcription start site[198]. In a digital circuit, reducing promoter efficiency attenuates the response of individual gates and reduces the output of the complete circuit (Fig 2-7c). Promoter attenuation can increase the amplitude of an oscillator and elongate the period, by reducing repressor expression. Insulator sequences can relieve some compositional context effects by standardizing the DNA sequences flanking promoters[214,215].

Context effects can also occur when promoters are fused to different RBSs. Promoters are sensitive to the DNA sequences near the transcription start site because that region can alter promoter melting and polymerase escape frequency[199]. Transcription start sites can also fluctuate based on the local sequence context[216,217], which can impact RBS strength by altering the length of the 5'-UTR and changing mRNA secondary structure. Tandem promoters can generate especially long 5'-UTRs that exacerbate this effect by base pairing with the RBS or sequences in the open reading frame[218–220]. Circuits can fail completely when mutations in the 5'-UTRs cause hairpins completely occlude the RBSs and prohibit translation (Fig 2-7d). To solve these problems, the 5'-UTR can be cleaved with ribozymes or CRISPR processing to standardize RBS accessibility[215,221]. Catalytic insulator elements serve dual functions by standardizing both the 5' end of mRNA and the promoter region downstream of the transcription start site. RBSs can be further insulated from the local context using bicistronic designs, which prime the mRNA for translation with an upstream RBS that keeps the mRNA unfolded[92].

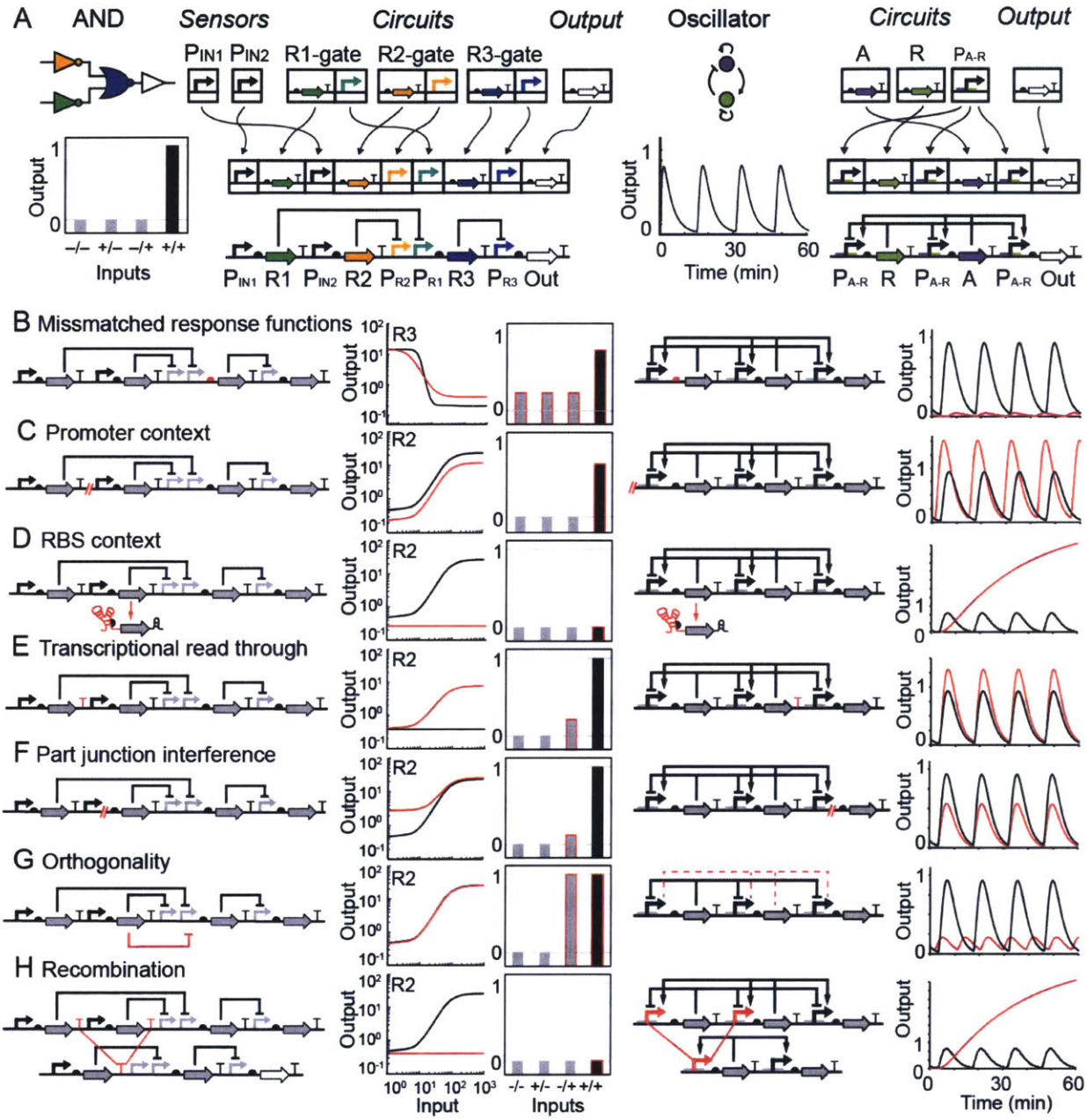

Figure 2-7: Common failure modes and their impact on circuit dynamics.
(a) An AND gate[114] and oscillator[113] are used as model systems to demonstrate the assembly of parts to build more complex circuits. Repression is indicated with a blunt ended connector and activation is indicated with an arrow. For the AND gate, the input promoters are P_{IN1} and P_{IN2} and the output promoter is P_{R3}. Promoters are named by the repressor to which it responds (*e.g.*, P_{R1} is repressed by R1). The steady-state response to different combinations of inputs is shown as a bar graph, where the OFF states are grey and the ON state is black. For the oscillator, the promoters P_{A-R} are repressed by R and activated by A. The impact of various failures (red lines) are shown for the AND gate (left) and oscillator (right) with expected dynamics shown in black. Models were used to simulate the R2 NOT gate, the AND gate[114], and oscillator. Oscillator model equations are identical to those used to create 2-5. Unless indicated otherwise, Input 2 is the input to the NOT gate transfer function. (b) Mismatched response functions. In the AND gate, R3 was modeled as a different repressor: Bet1 (k_d = 0.2, n = 2.4, max = 13, min = 0.4) instead of Orf2 (k_d = 0.4, n =

6.1, max = 16, min = 0.2)[114]. R2 is the input for the R3 transfer function. In the oscillator, the R translation rate is increased ten-fold. (c) Promoter context. Strength of the indicated promoters is reduced by 50% in both circuits. (d) RBS context. The translation rates of R2 (AND gate) and R (oscillator) are set to zero. Input 1 is the input for the R2 transfer function (e) Transcriptional read-through. 30% read-through from upstream operons through the red terminator is simulated in both circuits. (f) Part-junction interference. A new constitutive promoter (AND gate) is simulated as approximately 20% of the strength of P_{IN2}. New terminator (oscillator) decreases transcription 40%. (g) Orthogonality. $R3_{max}$ is set as $R2_{min}$ to simulate repression of P_{R3} by R2. Additional equations are added to the oscillator model to simulate repressor-activator complex formation. (h) Recombination. R2 and R were removed from the AND gate and oscillator models, respectively.

Transcriptional read-through can be a problem in genetic circuits with monocistronic designs, where every gene has its own promoter and terminator. These designs require strong terminators to insulate against read-through from neighboring promoters. Failure to fully insulate each cistron can link the expression of genes that are supposed to be regulated independently (Fig 2-7e) and can contribute to the leaky expression of uninduced genes. Strong, tandem terminators can be placed on either side of each gene to ensure isolated expression of individual operons[222]. Large libraries of rho-independent terminators were recently built and characterized to enable the construction of large circuits that are robust to read-through and homologous recombination (described below)[93,222].

DNA sequences are information rich, therefore connecting two parts can create a new functional sequence at the junction[223]. New regulatory elements, such as promoters or terminators, can be generated at a part junction if the combination creates a sequence of DNA that resembles a regulatory element. For large circuits, many parts have to be combined in a new order and unexpected parts that interfere with gene expression can be generated (Fig 2-7f). One way to scan for unintended functional sequences is to use computer algorithms that search for various regulatory elements[91,222,224–230].

Crosstalk, which occurs when regulators interact with each other's targets, can change the topology of a circuit and can lead to errors in the desired operation[98]. For example, crosstalk between a repressor and non-cognate promoter can inappropriately decrease expression of a gene and cause a circuit to fail (Fig 2-7g). Avoiding crosstalk requires that parts be screened for orthogonality via combinatorial experiments that test every combination of promoter and regulatory element[114,124,126,141,143,231].

Many of the circuits built to date re-use the same regulatory parts, which can lead to homologous recombination. Homologous recombination deletes DNA between repeated

sequences and can result in the loss of circuit components and circuit failure (Fig 2-7h)[222]. In general, the rate of recombination increases with circuit toxicity[232] and homolgous DNA length, with the threshold occurring between 20-30bp[233]. Homologous recombination can be avoided with large libraries of parts with redundant functions that have enough sequence diversity to avoid recombination[222,234].

2-5. INTERACTIONS BETWEEN SYNTHETIC CIRCUITS AND THE HOST ORGANISM

Genetic circuits are based on biochemical interactions within living cells. Most circuits use host resources to function, including transcription/translation machinery (*e.g.*, ribosomes and RNAP), DNA replication equipment, and metabolites (*e.g.*, amino acids). The availability of these resources and the details of the intracellular environment change significantly in different strain backgrounds, environmental conditions, media, growth rate, and cell density. When the first synthetic circuits were built, they were fragile and it was unclear why they would only work in specific conditions[20,21]. Now, there is a more precise understanding of the ways in which circuits break due to interactions with the host[104]. A better understanding these failure modes are and the methods natural systems use to overcome them will lead to new design rules for composing synthetic circuits.

A common observation is that some synthetic regulators can cause growth defects. Yet it remains unclear why certain regulators can be expressed at high levels with no noticeable impact whereas others in the same class are very toxic. This was evident in analyzing large libraries of TetR and s factor homologues sourced from diverse organisms and transferred into *E. coli* [114,143]. Expression of some of regulators slowed *E. coli* growth, but the origin of this effect is unclear as it does not correlate with the number of predicted binding sites in the genome or off-target gene expression measured using RNA-seq. T7 RNAP is another part that can be very toxic when combined with a strong T7 promoter[145]. It is also unclear how this toxicity arises, but it could be due to the difficulty terminating T7 RNAP, which could cause circular transcription on a plasmid or expose mRNA by decoupling RNAP and ribosome progression. Circuits based on protein-protein interactions can also exhibit toxicity when the proteins bind to off-target partners. We observed this with anti-s factors, which appear to bind and titrate native s factors[143]. Small RNA with RBS-like sequences can also cause toxicity by titrate ribosomes, increasing expression

variability, and reducing growth (Fig 2-8a)[190]. Larger circuits are particularly sensitive to the toxicity that can arise from individual regulators because their effects are compounded when they are expressed together[235].

Circuits can also decrease growth rate by monopolizing host resources and slowing essential protein/RNA production (Fig 2-8a)[236]. A small reduction in the growth rate can be a problem when using a circuit for industrial applications that rely on high product yields. A decrease in growth rate can reduce the dilution rate of circuit components and lead to unintended build up of proteins or RNA that can cause a circuit to fail. In fact, circuits can appear to function better when growth is impeded because slow dilution increases the observed concentration of transcription factors and reporters. Slow growth can also put pressure on the host organism to evolve away the burdensome circuit, either via homologous recombination, point mutations/deletions, or copy number reduction.

Circuits can diverge from their expected behavior when they use a limited resource that is shared with other cellular processes. Overburdening resources causes queuing, which results in a delay or reduction in circuit activity[237]. For example, when s factors are overexpressed, they can occupy the entire pool of free core RNAP. When this happens, sigma factors must compete to bind to the core, which indirectly couples their activity and can disrupt host processes[238]. Native s factors are able to avoid queuing by pulsing their expression such that they alternate the usage of core RNAP over time[239]. A similar coupling effect has been observed when the ClpXP protease is shared by regulators that have been modified to contain C-terminal tags for fast degradation. If too many proteins are targeted for degradation, the enzymatic machinery can become overwhelmed and force substrates to wait for processing[211]. The rapid degradation of regulators is important for dynamic circuits, such as oscillators, which will fail if the regulatory proteins accumulate (Fig 2-8b).

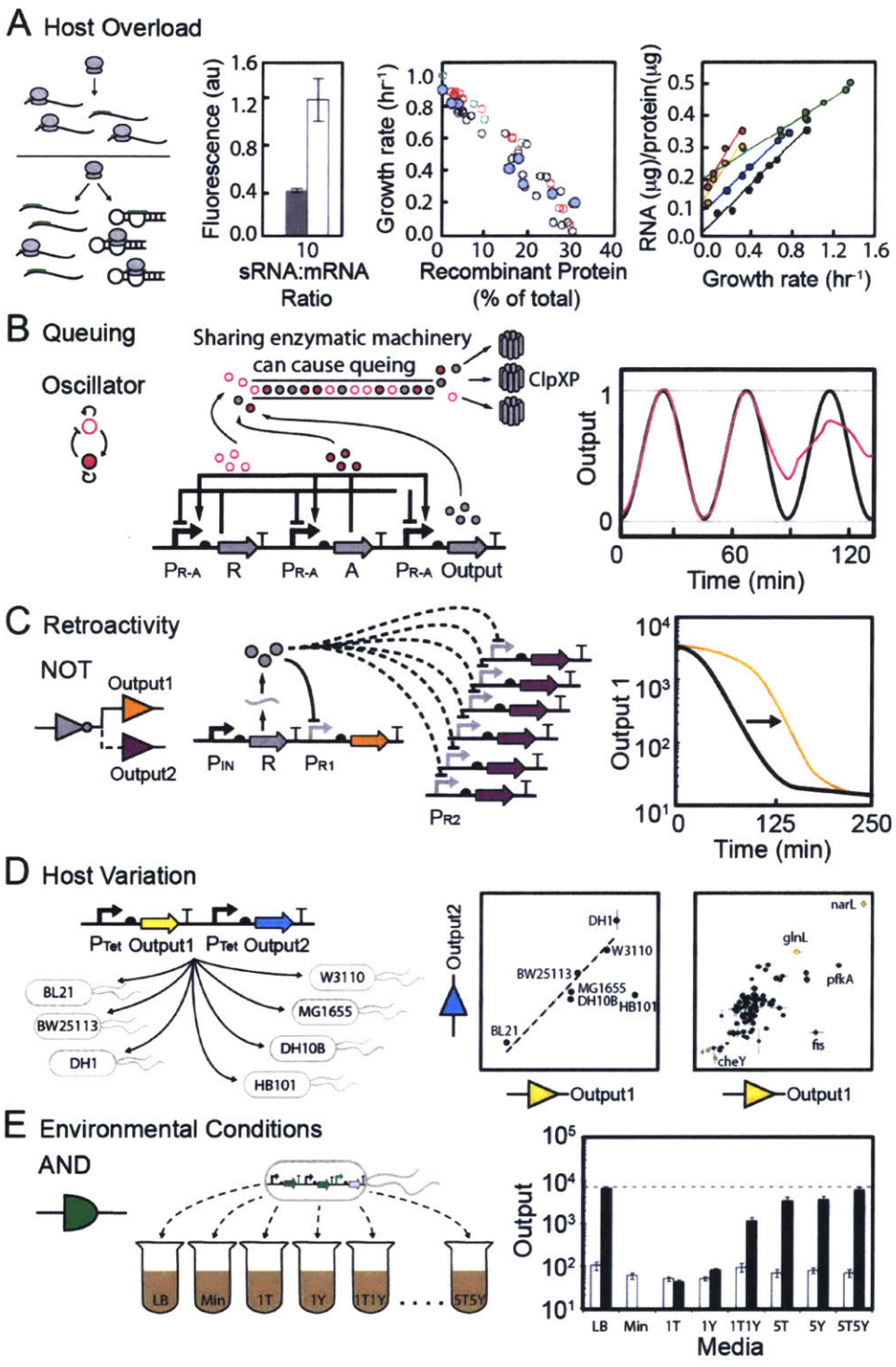

Figure 2-8: Circuit performance within the context of a living cell.
(a) Recombinant protein expression can cause a growth defect by reducing the availability of host resources (*e.g.*, RNAP and ribosomes). Here, synthetic sRNAs compete with mRNA for ribosomes to illustrate the impact of exogenous protein expression on host resource allocation. When sRNAs are produced (left graph, grey bars), ribosomes are titrated away from fluorescent protein mRNA and observed fluorescence is

reduced relative to no sRNA (left graph, white bars)[151]. Center graph, colored circles represent the overexpression of different proteins in *E. coli* (blue: *Pu* promoter b-Galactosidase, red: T7 promoter b-Galactosidase, black: *tac* promoter DEF-Tu, green: *bla* promoter b-Lactamase)[240]. Right graph, colored circles represent growth of different bacterial strains as a function of rRNA supply (blue: E. coli 30^{0}C, green: A. aerogenes 37^{0}C, red: C. utilis 25^{0}C, orange: C. utilis 30^{0}C, black: N. crassa 30^{0}C)[240]. (b) Queuing as a result of overloading the ClpXP protease machinery with proteins from a synthetic oscillator. The graph shows the difference between expected (black) and measured (red) dynamics for an oscillator affected by queuing[211]. (c) An additional output (P_{R2}) on a high copy plasmid is added to the NOT gate. This causes retroactivity, which alters the activation dynamics of the original output (P_{R1}) (black line: original dynamic response, orange line: retroactive effect)[241]. (d) One plasmid with two reporter proteins is transformed into different *E. coli* strains. The ratio of expression varies in some strains (left graph: wild type E. coli strains, right graph: KEIO collection knockouts)[242]. (e) Different media impact the performance of an AND gate based on T7 RNAP[40,145]. Data are shown for the circuit in the absence (white) and presence (black) of both inputs in different medias (LB: luria broth, Min: minimal media, #T and/or #L: minimal media supplemented with tryptone (#T = #g/L) or yeast extract (#Y = #g/L).

Retroactivity can also interfere with circuit activity. Retroactivity is defined as the influence that a downstream genetic element can have on an upstream one and it describes the changes in circuit behavior that result from connecting new downstream modules to a circuit[180]. Downstream modules may affect the performance of upstream circuits by titrating regulators away from the original circuit. For example, connecting a second output to a NOT gate may cause retroactivity by titrating the repressor away from the original output promoter (Fig 2-8c). Retroactivity will impact the NOT gate's dynamics by increasing the time it takes to build up an adequate amount of protein to repress promoter activity[241]. Retroactivity that delays a circuit's response to input stimulation can be alleviated by increasing expression of the problematic circuit component; however, increasing expression can lead to other trade-offs, including toxicity.

Strain variation can affect circuit performance in different ways. Differences in growth rate, ribosome concentration, and induction lag time have been identified as the main contributors to strain dependent variations in circuit performance[242]. In recent studies, these phenotypes have been correlated with specific genes by studying growth and circuit performance across single gene knockouts (Fig 2-8d)[242243]. Media and growth conditions can also impact circuit performance by altering promoter activity, protein stability, and regulator dilution[244,245]. These effects can be so severe that switching from LB to minimal media can cause circuits to fail (Fig 2-8e)[40] .

One approach to reduce strain- and media-based variation is to use reference standards to report circuit performance. To this end, the Relative Expression Unit (REU) was introduced as a

standard for reporting promoter measurements[40,246]. REUs report the promoter activity by normalizing measurements to a constitutive promoter standard in a strain that is treated identically and measured simultaneously. REU measurements have yielded reliable, reproducible data when compared across labs, strains, and media, which is important for transcriptional circuits that use promoters as inputs and outputs. In the future, this will facilitate the computer-aided design of large circuits.

2-6. CONCLUSIONS

The first circuits were built by repurposing a small number of regulators and genetic parts from other areas of genetic engineering. After early success[90,131], these parts were put together in different combinations to explore the range of circuit functions that could be performed in the cell. We are now in a phase where there are >100 new regulators[98,114,121–123,125,126,144,191,225] that are orthogonal and could theoretically be used to build synthetic regulatory networks at the scale of natural networks in bacteria[247].

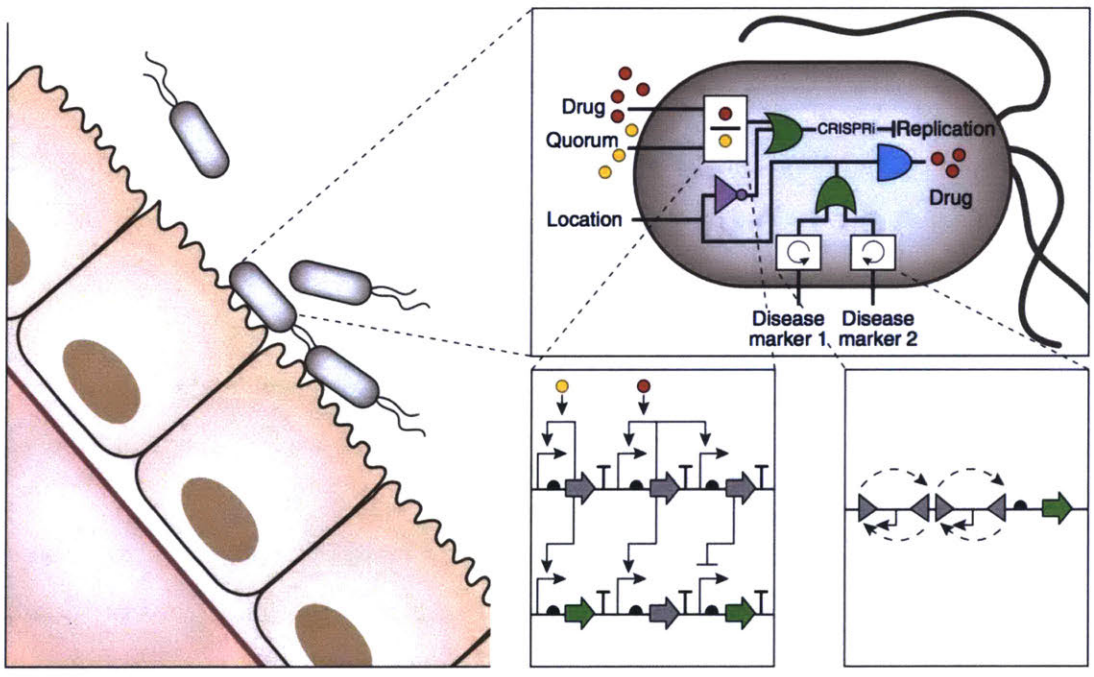

Figure 2-9: Conceptual circuit for a therapeutic bacterium.
This therapeutic bacterium colonizes a niche in the human microbiome and delivers a drug. This circuit demonstrates how the different classes of regulators and circuits described in this review could be combined into a single system. The leftmost panel shows genetically modified bacteria that have colonized the interior of a human gastrointestinal tract. The upper right panel focuses on the conceptual circuit that

the bacteria use to regulate their growth and deliver drugs to the human patient. An analog circuit[146] (left) and irreversible recombinases (right) are highlighted in the insets to emphasize the diverse biochemistries used to build this circuit.

There are several key advances that have to happen before we can build and debug genetic circuits this large. First, computational tools have to be developed to aid the design process. These programs need to be able to simulate the dynamics of a circuit and convert the designs into a linear assembly of genetic parts[111,248–251]. Insulating DNA sequences will be critical in future circuits because the majority of parts will be in new contexts[252,253]. Second, new approaches to whole cell omics measurements have to be integrated into the debugging cycle. Currently, there is an over-reliance on fluorescent proteins as the output of circuits. However, transcriptomics is now sufficiently inexpensive such that it could be used to infer polymerase flux on many of the parts internal to a circuit[254]. Other single molecule approaches, such as ribosome and RNAP mapping, will become powerful when the experiments become more routine[255,256]. Third, new approaches need to be developed that can rapidly test circuits under conditions that are difficult to control in the cell. Circuits are sensitive to parameters like the number of ribosomes, RNAP, redox, temperature, and ATP all of which change in different cell types and conditions. However, these are difficult to measure in the cell without broadly impacting the host. To this end, the development of *in vitro* cell-free methods to debug circuits will be valuable for designing circuits that are robust to these changes[257,257–266].

New biochemistries, tuning knobs, and troubleshooting methods are now converging for the sophisticated design and construction of genetic circuits. In these circuits, different classes of regulators can be used in a single circuit to fulfill specialized functions. In this vision, each regulator has found a niche within the larger circuit that exploits their strengths (Fig 2-9). For example, digital circuits can be used to integrate sensors and respond to a particular set of conditions, whereas analog circuitry can perform arithmetic function functions with a small number of regulators[146]. Integrases can store memory or cause an irreversible commitment. CRISPRi can regulate essentially any gene in the genome. A vision of this marriage is shown in Figure 6, which is an example of a commensal bacterium that has been engineered to produce a pharmaceutical while colonizing the gut. In it, repressor-based logic gates respond dynamically to environmental states and invertases record these observations. Analog circuits can be used to

calculate a dosage rate and, if surpassed, CRISPRi knocks down specific host genes to arrest growth and avoid overmedication. Collectively, these new circuits and the tools and knowledge to connect and debug them will enable a new era of cellular programming and the applications that come with this capability.

Chapter 3: Antisense transcription as a tool to tune gene expression

A surprise that has emerged from transcriptomics is the prevalence of antisense transcription, which occurs counter to gene orientation in genomes. While frequent, the roles of antisense transcription in regulation are poorly understood. We built a synthetic system in *Escherichia coli* to study how antisense transcription can change gene expression and tune the response characteristics of a regulatory circuit. We developed a new genetic part that consists of a unidirectional terminator followed by a constitutive antisense promoter and demonstrate that antisense transcription represses gene expression and the magnitude of repression is proportional to the antisense promoter strength. Chip-based oligo synthesis was applied to build a large library of 5,668 terminator-promoter combinations. This library was used to control the expression of three repressors (PhlF, SrpR, and TarA) in a simple genetic circuit (NOT gate). Using the library, we demonstrate that antisense promoters can be used to tune the threshold of a regulatory circuit without impacting other properties of its response function. We also determined the relative contributions of antisense RNA and transcriptional interference to repressing gene expression. Finally, we constructed a biophysical model of transcriptional interference to capture the impact of RNA polymerase collisions on gene repression. This work quantifies the role of antisense transcription in regulatory networks and introduces a new method for controlling gene expression that has been previously overlooked in genetic engineering.

3-1. Introduction

Genetic engineers typically follow a simple scheme for controlling gene expression: a promoter and terminator flank the gene and all parts are orientated in the same direction. Larger designs consisting of multiple genes and promoters often are organized similarly, where transcription is designed to proceed in one direction. This organization avoids potential interference between promoters that can arise due to RNA polymerase (RNAP) collisions[267], supercoiling[268,269], non-coding RNAs[270], and conflicts with the replication machinery[271–273]. Promoters that are oriented in the opposite direction of genes produce antisense transcription and although this is generally regarded as a nuisance[274,275], it can be useful for reducing leaky

expression of toxic proteins[276–279] and generating genetic switches[280–282]. Here, we propose that promoters oriented opposite to a gene can be used to reliably tune gene expression and control the input threshold of genetic switches.

Increased use of transcriptomics has demonstrated that antisense transcription is surprisingly common across all organisms, including archaea[283], prokaryotes[284–288], and eukaryotes[289–291]. For example, in *E. coli* ~30% of all transcription start sites were found to be antisense and internal to, or just after, genes[292–294]. Similarly in *H. pylori*, about half the genes have at least one antisense promoter[274]. Although some of this antisense transcription is the result of inefficient termination by intrinsic and rho-dependent terminators[295], it is often driven by promoters with well-defined regulatory motifs, such as housekeeping sigma factor binding sites[284,294,296]. Depending on the organism, antisense transcription can be constitutive or regulated under different environmental conditions[297,298].

While prevalent, the role of most antisense transcription in regulation is unclear[299,300]. Some have postulated that the majority of antisense transcription is non-functional and is background due to pervasive transcription[296]. However, antisense transcription is known to be an important component of the genetic switches that control bacterial competence[280] and virulence[301], as well as *Saccharomyces cereviseae*'s entry into meiosis[282]. Antisense transcription also occurs frequently for genes that require tight expression control under defined conditions, such as toxic or virulence proteins[279,302–304]. One role of antisense transcription may be to impact the threshold of a switch [305], defined as the amount of input signal required to reach half maximal activity. As outlined in Chapter 2, the input thresholds of synthetic systems can be tuned by mutating ribosome binding sites, adding small regulatory RNAs, or sequestering the proteins using dummy operators or protein-protein interactions[306–309]. There is evidence that antisense promoters can similarly change regulatory circuits by controlling expression of repressors, activators, s factors and anti-s factors [310,311].

There are two classes of mechanisms for how antisense promoters regulate gene expression. The first involves the antisense RNA (asRNA) that is generated, which can regulate gene expression by binding to the mRNA to change its stability or translation, or act as a transcriptional regulator[312,313]. The second is transcriptional interference, where the sense and antisense promoters interact directly or via the RNAPs to cause the down regulation of a gene[305].

There are four mechanisms by which transcriptional interference can occur[267]: 1. competition (promoters overlap and only one RNAP can bind at a time), 2. sitting duck (an RNAP that is slow to elongate is dislodged), 3. occlusion (one RNAP elongates over a promoter transiently blocking the other), and 4. collision (two actively transcribing RNAPs collide)[314–316]. Of these, modeling suggests that when promoters are >200 bp apart and oriented convergently, the dominant mechanism of interference is collision[316]. Regulation by asRNA and transcriptional interference are not mutually exclusive. Examples have been described where regulation occurs due to only one mechanism[305,310,317] or they work in concert[280,303].

In this work, we harness antisense transcription as a reliable "tuning knob" for the construction of genetic circuits. We introduce a new composite part to the 3'-end of the gene of interest that consists of a unidirectional terminator followed by a reverse constitutive promoter. We demonstrate that the antisense promoter represses gene expression in accordance with its strength and that the antisense transcription can cause a change in the threshold of inducible systems. A large library of promoter-terminator combinations was constructed via chip-based oligo synthesis[318] and screened using flow-seq to identify terminators and promoters that can be used to construct reliable antisense regulation[319,320]. This approach has been used previously to elucidate how translation rates affect mRNA stability[320] and codon bias influences RNA structure and translation[321]. Finally, we determined the relative contributions of antisense RNA and transcriptional interference to repressing gene expression and introduce a biophysical model to parameterize RNA polymerase collisions. This work contributes to a larger effort to expand on the classic concept of an "expression cassette" to include additional parts that utilize genetic context to fine-tune expression levels[26,322,323].

3-2. RESULTS

3-2-1. Repression correlates with the strength of the antisense promoter

A simple system was designed to quantify the impact of an antisense promoter on gene expression (Fig 3-1). The isopropyl b-D-1-thiogalactopyranoside (IPTG) inducible promoter P_{tac} was used as the forward promoter to drive the expression of red fluorescent protein (RFP). Downstream of *rfp*, there is a constitutive antisense promoter P_R whose RNAPs will be fired

toward P_{tac}. Four constitutive promoters of different strength[26,320] were selected to serve as the antisense promoters and the impact on RFP expression was quantified. The strengths of P_R and P_{tac} were determined using a separate plasmid system and normalized by a reference standard to estimate promoter strengths as polymerase firing rates (Fig 3-2ab) (Methods). The full cassette was placed on a plasmid containing the p15A origin.

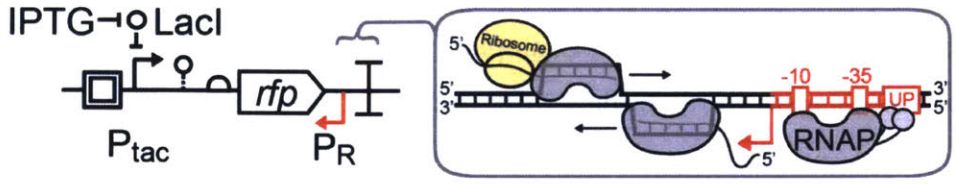

Figure 3-1: A schematic showing the antisense transcription reporter system.
A constitutive promoter (red) at the 3'-end of *rfp* represses gene expression by firing polymerases at the forward promoter P_{tac} (black).

The addition of an antisense promoter changes the response curve for the IPTG induction of RFP in three ways (Fig 3-2c). First, expression is reduced over the entire range of inducer concentration. We defined a parameter θ that captures the magnitude of repression,

$$\theta = \frac{[RFP]_0}{[RFP]_+} \qquad (3\text{-}1)$$

where the subscripts +/0 represent the presence/absence of an antisense promoter. Plotting θ versus inducer concentration shows that the impact of the antisense promoter is stronger when P_{tac} is less active (Fig 3-2d). This biased repression is consistent with previous findings that weak promoters are more susceptible to repression via transcriptional interference and asRNAs than strong promoters[267,314,316]. Second, the maximum repression increases with the strength of the antisense promoter (Fig 3-2e, exponential regression; R^2 = 0.99737). Notably, the strongest promoter tested (apFAB96) is unable to completely repress expression. This promoter is among the strongest from a large synthetic library[320] and of comparable strength to the *E. coli* rrn promoters[324]. Finally, the threshold for induction increases as a function of the strength of antisense promoter (Fig 3-2e, linear regression; R^2 = 0.9876).

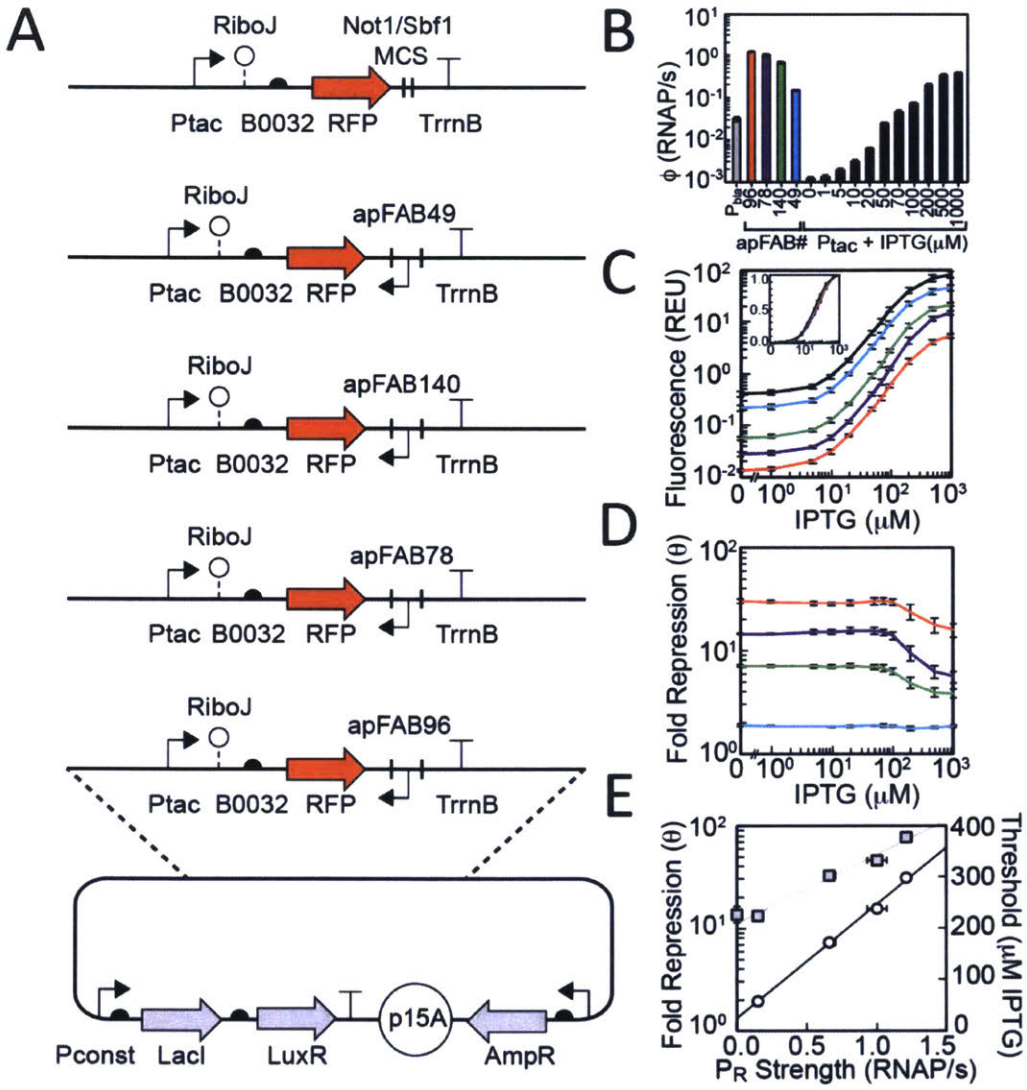

Figure 3-2: Impact of antisense transcription on gene expression
(A) Antisense transcription reporter plasmids used to quantify repression. Not1/Sbf1 multiple cloning sites were used to digest the plasmid and insert promoters (apFAB49, apFAB140, apFAB78, apFAB96) at the 3'end of rfp (B) Strengths of the forward and antisense promoters. The reference promoter (Pbla) used to calculate promoter strength in units of RNA polymerase firings per second is shown in grey. Strengths of the constitutive promoters used as PR (colors) and forward promoter (Ptac) at different inducer concentrations (black). (C) Response functions for P_{tac} with different antisense promoters located at the 3'end of RFP: no promoter, black; antisense promoters of different strength, colors as in (B). The inset is the $\log_{10}$ transform of the same data normalized by min and max. (D) The fold repression (Equation 3-1) is shown as a function of the induction of the forward promoter. The colors correspond to antisense promoters of different strength. (E) Maximum fold repression and threshold as a function of antisense promoter strength. The induction threshold K was calculated by fitting Equation 2 to the data in part C. The lines are linear and exponential fits to the threshold ($R^2 = 0.9876$) and repression ($R^2 = 0.99737$) data, respectively. In all panels, data represents the mean of three experiments performed on different days and the error bars are the standard deviation of these replicates.

Interestingly, the shape of the induction curve remains similar for the different antisense promoters and the overall response to IPTG follows approximately the same dynamic range (Fig 3-2c, inset). The cooperativity (n = 1.7 ± 0.12) is also unaffected by different antisense promoters. Thresholds and Hill coefficients were calculated by fitting the response functions in Fig 3-4a to the Hill Equation:

$$y = y_{min} + (y_{max} - y_{min})\frac{x^n}{K^n + x^n} \quad (3\text{-}2)$$

where x is the concentration of IPTG, y is the activity of P_{OUT}, n is the Hill coefficient, and K is the threshold level of input where the output is half-maximal. *E. coli* growth rates were unaffected by the addition of antisense promoters (data not shown). Fluorescence histograms of bacteria harboring the antisense reporter plasmids are also very narrow (Fig 3-3). This indicates very little variation between cells harboring the same genetic construct, therefore antisense transcription is affecting gene expression in all cells nearly equally. Thus, altering gene expression with constitutive antisense promoters does not cause a growth defect or alter RFP expression profiles.

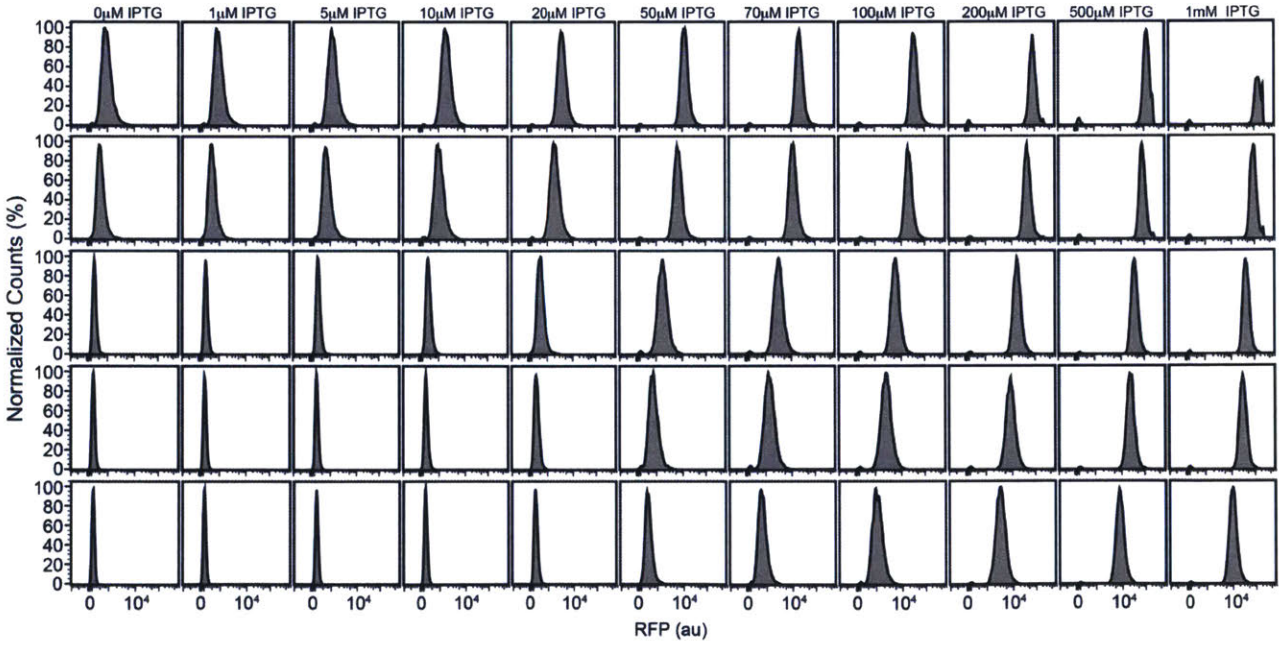

Figure 3-3: Fluorescence histograms corresponding to the data in Fig 3-2.
Antisense promoter: none (JBTI241 – top row), apFAB49 (JBTI384 – second row), apFAB140 (JBTI386 – third row), apFAB78 (JBTI244 – fourth row), apFAB96 (JBTI387 – bottom row).

3-2-2. Multiplexed characterization of antisense promoters

Experiments were designed to quantify the impact of an antisense promoter on the function of a simple genetic circuit. We chose to characterize NOT gates, where an input promoter drives the expression of a repressor that turns off an output promoter[325]. This creates a response function that is inverted, as compared to an inducible system alone. Our design adds antisense promoters to the 3'end of the repressor (Fig 3-4), which reflects natural motifs where regulatory proteins are controlled by antisense promoters[326]. The design also adds unidirectional terminators between the 3'end of the repressor and the antisense promoter to demonstrate that antisense promoters can alter gene expression when added to outside of complete expression cassettes.

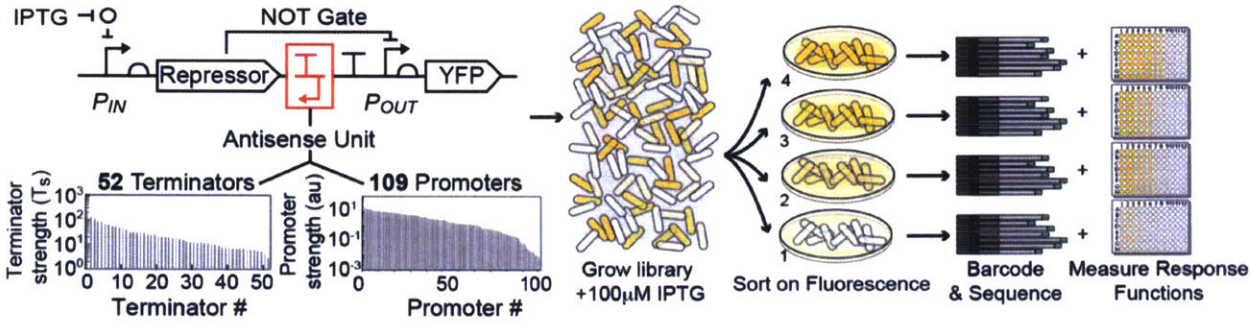

Figure 3-4: Construction and characterization workflow of a library of terminator/antisense promoter pairs.
Library construction and flow-seq screening used to measure the impact of terminator/antisense promoter pairs on regulatory circuit performance. All combinations of 52 unidirectional terminators and 109 promoters were constructed to create a library of 5,668 transcriptional interference constructs. The terminator and promoter strengths shown were measured previously [26,27,320]. The library was synthesized as oligonucleotides then cloned into genetic NOT gates at the 3' end of the repressor gene (red box). Each library was transformed into *E. coli*, grown with 100mM IPTG, and sorted into bins of varying YFP fluorescence to find constructs with increased induction thresholds. Bacteria from each bin were plated on solid media and individual colonies were selected to measure the full response function of sorted variants. Plates were scraped to isolate plasmids from bacteria in each bin and plasmid DNA was barcoded and deep sequenced (Methods).

Advances in chip-based DNA synthesis have made it possible to simultaneously synthesize 10,000s of unique ~200 bp oligos[327]. This length is appropriate to encode a terminator and antisense promoter. A library was constructed based on 52 terminators[27] and 109 constitutive promoters[26,320], paired combinatorially to produce 5,668 unique composite parts (Fig 3-4). All of the promoters are synthetic and their strengths fall within a range of 0.0047 au to 21 au, with an average of 3.6 au[320]. All of the terminators are naturally occurring sequences from the *E. coli*

K12 genome and were selected to encompass a wide range of terminator strengths. The majority of these terminators are unidirectional and allow RNAPs fired from the antisense promoter to proceed while blocking those from the forward promoter[27]. The composite parts were synthesized and cloned into three NOT gates made from TetR homologues (Fig 3-5a) (Methods)[328]. The NOT gate repressors (PhlF, SrpR, TarA) were selected to represent different response function shapes (Fig 3-5b).

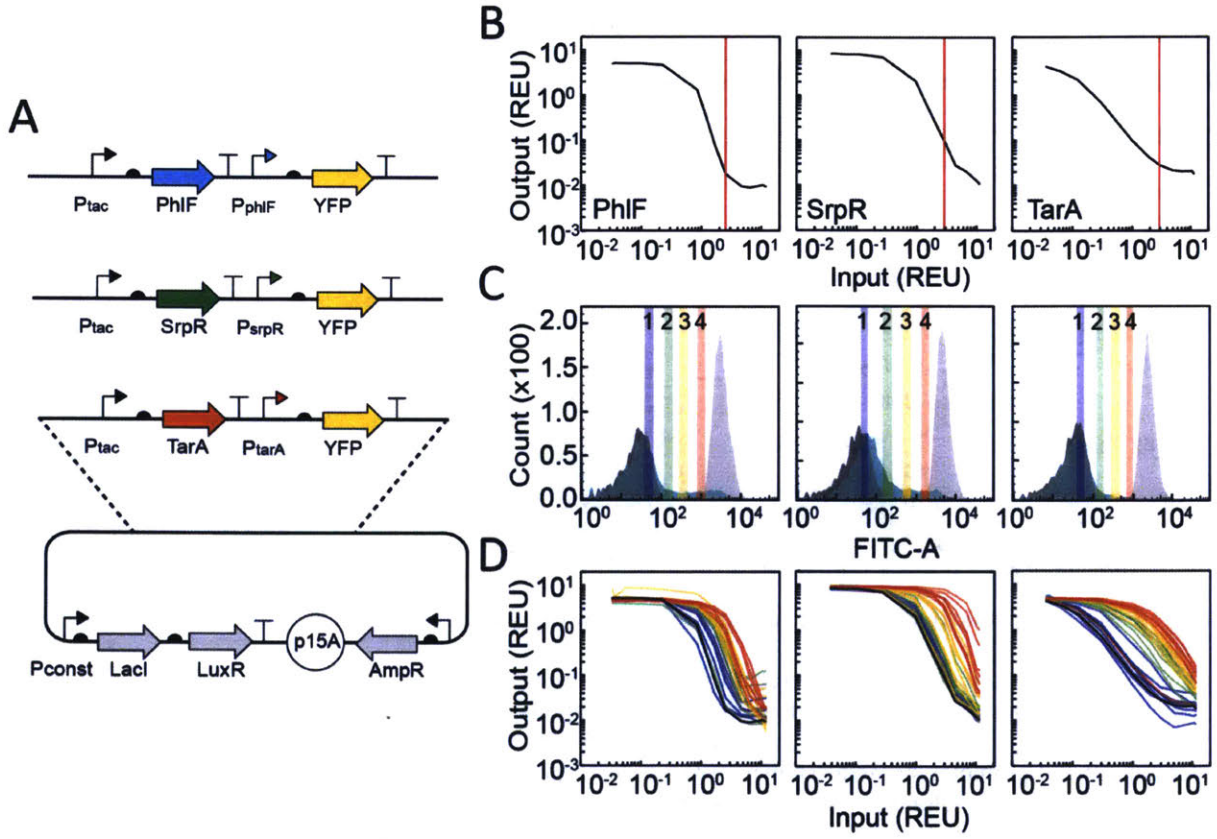

Figure 3-5: Response functions of NOT gates with antisense promoters.
(A) NOT gate plasmids used in this study. All NOT gate plasmids have Not1/Sbf1 multiple cloning sites used to insert terminator/antisense promoter oligonucleotide library. (B) Response functions of NOT gates built with TetR homologs: PhlF (left), SrpR (center), TarA (right). Response functions are measured using P_{tac} activity as the input and YFP as the output. Fluorescence measurements are converted into REU with a reference standard (methods). Vertical lines (red) demarcate P_{tac} activity with 100 μM IPTG, the inducer concentration used to sort libraries. (C) Fluorescence histograms of the starting NOT gates and libraries before sorting. Each NOT gate was grown with 0 μM (light grey) and 100 μM (dark grey) IPTG to set upper and lower bounds for sorting, respectively. Libraries (green) were sorted into four bins, shown as colored vertical bars and numbered by increasing fluorescence. (D) Response functions of randomly selected clones from the sorted libraries. Response functions are colored by the bin in which the clones were found: Bin 1: blue, Bin 2; green, Bin 3; yellow, Bin 4; red. The response functions of the NOT gates lacking antisense promoters are shown in black.

The NOT gate libraries were screened using flow-seq, a technique where fluorescence-activated cell sorting (FACS) is used to sort the cells into bins, the contents of which are determined using next-generation sequencing (Methods)[319,329]. Here, we sorted the cells by NOT gate threshold, *i.e.* the input promoter activity at which the output fluorescence is reduced to half-maximum. To do this, each library was grown with 100μM IPTG and sorted by fluorescence into four log-spaced bins (Fig 3-5c). At 100μM IPTG, all of the gates lacking antisense promoters are OFF (Fig 3-5b) and library members that are ON are likely to have increased induction thresholds. NOT gates without antisense promoters were grown with 0 μM and 100 μM IPTG to set upper and lower bounds for sorting, respectively. 6.5%, 18.3%, and 4.9% of the cells from the PhlF, SrpR, and TarA libraries have increased fluorescence relative to the gates without antisense promoters.

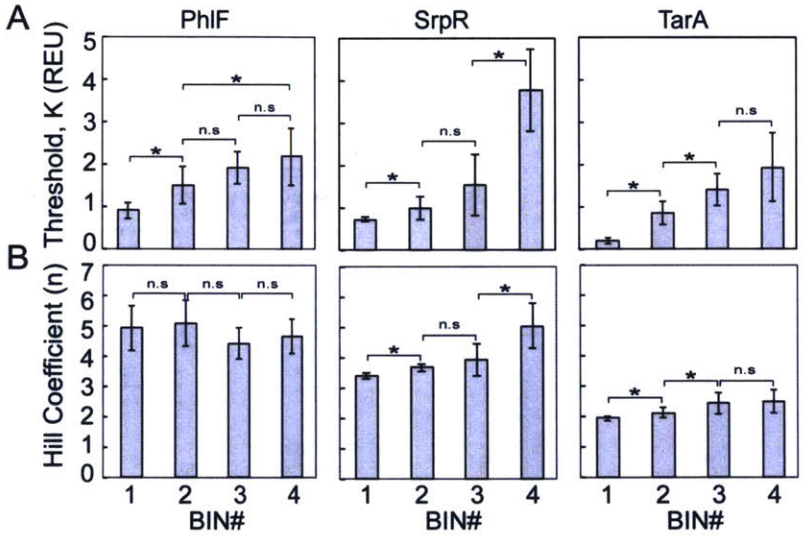

Figure 3-6: Thresholds and Hill coefficients of sorted NOT gate constructs.
(a) Average thresholds *K* and (b) Hill coefficients *n* for NOT gates sorted into each bin of the PhlF (Left), SrpR (Middle), and TarA (Right) libraries. Thresholds and Hill coefficients were calculated by fitting NOT gate response functions to the repressor Hill Equation (Eq. 3). Error bars are the standard deviation between the thresholds or Hill coefficients of the eight NOT gates characterized from each bin. Brackets indicate two-sample Student t tests with P values <0.05 (*).

After sorting, cells were plated onto solid agar medium and eight colonies from each bin were randomly selected and their full response functions were measured (Fig 3-5d). As expected, the response functions cluster according to bins and the most fluorescent bins (bins 3 & 4)

captured gates with the largest increase in threshold (Fig 3-5d, yellow & red). To quantify the effects of the antisense promoter/terminator parts, each response function was fit to a repressor Hill Function

$$y = y_{min} + (y_{max} - y_{min})\frac{K^n}{K^n + x^n} \quad , \tag{3-3}$$

where x is the activity of P_{IN}, y is the activity of P_{OUT}, n is the Hill coefficient, and K is the threshold level of input where the output is half-maximal. High fluorescence bins have increased thresholds K relative to lower fluorescence bins, but the Hill coefficient n does not show a consistent trend across bins (Fig 3-6). After individual colonies were analyzed, we pooled the remaining cells from each bin (50,000 – 200,000 colonies) and measured their response functions in aggregate. Analysis of pooled constructs shows that the ON and OFF states of all library members are essentially constant and threshold differences between the bins are statistically significant (Fig 3-7, one-way ANOVA).

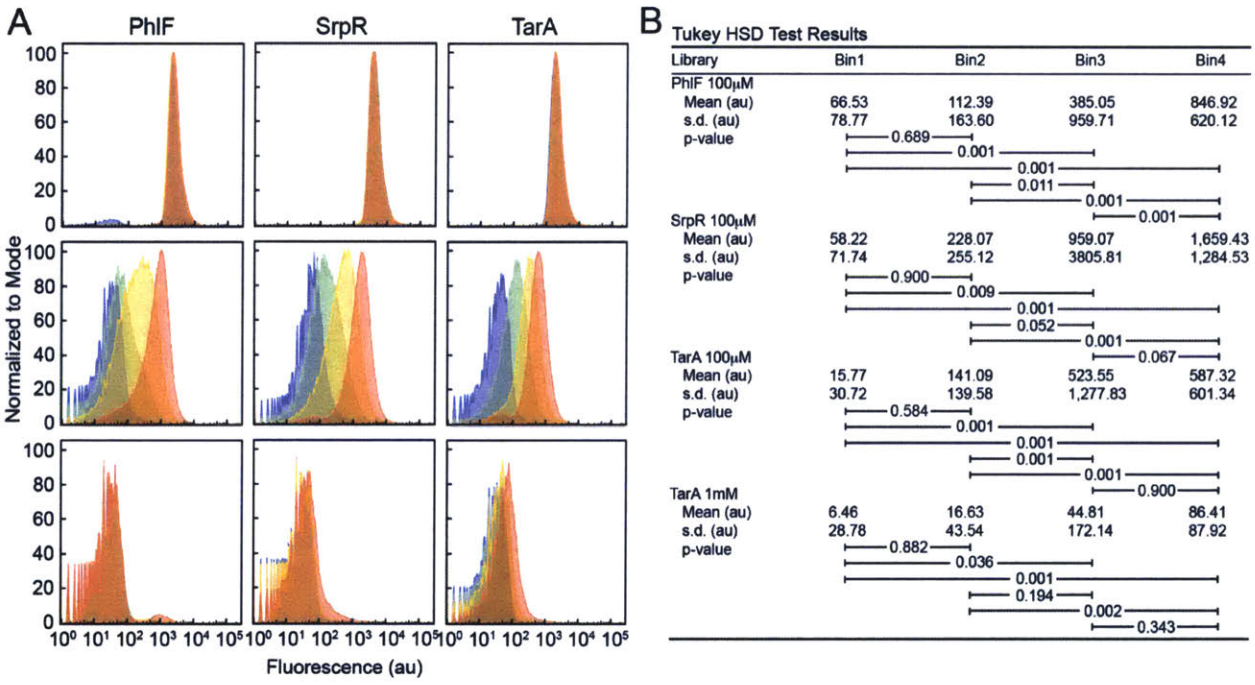

Figure 3-7: Fluorescence histograms and statistics of sorted bins.
(a) The fluorescence histogram of cells from the PhlF (Left), SrpR (Center), and TarA (Right) libraries. Fluorescence was measured using flow cytometry in ON (0mM IPTG - top), threshold (100mM IPTG - middle), and OFF (1000mM IPTG - bottom) conditions. Histograms are colored by bin: 1-purple, 2-green, 3-yellow, 4-orange. One hundred randomly selected individual cells from each data set were used as input to one-way ANOVA tests. There were no statistically significant differences between group means for the PhlF ON ($F(3,396) = 1.06$, $p = 36.67$) and OFF ($F(3,396) = 1.54$, $p = 20.48$) conditions, the SrpR ON

($F(3,396) = 1.22, p = 30.04$) and OFF ($F(3,396) = 1.41, p = 23.87$) conditions, or the TarA ON ($F(3,396) = 1.25, p = 29.19$) condition. There are significant differences between group means for the threshold conditions: PhlF ($F(3,396) = 34.67, p = 5.42E\text{-}18$), SrpR ($F(3,396) = 13.3, p = 2.65E\text{-}8$), TarA ($F(3,396) = 15.71, p = 1.13E\text{-}9$) and the TarA OFF condition ($F(3,396) = 7.85, p = 0.0042$). (b) Post hoc comparisons using the Tukey HSD test for the PhlF threshold, SrpR threshold, TarA threshold, and TarA OFF conditions. Mean fluorescence values of the groups are shown in arbitrary units (au). Brackets indicate Tukey HSD comparisons with P-values listed.

After sorting, the bins were sequenced to identify the promoter/terminator combinations responsible for shifting gate thresholds. Briefly, plasmid DNA was isolated from bacteria in each bin, the composite parts were amplified from the plasmids and barcoded for multiplexed sequencing. Paired end reads were used to ensure complete sequencing of each promoter/terminator pair (Methods). For the following analysis, we removed any sequencing reads that did not perfectly match the designed promoter/terminator pairs. The percent of perfect sequences from the pools was 32.38%, which is consistent with the error rate of chip-based oligo synthesis (Table 3-1)[327].

	Total	Unpaired		Paired		Not1/Sbf1 (-)		Not1/Sbf1 (+)		Perfect	
		Reads	%	Reads	%	Reads	%	Reads	%	Reads	%
PhlF											
Unsorted	2,187,634	258,535	11.82%	1,929,099	88.18%	79,033	4.10%	1,850,066	84.57%	607,711	32.85%
BIN1	2,004,210	239,025	11.93%	1,765,185	88.07%	72,371	4.10%	1,692,814	84.46%	548,928	32.43%
BIN2	167,601	12,946	7.72%	154,655	92.28%	5,204	3.36%	149,451	89.17%	46,927	31.40%
BIN3	125,143	13,686	10.94%	111,457	89.06%	4,628	4.15%	106,829	85.37%	34,268	32.08%
BIN4	106,160	10,798	10.17%	95,362	89.83%	3,369	3.53%	91,993	86.66%	35,642	38.74%
SrpR											
Unsorted	1,663,731	198,112	11.91%	1,465,619	88.09%	75,217	5.13%	1,390,402	83.57%	431,538	31.04%
BIN1	1,872,389	225333	12.03%	1647056	87.97%	63961	3.88%	1583095	84.55%	497,718	31.44%
BIN2	122,744	12,629	10.29%	110,115	89.71%	5,097	4.63%	105,018	85.56%	39,937	38.03%
BIN3	153,442	13,987	9.12%	139,455	90.88%	4,774	3.42%	134,681	87.77%	37,026	27.49%
BIN4	163,519	12,563	7.68%	150,956	92.32%	4,770	3.16%	146,186	89.40%	49,411	33.80%
TarA											
Unsorted	2,171,123	198,140	9.13%	1,972,983	90.87%	76,631	3.88%	1,896,352	87.34%	622,713	32.84%
BIN1	1,910,985	184,335	9.65%	1,726,650	90.35%	77,528	4.49%	1,649,122	86.30%	517,513	31.38%
BIN2	180,667	18,879	10.45%	161,788	89.55%	6,060	3.75%	155,728	86.20%	57,696	37.05%
BIN3	157,354	23,898	15.19%	133,456	84.81%	6,502	4.87%	126,954	80.68%	51,839	40.83%
BIN4	143,926	17,630	12.25%	126,296	87.75%	4,831	3.83%	121,465	84.39%	47,742	39.31%

Table 3-1: Illumina sequencing results.

3-2-3. The response threshold correlates with antisense promoter strength

Previously measured values of the promoter and terminator strengths were used to analyze the parts identified in each bin during deep sequencing (Fig 3-4). In all three libraries, antisense promoter strength increases as a function of bin fluorescence (Fig 3-8a, linear regressions; $R^2 = 0.87421 - 0.96112$). The high fluorescence bins (bins 3 & 4) contain constructs with greater median antisense promoter strength than lower fluorescence bins (bins 1 & 2) (Fig

3-8b). In contrast, there is no consistent trend in the forward or antisense terminator strengths across the sorted libraries (Fig 3-8ab).

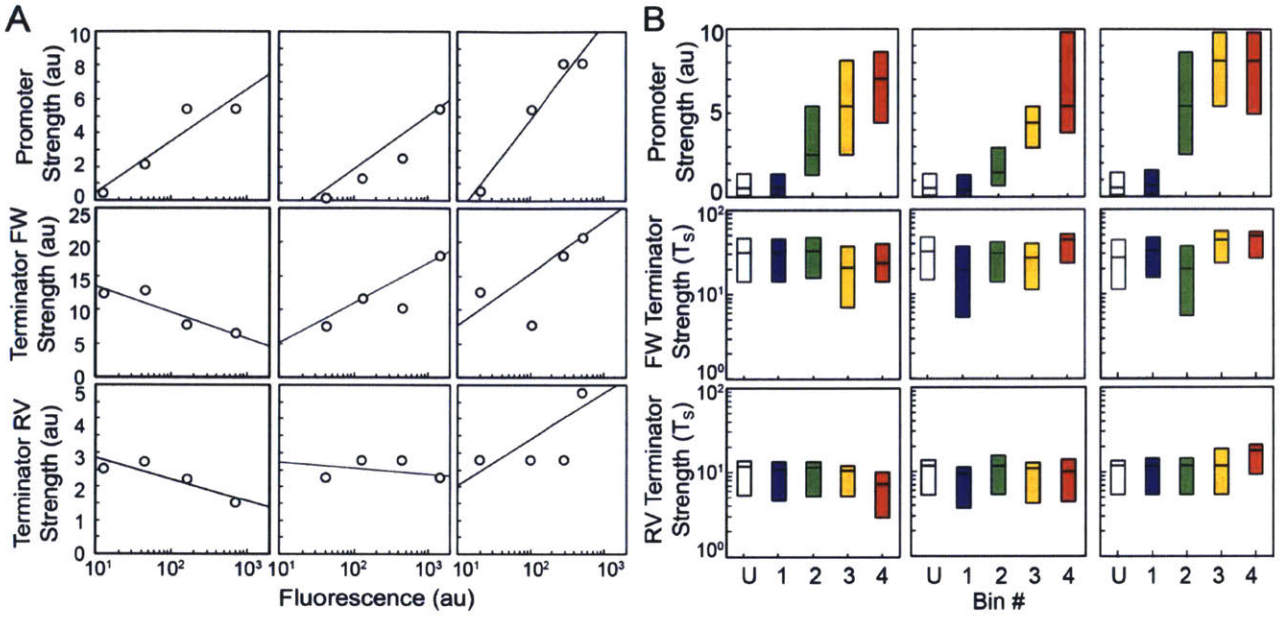

Figure 3-8: Correlations between bin fluorescence and part strength.
(A) Median strengths of the composite parts sorted into bins in the PhlF (left), SrpR (center), and TarA (right) libraries as a function of median bin fluorescence. Antisense promoter strengths correlate positively with bin fluorescence; PhlF $R^2 = 0.87421$, SrpR $R^2 = 0.96061$, TarA $R^2 = 0.96112$ (logarithmic regression). Forward and reverse terminator strengths do not show consistent trends across the libraries; PhlF FW $R^2 = 0.84079$, PhlF RV $R^2 = 0.78599$, SprR FW $R^2 = 0.8628$, SrpR RV $R^2 = 0.16902$, TarA FW $R^2 = 0.71405$, TarA RV $R^2 = 0.74852$ (logarithmic regression). (B) Strengths of antisense promoters and terminators sorted into each bin. Previously measured values (Fig 3-4) were used to calculate the strength of parts sorted into each bin. Box plots display the median, with hinges indicating the first and third quartiles. The unsorted library is marked "U."

The parts responsible for shifting gate thresholds were further explored by enrichment analysis. Enrichment identifies the parts that are selected for, or against, in each bin during sorting (Methods). Since high fluorescence bins (bins 3 & 4) have an increased threshold relative to lower fluorescence bins (bins 1 & 2) (Fig 3-7), composite parts that are enriched in high fluorescence bins are more likely to generate large shifts in gate thresholds than those enriched in lower fluorescence bins. To visualize trends in the data, enrichment was used to assign each composite part to the one bin (1-4) that best reflects its ability to shift gate thresholds (Fig 3-9). Most composite parts with strong antisense promoters are maximally enriched in bins 3 and 4 (Fig 3-9, top). However when strong promoters are paired with bidirectional terminators that

have significant antisense termination efficiencies (T_s antisense > 10), the composite parts are incapable of shifting circuit thresholds (Fig 3-9, left). In contrast, strong antisense promoters are sorted into high fluorescence bins when paired with unidirectional terminators.

These terminators (T_S antisense <10) most likely facilitate greater shifts in gate thresholds than bidirectional terminators by allowing more RNAPs fired from the antisense promoter to interfere with gene expression. In addition, terminators that are predicted to destabilize mRNA or contain cryptic antisense promoters also facilitate large shifts in gate thresholds (Fig 3-9, right). Changes in mRNA stability can result in a large shift because the mRNA produces less protein before it is degraded[27], thus more transcripts are required to produce the threshold amount of repressor protein. Similarly, terminators with cryptic antisense promoters increase the gate threshold by increasing the basal level of antisense transcription.

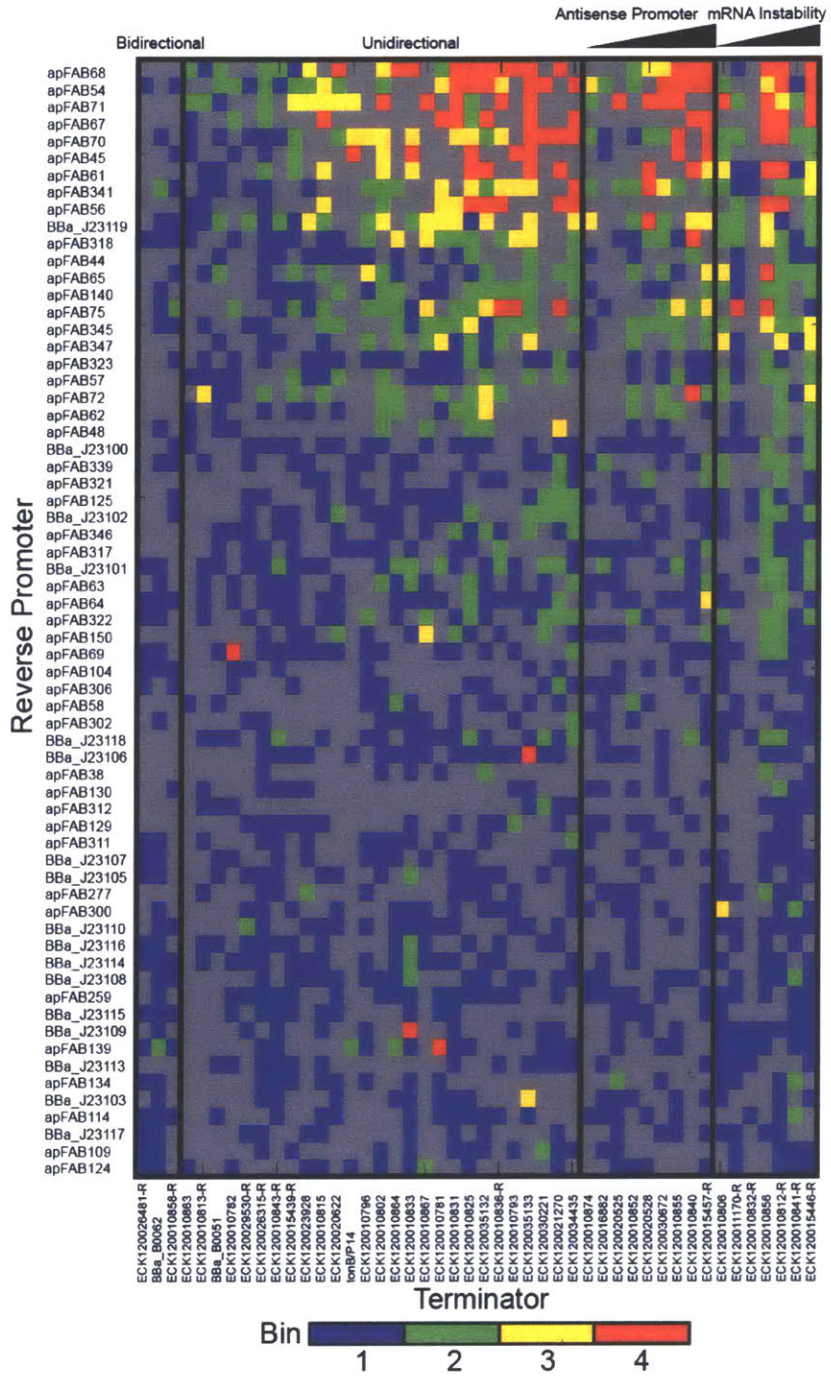

Figure 3-9: Heat map of the bins in which terminator/promoter pairs were most enriched. Promoters are rank ordered by their strength (Part A), with the strongest at the top. The terminators were grouped based on known or predicted terminator features and sorted by the predicted strength of a cryptic antisense promoter or impact on mRNA stability, if relevant (Methods). Unidirectional terminators were sorted based on similarity in their profiles across the promoter set. Terminator/promoter combinations that are not enriched in any of the bins are colored grey. Columns and rows >90% grey were removed from the enrichment grid.

3-2-4. Characterization of terminator/promoter pairs as "parts"

One of our goals is to use antisense transcription to reliably change the expression level of a gene or shift the threshold of a genetic circuit. Ideally, the impact of an antisense promoter on these functions would be predictable and a set of promoters of different strengths could be used to tune expression. When building multi-gene systems, it is desirable to use different terminators to control each gene in order to avoid homologous recombination[330]. Therefore, we sought to identify a set of strong terminators that could be used in conjunction with a set of antisense promoters to reliably tune gene expression. Predictability would require that the promoters impart their effect independent of the terminator to which they are paired.

Some terminators may have mechanisms that impact the effectiveness of the antisense promoter. As such, we eliminated those with known features (cryptic promoters, bidirectional termination, hairpins that impact mRNA stability) from the set (Fig 3-9). Then, terminators and promoters were systematically removed until there remained a core set of both where the promoters produced a reliable response when combined with any of the terminators. This set, shown in Figure 3-10a, provides nine strong terminators that can be fused to different genes or operons and twenty antisense promoters that can be added to control their expression (Fig 3-10b). To confirm predictability, several terminator-promoter pairs were tested in the reporter construct (Fig 3-10a). The repression produced by these pairs collapse onto a single curve, independent of the identity of the terminator used (Fig 3-10c, exponential regression; R^2 = 0.9275)(Table 3-2).

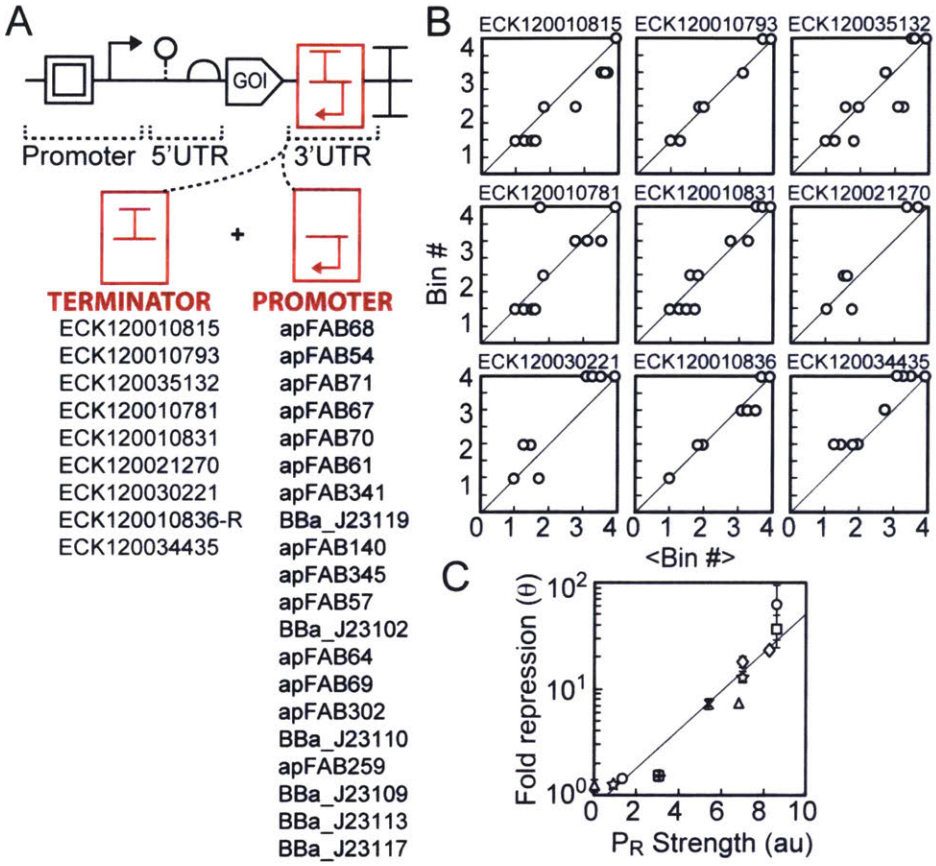

Figure 3-10: Composability of unidirectional terminators and antisense promoters.
(A) The flow-seq data was used to identify a subset of promoters and terminators that could be combined to obtain a reliable reduction of gene expression. (B) Each graph shows a terminator (name at top) and each point is a promoter from the list in Part A. The x-axis (<Bin#>) is the average for the promoter across the complete terminator set and the y-axis (Bin#) is the bin for the specific terminator. The Bin# is calculated as described in Fig 2F. (C) Repression was explicitly measured for a subset of terminator/promoter pairs selected from Part A. The pairs were cloned into the plasmid from (a) and fold-repression (Equation 1) was measured as a function of the forward promoter activity (see Table 3-2 for terminator/promoter combinations tested). Maximum fold repression is plotted against the previously measured promoter activities[320]; R^2 = 0.9275. Composite parts are marked by terminator (ECK120035132, circle; ECK120010831, square; ECK120034435, triangle; ECK120021270, diamond; ECK120010793, star; ECK120030221, x; ECK120010815, +).

Terminator	Promoter	Maximum repression ± s.d.
ECK120035132	apFAB71	61.86 ± 33.25
ECK120010831	apFAB71	36.60 ± 12.32
ECK120021270	apFAB67	23.08 ± 1.50
ECK120021270	apFAB61	17.94 ± 2.29
ECK120010793	apFAB61	12.84 ± 1.54
ECK120034435	apFAB341	7.05 ± 0.63
ECK120030221	Bba_J23119	7.16 ± 0.75
ECK120010815	apFAB345	1.49 ± 0.02
ECK120010831	apFAB345	1.51 ± 0.18
ECK120035132	Bba_J23102	1.44 ± 0.06

ECK120010793	apFAB69	1.26 ± 0.03
ECK120034435	apFAB259	1.23 ± 0.17
ECK120030221	apFAB259	0.98 ± 0.04
ECK120010836-R	Bba_J23109	0.87 ± 0.03

Table 3-2: Terminator promoter pairs tested in Figure 3-10.
Maximum repression is the average of three replicates collected on different days.

3-2-5. Repression occurs due to a combination of asRNA activity and transcriptional collisions

To determine the relative contributions of asRNA and transcriptional interference to repression generated by antisense promoters, we built a set of plasmids to express asRNA corresponding to the reporter gene (Fig 3-11a). These plasmids each have one of the four antisense promoters (apFAB49, apFAB140, apFAB78, or apFAB96) driving expression of the reverse compliment of RFP followed by a strong bi-directional terminator (ECK120034435). The cassette is placed on a plasmid containing the colE1 origin, which is maintained at a copy number approximately two times higher than p15A[331]. The asRNA plasmids were co-transformed into *E. coli* along with the original RFP reporter plasmid (pJBTI241, Fig 3-2a).

Using these data, the fold-repression due to antisense RNA θ_{asRNA} is calculated as in Equation 1, where the subscripts +/0 now represent the presence or absence of *trans* encoded asRNA. θ_{asRNA} should be viewed as an upper bound on the contribution from asRNA to total fold repression θ. This is because its expression in *trans* causes the asRNA to be longer (relative to the asRNAs generated at the 3'-end by RNAP collisions) and expressed at a higher level (due to absence of RNAP collisions and the higher copy number plasmid). Fold repression generated by transcriptional interference θ_{TI} can then be determined by dividing the total fold repression q by θ_{asRNA}. Thus, θ_{asRNA} and θ_{TI} reflect the relative contributions of asRNA and transcriptional interference to total repression generated by antisense transcription. Plotting θ, θ_{asRNA}, and θ_{TI} versus inducer concentration shows that asRNA and transcriptional interference generate approximately equivalent contributions to repression (Fig 3-11b). When the forward promoter is strongly induced, the predicted *cis* contribution declines, which is consistent with models of transcriptional interference[316] (Fig 3-11b).

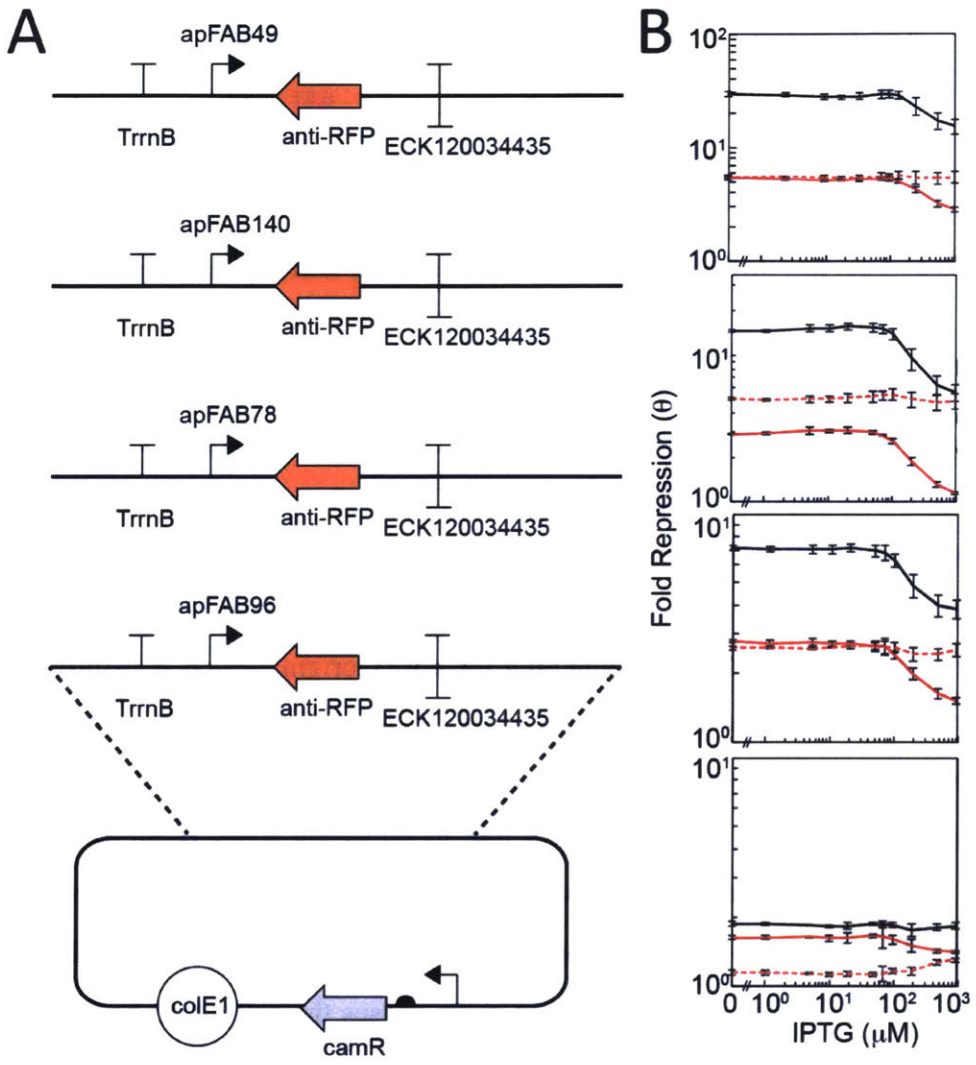

Figure 3-11: Repression generated by asRNA.
(A) Plasmids to measure repression generated by asRNA produced in *trans*. These plasmids were co-transformed with pJBTI241 and RFP fluorescence was measured to test repression generated by asRNA in *trans*. (B) Total fold-repression θ generated by apFAB96 (a), apFAB78 (b), apFAB140 (c), apFAB49 (d) antisense promoters are shown as a function of forward promoter activity using the characterization system in Fig 4a (black line). This is compared to the repression observed when the same promoter is used to drive the transcription of asRNA in *trans* from a separate plasmid (dashed red line). The repression due to transcriptional interference θ_{TI} (solid red line) is inferred from the total and *trans* asRNA repression data (see text for details).

It is noteworthy that strong repression cannot be achieved through either transcriptional interference or asRNA alone. They each contribute equally in a multiplicative way to the total repression that can be achieved using antisense promoters. The *cis* regulation is likely most important for achieving maximal repression when the asRNA does not have specific regulatory

qualities, such as RNA-RNA or RNA-DNA interaction hairpins[332,333], Hfq binding motifs[334], or RNase processing sites[304,335].

A differential equation model was developed to explore collision interference and parameterize the repression that arises for different forward and antisense promoter firing rates (ϕ_F and ϕ_R) and gene length N. The model and parameters are shown in Figure 3-12 and Table 3-3. Polymerases that originate from the forward promoter P_F transcribe a gene at a constant velocity v unless they collide with polymerases from the interfering promoter P_R on the opposing strand of DNA. In the event of a collision, polymerases dissociate from the DNA.

Collision interference can be captured by two differential equations that track the steady-state concentration of polymerases on the forward C_F and reverse C_R strands as a function of the distance from the start site x (in bp):

$$\frac{dC_F}{dx} = -\varepsilon_F C_F C_R \tag{3-4}$$

$$\frac{dC_R}{dx} = \varepsilon_R C_F C_R \tag{3-5}$$

Here, the ε_F and ε_R are parameters that reflect the possibility that RNAPs fired from the forward and antisense promoters could have different propensities to encounter collisions and dissociate from the DNA. Transcriptional bursting could prevent RNAPs from encountering head on collisions by increasing the time between transcription events initiated at the opposing promoter. In addition, RNAPs fired from the forward and antisense promoters could have different propensities to dissociate from the DNA following collision. *In vitro* experiments show that head-on RNAP collisions result in stalling and backtracking of the enzymes[336], which leaves them vulnerable to clearance[337,338]. However, co-translating ribosomes[339] and actively transcribing RNAPs[340,341] have been shown to rescue stalled/backtracked complexes by realigning the 3'termini of the RNA transcript with the enzyme's active site. In addition, there are active mechanisms that favor the termination and release of RNAPs transcribing non-coding RNA[342]. Therefore the model was built to accommodate differences in dissociation for RNAPs fired from forward and antisense promoters of different strengths.

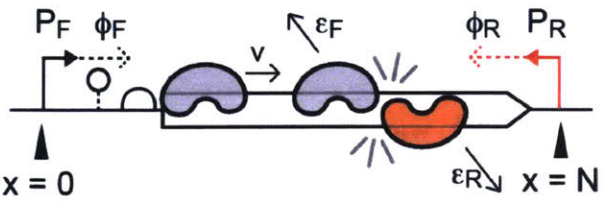

Figure 3-12: Schematic of the transcriptional interference model.
A forward P_F and antisense P_R promoter are located on either side of a gene, N bases apart. The promoters fire at rates ϕ_F and ϕ_R RNAP/second. Polymerases transcribe at a constant velocity v unless they collide with polymerases fired from opposing promoter. Polymerases collide and dissociate from the DNA with a probability ε.

Parameter	Meaning	Values	References
N	Distance between transcription start points (bp)	841	
v	Speed of transcription (bp/s)	40	Sneppen *et al*, 2005
ϕ_F	Rate of RNAP production from P_F (s-1)	pBla: 0.031	Liang *et al*, 1999
		pTac [1]: 0.001439	this study
		pTac [5]: 0.001548	this study
		pTac [10]: 0.002232	this study
		pTac [20]: 0.003643	this study
		pTac [50]: 0.006778	this study
		pTac [70]: 0.025262	this study
		pTac [100]: 0.03976	this study
		pTac [200]: 0.18756	this study
		pTac [500]: 0.35764	this study
		pTac [1000]: 0.41141	this study
ϕ_R	Rate of RNAP production from P_R (s-1)	apFAB49: 0.15345	this study
		apFAB140: 0.64814	this study
		apFAB78: 1.01410	this study
		apFAB96: 1.23185	this study
ε_F	Ejection factor for RNAP fired from P_F	0.104 ± 0.031	fitted
ε_R	Ejection factor for RNAP fired from P_R	0.494 ± 0.027	fitted

Table 3-3: Parameters in the transcriptional interference model.

Boundary conditions are defined by the rates that polymerases are fired, *i.e.* begin elongating, at the forward $C_F(x=0) = \phi_F/v$ and reverse $C_R(x=N) = \phi_R/v$ promoters. The polymerase velocity v = 40 bp/s is held constant[316,343]. The equations are numerically solved for each ϕ_F and ϕ_R combination (Methods). We chose to model P_F as P_{tac} with ten different IPTG concentrations and P_R as four different constitutive promoters: apFAB49, apFAB140, apFAB78, and apFAB96 (Fig 3-2). Simulated repression can be calculated as the ratio of full-length ($x = N$) transcripts produced from P_F with and without an antisense promoter

$$\theta = \frac{C_F|_{\varepsilon_F=0}}{C_F|_{\varepsilon_F>0}} \quad , \tag{3-6}$$

providing a prediction that can be compared with measurements (Equation 3-1). For each combination of forward and antisense promoters, θ is calculated and compared with that derived from experiments.

Repression due to asRNA is not included in the model. Rather, our approach was to fit the model predictions to the two bounds that we measured. First, results were fit to total repression θ, which assumes there is no contribution from asRNA to the observed repression. Next results were fit to θ_{TI}, which represents the minimum amount of repression attributable to transcriptional interference. These bounds were then used to fit the underlying biophysical parameters and provide a range of values that reflect the possible contribution of transcriptional interference to gene repression.

We first simulated collision interference with $\varepsilon_F = \varepsilon_R = 1$, since previous models of transcriptional interference assume that actively transcribing polymerases never survive head-on collisions [280,281,316] (Fig 3-13a). However, assuming $\varepsilon_F = \varepsilon_R = 1$ predicts too much repression and results in a poor fit to our experiments.

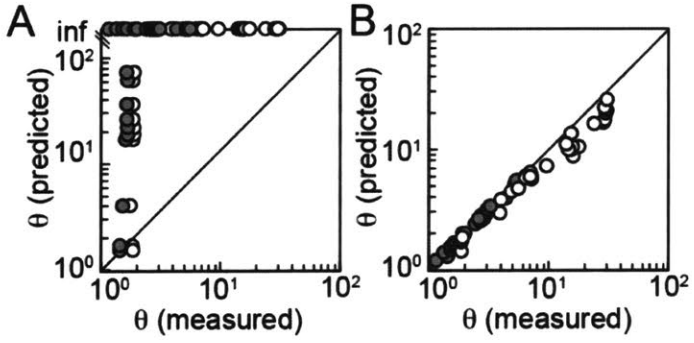

Figure 3-13: Comparison of model predictions to experimental data.
Each data point shows experimentally measured repression (θ – white, θ_{TI} - grey) for each P_F/P_R pair plotted against model predictions generated with the same promoter combinations. Repressions predicted by the model were calculated using Equation 6 with (a) $\varepsilon_F = \varepsilon_R = 1$ or (b) optimal ε_F and ε_R. The optimal ε_F and ε_R values simulate repression closest to the experimentally measured value (Methods).

To optimize ε_F and ε_R, the model was solved where these parameters are varied in the range [0,1] in increments of 0.01 (Fig 3-14). This was repeated for each value of ϕ_F and ϕ_R in our data set and the values of ε_F and ε_R that generate repression closest to the experimentally

measured θ_{TI} were determined. This yielded a set of 88 optimal values of ε_F and ε_R, corresponding to all of combinations of forward and antisense promoter activities (44 pairings) fit to either θ or θ_{TI}. The optimal values of ε_F and ε_R produced behavior that closely matches the experimental data (Fig 3-13b, linear regression; m = 1, R^2 = 0.84082). Values of $\varepsilon < 1$ are interpreted as cases where polymerases either avoid or survive collision and continue transcribing. Importantly, the model does not assume that one RNAP must dissociate in order for the RNAP on the opposing strand of DNA to survive collision. This assumption that polymerases can bypass is supported by *in vitro* experiments done with viral[344] and yeast[345] RNA polymerases.

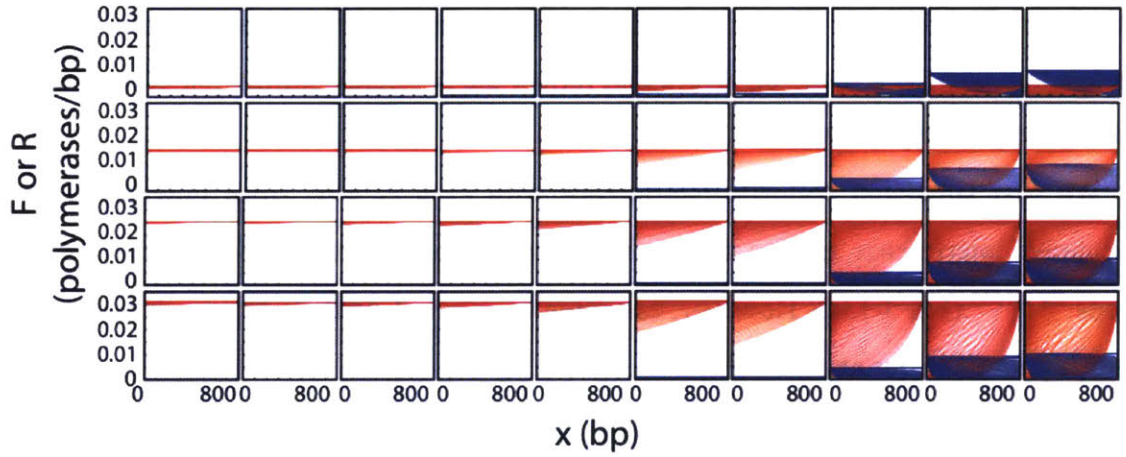

Figure 3-14: Model results when ε_F and ε_R are varied between 0 and 1 at increments of 0.01.
Graphs show density of polymerases fired from either the forward (F - blue) or interfering promoter (R - red) along the DNA. Forty different forward/interfering promoter combinations were simulated, which model P_F as pTac + ten IPTG concentrations (1, 5, 10, 20, 50, 70, 100, 200, 500, 1000 μM; across) and P_R as apFAB49 (top row), apFAB140 (second row), apFAB78 (third row), or apFAB96 (bottom row).

The optimum values of ε_F and ε_R are surprisingly constant across the dataset and are independent of the identity of the antisense promoter or the firing rate of the forward promoter (Fig 3-15a). ε values fit using θ represent the probabilities that polymerases collide and dissociate when transcriptional interference is assumed to be the sole mechanism for gene repression. In contrast ε values fit using θ_{TI} reflect the smallest possible contribution of transcriptional interference to gene repression. Thus we find a range of ε values that reflect the potential contributions of collision interference to gene repression. We find that the probability that a polymerase dissociates due to a competing enzyme is significantly larger for RNAPs fired from

the antisense promoter (<ε_R> = 43-54%) as compared to those from the forward promoter (<ε_F> = 7-15%) (Fig 3-15b). In addition, the model predicts that fold repression due to transcriptional interference increases exponentially as a function of the antisense promoter strength and distance between the two promoters (Fig 3-16). To test repression as a function of distance between the forward and antisense promoters, we modified our antisense reporter system and inserted *yfp* between the 3'end of *rfp* and the antisense promoter (Fig 3-25b). This increases the distance between the two promoters from 850bp to 1,500bp. Measuring repression of RFP and YFP using this system shows that repression increases as the distance between the two promoters grows.

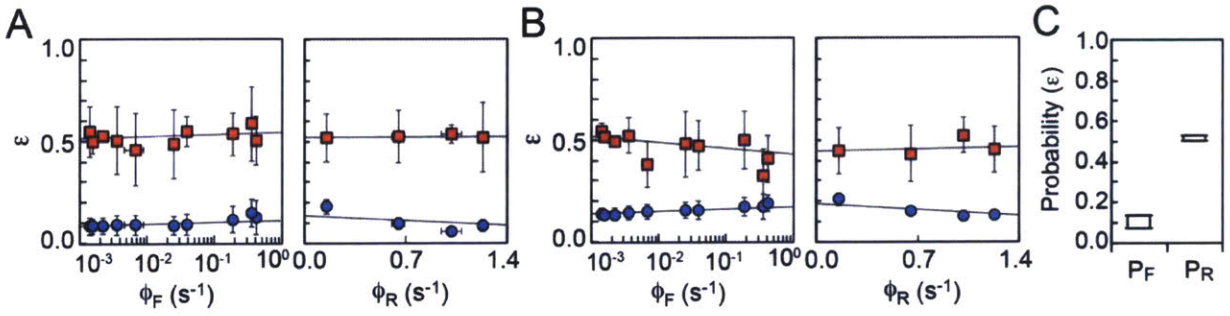

Figure 3-15: Best fit values of ε as a function of forward P_F and antisense P_R promoter strength. Relationships between forward promoter strengths, ϕ_F and ϕ_R, and the probabilities that polymerases fall off the DNA after collision, ε_F (blue circles) and ε_R (red squares). Graphs show the best ε_F and ε_R for each promoter combination from the ε parameter sweep and experimentally measured ϕ_F and ϕ_R. ε_F and ε_R were fit to experimental data of fold repression generated by transcriptional interference alone θ_{TI} (a) or by maximum fold repression (b). The highest scoring ε_Fs and ε_Rs for each ϕ_F and ϕ_R are averaged, y-error bars show the s.d. between these values. x-error bars show the s.d. of three replicates collected on different days. (C) Range of optimal ε values that result from fitting the model to θ or θ_{TI}. Box plot extends from the median optimal ε_F and ε_R when the model is fit to θ_{TI} (<ε_F> = 0.07, <ε_R> = 0.52) to median optimal ε_F and ε_R when the model is fit to θ (<ε_F> = 0.14, <ε_R> = 0.47). In all figure parts, the data represents the mean of three experiments performed on different days and the error bars are the standard deviation of these replicates. In all panels, fold repression θ is the magnitude of gene expression produced by forward promoter in the absence of an antisense promoter divided by the amount of gene expression produced in the presence of an antisense promoter.

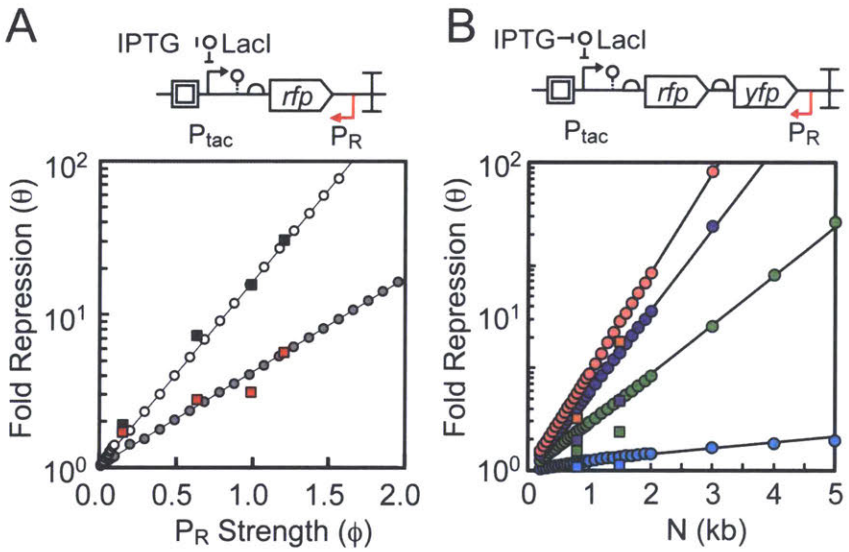

Figure 3-16: Model predicts an exponential increase in repression.
The collision interference model was used to predict fold repression with wide range of antisense promoter strength P_R values and different distances between promoters. (a) Top: schematic showing the antisense transcription reporter system. Bottom: model simulations where ϕ_R is varied from 0 to 2 at increments of either 0.01 or 0.1 (1/s). ε_F is set to the median value fit by θ_{TI} (ε_F = 0.07, white circles) and θ (ε_F = 0.14, grey circles). For these simulations, ε_R was held constant at 0.515. Experimental results for θ (black squares) and θ_{TI} (red squares) are also shown as a function of antisense promoter strength. Data from Fig 3-19. (b) Top: Schematic showing the antisense transcription reporter system with longer distance between P_F and P_R. A second fluorescent protein *yfp* was added between the 3'end of *rfp* and PR to increase the distance between promoters from 841bp to 1,500 bp. Plasmid maps in Appendix Fig S8. Bottom: model simulations where N is varied from 200 – 5,000bp in increments of 50, 100, or 1,000 bp. P_R was simulated as apFAB49, apFAB140, apFAB78 and apFAB96 by setting ϕ_R to values of 0.15345 RNAP/s (blue circles), 0.64814 RNAP/s (green circles), 1.014108 RNAP/s (purple circles), and 1.231852 RNAP/s (red circles), respectively. For all simulations in this figure, ϕ_F was held constant at 0.03976 (to simulate P_{tac} + 100μM IPTG) and ε_F and ε_R were set at the median optimum values: 0.07 and 0.515, respectively. Experimental results show RFP and YFP repression with four different antisense promoters: apFAB49 (blue squares), apFAB140 (green squares), apFAB78 (purple squares), and apFAB96 (red squares). For all figure panels, repression is measured when Ptac is induced with 100μM IPTG.

3-3. CONCLUSIONS

This work demonstrates that antisense promoters can be reliably used to tune gene expression. The degree of repression is proportional to the strength of the antisense promoter over a >30-fold range. This builds on the modern revisiting of the classical "expression cassette" to incorporate additional non-canonical parts to tune expression, insulate against context, and provide for rapid debugging via –omics techniques[26,346]. In this paradigm, there are alternative means to control the expression level of a gene, such as by changing the forward promoter[26], the RBS[347], small RNA[332,333,348], and 3'-hairpins for mRNA stability[349]. While some of these

approaches can achieve greater ranges of expression control, antisense promoters have some unique features that are advantageous for some applications. Notably, they offer a means of control external to the expression cassette. This is particularly valuable when the forward control elements (promoter and RBS) have been engineered to integrate additional regulatory information[328,350–352]. In these cases, it is not simple to adjust the overall expression level without interfering with how the signals are integrated. Exploiting antisense transcription allows for control without changing the inputs to the system or the sequence of the forward transcript. The flow-seq data demonstrate that the impact of the antisense promoter is largely context-independent. From this, we derive a set of unidirectional terminators that can be combined with the antisense promoter in a modular manner. Thus, implementing this control is simple and modular and can be done with existing promoter libraries.

The performance of antisense transcription is derived from its unique synergy between its impact on transcription from the forward promoter and post-transcriptional impact on protein expression. For our system, we find that antisense RNA and collision between actively transcribing RNA polymerases contribute roughly equally to repress gene expression. This synergy is important because transcriptional interference implements its control at the transcriptional level and would not be able to repress mRNAs made by RNAPs that avoid collision. Antisense RNAs prohibit escaped mRNA transcripts from being translated. For asRNAs with weak affinity for the target mRNA, cohesion between the two mechanisms may facilitate greater repression of target genes than the asRNA alone.

The model predicts some mechanistic details about collision interference. Most strikingly, polymerases transcribing translated mRNA survive or avoid ~85% of their head on collisions, which may explain the inability to completely abolish gene expression with transcriptional interference alone, despite the use of very strong interfering promoters. Polymerase survival rates are not as high if the polymerase is fired from the antisense promoter (~50% survival). This imbalance may be due to differences in the kinetic properties of the two promoters, *e.g.* burstiness, or differential dissociation of the RNAPs. Single molecule experiments that measure polymerase survival rates directly could be done to differentiate between these two mechanisms. Additional experiments can also refine our understanding of antisense transcription by parameterizing additional mechanisms. The model presented here is

limited to collision interference as the mechanism of repression in *cis*, however several additional factors, such as r-loop formation[353], changes in DNA topology[269], differences in local asRNA concentration[354], and occlusion interference[315], could also be considered. Direct measurement of asRNA activity in *cis* and *trans*, as well as measurement of RNA duplex formation and degradation rates, r-loop formation, and other modes of transcriptional interference would facilitate the construction of a more detailed mechanistic model of repression mediated by antisense transcription.

Considering natural genomes, antisense promoters could be a simple evolutionary mechanism to reduce gene expression. This is especially true of housekeeping sigma factors – such as s_{70} – whose binding sites are relatively information poor, and mutagenesis is expected to cause promoters to frequently arise spontaneously during evolution[296]. The particular location of the antisense promoter would not matter and it could even occur within the gene itself, as is often observed[285,292,294,305,355]. Constitutive promoters arise quickly and this would provide a simple mechanism to reduce gene expression that could be rapidly discovered during an evolutionary search. This is consistent with the lack of conservation of antisense promoters between species[296,297], where they may appear and disappear quickly in evolutionary time to fine tune expression. This would be an easier solution to find than making mutations to the sense promoter, which can be significantly constrained by needs of regulatory signal integration[356]. Thus, the total expression of a gene could be tuned without disturbing the integration of signals. The rate of evolution around any individual promoters may not be high[296] because there are many similar solutions that can be found.

Here, we show how antisense transcription can be integrated into a simple NOT gate, that has been a common motif in building larger synthetic genetic circuits. This provides a mechanism where the switching threshold can be tuned without impacting other characteristics of the gate, such as the cooperativity. More complex circuits could be built by exploiting antisense transcription in both prokaryotic and eukaryotic systems, where antisense regulation is known to occur[280,282,301] and additional tuning knobs can help improve performance of synthetic systems[323,357,358]. The simple constitutive promoters we employ here could be exchanged for dynamic promoters that respond to inducers or cellular/environmental conditions or implement negative feedback. This motif occurs in natural regulatory networks, for example many of the

antisense promoters in *B. subtilis* are regulated by alternative sigma factors that respond to different environmental conditions[297]. This gets more complex as the sigma factors themselves are regulated by antisense transcription[326]. Even more interesting architectures have been observed in nature, for example there are many that involve overlapping 5'- and 3'-UTRs. The overlap can include entire genes; for example, divergent operons have been observed where the promoter for each occurs one gene into the other[280,359]. These motifs would enable mutually exclusive switch-like changes between the sets of genes that are expressed[300]. Collectively, this points to antisense transcription as something that should be routinely incorporated into engineered systems, as opposed to being avoided.

3-4. Materials and Methods

3-4-1. Strains and media.

E. coli strains NEB10β (Δ(ara-leu)7697 araD139 fhuA ΔlacX74 galK16 galE15 e14-f80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ (mrr-hsdRMS-mcrBC)) and DH10B (F- D(ara-leu)7697 araD139 DlacX74 galE15 f80dlacZDM15 recA1 endA1 nupG rpsL mcrA D(mrr-hsdRMS-mcrBC) l-) were used for all experiments. Cells were grown in either LB Miller Broth (Becton Dickinson 244630) or M9 minimal medium supplemented with glucose ((6.8 g/L Na2HPO4, 3 g/L KH2PO4, 0.5 g/L NaCl, 1 g/L NH4Cl; Sigma M6030), 2mM MgSO4 (Affymetrix 18651), 100 mM CaCl2 (Sigma C1016), 0.4% glucose (Fisher scientific M10046), 0.2% casamino acids (Becton Dickinson 223050), 340 mg/mL thiamine (vitamin B1) (Alfa Aestar A19560). Carbenicillin (100 mg/mL) (Gold Bio C-103), kanamycin (50 mg/mL) (Gold Bio K-120) and/or chloramphenicol (35 mg/mL) (USB Corporation 23660) were added to growth media to maintain plasmids when appropriate. Isopropyl b-D-1-thiogalactopyranoside (IPTG) (Gold Bio I2481C) was used as the inducer for all constructs.

3-4-2. Measurement of response functions.

E. coli strains were grown for 16 h in LB media containing antibiotics in 96-deep well blocks (USA Scientific 1896-2000) at 37°C and 250 r.p.m in an INFORS-HT Multitron Pro. After 16 h, the cultures were diluted 1:200 into M9 medium with antibiotics and grown for 3 h with the same shaking and temperature settings as the overnight growth. Next, the cultures were diluted

1:700 into fresh M9 medium with antibiotics and different concentrations of isopropyl β-D-1-thiogalactopyranoside (IPTG). These cultures were grown for 6 h then diluted 1:5 into phosphate buffered saline (PBS) containing 2mg/mL kanamycin or 35μg/mL chloramphenicol to arrest protein production and fluorescence was measured using a flow cytometer.

3-4-3. Cytometry measurement and data analysis.

Cells were analyzed by flow cytometry using a BD Biosciences Fortessa flow cytometer with blue (488 nm) and red (640 nm) lasers. An injection volume of 10 μL and flow rate of 0.5 μL/s were used. Cytometry data was analyzed using FlowJo (TreeStar Inc., Ashland, OR) and populations were gated on forward and side scatter heights. The gated populations consisted of at least 30,000 cells. The median fluorescence of the gated populations was used calculated using FlowJo and used for all reporting. Auto-fluorescence of white cells (NEB10β without plasmids) was subtracted from all fluorescence measurements.

3-4-4. Promoter strength calculations.

Promoter firing rates (RNAP/second) were estimated using NEB10β cells harboring one of the following plasmids: pJBTI26, pJBTI264, pJBTI265, pJBTI266, pJBTI267, pJBTI136 (Fig 2-17). Fluorescence of each strain was measured as described above. Fluorescence produced by the strain harboring plasmid pJBTI136 (<YFP> = 528 au) was used to define a promoter-firing rate of 0.031 RNAP/second, which has been reported for promoter P_{bla}[59]. Fluorescence of strains carrying the other plasmids were divided by fluorescence produced by the strain harboring pJBTI136 and multiplied by 0.031 RNAP/second to obtain promoter firing rates. The hammerhead ribozyme insulator RiboJ[56] was used to standardize the 5'UTR of YFP mRNA so that changes in fluorescence could be attributed solely to differences in polymerase firing. To convert promoter-firing rates (RNAP/second) back to arbitrary units reported by the cytometer, multiply the firing rates by 17,032.

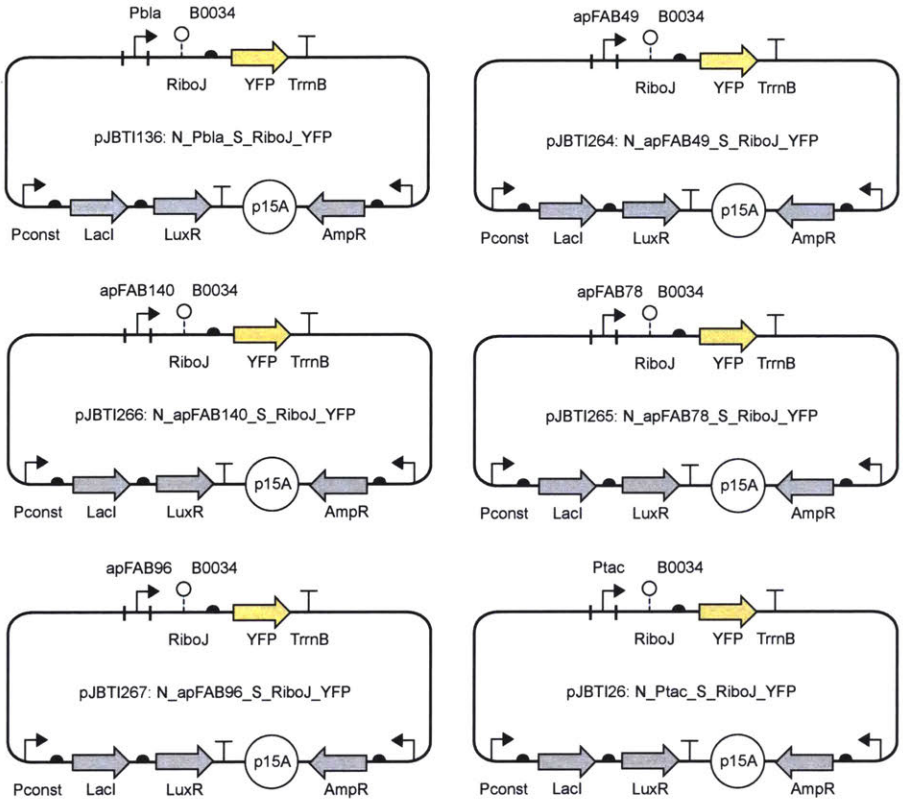

Figure 3-17: Plasmids to measure promoter firing rates (RNAP/second).
Promoter P_{bla} was used to define a polymerase firing rate of 0.031 s^{-1}. NotI/SbfI multiple cloning sites were used to digest the reference plasmid and insert other promoters (apFAB49, apFAB140, apFAB78, apFAB96, P_{tac}) for promoter strength measurement.

Relative expression units (REUs) were calculated using DH10B cells harboring one of the following plasmids: 0RFP2, pAN1717 (Fig 3-18a). Strains harboring 0RFP2 and pAN1717 were grown and measured in parallel with experimental strains. To convert raw RFP fluorescence measurements into REU, RFP produced by experimental strains was divided by red fluorescence produced by the strain harboring 0RFP2. To convert our reported RFP measurements (REU) back to arbitrary units, multiply the REU value by 2295. To convert raw YFP fluorescence measurements into REU, YFP produced by experimental strains was divided by YFP produced by pAN1717. To convert our reported YFP measurements (REU) back to arbitrary units, multiply the REU value by 550. When measuring NOT gate response functions, input promoter (P_{tac}) activity was measured using plasmid pJBTI26 (Fig 3-18b) and converted into REUs as described here.

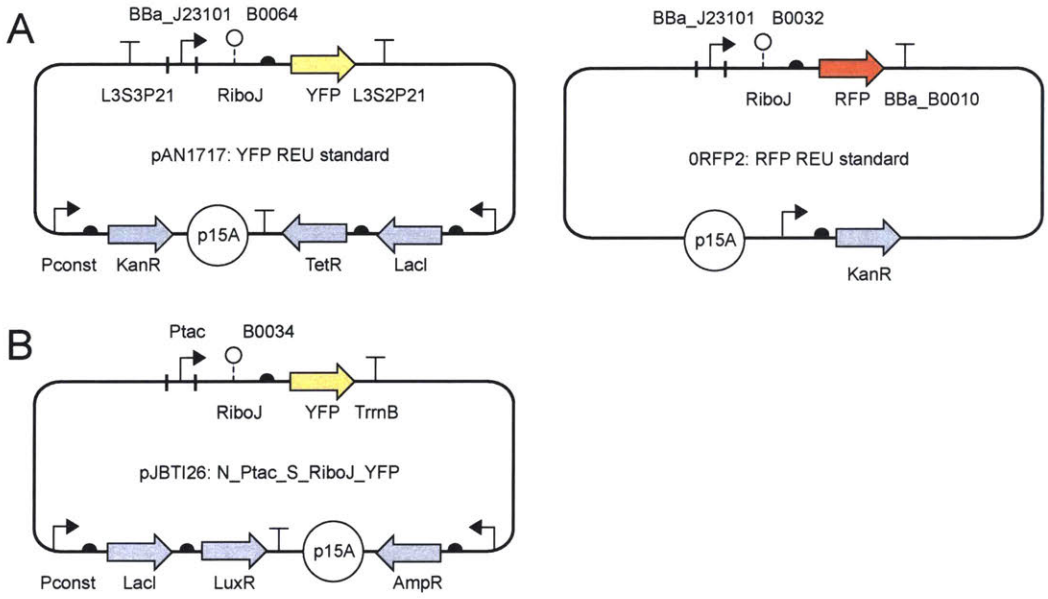

Figure 3-18: REU standard plasmids used in this study.
(a) Plasmids used to convert YFP (pAN1717) and RFP (0RFP2) fluorescence measurements into Relative Expression Units (REUs). (b) Plasmid used to measure input (P_{tac}) promoter activity for NOT gate response functions.

Promoter strength measurements reported throughout the paper as au are from the RNA-seq experiments of Kosuri *et. al*, 2013[54]. In these experiments, promoter strengths are calculated using RNAseq read depth of mRNA produced by each promoter driving expression of green fluorescent protein.

3-4-5. Classification of terminators.

Terminators that encode cryptic antisense promoters or destabilize mRNA when placed at the 3'end were identified by analyzing data from Chen *et al.*'s study of *E. coli* intrinsic terminators[63]. In this study, termination strength was measured by observing the changes in GFP and RFP expression that occur when a terminator is placed between two fluorescent proteins (5' GFP and 3'RFP). Strong terminators resulted in a large drop in RFP fluorescence relative to a control plasmid with no terminator (pGR). Chen et al. measured several terminators in both the forward and reverse orientation, which allowed us to identify unidirectional terminators for this study. We classified terminators as unidirectional if they have termination strength <10 in the reverse orientation and >10 in the forward direction.

Average levels of GFP and RFP fluorescence produced by plasmids carrying Chen *et. al*'s library of terminators were used to identify terminators that encode cryptic antisense promoters or destabilize mRNA. Terminators that decreased GFP expression relative to the average were assumed to destabilize mRNA. Similarly, terminators that increased RFP expression when measured in the reverse direction, were assumed to encode cryptic antisense promoters. We classified terminators that decreased GFP expression more than one standard deviation below the mean as destabilizing mRNA and terminators that increased RFP expression more than one standard deviation as encoding cryptic antisense promoters.

3-4-6. Library design and construction.

The terminator/antisense promoter library was built as described previously[320]. The library was constructed based on 52 terminators[27] and 109 constitutive promoters[26,320], paired combinatorially to produce 5,668 unique composite parts. We used 90 promoters from an existing library[26] and 19 from the Anderson promoter library on the BioBricks Registry. The terminators are naturally occurring sequences from the *E. coli* K12 genome that were previously characterized by Chen *et al.*[27] and selected to encompass a wide range of terminator strengths. The composite parts were checked for restriction sites (NotI and SbfI) and none were found. To generate the final library, all sequences were flanked by restriction enzyme sites (NotI and SbfI) and PCR primer binding sites: (1) ATATAGATGCCGTCCTAGCG and (2) AAGTATCTTTCCTGTGCCCA.

The oligonucleotide library was constructed by CustomArray, Inc. using their CMOS semiconductor technology. The library was delivered as a 1 fM oligonucleotide pool and amplified using specific PCR primers: oj1299 and oj1300 (Table 3-4). The PCR products were then digested with NotI (New England Biolabs R3189) and SbfI (New England Biolabs R3642) restriction enzymes and cleaned with DNA Clean & Concentrator columns (Zymo Research C1003). Plasmid backbones encoding repressor-protein based NOT gates (PhlF, SrpR, TarA; maps in Fig 2-9) were amplified by PCR with primers to add NotI and SbfI restriction sites to the 3'end of the repressor gene. Plasmid backbones were then digested with the same restriction enzymes and cleaned using DNA Clean & Concentrator columns. After digestion, the library inserts and plasmid backbones were ligated using T4 DNA Ligase (New England Biolabs M0208) and cloned into *E. coli* NEB10β electrocompetent cells (New England Biolabs

C3020K), resulting three libraries (PhlF, SrpR, and TarA) of ~160,000 clones each and >20 fold coverage of the designed sequence space. Each library was scraped from solid media plates and frozen at -80°C in 200 μL aliquots with 15% glycerol for subsequent analysis.

3-4-7. Library growth and fluorescence activated cell sorting (FACS).

To grow libraries for flow cytometry analysis or cell sorting, one aliquot of each library was thawed and 10μL of the sample were added to 3 mL of LB media supplemented with carbenicillin in 15mL culture tubes (Fischer Scientific 352059). Once thawed, the remaining library aliquot was discarded to avoid cell death from repeated freeze-thaw cycles. The inoculated libraries were grown for 12 h at 30°C and 250 r.p.m. in a New Brunswick Scientific Innova 44. NEB10β control strains, containing unmodified NOT-gate plasmids (Fig 3-2a) were inoculated from single colonies into 3 mL of LB supplemented with carbenicillin and also grown at 30°C and 250 r.p.m. After 12 h, both library and control strain cultures were diluted 1:200 into 25 mL of M9 medium with carbenicillin in 250mL Erlenmeyer flasks (Corning 4450-250) and grown at 37°C and 250 r.p.m. for 3 h. Next, the cultures were diluted to 0.001 OD_{600} in 25 mL of M9 medium with carbenicillin and either 0μM or 100μM IPTG. These cultures were grown for 6 h to obtain exponential phase growth. At the end of 6 h, cultures were diluted to OD_{600} ~0.05 into PBS containing 35 μg/mL chloramphenicol to arrest cell growth and protein production until sorting. Aliquots of each library were also frozen at -80°C with 15% glycerol (VWR BDH1172) to serve as 'unsorted' controls.

Cell sorting was done on a BD Biosciences FACSAria II with a blue (488 nm) laser. Each NOT gate library was sorted into four non-adjacent log-spaced bins based on YFP fluorescence. Control strains grown with 0μM and 100μM IPTG defined the upper and lower boundaries for bin placement, respectively. One million cells were sorted into the lowest fluorescence bin (Bin 1; Fig 3-10, blue), which captured 9.8-13.9% of each library. 50,000 – 200,000 cells were sorted into all other bins, which captured 0.2-4.4% of the cells in each library. After sorting, cells were plated on solid media to minimize the effect of growth rate differences on library representation. Each bin was then scraped from the solid media plates and frozen at -80°C in 200 μL aliquots with 15% glycerol (VWR BDH1172) for subsequent analysis.

3-4-8. Sorted library sequencing.

Plasmids were isolated from cells in each bin using a miniprep kit (Qiagen 1018398) by thawing one aliquot of each frozen sorted bin and using the entire sample as input to the kit. For deep sequencing, 30ng of each miniprepped sample was amplified for thirty cycles of PCR with Phusion High-Fidelity Polymerase Master Mix (New England Biolabs M0531). This PCR step added barcodes to each sample using primers oj1302, 1334 - oj1348 (Table 3-4). Amplification of samples was verified with gel electrophoresis and quantified using a NanoDrop spectrophotometer (ND-1000). Unsorted control samples were identically processed and sequenced. 13.1M constructs were sequenced in a single MiSeq 150 paired-end lane with the sequencing primers oj1301, oj1303, and oj1356 (Table 3-4). To correct for the fact that fewer cells are sorted into the later bins (BIN2-4), the samples were mixed such that the 'unsorted' and 'BIN1' samples were present in equimolar ratios and made up 90% of the final sequenced mixture. The 'BIN2', 'BIN3', and 'BIN4' samples, which were also mixed in equimolar ratios, constituted the last 10% of the final sequenced mixture. This resulted in 1.7-2.2 million sequencing reads from the each of the 'unsorted' and 'BIN1' samples and 100,000 – 180,000 sequencing reads from each of the 'BIN2', 'BIN3', and 'BIN4' samples (Table 3-1).

Amplify oligonucleotides from chip synthesized pool:

oj1299	FW	ATATAGATGCCGTCCTAGCGGCG
oj1300	RV	TGGGCACAGGAAAGATACTTCCTG
Add deep sequencing barcodes:		
oj1302	FW amplification primer:	AATGATACGGCGACCACCGAGATCTACACgtgacatT AACTAGGGCGCGGCCGC
oj1334	RV PhlF BIN1 amplification primer	CAAGCAGAAGACGGCATACGAGATGTGTGTgcttctcg ccaTGGGGGTATGGCCTGCAGG
oj1335	RV PhlF BIN2 amplification primer	CAAGCAGAAGACGGCATACGAGATAGGTGTgcttctc gccaTGGGGGTATGGCCTGCAGG
oj1336	RV PhlF BIN3 amplification primer	CAAGCAGAAGACGGCATACGAGATTCGTGTgcttctcg ccaTGGGGGTATGGCCTGCAGG
oj1337	RV PhlF BIN4 amplification primer	CAAGCAGAAGACGGCATACGAGATCAGTGTgcttctcg ccaTGGGGGTATGGCCTGCAGG
oj1338	RV PhlF Unsorted amplification primer	CAAGCAGAAGACGGCATACGAGATGTAGGTgcttctc gccaTGGGGGTATGGCCTGCAGG
oj1339	RV SrpR BIN1 amplification primer	CAAGCAGAAGACGGCATACGAGATGTGTTCgcttctcg ccaTGGGGGTATGGCCTGCAGG
oj1340	RV SrpR BIN2 amplification primer	CAAGCAGAAGACGGCATACGAGATCATGACgcttctc gccaTGGGGGTATGGCCTGCAGG
oj1341	RV SrpR BIN3 amplification primer	CAAGCAGAAGACGGCATACGAGATGTTCTCgcttctcg

		ccaTGGGGGTATGGCCTGCAGG
		CAAGCAGAAGACGGCATACGAGATAGTCTCgcttctcg
oj1342	RV SrpR BIN4 amplification primer	ccaTGGGGGTATGGCCTGCAGG
		CAAGCAGAAGACGGCATACGAGATTCTCTCgcttctcg
oj1343	RV SrpR Unsorted amplification primer	ccaTGGGGGTATGGCCTGCAGG
		CAAGCAGAAGACGGCATACGAGATTGCAGAgcttctc
oj1344	RV TarA BIN1 amplification primer	gccaTGGGGGTATGGCCTGCAGG
		CAAGCAGAAGACGGCATACGAGATGTATCAgcttctcg
oj1345	RV TarA BIN2 amplification primer	ccaTGGGGGTATGGCCTGCAGG
		CAAGCAGAAGACGGCATACGAGATTCTCCAgcttctcg
oj1346	RV TarA BIN3 amplification primer	ccaTGGGGGTATGGCCTGCAGG
		CAAGCAGAAGACGGCATACGAGATCATCCAgcttctcg
oj1347	RV TarA BIN4 amplification primer	ccaTGGGGGTATGGCCTGCAGG
		CAAGCAGAAGACGGCATACGAGATGTCACAgcttctc
oj1348	RV TarA Unsorted amplification primer	gccaTGGGGGTATGGCCTGCAGG
Sequencing primers:		
oj1301	FW seq primer:	gtgacatTAACTAGGGCGCGGCCGC
oj1303	RV seq primer:	gcttctcgccaTGGGGGTATGGCCTGCAGG
oj1356	Barcode read:	CCTGCAGGCCATACCCCCAtggcgagaagc

Table 3-4: Oligonucleotides used in this study.

3-4-9. Deep sequencing analysis.

Custom software ('IlluminaSeqAnalysis.m') was written to combine paired end reads and identify composite parts with perfect sequence identity to designed constructs. Each set of paired 150bp reads was aligned and merged into a contig based on overlapping sequence. Not1 and Sbf1 restriction enzyme sites were identified and all sequences (including adapter and constant primer sequence) outside the restriction sites were trimmed from both ends of the contig. Reads that did not pair or did not have both restriction sites were discarded since all composite parts were under 200bp, thus paired reads should have overlapping sequence and yield contgis with both restriction sites. Of the 13.1M constructs sequenced, 11.2M (85.30%) yielded paired reads with overlapping sequence and both restriction sites. Once paired, all remaining contigs with mismatches (insertions, deletions, or substitutions) to designed constructs were discarded. Of the 11.2M contigs, 3.6M (32.28%) are perfect matches to the designed library. This is consistent with the error rate of chip-based oligo synthesis[327].

Analysis of the perfect sequences shows that 70.0 – 77.5% of the composite parts appear at least once in each of the unsorted libraries (Fig 3-19, Table 3-1). When we select for library members that alter the NOT gate response functions, coverage of the library decreases to 50.2 - 74.8% in Bin 1, 38.5 – 50.0% in Bin 2, 25.0 – 26.1% in Bin 3 and 15.8 – 25.3% in Bin 4 (Fig 3-

19). This is expected since a limited subset of the constructs will be capable of shifting the gate thresholds. Indeed, >50% of the composite parts encode promoters weaker than apFAB49, which generated less than a two-fold change in RFP expression in our original experiments (Fig 3-4).

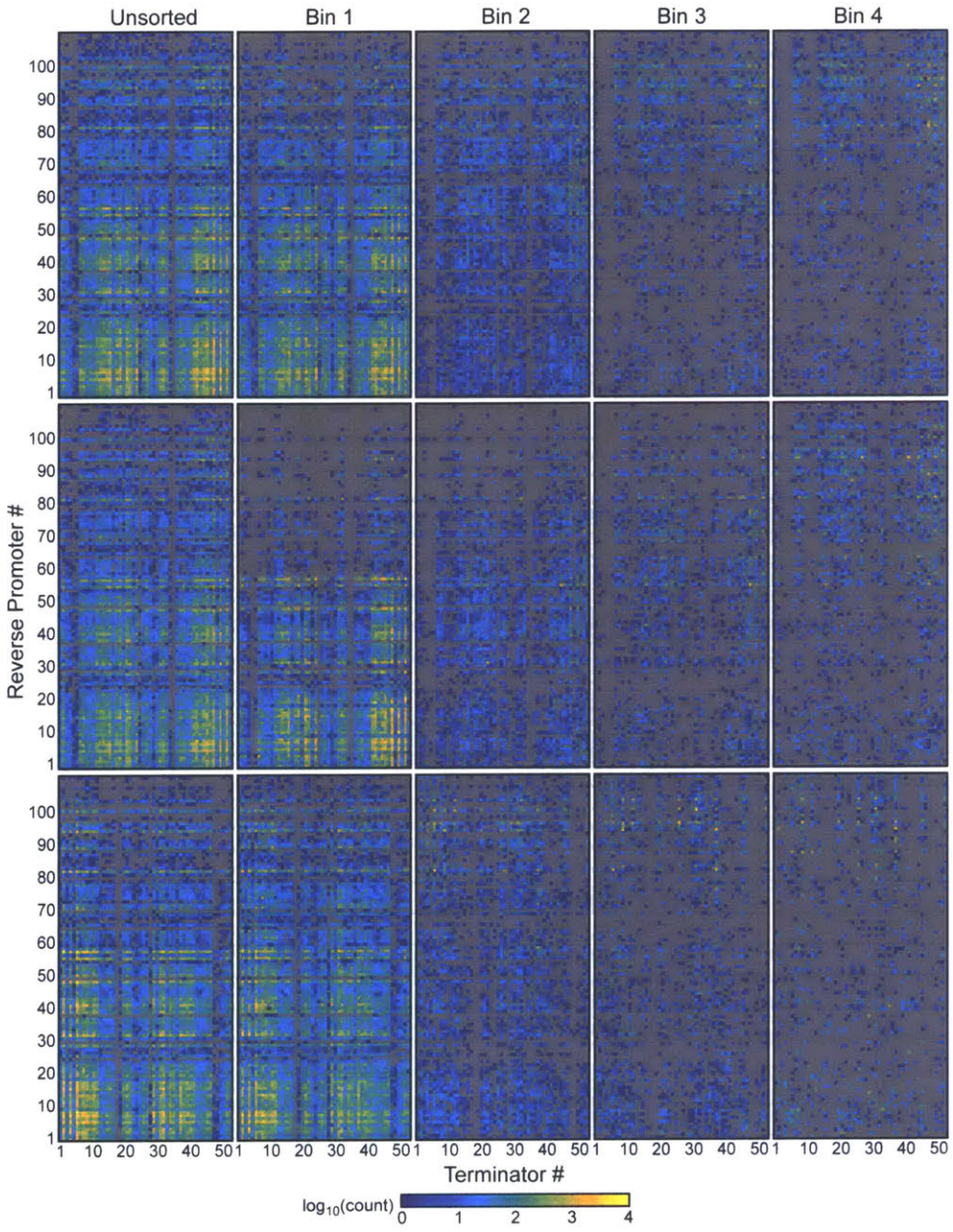

Figure 3-19: Composite part frequencies in each library by bin.
Sequencing read counts for each composite part in the PhlF (top), SrpR (middle), and TarA (bottom) libraries. Only reads that are perfect matches to the designed sequences are shown. Promoter and terminator ordering is the same for all heatmaps. Promoters are ordered by strength from 1 (weakest) to 109 (strongest). Reference spreadsheet is Table 3-5.

Terminator number	Terminator Name	Promoter number	Promoter Name
1	ECK120010843-R	1	apFAB46
2	ECK120010802	2	apFAB101
3	ECK120010813-R	3	apFAB96
4	ECK120010780	4	apFAB47
5	ECK120029530-R	5	apFAB42
6	ECK120015446-R	6	apFAB95
7	BBa_B0051	7	apFAB36
8	ECK120030672	8	apFAB68
9	ECK120035132	9	apFAB31
10	ECK120015439-R	10	apFAB93
11	ECK120026315-R	11	apFAB54
12	ECK120010831	12	apFAB92
13	ECK120010841-R	13	apFAB52
14	ECK120020528	14	apFAB81
15	ECK120010796	15	apFAB71
16	ECK120010874	16	apFAB67
17	ECK120020622	17	apFAB70
18	ECK120010857	18	apFAB79
19	ECK120030802	19	apFAB100
20	ECK120021270	20	apFAB45
21	ECK120011170-R	21	apFAB61
22	ECK120010852	22	apFAB341
23	ECK120020525	23	apFAB80
24	ECK120010832-R	24	apFAB29
25	ECK120010833	25	apFAB32
26	ECK120010812-R	26	apFAB40
27	ECK120010836-R	27	apFAB30
28	ECK120010871	28	apFAB56
29	ECK120010806	29	j23119
30	ECK120030221	30	apFAB50
31	ECK120010825	31	apFAB85
32	ECK120015457-R	32	apFAB53
33	ECK120010840	33	apFAB318
34	ECK120010781	34	apFAB44
35	ECK120010864	35	apFAB76
36	ECK120010793	36	apFAB65
37	ECK120010856	37	apFAB140
38	ECK120010867	38	apFAB66
39	ECK120010815	39	apFAB75
40	ECK120010863	40	apFAB345
41	ECK120010782	41	apFAB103
42	ECK120016882	42	apFAB347
43	ECK120023928	43	apFAB39
44	tonB/P14	44	apFAB94
45	ECK120035133	45	apFAB323
46	ECK120010855	46	apFAB97
47	ECK120017009	47	apFAB78
48	ECK120015170	48	apFAB57
49	ECK120026481-R	49	apFAB72

50	BBa_B0062	50	apFAB62
51	ECK120010858-R	51	apFAB48
52	ECK120034435	52	apFAB55
		53	j23100
		54	apFAB77
		55	j23104
		56	apFAB33
		57	apFAB339
		58	apFAB321
		59	apFAB115
		60	apFAB125
		61	apFAB82
		62	j23102
		63	apFAB346
		64	apFAB317
		65	apFAB63
		66	apFAB64
		67	apFAB322
		68	apFAB150
		69	apFAB69
		70	j23101
		71	apFAB104
		72	apFAB89
		73	apFAB306
		74	apFAB338
		75	apFAB58
		76	apFAB302
		77	apFAB73
		78	j23118
		79	apFAB342
		80	apFAB87
		81	j23106
		82	apFAB49
		83	apFAB38
		84	apFAB130
		85	apFAB98
		86	apFAB312
		87	apFAB129
		88	apFAB311
		89	j23107
		90	j23105
		91	apFAB277
		92	apFAB300
		93	j23110
		94	j23116
		95	j23114
		96	j23108
		97	apFAB259
		98	j23115
		99	j23109
		100	apFAB139

		101	j23113
		102	j23112
		103	apFAB134
		104	j23103
		105	apFAB114
		106	j23117
		107	apFAB109
		108	apFAB149
		109	apFAB124

Table 3-5: Promoter and Terminator ordering for Figure 3-28.

3-4-10. Sorted-parts strength analysis.

Custom software ('IlluminaPerfSeqAnalysis.m') was written to analyze the perfect oligonucleotides, *i.e.* sequences that are perfect matches to the designed library, sorted into each bin. Analysis relied on previously measured terminator[27] and promoter [26,320] strengths, therefore all sequences with mutations were disregarded because they could change a part's activity and convolute the analysis. Occurrences of each promoter and terminator were counted per bin and used to calculate the median promoter, forward and reverse terminator strengths for each bin.

3-4-11. Enrichment calculation.

To calculate enrichment for each composite part, we normalized the counts of each composite part in a bin to the total number of perfect sequences in that bin. We defined the frequency of a composite part f_{ijx} in a bin as

$$f_{ijx} = \frac{c_{ijx}}{\sum_i c_{ijx}}$$

where c_{ij} is the number of occurrences of composite part i in bin j for library x, where x = PhlF, SrpR or TarA. Then we defined enrichment E_{ijx} as the ratio of the frequencies of a composite part i in a sorted bin j to the frequency of that composite part in the unsorted pool (f_{iux}).

$$E_{ijx} = \frac{f_{ijx}}{f_{iux}}$$

If a composite part did not appear in the unsorted library at least once ($f_{iux} = 0$), c_{iux} was set to one, indicating one count of the part in the unsorted library. This correction was used to ensure that none of the enrichments were infinite. Enrichment E_{ijx} for each composite part was then

averaged by bin across all three libraries. We defined the average enrichment $\overline{E_{ij}}$ for a composite part as

$$\overline{E_{ij}} = \frac{\sum_x E_{ijx}}{N}$$

where $N = 3$, the total number of libraries. To ensure that composite part behavior is consistent across all three libraries, any composite parts that did not appear in all three libraries for a given bin were assigned an enrichment of zero for that bin, *i.e.* if $E_{ijx} = 0$ for any x, $\overline{E_{ij}} = 0$.

Next each composite part was assigned to the bin where its average enrichment was highest. Maximum average enrichment *Emax* for each composite part *i* was calculated as

$$Emax_i = \max_{1 \le j \le 4} (\overline{E}_{ij})$$

Then the composite part *i* is assigned to the bin *j* where $\overline{E_{ij}} = Emax_i$. If the maximum enrichment is less than one, the composite part is depleted in the sorted library and is not assigned to a bin (Fig 3-16, grey). Depletion of a composite part in all of the sorted bins relative to the unsorted pool may be the result of biases in cell recovery after sorting or in amplification of the DNA for deep sequencing. The matrix of bin assignments was generated using custom software ('IlluminaEnrichmentGrid.m') and used to create Fig 3-16.

3-4-12. Measurement of growth curves.

E. coli strains were grown for 16 h in LB media containing antibiotics, when appropriate, in 96-deep well blocks (USA Scientific 1896-2000) at 37°C and 250 r.p.m in an INFORS-HT Multitron Pro. After 16 h, the cultures were diluted 1:200 into M9 medium with antibiotics and grown for 3 h with the same shaking and temperature settings as the overnight growth. Next, the cultures were diluted to $OD_{600} = 0.001$ into fresh M9 medium with antibiotics and 100μM IPTG. 150 μL of these cultures were grown in black 96-well optical bottom plates (Thermo scientific 165305) at 37°C and 1mm orbital shaking in a BioTek Synergy H1 plate reader. Optical density measurements at 600nm wavelength (OD_{600}) were made every 20 minutes for 12 h.

3-4-13. Construction and testing of the transcriptional interference model.

Custom MATLAB software was written to solve the model ODEs (Equations 3-7 & 3-8) with mixed boundary conditions. Initial mesh for the MATLAB boundary solver bvp4c were formed using MATLAB function bvpinit with general solutions for the model ODEs derived in Wolfram Alpha's Mathematica:

$$C_F(x) = \frac{C_1}{\varepsilon_R + C_2 e^{C_1 x}} \tag{3-7}$$

$$C_R(x) = \frac{C_1}{\varepsilon_F}\left(\frac{C_2 e^{C_1 x}}{\varepsilon_R + C_2 e^{C_1 x}}\right) \tag{3-8}$$

Integration constants were approximated using boundary conditions $C_F(x=0) = \phi_F/v$ and $C_R(x=N) = \phi_R/v$. ε_F and ε_R were input directly into the model for $\varepsilon_F = \varepsilon_R = 1$ or parameter sweep experiments. Model results were reported as polymerase concentrations $C_F(x)$ and $C_R(x)$. Full-length transcript production is assumed to be proportional to $C_F|_{\varepsilon_F=0}$, which should be a measure of polymerases fired from P_F that successfully transcribe the entire stretch of DNA between promoters. Fold repression θ is calculated using Eq. 3-6, which compares C_F in the absence of interference ($C_F|_{\varepsilon_F=0}$) to C_F with an interfering promoter ($C_F|_{\varepsilon_F>0}$). Model results for each forward/interfering promoter pair were scored by simple comparison to experimental data:

$$s = [\mathrm{abs}(\theta_{TI\_m} - \theta_{TI\_p})]^{-1} \tag{3-9}$$

where s is the score for a specific promoter pairing, θ_{TI_m} and and θ_{TI_p} are measured and predicted repression, respectively, for that pair. The best ε_F and ε_R values were calculated for each forward/interfering promoter pair using a weighted average, were each ε was weighted by its score.

3-4-15. Data availability

Brophy, JAN, Voigt, CA (2015). Antisense library. NCBI Sequence Read Archive SRP065456

Chapter 4: Stable Engineering of Undomesticated Bacteria using a Miniaturized Integrative Conjugative Element (ICE)

Engineering probiotics to optimize human or plant health will require robust bacteria that are capable of surviving in harsh, competitive environments. Here, we present an Integrative Conjugative Element from *Bacillus subtilis* (ICE*Bs1*) as a tool to deliver synthetic programs to diverse collections of undomesticated bacteria isolated from valuable microbiomes. We demonstrate that ICE*Bs1* can be used to deliver synthetic programs to >X species of Gram-positive bacteria isolated from the human gut and soil microbiomes (*e.g.*, *Bacillus sp.*, *Enterococcus sp.*, *Staphylococcus sp.*). We develop a miniature ICE, which irreversibly integrates DNA into recipient bacteria's genomes, to deliver therapeutically relevant biosynthetic pathways (*e.g.*, nitrogen fixation, indole-3-acetic acid production) to the bacteria. We measure the performance of these pathways across species and demonstrate that species variation can be used to optimize the desired phenotypes. This work produces an easy to use tool for stably engineering diverse species of undomesticated bacteria.

4-1. Introduction

Advancements in microbiome analysis techniques have revealed the importance of microbial communities for human, plant, and animal health[360]. The bacteria that inhabit these ecosystems help maintain fitness by producing molecules that eliminate pathogens[12,13,34], function as nutrients[35] or hormones[36–38], or educate the immune system[361]. Importantly, the strains of bacteria that have essential roles in human, plant, and animal health produce their effect while competing with tens to thousands of other bacterial species. Deep sequencing studies find that the human gut, oral, and skin microflora are composed of 500 - 25,000 species worldwide[362]. Similarly, a gram of soil may contain 2,000 – 8.3M microorganisms representing 5,000 – 20,000 different operational taxonomic units (OTU)[363]. In addition to being diverse, bacteria in these environments are highly variable. The most prevalent bacteria vary significantly from sample to

sample, indicating a great flexibility in the combinations of species that make up a functional microbiome[364,365].

Bacteria's ability to alter human, plant, and animal health has inspired researchers to create designer probiotics that leverage genetic engineering to optimize the fitness of these systems. These efforts have largely been focused on moving simple phenotypes, such as therapeutic protein secretion, into a few well-established probiotic strains (*e.g., Lactococus lacis* and *Escherichia coli Nissle* 1917 (EcN))[366]. These strains, though effective for specific applications, such as secreting interleukin 10 (IL-10) to treat irritable bowl disease (IBD)[367], have limited potential in applications that require different colonization patterns, immune interactions, or other innate behaviors. Consequently, researchers are developing toolkits for new strains that are well adapted for specific applications, such as long-term human gut colonization[368].

However, broadly applicable tools that can be used to engineer multiple species are lacking. These tools could be used to accelerate probiotic development by eliminating the need to generate specific toolsets for each new organism of interest. They could also impact industrial bio-production by allowing metabolic engineers to work with bacteria that have advantageous features, such as solvent tolerance, leading to increased yields[369] or streamlined manufacturing[370]. Unfortunately, even the basic tools required for genetic engineering, such as transformation methods, growth medias, and functional plasmids, are underdeveloped or unknown for the vast majority of bacterial species worldwide. Transformation is especially difficult for Gram-positive bacteria, which are often electroporated despite poor performance[30,371]. Electroporation creates stable tears in the peptidoglycan wall of Gram-positive bacteria, which leads to cell death and transformation efficiencies at least four orders of magnitude worse than commercially available Gram-negative cloning strains[372,373]. These low transformation efficiencies (10^2 - 10^6 transformants/μg DNA) prohibit library construction and deter high throughput engineering in Gram-positive bacteria. Vectors for introducing new DNA to bacteria, such as plasmids, are also unknown for most bacterial species. Plasmids often have narrow host ranges[374] and can vary in copy number/stability[375], which make them difficult to use without extensive species-specific characterization. Additionally, plasmids are prone to horizontal gene transfer and often require constant selection for maintenance[376], which make them inappropriate for *in vivo* applications

where a synthetic program can be unintentionally lost or spread to other bacteria in the target environment[377].

Integrative conjugative elements (ICEs) have great potential as tools for engineering diverse bacteria. ICEs are unique mobile genetic elements (MGE) that reside in bacterial genomes and encode all of the proteins needed to excise from the chromosome, replicate, conjugate into a neighboring cell, and integrate into the recipient's genome (Fig 4-1). ICEs are found across the bacterial domain and a single element can often transfer into a diverse range of recipient species[378]. Tn*916*, an ICE first identified in *Enterococcus faecalis*, appears in tetracycline resistant clinical isolates, as well as unrelated species of soil and skin bacteria[379]. Similarly, ICE*Bs1*, an ICE from the Gram-positive model organism *Bacillus subtilis*, was shown to conjugate into four different bacterial species in a laboratory setting (*B. subtilis, B. anthracis, B. licheniformis,* and *L. monocytogenes*)[380]. The success of these elements across species and their ability to integrate into the host genome make them attractive candidate tools for introducing DNA to a broad range of bacterial species for advanced engineering applications.

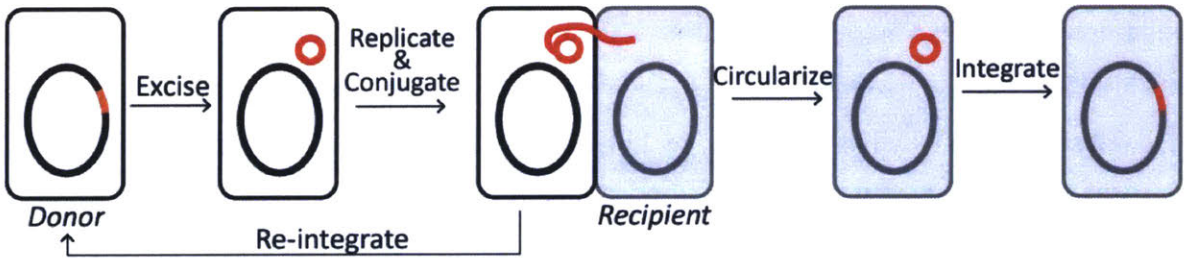

Figure 4-1: Typical Integrative Conjugative Element (ICE) life cycle.
ICEs reside in bacterial chromosomes bound by specific sequences on the right (*attR*) and left (*attL*) ends. When stimulated, ICEs will excise from the chromosome to form a covalently closed circle. Circularized ICE will then be nicked by a relaxase protein and will replicate via a rolling circle mechanism. If an ICE-cell is in contact with the donor, a single strand of ICE DNA will be transferred to the recipient during replication. Once in the recipient, the ICE will circularize and the complementary strand will be synthesized to regenerate double-stranded ICE. This circularized form will then integrate into the host chromosome via a recombination event between *attP* (in the ICE) and *attB* (in the chromosome).

Natural ICEs (natty ICEs) carry a wide range of useful phenotypes between species. Several natty ICEs facilitate bacteria survival by moving antibiotic[381] and heavy metal resistance cassettes[382] into new species. Other natty ICEs carry more complex traits, such as the ability to colonize a eukaryotic host[383] or promote virulence and biofilm formation[384]. ICE*Ml*Sym^{R7A} is one

notable example that encodes 0.5 Mb of pathways sufficient for transferring plant symbiosis, including root nodule formation and nitrogen fixation, between cells[383].

ICE*Bs1*, the integrative conjugative element from *Bacillus subtilis*, is unusual because it conjugates at an extremely high efficiency (~0.2 *Bacillus subtilis* transconjugates/donor) compared to other Gram-positive conjugation systems (*e.g.*, $\sim 10^{-5}$ transconjugates/donor (Tn*916*), 10^{-7} – 10^{-8} transconjugates/donor (pRK212)[385]) (Fig 4-2). During growth, ICE*Bs1*-encoded repressor protein ImmR prohibits transcription of the ICE conjugation proteins. However, when ICE+ *B. subtilis* encounter DNA damage or a high density of ICE- cells, anti-repressor protein ImmA cleaves ImmR and stimulates expression of the ICE conjugation and excision proteins[386]. Excision is mediated by tyrosine recombinase Int and directionality factor Xis[387]. After excision, ICE circularizes and the relaxase protein NicK nicks ICE*Bs1* DNA at the origin of transfer oriT[388]. This is followed by rolling circle replication, facilitated by the ICE*Bs1*-encoded helicase HelP[389]. During replication, the relaxase associates with the coupling protein (putatively ConQ) and a single strand of ICE*Bs1* DNA is conjugated into a neighboring cell. The structure of the conjugation apparatus is unsolved, however ATPase ConE, peptidase/muramidase CwlT, and putative transmembrane proteins ConB, ConC, ConD and ConG are known to be essential for conjugation[390]. Once in the recipient cell, ICE re-circularizes, its second strand of DNA is synthesized by host factors from a single stranded origin (sso)[391]. In the recipient, Int and ImmR are expressed to integrate ICE into the recipient chromosome and repress further conjugation/superinfection. Benefical phenotypes encoded in ICE*Bs1* are currently unknown, however the high conjugation efficiency, potentially broad host range, and integrative phenotype make it an attractive tool to deliver DNA to diverse bacterial species.

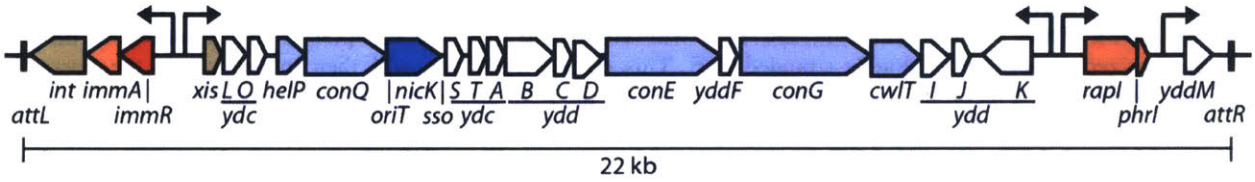

Figure 4-2: Schematic of wild type ICE*Bs1*.
Proteins are color coded by function: regulatory proteins, red; integration/excision proteins, beige; conjugation proteins, blue; unknown, white.

Here, we develop ICE*Bs1* as a tool for engineering a diverse range of undomesticated bacteria. First, we demonstrate control over ICE*Bs1* conjugation by creating a miniature element that is incapable of self-transfer. Second, we broaden the known host range of ICE*Bs1* by compiling a collection of 82 strains of Gram-positive bacteria isolated from several important environments (*e.g.*, human gut, human skin, organic fertilizer, crop rhizospheres). Using the collection, we show that a miniature ICE can be used to introduce heterologous DNA into 16 different bacterial species from 5 different genera. Finally, we utilize miniature ICE to deliver a functional heterologous nitrogen fixation pathway to the soil-dwelling species. These engineered soil bacteria should promote plant growth by delivering bioavailable nitrogen to plant roots. This work contributes to a larger effort to facilitate engineering of real-world applicable organisms for the transition of genetically modified probiotics from lab concepts to useable and effective products.

4-2. RESULTS

4-2-1. Engineering control over ICE*Bs1* conjugation: small molecule inducers

Controlling ICE*Bs1* conjugation will make it easier to use ICEs as an engineering tools by allowing researchers to dictate the conditions under which DNA is delivered to recipient species. ICE*Bs1* conjugation is stimulated by high concentrations of ICE- *B. subtilis* cells and DNA damage[380]. Over expression of the quorum sensing protein RapI using a heterologous inducible promoter can also cause ICE*Bs1* excision and conjugation[380]. We carefully tested several different RapI induction cassettes to identify the inducible promoter that gave us low uninduced and high induced conjugation rates. For these experiments, we used a *ΔrapIphrI* ICE*Bs1* and over expressed RapI using inducible promoters integrated at *amyE* (Fig 4-3).

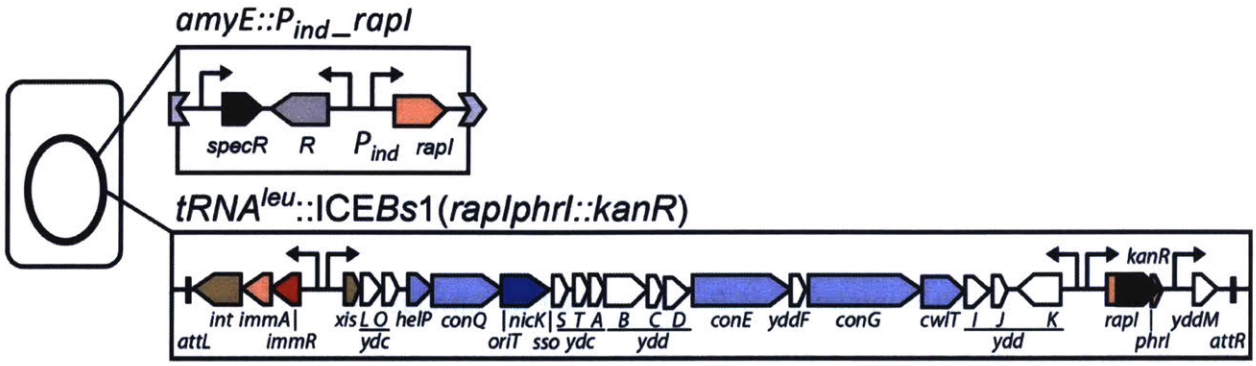

Figure 4-3: Schematic for inducible ICE*Bs1* strains.
Quorum sensing proteins RapI and PhrI are removed from wild type ICE*Bs1*. RapI induction cassettes are integrated at *amyE* in the *B. subtilis* chromosome with a spectinomycin cassette. Here *R* and P_{ind} are generic repressor and inducible promoter and the dark circle represents the B. subtilis chromosome with lines pointing to approximate locations of ICE*Bs1* and *amyE*.

First, we expressed RapI using isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible promoters P_{spank} and $P_{spank(hy)}$ (Fig 4-4a). IPTG is a non-hydrolyzable molecular mimic of allolactose that is used to induce expression of genes under control of lac repressor (LacI). The $P_{spank(hy)}$ promoter is ~10 fold stronger than P_{spank} in *B. subtilis* (Fig 4-4a). Full induction of P_{spank} and $P_{spank(hy)}$ with 1mM IPTG lead to high levels of conjugation (~0.2 transconjugates/donor) (Fig 4-4c). However, uninduced (0mM IPTG) conjugation frequencies were at least 50 fold higher than the *ΔrapI* control.

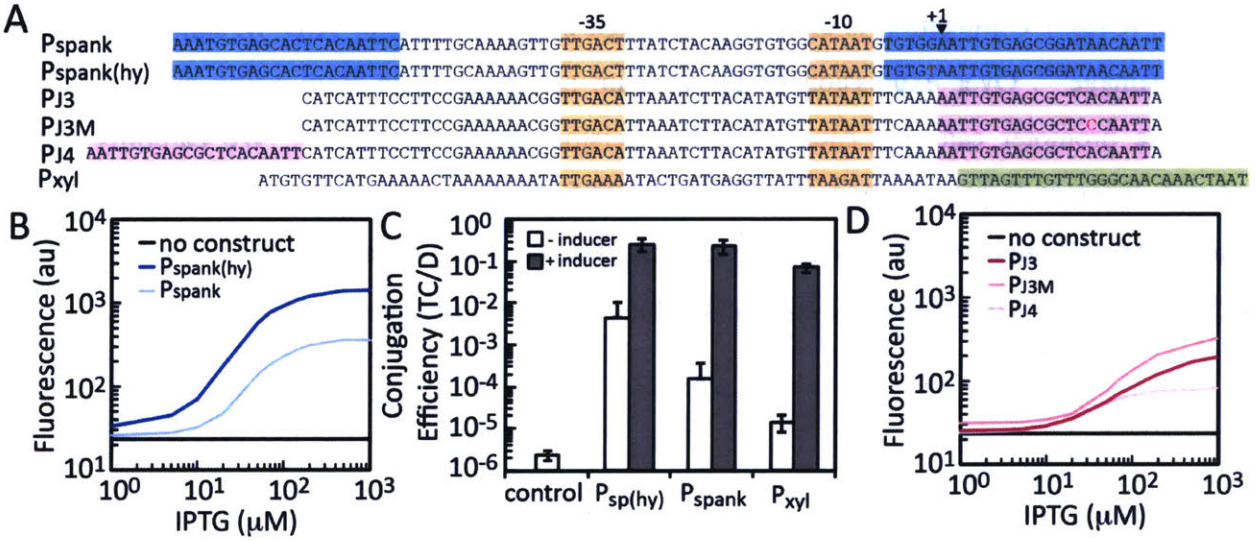

Figure 4-4: Inducible *B. subtilis* promoters for RapI expression/conjugation.
(A) Promoter sequences. Lac operators shown in blue (wt) or pink (synthetic inverted repeats), xylose operator in green, -10 and -35 recognition sequences in orange. Transcription start site (+1) indicated with an arrow. (B) P_{spank} and $P_{spank(hy)}$ transfer functions measured using GFPmut2. (C) Conjugation efficiencies

+/- 1mM IPTG for $P_{spank(hy)}$ (JMA168) and P_{spank} (JAB221) and +/- 2% xylose for P_{xyl} (JAB785). Control strain is *ΔrapI* (JMA8). (D) New IPTG-inducible promoter transfer functions measured using GFPmut2.

To reduce leaky expression of RapI, we attempted to build new IPTG-inducible promoters by replacing the wild type lac operators with perfect inverted repeats ($lacO^{syn}$) (Fig 4-4a, pJ3, pJ3m, pJ4). These synthetic lac operators have been shown to produce tighter OFF states in *Escherichia coli* (data not shown). We added these lac operators to a strong constitutive promoter from *Bacillus licheniformis* P_{pen}[392]. Unfortunately, our best inducible promoters did not provide substantial improvement in leaky expression when compared to the P_{spank} promoter (Fig 4-4d).

Next, we tried using xylose inducible promoter P_{xyl} to express RapI. P_{xyl} is a modified version of the *xylA* promoter from *Bacillus subtilis* subsp. *spizizenii* W23 (Fig 4-4a)[393]. This promoter is catabolite repressed in rich media, which reduces leaky expression[393]. Unfortunately, catabolite repression can also inhibit full induction. We quantified xylose-inducible conjugation in rich media (LB) using a *P_{xyl}_rapI* cassette and found ~4,500 fold induction of conjugation (Fig 4-4b). P_{xyl} results in a much tighter OFF state than the IPTG-inducible promoters (>10 fold less leaky than P_{spank}) with only ~3.5 fold reduction in transconjugates/donor when fully induced. Thus we chose to use P_{xyl} to drive expression of RapI and control ICE*Bs1* conjugation.

4-2-2. Engineering control over ICE*Bs1* conjugation: miniature elements

We gained further control over conjugation by building miniature ICEs (mICEs) that are incapable of self-transfer. These ICEs are designed to conjugate once (from *B. subtilis* donor into neighboring recipient) and then become stuck in the recipient cell's genome. mICEs can be when used for stabling engineering target strains and should reduce the unintended spread of synthetic DNA through microbial populations. mICEs do not encode all of the proteins necessary to excise and conjugate themselves into neighboring cells. Instead, the proteins that mediate ICE*Bs1* excision and conjugation are expressed ectopically from the threonine synthase locus *thrC* (Fig 4-5a). These genes, collectively referred to as locked-in ICE*Bs1* (liICE), complement the missing components of mICE in the *B. subtilis* donor strain[388]. Consequently, mICE can conjugate from a *B. subtilis* strain containing liICE into any recipient cell, but cannot conjugate further.

We built two different mICE/liICE systems. In the first system (1.0) (Fig 4-5a), liICE encodes all of the essential conjugation proteins except the relaxase NicK. The *nicK* gene was not included in liICE 1.0 because its coding region includes the origin of transfer *oriT*. During a normal ICE life cycle (Fig 4-1), the relaxase nicks the ICE DNA at *oriT* after it excises from the chromosome to initiate replication and transfer. However, when *oriT* is suck in the genome, nicking leads to single-stranded breaks and donor cell death (data not shown). Thus, NicK and oriT were encoded in mICE 1.0 instead of liICE (Fig 4-5a). All genes upstream of *nicK* in wild type ICE*Bs1* were kept in mICE 1.0 to maintain the original context for expression of NicK. In total, mICE 1.0 is 8.7 kb (reduced from 22 kb of wild type ICE*Bs1*) and its conjugation efficiency of is ~50 fold less than wild type ICE*Bs1* under identical conditions (Fig 4-5b). This drop in efficiency is probably due to the elimination of secondary transfer, which is known to occur for wild type ICE*Bs1* in chains of *B. subtilis* cells[394], and not a deficiency expression of ICE conjugation proteins.

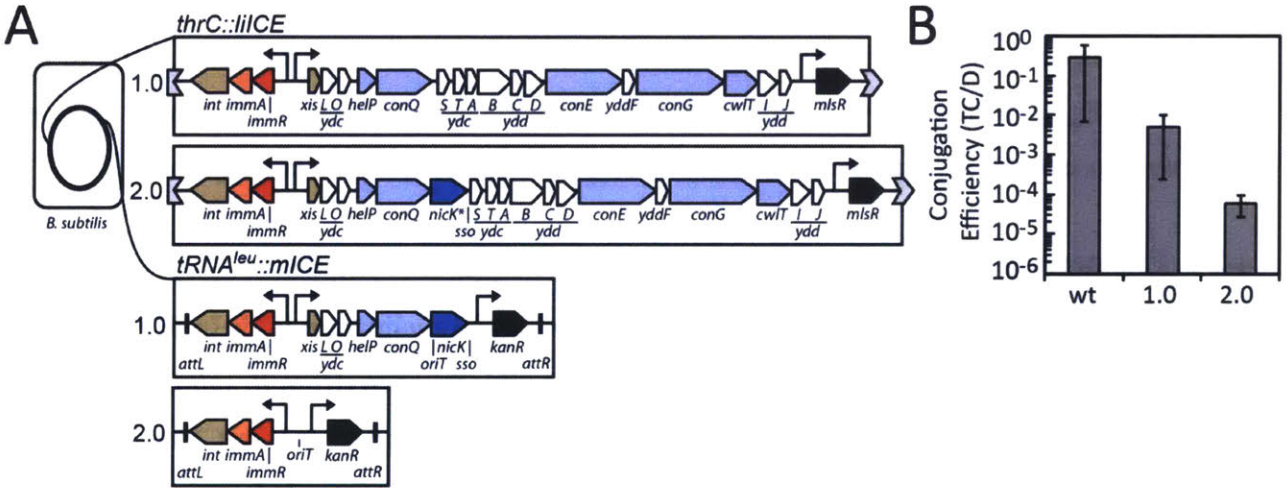

Figure 4-5: Miniature ICE schematics and conjugation efficiencies.
(A) Schematic for mICE and liICE versions 1.0 and 2.0 Here the dark circle represents the *B. subtilis* chromosome with lines pointing to approximate locations of the liICE/mICE. (C) Conjugation efficiencies for mICE 1.0, 2.0, and wild type ICE*Bs1*.

In the second system (2.0) (Fig 4-5a), liICE includes a recoded version of *nicK* that does not contain *oriT*. mICE 2.0 is reduced to 2.4 kb and only encodes ImmR, to silence P_{int} in the recipient, and parts that are essential for transfer: *oriT* for recognition by the relaxase, Int for integration in the recipient cell, and kanR for selection of transconjugates. This tiny ICE is

approximately 3.5 orders of magnitude less efficient than wild type ICE*Bs1* (Fig 4-5b). The severe drop in efficiency is likely attributable to the lack of ICE components that enable replication in the recipient, such as NicK, the sso, and HelP. ICE DNA can get lost during cell division if it is unable to replicate, thereby reducing the conjugation efficiency. Truncations of important, yet ill defined, sequences such as attL, attR, or oriT, may also contribute to the poor conjugation efficiency of mICE 2.0. Int and Xis may recognize more than the 60 bp att sites in mICE 2.0 to catalyze excision, circularization, or integration of ICE*Bs1*. Similarly, the 100 bp fragment of *nicK* that we used as oriT may be insufficient for NicK recognition/cleavage. Additional mICE variants will need to be constructed to test these hypotheses and identify the minimum number of components needed for efficient conjugation.

4-2-3. High throughput mating conditions

One goal of this project is to find conjugation conditions that enable efficient DNA delivery. ICE*Bs1* conjugations are typically performed using cells grown in large volume flasks that are concentrated onto analytical filters and incubated for 3 hours on agar plates (Fig 4-6a, filters). These mating conditions simulate biofilms and increase cell-cell contact for maximal mating efficiency. However, the protocol is not ideal for high throughput screening. Several steps in the mating protocol, including incubation time, media, temperature and donor/recipient growth phases, could be tuned to increase throughput and alter mating efficiencies. We compared conjugation efficiencies using different medias for the solid agar support in filter matings, *e.g.*, Spizizen minimal media salts[395], TSS salts[396], +/- magnesium chloride ($MgCl_2$). Then we used the best medias to compare conjugation techniques, *e.g.*, liquid, spots, filters, with different potentials for high-throughput adaptation (Fig 4-6a).

While screening medias, we confirmed previous results that $MgCl_2$ increases conjugation efficiency (Fig 4-6b)[397]. We compared TSS + $MgCl_2$ to Spiz + $MgCl_2$ and found similar conjugation efficiencies with the two salt solutions (Fig 4-6b). Although the previous study hypothesized that $MgCl_2$ increases ICE*Bs1* conjugation efficiency by affecting the activity of cell surface components involved in conjugation or stabilizing mating pairs, it is possible that $MgCl_2$ increases conjugation efficiency by neutralizing unfavorable interactions between the DNA (negatively charged) and charged portions of macromolecules on the bacteria's outer surface.

This is the mechanism by which another multivalent cations (calcium) increase transformation efficiency[398]. Thus, we tested a conjugation null (*ΔconG*) strain to make sure that the $MgCl_2$ mediated increase in conjugation efficiency was not the result of increased donor cell lysis and natural competency. Based on our results, $MgCl_2$ specifically increases conjugation efficiency and does not facilitate recipient acquisition of ICE*Bs1* DNA via donor cell lysis and natural competency (Fig 4-6b).

To streamline ICE*Bs1* conjugation, we sought to eliminate filters and reduce the volume of cell culture needed for matings by conjugating in spot or liquid conditions. Spot matings are commonly for Gram-negative conjugations and, like filter matings, spots increase donor-recipient contacts by concentrating a large number of cells into a small area. Spot matings were performed by concentrating 150 μL each of donor and recipient cell culture ten-fold and spotting the dense cell mixture onto agar plates. Spots were allowed to dry before incubation. Liquid matings were performed by mixing 150 μL of donor and recipient cell culture. The cells were either allowed to conjugate in LB, or were harvested and resuspended in an equivalent volume of Spiz + $MgCl_2$. For all comparisons, matings were incubated for 3 h at 37°C.

Spot matings were just as efficient as filter matings, whereas liquid matings were >1,000 fold less efficient than filter matings (Fig 4-6c). These findings are unsurprising since matings require direct contact between donor and recipient cells. Spots, like filters, concentrate cells into a small area and yield high conjugation efficiencies by increasing cell-cell contact. In liquid, cells are too disperse for donor strains to encounter as many recipients. To fix this 'large volume' problem, *Enterococcus faecalis* conjugative plasmid pCF10 encodes a surface protein called aggregation substance (PrgB) during conjugation that increases cell stickiness and helps donors bind to recipient cells to increase conjugation efficiency[399]. As far as we know, ICE*Bs1* does not encode a similar surface protein, therefore liquid matings, which likely decrease cell-cell contacts are less efficient than spot or filter matings. Although spot matings were only X fold less efficient than filter matings, the hit in conjugation efficiency means that low frequency conjugation events (such as those between B. subtilis and unrelated recipient strains) may be missed. Thus we will use filter matings to define a baseline for conjugation into diverse recipient

species. Once conjugation is established, we can use spot matings for more economical engineering.

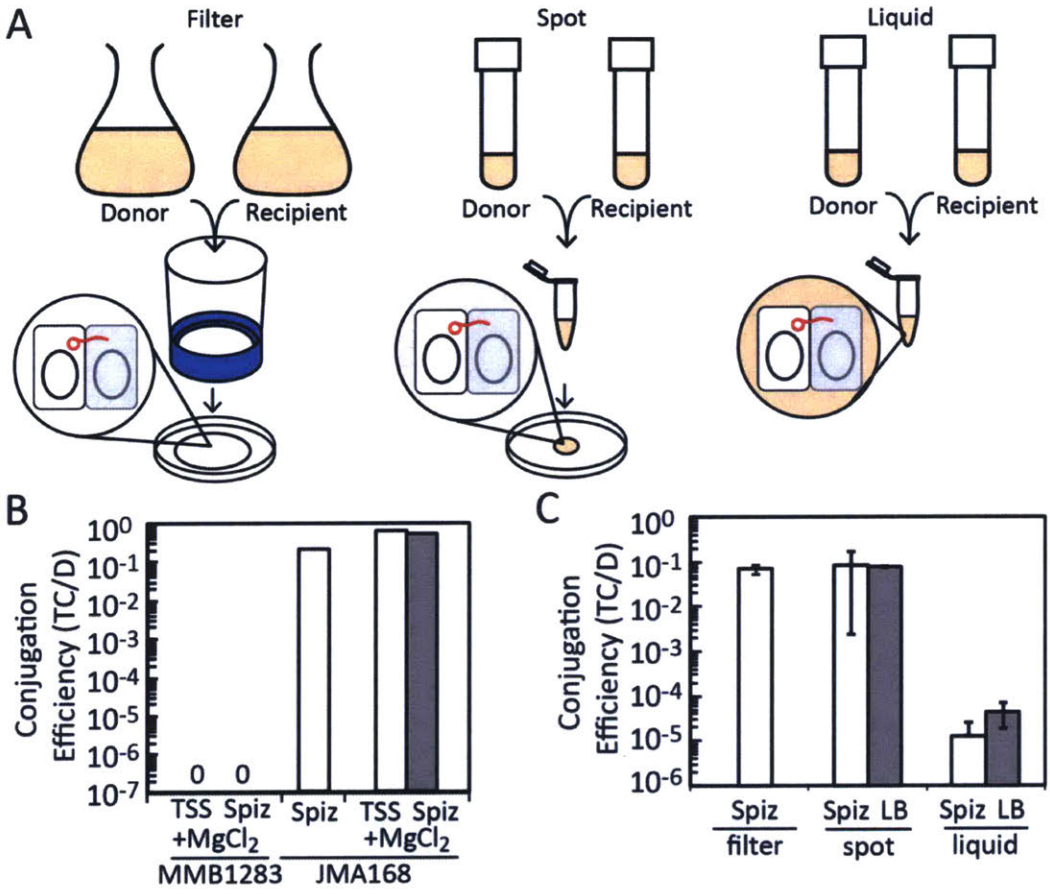

Figure 4-6: High throughput conjugation methods optimization.
(A) Schematics for each conjugation method. (B) The effects of different salt solutions (Spiz or TSS +/- 125 mM $MgCl_2$) in the solid agar support were compared using filter matings. (C) Conjugation method variations. Liquid and spot matings were compared to filter matings with LB or Spiz + 125 mM $MgCl_2$. For all assays, donor strain is JAB785 and recipient is JAB299.

4-2-4. D-alanine auxotrophy as a counter selection to streamline transconjugate isolation

Isolating transconjugates requires a unique selectable marker in the recipient bacterium that is used to kill donor cells after mating (Fig 4-7a). Typically researchers rely on identifying a unique naturally-occurring antibiotic resistance in the recipient to select for transconjugates. If no suitable antibiotic resistance can be found, resistance to spectinomycin can be generated by plating a large number of cells on spectinomycin selective plates. Specinomycin resistance frequently arises via mutations in the 16S rRNA and resistant strains can often be isolated

quickly[400,401]. If no spectinomycin resistant recipients can be created, medias or growth temperatures that are inhospitable to the donor strain can also be used to isolate transconjugates after mating.

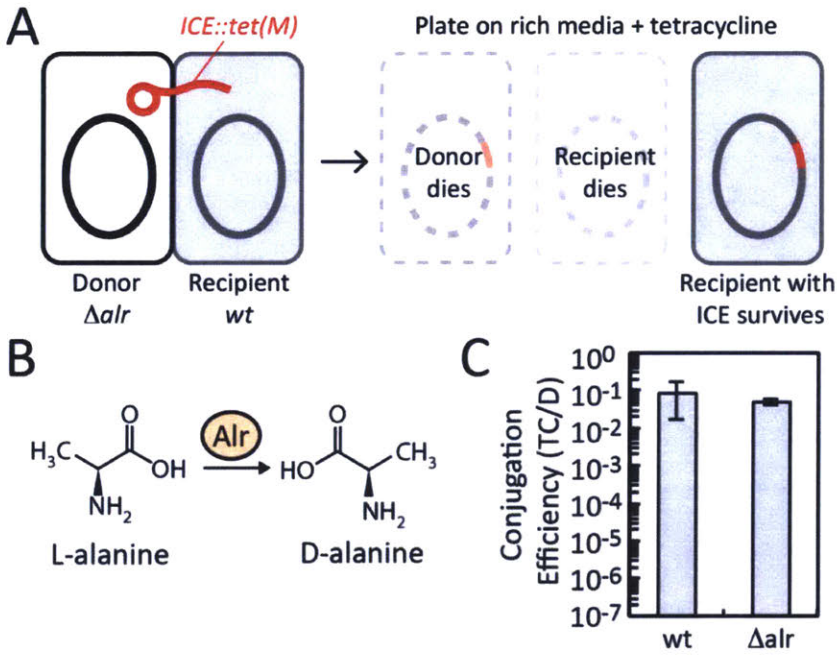

Figure 4-7: Transconjugate isolation using a D-alanine auxotrophic donor.
(A) Typical mating selection scheme. Donor is D-alanine auxotroph (*Δalr*), recipient is wild type, ICE carries a tetracycline resistance marker (*tet(M)*). After mating, cells are plated on rich media with tetracycline. Only the recipients that received ICE (transconjugates) survive. (B) Alanine racemase schematic. Alr is the alanine racemase in *B. subtilis*. (C) Conjugation efficiencies are the same for wildtype (JMA168) and *Δalr* (JABXX) donor strains. Error bars are the s.d. of three replicates collected on different days.

Each selection method requires a significant amount of recipient strain characterization and often yields mediocre results. Innate resistance phenotypes can be unreliable, making it difficult to select for transconjugates using dual antibiotic selections. Generating spectinomycin resistance becomes more difficult as the number of rRNA operons increases. Plus, 16S rRNA mutants can are often much less robust than their spectinomycin sensitive parents. Crafting new growth medias to isolate specific strains is tedious and may lower the observed conjugation efficiency if it is difficult for the recipient strain to survive in the selective media. Ideally, one would be able to isolate transconjugates using rich media and a single antibiotic to select for acquisition of the conjugated element.

We introduced a D-alanine auxotrophy to our *B. subtilis* donor strain to streamline transconjugate isolation. D-alanine is an essential cell wall component that is either taken directly from the environment or made by isomerizing L-alanine using a dedicated racemase[402,403] (Fig 4-7b). Alanine racemases have been used previously for cloning plasmids into auxotrophic strains[404]. Here, we knocked out the *B. subtilis* D-alanine racemase encoded by *alr* to generate donor strains that require D-alanine to grow. These strains do not grow in any of the rich medias tested (luria broth (LB), nutrient broth (NB), tryptic soy broth (TSB), brain heart infusion (BHI), or deMan, Rogosa and Sharpe media (MRS)) without 100 μg/mL D-alanine. This auxotrophy simplifies transconjugate isolation by allowing us to select for transconjugates on rich media with a single antibiotic. *Δalr* donor strains conjugate with the same efficiency as *alr+* strains (Fig 4-7c).

4-2-5. Compilation and characterization of recipient bacteria collection

Bacteria were isolated from several different environments to test the ICE*Bs1* host range. In total, 82 strains were collected from >6 different sources, including the American Type Culture Collection (ATCC), Bacillus Genetic Stock Center (BGSC), research groups at MIT and the Broad Institute, and Epsoma soil products (Table 4-1). All potential recipient strains are Gram-positive aerobic (or facultatively aerobic) bacteria. We focused on collecting commensal strains from humans and agricultural soils. These organisms are of particular interest for generating human probiotics (human isolates) and modulating plant growth (soil isolates). Several of the strains in our collection have been well characterized as plant growth promoting bacteria (*e.g., Paenibacillus macerans, Bacillus amyloliquefaciens* subsp. *plantarium FZB42)*, human probiotics (*e.g., Bacillus coagulans GBI-30, Bacillus clausii domuvar, Lactococcus lactis)*, insecticide producers (*e.g., Bacillus thuringensis*), or living concrete additives (*Bacillus cohnii)* (Table 4-1). Others were added to the collection because they are known colonizers of interesting microbiomes (*e.g., Lactobacillus reuteri* – human gut, *Staphylococcus epidermis* – human skin, *Bacillus megaterium* - soil). Finally, some bacteria were simply isolated from the target environments (*e.g.,* commercially available organic fertilizer), though their rolls and persistence in these settings are unknown. Table 4-1 lists some notable features of the recipient strains. 16S rRNA sequencing was used to identify bacterial strains isolated from commercial

products during this study (strain names with BT# or CP#, Table 4-1). A phylogenetic map of the recipient collection was also generated using the 16S rRNA sequences (Fig 4-8). Altogether, the collection spans 56 distinct species from 15 different genera.

	Strain	Source	Isolation	Known Functions	Ref
1	Enterococcus_mundtii_MOI-7_JAB432	Alm	Feces	-	
2	Lactobacillus_sp.CD12_JAB487	Alm	Feces	-	
3	Lactobacillus_sp.DLS1_JAB488	Alm	Feces	-	
4	Pediococcus_pentosaceus_JAB482	Alm	Feces	-	
5	Bacillus_cereus_55000_JAB564	ATCC	Soybean	Biological control *R. solani*	
6	Corynebacterium_ammoniagenes_6871_ JAB511	ATCC	Feces	Industrial chemical production	405,406
7	Enterococcus_caccae_BAA-1240_JAB509	ATCC	Feces	-	407
8	Enterococcus_durans_6056_JAB518	ATCC	Feces	-	
9	Enterococcus_faecalis_19433_JAB510	ATCC	Feces	Gut commensal, probiotic	408
10	Lactobacillus_brevis_14869_JAB513	ATCC	Feces	Gut commensal, Biological control *B. cereus*	409
11	Lactobacillus_gasseri_33323_JAB514	ATCC	Feces	Gut commensal, Immunostimulant	410,411
12	Lactobacillus_paracasei _27092_JAB515	ATCC	Feces	Gut commensal, Beverage production	412
13	Lactobacillus_reuteri_23272_JAB512	ATCC	Feces	Modulate cytokine responses	413
14	Streptococcus_infantarius_coli _BAA-103_JAB517	ATCC	Feces	-	414
15	Streptococcus_infantarius_infantarius_BAA-102_JAB516	ATCC	Feces	-	414
16	Lactobacillus_planarium_BAA-793_JAB480	ATCC	Saliva	Gut commensal, probiotic	415
17	Paenibacillus_rhizohabitans_BAA-94_ JAB560	ATCC	Maize soil	Nitrogen fixation	416
18	Paenibacillus_tropicalis_BAA-414_JAB559	ATCC	Maize soil	-	417
19	Bacillus_sp_53935_JAB563	ATCC	Maize soil	Biological control of root rot	418
20	Paenibacillus_brasilensis_BAA-413_JAB558	ATCC	Maize soil	Nitrogen fixation	419
21	Paenibacillus_graminis_BAA-95_JAB561	ATCC	Maize soil	-	420
22	Paenibacillus_polymyxa_39564_JAB562	ATCC	Potato roots	Biological control of *Verticillium* sp.	421
23	Lactococcus_lactis_JAB481	ATCC	Probiotic	Food fermentation	422,423
24	Paenibacillus_macerans_8244T_JAB801	ATCC	Soil	Nitrogen fixation	424
25	Paenibacillus_polymyxa_842_JAB508	ATCC	Soil	Nitrogen fixation	425,426
26	Bacillus_amyloliquefaciens_BAA-390_JAB557	ATCC	Wheat soil	Biological control of blight/tan spot	
27	Paenibacillus_durus_35681_JAB802	ATCC	Wheat roots	Nitrogen fixation	424,427
28	Bacillus_circulans_16A1_JAB252	BGSC	-	Industrial ab production	
29	Bacillus_megaterium_7A16_JAB768	BGSC	Soil	Industrial enzyme production	428
30	Bacillus_thuringiensis_ alesti_4C1_JAB240	BGSC	-	Agricultural insecticide	
31	Bacillus_thuringiensis_andalousiensis_4AW1 _JAB767	BGSC	-	Agricultural insecticide	
32	Bacillus_thuringiensis_finitimus_ 4B1_JAB239	BGSC	-	Agricultural insecticide	
33	Bacillus_thuringiensis_israelensis_4Q1_ JAB550	BGSC	-	Agricultural insecticide	
34	Bacillus_thuringiensis_subsp_kurstaki_4D1_ JAB549	BGSC	-	Agricultural insecticide	
35	Bacillus_thuringiensis_mexicanensis_4AC2 JAB551	BGSC	-	Agricultural insecticide	
36	Bacillus_thuringiensis_thuringiensis_4B1_ JAB239	BGSC	-	Agricultural insecticide	

37	Bacillus_subtilis_BSn5_3A35_JAB547	BGSC	Plant tissue	Biological control of *Erwinia* sp.	429
38	Bacillus_subtilis_inaquasorum_3A28T_ JAB766	BGSC	Desert soil	-	430
39	Bacillus_subtilis_GB03_3A37_JAB548	BGSC	Douglas fir foliage	Enhances vegetable growth, photosynthesis, iron uptake, diease resistance	431
40	Brevibacillus_laterosporus_40A4_JAB556	BGSC	Maize soil	Biological control corn rootworm	432
41	Bacillus_amyloliquefaciens_plantarium_10A6 _JAB552	BGSC	Maize roots	Auxin production	433,434
42	Bacillus_pumilus_Biosubtyl_8A1_JAB171	BGSC	Probiotic	Immunostimulant, diarrhea treatment	435,436
43	Bacillus_methylotrophicus_10A23_JAB553	BGSC	Rice soil	ACC deaminase production	437
44	Bacillus_glycinifermentans_51A1_JAB770	BGSC	Soil	-	438
45	Bacillus_lentus_60A1_JAB771	BGSC	Soil	-	
46	Bacillus_licheniformis_5A2_JAB242	BGSC	Soil	Glutamic acid production	439
47	Bacillus_licheniformis_5A24_JAB249	BGSC	Soil	Biological control of *Sclerotinia* sp.	440
48	Bacillus_megaterium_7A1_JAB170	BGSC	Soil	-	
49	Bacillus_subtilis_Nm-1_3A25_JAB546	BGSC	Soil	Oozone resistance in *Brassica* sp.	441
50	Lysinibacillus_spaericus_13A1_JAB253	BGSC	Soil	Mosquito larvae toxin	442,443
51	Paenibacillus_macerans_22A1_JAB174	BGSC	Soil	-	
52	Paenibacillus_vorticalis_31A1_JAB555	BGSC	Soil	Pattern formation	444
53	Rummeliibacillus_pycnus_24A1_JAB769	BGSC	Soil	-	445
54	Bacillus_shakletonii_102A1_JAB772	BGSC	Volcanic soil	Thermophile	446
55	Bacillus_clausii_Domuvar _17A1_JAB554	BGSC	Probiotic	Antidiarrehal, B vitamin production	436,447
56	Arthrobacter_creatinolyticus_BT9_JAB452	Biotone	Soil	-	
57	Bacillus_amyloliquefaciens_BT3_JAB397	Biotone	Soil	-	
58	Bacillus_amyloliquefaciens _BT16_JAB570	Biotone	Soil	-	
59	Bacillus_cereus_BT1_JAB395	Biotone	Soil	-	
60	Bacillus_flexis_BT7_JAB409	Biotone	Soil	-	
61	Bacillus_licheniformis_BT2_JAB396	Biotone	Soil	-	
62	Bacillus_megaterium_BT12_JAB566	Biotone	Soil	-	
63	Brevibacilius_laterosporus_BT8_JAB451	Biotone	Soil	-	
64	Oceanobacillus_sojae_BT6_JAB400	Biotone	Soil	-	
65	Rummeliibacillus_stabekisii_BT20_JAB574	Biotone	Soil	-	
66	Sporosarcina_sp_BT19_JAB573	Biotone	Soil	-	
67	Compost_bacterium_BT10_JAB453	Biotone	Soil	-	
68	Bacillus_anthracis_UM44-1C9	Grossman	Lab strain	Non-virulent	448
69	Bacillus_subtilis_AG174	Grossman	Lab strain		449
70	Bacillus_subtilis_PY79	Grossman	Lab strain	-	450
71	Bacillus_coagulans_GBI-30_JAB565	DA	Probiotic	Immunomodulatory, *C. difficile* treatment	451
72	Enterococcus_faecalis_(JMJ28)_JAB790	Vlamakis	Feces	-	
73	Enterococcus_faecalis_(JMJ29)_JAB791	Vlamakis	Feces	-	
74	Enterococcus_faecium_(JMJ27)_JAB789	Vlamakis	Feces	-	
75	Pediococcus_acidlactici_(JMJ36)_JAB798	Vlamakis	Feces	-	
76	Streptomyces_laendulae_(JMJ39)_JAB800	Vlamakis	Feces	-	
77	Streptococcus_salivarius_(JMJ33)_JAB795	Vlamakis	Mouth	-	
78	Streptococcus_vestibularis (JMJ34)_JAB796	Vlamakis	Mouth	-	
79	Staphylococcus_captitis_(JMJ32)_JAB794	Vlamakis	Skin	-	
80	Staphylococcus_epidermis_(JMJ31)_JAB793	Vlamakis	Skin	-	
81	Streptococcus_dyscalactiae_(JMJ30)_JAB792	Vlamakis	Skin	-	
82	Escherichia_coli_MG1655_JAB410	Voigt	Lab strain	Cloning	

Table 4-1: Complete list of potential recipient bacteria.
Strains source abbreviations: American Type Culture Collection (ATCC), *Bacillus* Genetic Stock Center (BGSC), MIT/Broad labs: Alm, Grossman, Vlamakis, Voigt. Biotone is an organic fertilizer produced by Epsoma. Digestive Advantage (DA) is a probiotic pill made by Schiff. All 'Mouth,' 'Skin,' and 'Feces' strains were isolated from humans.

We began characterizing the recipient collection by measuring their growth rates in different commercially available medias. We chose four medias (LB, TSB, BHI, and MRS) and two temperatures (30°C or 37°C) for the growth experiments. The human isolates were grown in BHI and MRS at 37°C, whereas the soil strains, which are typically more tolerant to diverse culture conditions, were grown in both LB and TSB at 30°C and 37°C to find conditions that maximize growth rate. Doubling times were measured using a plate reader and calculated as the slope of the natural log of OD_{600} *vs.* time (Tables 4-2 & 4-3). From the growth experiments, we found that *Bacillus* soil strains grew fastest and to the highest optical density (OD_{600}) of all strains. Unsurprisingly, MRS, which contains sodium acetate to suppress the growth of non-*Lactobacilli*, was the least hospitable growth medium for all strains tested except *Lactobacillus* sp. and *Pediococcus* sp. A few strains (JAB453, JAB514, JAB796) grew very poorly in all conditions tested and were not analyzed further.

Next, the strains were tested for natural resistance to tetracycline, chloramphenicol, and spectinomycin (Tables 4-2 & 4-3). Tetracycline and chloramphenicol resistance cassettes, *tet(M)* and *cat*, were integrated into wild type ICE*Bs1* to create a dual-marker element for screening conjugation into the recipient collection. *Tet(M)* and *cat* were cloned from broad host range vectors Tn*916* and pC194, respectively. These cassettes were chosen because they have been identified during deep sequencing studies or used for cloning in several different Gram-positive genera (*e.g., Bacillus, Enterococcus, Staphylococcus, Clostridium, Lactobacilus*)[452–454]. Since isolating transconjugates requires recipients to become resistant to an antibiotic upon receiving ICE*Bs1*, it is nearly impossible to differentiate between strains that have not received ICE and those that received the element, but were unable to utilize the antibiotic resistance marker. Thus, we chose to use two antibiotic resistance markers to enhance our ability to successfully screen for transconjugates. Spectinomycin was used to enumerate donors after conjugation. Each strain was tested for antibiotic resistance by plating serial dilutions of recipient cultures onto selective

media plates. Strains were deemed resistant when bacteria formed colonies on the ten-fold dilution.

	Human Isolates	Doubling time (m)		Antibiotic resistance		
		BHI	MRS	Spec	Tet	Cam
1	Enterococcus_mundtii_MOI-7_JAB432	27.20		S	S	S
2	Enterococcus_faecium_(JMJ27)_JAB789	30.45	34.09	ND	S	ND
3	Enterococcus_durans_6056_JAB518	38.26		R	S	S
4	Enterococcus_caccae_BAA-1240_JAB509	63.70		S	S	S
5	Enterococcus_faecalis_19433_JAB510	31.48		R	S	S
6	Enterococcus_faecalis_(JMJ28)_JAB790	41.54		ND	R	ND
7	Enterococcus_faecalis_(JMJ29)_JAB791	37.97		ND	R	ND
8	Corynebacterium_ammoniagenes_6871_JAB511	62.53		S	S	R
9	Escherichia_coli_MG1655_JAB410	24.62		S	S	R
10	Lactococcus_lactis_JAB481	31.29	43.01	S	S	S
11	Streptococcus_salivarius_(JMJ33)_JAB795	23.39		ND	S	ND
12	Streptococcus_vestibularis (JMJ34)_JAB796			ND	ND	ND
13	Streptococcus_infantarius _infantarius_BAA-102_JAB516	26.75		R	S	S
14	Streptococcus_infantarius_coli_BAA-103_JAB517	44.37		R	S	S
15	Streptococcus_dyscalactiae_(JMJ30)_JAB792	36.34		ND	S	ND
16	Lactobacillus_gasseri_33323_JAB514			ND	ND	ND
17	Lactobacillus_paracasei_paracasei_27092_JAB515	87.99	40.84	S	S	S
18	Lactobacillus_brevis_14869_JAB513	35.17		R	R	S
19	Pediococcus_pentosaceus_JAB482	33.54	33.60	R	S	S
20	Pediococcus_acidlactici_(JMJ36)_JAB798	93.30	33.12	ND	S	ND
21	Streptomyces_laendulae_(JMJ39)_JAB800		64.32	ND	ND	ND
22	Lactobacillus_reuteri_23272_JAB512		60.61	ND	ND	ND
23	Bacillus_clausii_Domuvar_17A1_JAB554	38.26		S	S	ND
24	Bacillus_coagulans_GBI-30_JAB565	61.78		S	S	S
25	Staphylococcus_captitis_(JMJ32)_JAB794	46.99	61.38	ND	S	ND
26	Staphylococcus_epidermis_(JMJ31)_JAB793	46.28		ND	S	ND
27	Lactobacillus_sp.CD12_JAB487	33.46	30.33	R	S	S
28	Lactobacillus_planarium_BAA-793_JAB480	32.39	53.74	R	S	S
29	Lactobacillus_sp.DLS1_JAB488	33.91	34.67	R	S	S
30	Bacillus_subtilis_PY79	22.12		S	S	S

Table 4-2: Doubling times and antibiotic resistance of human isolates.
Antibiotic concentrations are 100 μg/mL spectinomycin (spec), 10μg/mL tetracycline (tet), and 10 μg/mL chloramphenicol (cam).

	Soil & Other Isolates	Doubling time (m)				Antibiotic resistance		
		LB		TSB				
		37°C	30°C	37°C	30°C	Spec	Tet	Cam
1	Arthrobacter_creatinolyticus_BT9_JAB452	56.44	48.12	77.09	58.96	S	S	S
2	Paenibacillus_brasilensis_BAA-413_JAB558	56.16	83.11	36.66		R	R	S

3	Paenibacillus_polymyxa_842_JAB508	36.52	35.89	43.98	47.67	S	S	ND
4	Paenibacillus_polymyxa_39564_JAB562	44.86	49.29	45.62	46.74	R	S	S
5	Paenibacillus_graminis_BAA-95_JAB561	57.09	42.93	59.00	51.57	R	S	S
6	Paenibacillus_tropicalis_BAA-414_JAB559	58.61	43.98	34.91	ND	R	R	S
7	Paenibacillus_rhizohabitans_BAA-94_JAB560				59.24	S	S	S
8	Paenibacillus_durans_35681_JAB802	ND	ND	ND	ND			
9	Paenibacillus_vorticalis_31A1_JAB555	46.27	62.55	51.20	32.64	S	R	ND
10	Paenibacillus_macerans_BKM-B-51_JAB174	16.66	37.84	22.71	41.28	S	S	S
11	Paenibacillus_macerans_8244T_JAB801	55.84	ND	33.24	ND	ND	ND	ND
12	Brevibacillus_laterosporus_BT8_JAB451	24.78	ND			S	S	S
13	Brevibacillus_laterosporus_53694_JAB556	26.85	38.62	25.46	ND	S	S	ND
14	Sporosarcina_sp_BT19_JAB573	67.22	ND	59.64	ND	R	S	S
15	Lysinibacillus_spaericus_13A1_JAB253	25.27	36.61	29.81	36.11	ND	ND	ND
16	Rummeliibacillus_stabekisii_BT20_JAB574	29.10	ND	27.21	ND	S	S	S
17	Rummeliibacillus_pycnus_24A1_JAB769	32.83	ND	32.67	ND	ND	ND	ND
18	Compost_bacterium_BT10_JAB453					ND	ND	ND
19	Oceanobacillus_sojae_BT6_JAB400	46.77	64.80	51.95	66.60	R	S	S
20	Bacillus_circulans_16A1_JAB252	33.54	49.31	32.03	57.96	ND	ND	ND
21	Bacillus_shakletonii_102A1_JAB772	30.16	ND	32.66	ND	ND	ND	ND
22	Bacillus_thuringensis_subsp_kurstaki_JAB549	23.16	33.38	23.31	35.73	R	S	ND
23	Bacillus_lentus_60A1_JAB771	25.15	ND	27.86	ND	ND	ND	ND
24	Bacillus_thuringensis_mexicanensis_JAB551	22.34	31.36	25.01	37.45	R	S	ND
25	Bacillus_thuringensis _israelensis_JAB550	24.19	32.76	25.21	36.03	S	S	ND
26	Bacillus_thuringensis_andalousiensis_JAB767	22.92	ND	23.46	ND	ND	ND	ND
27	Bacillus_cereus_BT1_JAB395	21.13	30.32	31.92	53.41	R	S	R
28	Bacillus_anthracis_UM44-1C9	26.50	39.88	30.06	44.96	S	S	S
29	Bacillus_thuringensis_subsp_finitimus_JAB239	17.06	24.44	20.92	31.89	R	S	S
30	Bacillus_thuringensis_thuringensis_JAB237	20.93	28.03	24.04	37.73	R	S	S
31	Bacillus_cereus_55000_JAB564	22.08	29.99	22.94	30.73	S	S	ND
32	Bacillus_thuringensis_alesti_JAB240	26.78	30.93	26.52	35.14	R	S	S
33	Bacillus_flexis_BT7_JAB409	21.90	32.69	19.01	31.09	R	S	S
34	Bacillus_megaterium_899_JAB170	20.76	20.79	20.81	30.73	ND	ND	ND
35	Bacillus_megaterium_BT12_JAB566	24.40	29.00	25.25	34.06	S	S	S
36	Bacillus_megaterium_7A16_JAB768	22.91	ND	24.59	ND	ND	ND	ND
37	Bacillus_pumilus_Biosubtyl_JAB171	29.92	45.47	28.83	45.44	S	S	S
38	Bacillus_licheniformis_5A2_JAB242	25.92	40.73	22.42	26.87	ND	ND	ND
39	Bacillus_licheniformis_BT2_JAB396	27.15	23.49	26.52	26.59	S	S	R
40	Bacillus_licheniformis_BGSC-5A24_JAB249	20.50	29.02	8.56	27.09	ND	ND	ND
41	Bacillus_subtilis_subsp_inaquasorum_JAB766	21.14	ND	14.80	ND	ND	ND	ND
42	Bacillus_subtilis_PY79	21.72	35.79	22.58	34.25	S	S	S
43	Bacillus_subtilis_BSn5_JAB547	21.78	36.23	18.93	21.96	ND	ND	ND
44	Bacillus_methylotrophicus_ JAB553	19.91	30.91	28.96	54.53	S	R	ND
45	Bacillus_sp_53935_JAB563	23.81	32.51	23.38	37.09	S	S	S
46	Bacillus_glycinifermentans_51A1_JAB770	26.50	ND	28.44	ND	ND	ND	ND
47	Bacillus_amyloliquefaciens_ JAB570	21.37		19.80	83.36	S	R	ND

48	Bacillus_subtilis _Nm-1_3A25_JAB546	25.05	33.77	18.20	32.91	S	S	ND
49	Bacillus_amyloliquefaciens_BT3_JAB397	20.42	28.15	22.74	26.73	S	S	S
50	Bacillus_subtilis_GB03_JAB548	21.34	29.87	22.60	27.48	S	S	ND
51	Bacillus_amyloliquefaciens_BAA-390_JAB557	25.97	33.64	21.53	16.66	S	S	S
52	Bacillus_amyloliquefaciens _10A6_JAB552	22.29	35.04	64.16	24.45	S	S	ND

Table 4-3: Doubling times and antibiotic resistance of soil and other isolates.
Antibiotic concentrations are 100 μg/mL spectinomycin (spec), 10μg/mL tetracycline (tet), and 10 μg/mL chloramphenicol (cam).

Using data from the antibiotic and growth screens, we loosely grouped the recipients into three different growth categories: fast (doubling times <30 min), medium (doubling times 30-50 min), and slow (doubling times >50 min). These groupings were used to divide strains into manageable units for the conjugation assays (see next section).

4-2-6. Conjugation of wild type and miniature ICE*Bs1* into diverse bacterial species

ICE*Bs1* conjugation has been tested in 44 of the 82 recipient strains (53%) (Fig 4-9b). Of those tested, 20 yielded transconjugates (45%), which were verified by re-streaking on the appropriate antibiotics and colony PCR using ICE-specific primers (Fig 4-9a). Interestingly, there is little correlation between species' 16S rRNA phylogenetic distance and successful conjugation. Several strains that are closely related to our *B. subtilis* donor (*e.g., Bacillus amyloliquefaciens FZB42, Bacillus subtilis GB03*) did not receive ICE*Bs1* during the conjugation assays. However, all of the *Enterococci* strains, which are the most distantly related strains in our collection, did receive ICE*Bs1*. Similarly, distinct strains of the same species have different mating abilities. ICE*Bs1* conjugates into *Bacillus thuringensis* subsp. *fimintus* and subsp. *kurstaki* strains, however we did not detect transconjugates when attempting to mate ICE*Bs1* into the *mexicanensis, israelensis, thuringensis,* and *alesti* subspecies strains.

In addition, some of the recipient strains showed mixed colony morphologies after conjugation (Fig 4-9, starred). We presume that ICE*Bs1* integrates randomly into the genome of these strains and occasionally disrupts genes that are necessary for robust growth leading to diverse colony sizes. In *B. subtilis,* ICE*Bs1* recognizes a specific 60-bp direct repeat sequence (*attB*) for integration[455]. When *attB* is missing, ICE*Bs1* will integrate randomly into the *B. subtilis* chromosome. The ICE*Bs1 attB* is in a leucine tRNA ($tRNA^{leu2}$) and although tRNAs are highly conserved across species, it may be missing from many of the recipient bacteria, causing

Figure 4-8: Phylogenetic tree and conjugation proficiencies for recipient bacteria collection.

Tree was constructed using 16S rRNA sequences. Bacteria that successfully receive ICE*Bs1* are shown in green. Unsuccessful conjugation in black. Untested in grey. *Bacillus subtilis* donor strain (AG174) in blue.

ICE*Bs1* to integrate randomly. Although genome sequences are unavailable for the vast majority of our recipient strains, BLAST was used to identify species with potential att sites. BLAST results show that 30 of the 56 recipient species (53.5%) contain sequences with at least some sequence identity to the ICE*Bs1* attachment site (Fig 4-9). Thus ICE*Bs1* is likely integrating randomly in the majority of recipient species.

```
ICEBs1_attB                    CTAGGTTGAGGGCCTAGTGGGT-GAATAACCCGTGGAGGTTCAAGTCCTCTCGGCCGCATC
Lactobacillus_paracasei        ------------CCTAGTGGGTGAAT-ATCCGGATGAGGT--------------------
Lactobacillus_planarium        -------------------------TAACCCGGTGGAAGTTCAAATCTTCTTGGCCGCA--
Pediococcus_pentosaceus        --------------------------------GTGGAGGTTCGAACCCTCTTGGCCGCATC
Streptococcus_salivarius       -------------------------------CGTGGGGGTTCAAGTCCCTTCACCCGCAT-
Lactococcus_lactis             -------------------------------CGTGGGGGTTCAAGTCCCTTCATCCGCAT-
Streptococcus_infantarius      -------------------------------CGTGGGGGTTCAAGTCC-------------
Paenibacillus_macerans         ---------------------------------TGGAGTTTCAAGTTCTCTC---------
Rummeliibacillus_pycnus        -------------------------------CCGTGGAGGGTCA-----------------
Streptomyces_lavendulae        ------------------------------------------AAGTCCTCTCGGACGC---
Paenibacillus_polymyxa         -----TTGAGGGGGTAGTGG-GCGTACG-CCCGTGGAGGTTCGAGTCCTCTCGACCGCAT-
Paenibacillus_graminis         -----TTGAGGGGGTAGTGG-GTGTATA-CCCGTGGAGGTTCGAGTCCTCTCTACCGCAT-
Paenibacillus_durus            ---GGTTCAGGTCCG-TGTGGGCTAACCCCCCGTGGAGGTTCGAGTCCTCTCGACCGCATC
Bacillus_anthracis             CTAGGTTGAGGGCCTAGTGGG--GGAAACCCCGTGGAGGTTCAAGTCCTCTCGGCCGCATC
Bacillus_cereus                CTAGGTTGAGGGCCTAGTGGG--GGAAACCCCGTGGAGGTTCAAGTCCTCTCGGCCGCATC
Bacillus_thuringiensis         CTAGGTTGAGGGCCTAGTGGG--GGAAACCCCGTGGAGGTTCAAGTCCTCTCGGCCGCATC
Staphylococcus_epidermis       CTAGGTTGAGGGCCTAGTGGGA-GTTAATCCCGTGGAGGTTCAAGTCCTCTCGGCCGCA--
Staphylococcus_captitis        CTAGGTTGAGGGCCTAGTGGGA-GTTAATCCCGTGGAGGTTCAAGTCCTCTCGGCCGCA--
Sporosarcina_sp                CTAGGTTGAGGGCCTAGTGGGG-GTTAACCCCGTGGAGGTTCAAGTCCTCTTTACCGCA--
Rummeliibacillus_stabekisii    CTAGGTTGAGGGCCTAGTCGGG-GCA-ACTCGGTGGAGGTTCAAGTCCTCTCGATCGCA--
Pediococcus_acidlactici        ------------------------AATAACCCGTGG-------------------------
Lysinibacillus_spaericus       CTAGGTTGAGGGCCTAGTGGGG-GCA-ACCCCGTGGAGGTTCAAGTCCTCTCGGCCGCA--
Bacillus_megaterium            CTAGGTTGAGGGCCTAGTGGG--GGCGACCCCGTGGAGGTTCAAGTCCTCTCGGCCGCA--
Bacillus_coagulans             CTAGGTTGAGGGCCTAGTGGGAGTTAAATCCCGTGGAGGTTCAAGTCCTCTCGGCCGCAT-
Bacillus_amyloliquefaciens     CTAGGTTGAGGGCCTAGTGGGT-GAATAACCCGTGGAGGTTCAAGTCCTCTCGGCCGCATC
Bacillus_licheniformis         CTAGGTTGAGGGCCTAGTGGGT-GAATAACCCGTGGAGGTTCAAGTCCTCTCGGCCGCATC
Bacillus_methylotrophicus      CTAGGTTGAGGGCCTAGTGGGT-GAATAACCCGTGGAGGTTCAAGTCCTCTCGGCCGCATC
Bacillus_pumilus               CTAGGTTGAGGGCCTAGTGGGT-GAATAACCCGTGGAGGTTCAAGTCCTCTCGGCCGCATC
Bacillus_subtilis              CTAGGTTGAGGGCCTAGTGGGT-GAATAACCCGTGGAGGTTCAAGTCCTCTCGGCCGCATC
Lactobacillus_reuteri          -------------------------------CCGTGGAGGTTCAAGTCCTCTCGGCCGCA--
Streptococcus_dyscalactiae     ----------------------------------GGAGGTTCAAGTCCTC-----------
```

Figure 4-9: ICE*Bs1* attB site in recipient bacterial species.
Sequences closely related to the ICE*Bs1* att site were identified through BLAST and aligned using Clustal Omega.

After confirming ICE*Bs1* conjugation into several recipient species, we began using mICE to deliver exogenous DNA to the bacteria. First, we added a P_{spank}*_GFPmut2* cassette to

mICE 1.0 and delivered it to the recipient bacteria that received wild type ICE*Bs1* (Fig 4-10a). We measured the response function of this cassette across species. Incredibly, the cassette is functional in all strains tested thus far. All of the bacteria express GFP in an IPTG-dependent manner, however the background fluorescence (Fig 4-10b, blue lines) and fold change in GFP expression +/- IPTG varies significantly across species. The variation in fold change across species is attributable to differences in both leaky expression of GFP and maximum fluorescence.

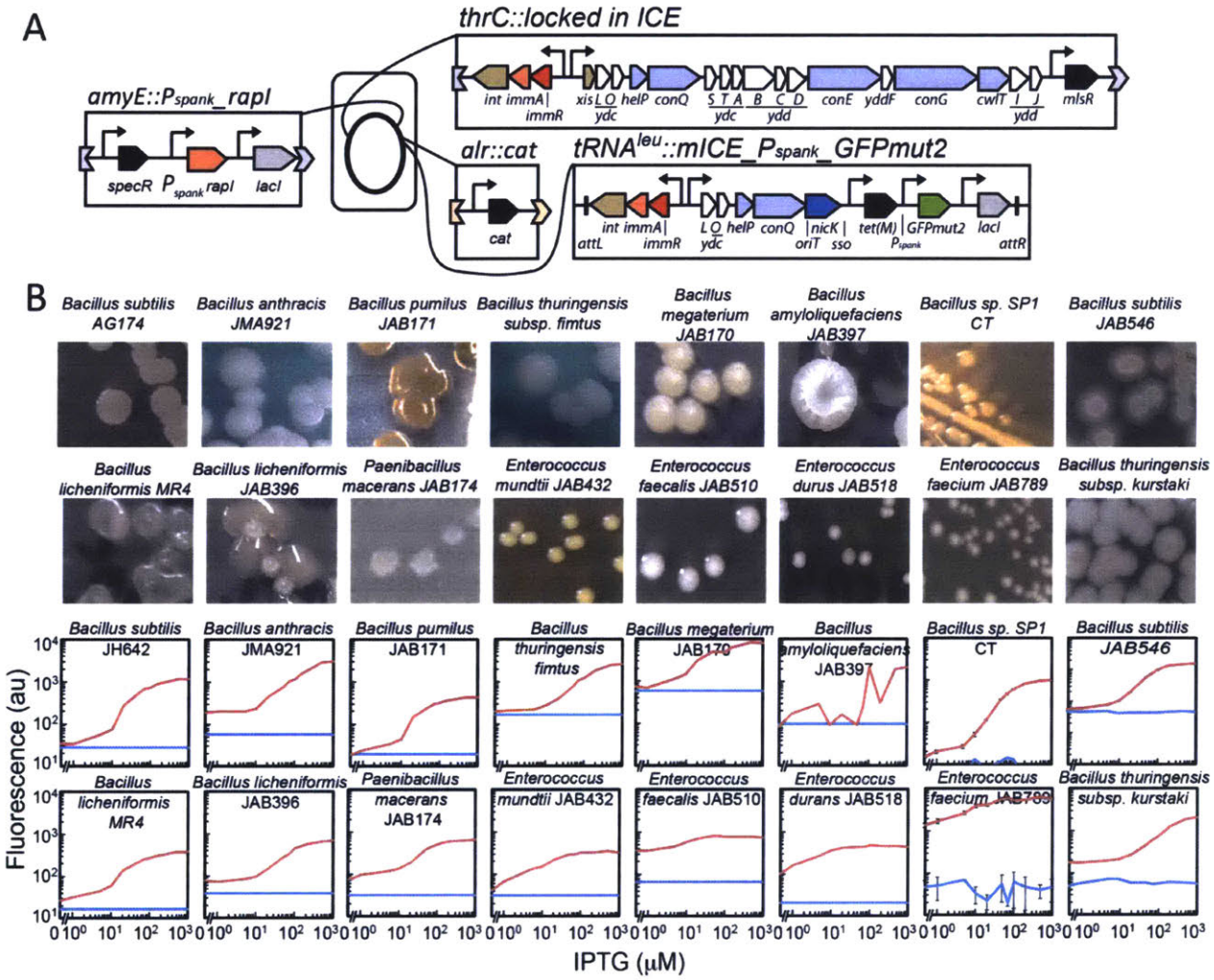

Figure 4-10: P_{spank}_*GFPmut2* response functions in diverse bacterial species.
(A) Schematic of B. subtilis donor strain (JAB402) used to deliver mICE carrying P_{spank}_*GFPmut2* to diverse bacterial species. (B) Recipient species engineered with mICE carrying P_{spank}_*GFPmut2* and the IPTG-response functions. Engineered strains in red, unmodified recipient bacteria in blue. Error bars are the s.d. of three separate transconjugates from the same conjugation experiment.

4-2-7. Delivery of a functional nif clusters to soil dwelling bacteria

The most famous plant growth promoting bacteria are nitrogen fixers that turn atmospheric nitrogen into a form that plants can easily use. These bacteria are typically endophytes that invade plant roots and, through a complex interaction with plant cells, form oxygen poor nodules in which the bacteria fix nitrogen in return for sugars[456]. Nitrogen-fixing bacteria can improve crop yields and reduce the application of nitrogen fertilizers, which are bad for the environment. However, nodule-forming nitrogen-fixing bacteria can only interact with specific plant species (*i.e.*, legumes). Thus, fertilizers must be used to produce many of the most prevalent food crops worldwide (*i.e.*, wheat, rice, maize). Fortunately, non-nodule forming nitrogen-fixing bacteria (diazotrophs) can also improve plant growth in nitrogen-limited environments[457,458]. Thus there is immense interest in identifying diazatrophic bacteria that can be used to improve growth of non-legume food crops.

Here, we use mICE to introduce nitrogen fixation capabilities to diverse soil bacteria and generate new diazatrophs. Most microbial nitrogen fixation is catalyzed by a molybdenum-dependent nitrogenase enzyme that converts N_2 into ammonia. The structural subunits of this enzyme are typically encoded in complex nitrogen fixation *nif* gene clusters along with proteins for co-factor biosynthesis and electron transport[459]. However, a simple single-operon *nif* cluster was recently discovered in *Paenibacillus* sp. WLY78[460]. This *nif* cluster successfully conferred nitrogen-fixing capabilities to *Escherichia coli* when over expressed in laboratory conditions. We cloned the 10 kb *nif* cluster into mICE in *B. sutbilis* AG174 under the control of three different promoters: P_{nifB} (the native *Paenibacillus* promoter), P_{spank}, and P_{T7} (a promoter recognized by bacteriophage promoter T7 RNA polymerase (T7RNAP)) (Fig 4-11a). Minimal ICE*Bs1* with the *nif* gene cluster will be referred to as nICE. Then nitrogenase enzyme production was measured using an acetylene reduction assay (methods). Of the three promoters, P_{T7} produced the most active nitrogenase (Fig 4-11b). In the *P_{T7}_nifWLY78* strain, T7RNAP is expressed from a non-ICE locus (*amyE*). However, this configuration is not useful for engineering nitrogen fixation into other bacterial species. Thus, we added T7RNAP to nICE. The cluster is only functional when transcription of all of the genes (*nifWLY78*, *T7RNAP*, and *lacI*) is co-directional. The first iteration of T7RNAP-nICE that we built had P_{spank} oriented convergently with P_{T7} (Fig 4-11c). However, this nICE produced significantly less functional nitrogenase than the original

P_{T7}_*nifWLY78* strain (Fig 4-11d). I speculate that this architecture leads to repression via antisense transcription and, although no experiments were done to prove this hypothesis conclusively, flipping the P_{spank}_*T7RNAP* cassette so that the promoters are no longer oriented convergently restores nitrogenase production.

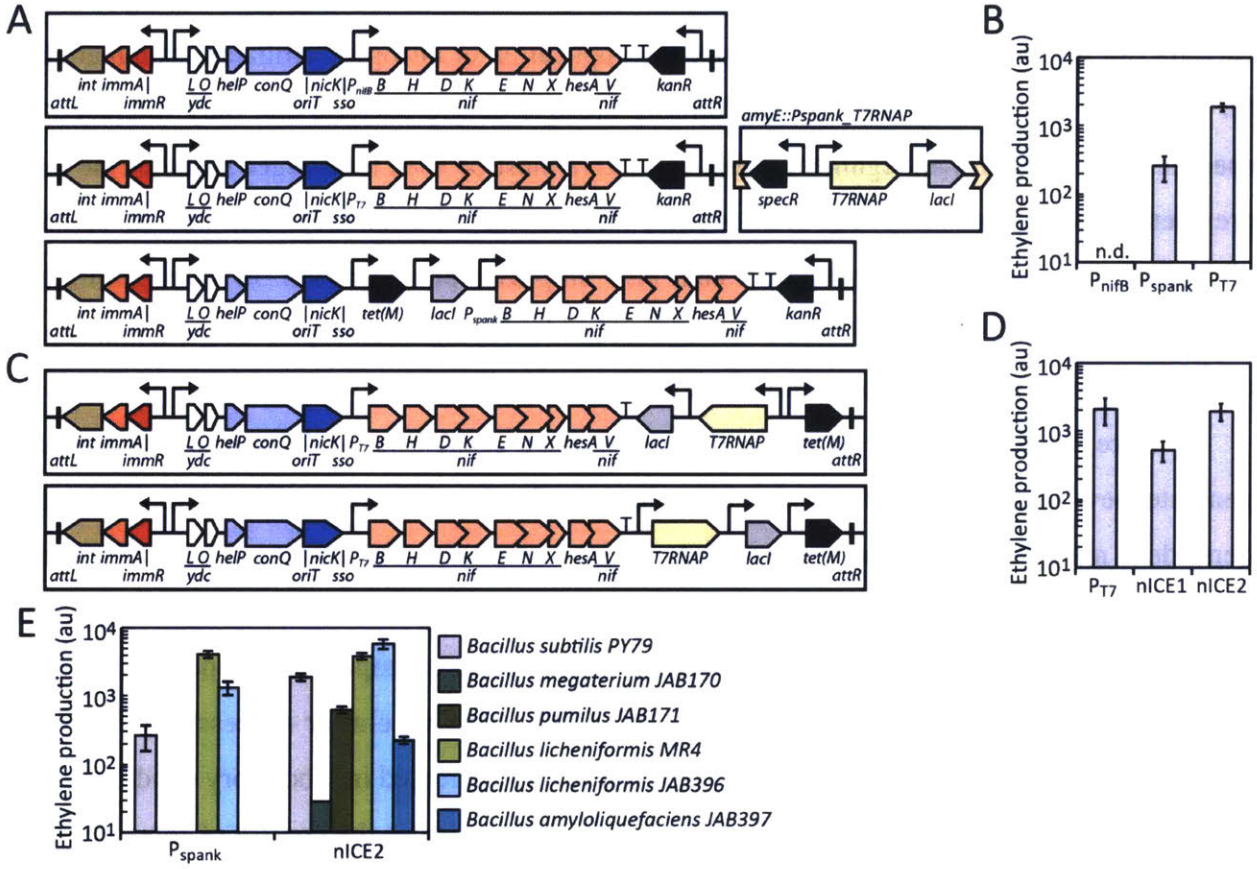

Figure 4-11: Activity of nitrogen fixation pathway *nifWLY78* across species.
(A) Schematics for mICE with *nifWLY78* driven by native promoter P_{nifB} (top), viral polymerase P_{T7} (middle), or IPTG-inducible promoter P_{spank} (bottom). B. subtilis strain with P_{T7}-driven *nif* cluster has IPTG inducible T7RNAP cassette at *amyE*. (B) Activity of the *nif* clusters from (A) in *Bacillus subtilis PY79*. (C) Schematics for mICE with *nifWLY78* and P_{spank}_*T7RNAP* oriented convergently (nICE1, top) or co-directionally (nICE2, bottom). (D) Activity of the *nif* clusters from (C) in *Bacillus subtilis PY79*. (E) Activity of nICE2 (panel (C), bottom) and P_{spank}-driven *nifWLY78* (panel (A), bottom) across soil bacterial species. In all panels, 1mM IPTG was added to induce expression of *nif* cluster genes. Error bars are the s.d. of three different transconjugates.

Once we had a functional T7RNAP-nICE donor, we conjugated the T7RNAP-driven *nif* cluster into several of the soil bacteria species and measured nitrogen fixation (Fig 4-11e). X out of X bacteria tested were able to fix nitrogen. Importantly, none of the strains fix nitrogen in the

absence of IPTG, therefore the diazatrophic phenotype is due to expression of the genes carried on nICE. We also tested the *P_{spank}-nifWLY78* nICE in the other bacterial species and found that although some of the *Bacillus licheniformis* strains are able to fix nitrogen using P_{spank}-nICE, addition of the orthogonal viral polymerase improves the nitrogen fixation phenotype.

4-3. DISCUSSION

Here, we present ICE*Bs1* as a tool for engineering diverse bacterial species. We created a miniature version of ICE*Bs1* that is incapable of self transfer and demonstrate that it can be used to introduce foreign DNA to at least twenty different strains spanning sixteen bacterial species and five genera (*Bacillus, Enterococcus, Paenibacillus, Streptococcus, Staphylococcus*). We use this minimal ICE to test an inducible promoter system across species and demonstrate that a single cassette is functional across all strains tested. Since our interest in developing this tool is to enable the construction of sophisticated probiotics, we demonstrate that mICE can be used to transform non-nitrogen fixing soil bacterium into diazatrophs with potential for improving plant growth. We use mICE to deliver a minimal *nif* cluster to six different soil bacteria and show that the best nitrogen fixation results are achieved when a viral polymerase (T7RNAP) is used to drive expression of the cluster. Future work will focus on testing the remainder of the recipient bacteria collection and modifying mICE 1.0 to improve reliability of the tool and potentially expand host range further.

Although our current data shows ICE*Bs1* conjugation >45% of the bacterial species tested, there are several modifications that could be made to the element that may expand its host range. There are several potential points of failure that may be easily addressed by modifying the mICE (or full length ICE) delivered to the recipient bacteria. Each modification should be preceded by simple assays to elucidate the point(s) of failure in the recipient. The first point of failure may be degradation of ICE DNA in the recipient bacteria by an incompatible restriction modification system. If the ICE DNA is digested by restriction enzymes in the recipient, it will be immediately lost before transconjugates can form. There are several ways to circumvent restriction modification (R/M) systems, however a straightforward method used by other mobile genetic elements is expression of antirestriction proteins. One (or more) of these antirestriction proteins could be added to ICE*Bs1* to prevent digestion of the delivered DNA by recipient R/M

systems. Alternatively, the element could be recoded to remove all known restriction enzyme recognition sequences or the *B. subtilis* donor strain could be engineered to express more DNA modification enzymes to methylate tDNA before it is transferred to the recipient cell.

A second point of failure could be integration of the ICE*Bs1* DNA into the recipient bacteria's genome. Failure to integrate is most likely due to insufficient expression of the Int protein in the recipient (since we know the protein itself is capable of integrating ICE DNA in the absence of attB)[455]. Promoter and RBS engineering could potentially be used to boost expression of the integrase in the recipient cell. Synthetic biologists have recently published mixed feedback control loops and software tools for developing 'universal' RBSs that may help drive expression of Int in organisms with distant transcription/translation machinery[461]. Failure to integrate may also be due to insufficient second strand synthesis in the recipient. tDNA needs to be double stranded in order for host RNAP to transcribe the *int* gene and for the Int protein to catalyze recombination. To ensure second strand synthesis in the recipient, additional single strand origins (sso) could be added to ICE*Bs1*. Ssos from broad host range rolling circle plasmids may increase the amount of double stranded DNA in the recipient and facilitate integration[462].

Failure to maintain ICE*Bs1* may be due to disregulation of *int* and *xis* in the recipient species. If too much excisionase is expressed, ICE*Bs1* will continually excise and eventually get lost. This can be fixed by eliminating *xis* from the transferred DNA so that once ICE has integrated into the recipient's genome it cannot excise and become lost. Integration of ICE DNA into essential genes may lower the observed conjugation efficiency. Random integration of ICE*Bs1* could be prevented by swapping the native integrase for one with a known recognition sequence in the target strain[463]. Unfortunately, this requires genome sequence information in the target, which may be difficult to aquire. CRISPR/Cas systems could also be used to target integration of ICE DNA to specific locations in the host chromosome[464], though this would also require knowledge of genome sequences and expression of more genes in the target strain.

Several additional strains, including anaerobics, thermophiles, and isolates from other important microbiomes (*e.g.*, aquatic bacteria, animal commensals), could be tested as ICE*Bs1* recipients in future studies. *B. subtilis* is capable of anaerobic growth and can withstand very high temperatures (it can still *grow* at 65°C). Thus, it may be amenable to conjugation experiments in more extreme environments. Being able to engineer anaerobic strains would be

especially useful for manipulating gut microbes. Several of the most important gut bacterial species, including *Bifidobacterium* and *Clostridia* are obligate anaerobes with very few tools available for genetic modification.

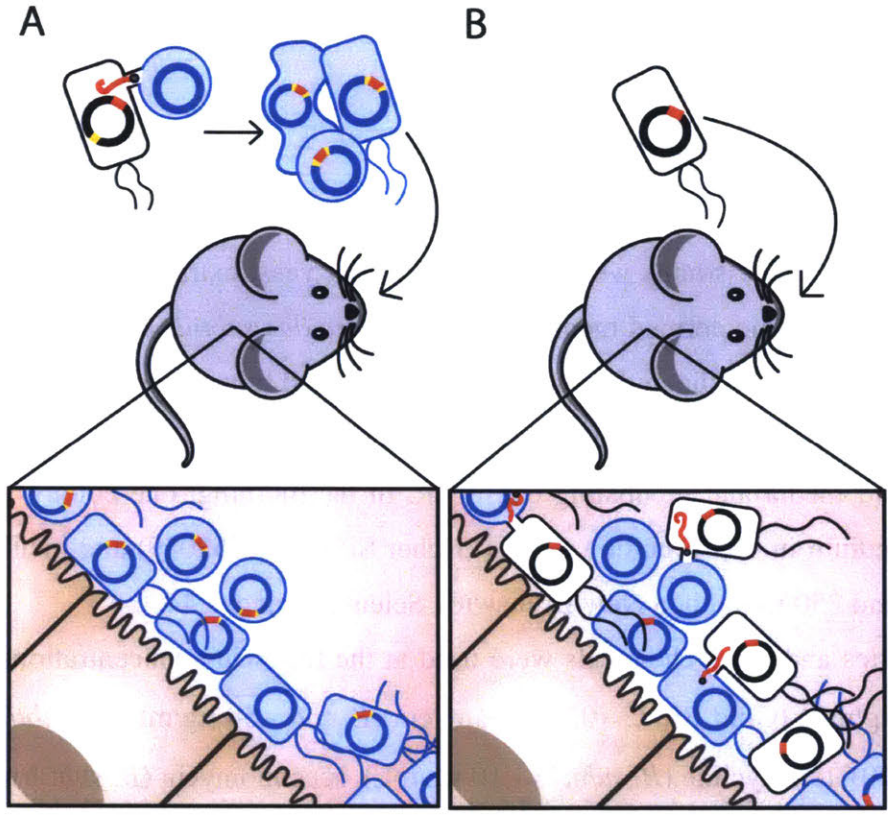

Figure 4-12: Methods for engineering bacteria with ICE.
ICEs could be used to genetically modify bacteria either *in vitro* (A) or *in situ* (B). Bacteria can be extracted from the target environment, modified, and reintroduced (A). Or a donor strain of bacteria can be introduced to the target environment to modify bacteria *in situ* (B).

Most excitingly, ICE could be used in the future to modify recipient bacteria *in situ*. During this project, we brought bacteria out of their environments and into the lab in order to modify them (Fig 4-12a). However taking bacteria out of their environment(s) and into lab is not necessary if the synthetic DNA is delivered to bacteria via conjugation. Horizontal gene transfer via conjugation is known to occur in almost all environments[465]. Thus, instead of bringing bacteria into the lab, donor strains of *B. subtilis* could introduced to the environment to spread target synthetic programs to the bacteria already living in these environments (Fig 4-12b). This would eliminate the need to find culture conditions for target bacterial species or clear out bacteria from specific niches to prevent colonization competition. *In situ* engineering may

instigate a new paradigm for treating disease in which donor bacteria are consumed or distributed, like probiotics to deliver 'smart' therapeutic pathways, instead of bioactive molecules, to the native cells.

4-4. Materials and Methods

4-4-1. Media and growth conditions.

Escherichia coli strains were grown at 37°C in Luria Broth (LB) (Becton Dickinson 244630). *Saccharomyces cerevisiae* strains were grown at 30°C in yeast extract peptone dextrose (YPD) or uracil dropout synthetic defined media (SD-URA). *Bacillus subtilis* donor strains were grown at 37°C in LB medium. Media and growth conditions for all other strains are listed in Table X. Starter cultures were generated for all assays by streaking out the appropriate strains from freezer stocks onto solid media and incubating overnight. In the morning, one colony was inoculated into 2mL of medium in 15mL culture tubes (Fischer Scientific 352059) and incubated for 2h at 30°C or 37°C and 250 r.p.m. in a New Brunswick Scientific Innova 44.

Antibiotics and other chemicals were used at the following concentrations: carbenicillin (*E. coli* - 100 μg/mL) (Gold Bio C-103), kanamycin (*E. coli* - 50 μg/mL, *B. subtilis* – 5 μg/mL) (Gold Bio K-120), tetracycline (*B. subtilis* - 10 μg/mL), spectinomycin (*B. subtilis* - 100 μg/mL), erythromycin (*B. subtilis* – 0.5 μg/mL), lincomycin (*B. subtilis* – 12.5 μg/mL), chloramphenicol (*B. subtilis* - 5 μg/mL), isopropyl β-D-1-thiogalactopyranoside (IPTG) (1mM) (Gold Bio I2481C), xylose (2% wt/vol), D-alanine (100 μM) (Sigma A7377). For the nitrogenase activity assay, all strains were incubated in modified BB medium (20 g sucrose, 0.25 g MgSO4*7H2O, 1 g NaCl, 0.1 g CaCl2*2H2O, 2.9 mg FeCl3, 0.25 mg NaMoO4*2H2O, 2.5 g Na2HPO4, 0.3 g KH2PO4, 1.5 mL 10% serine; per liter).

4-4-2. Strains.

Escherichia coli strain NEB10β (Δ(ara-leu)7697 araD139 fhuA ΔlacX74 galK16 galE15 e14-f80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC)), *Escherichia coli* strain XL10Gold (endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tetR F'[proAB $lacI^qZ\Delta M15$ $Tn10(Tet^R$ Amy

Cm^R)])(Agilent Technologies 200314), and *Saccharomyces cerevisiae* were used for all plasmid constructions. *Bacillus subtilis* subsp. *subtilis* strain JH642 was used as the base strain for all donors. All recipient strains in this study are listed in Table 4-1.

4-4-3. 16S rRNA sequencing.

Colony PCR was performed to amplify and sequence 16S rRNA. To perform colony PCR, single colonies were inoculated into 35 μL of Bio-Rad InstaGene Matrix (Bio-Rad 7326030) and incubated at X for X min, then boiled at X for X min. InstaGene Matrix beads were removed via centrifugation and 1 μL of supernatant was added to a PCR reaction using Integrated DNA Technologies ReadyMade Primers for 16S rRNA amplification (IDT 51-01-19-06 and 51-01-19-07).

4-4-4. Growth rate measurement.

Strains are struck out on the appropriate solid media and incubated at 25°C, 30°C, or 37°C overnight. The next day, colonies are picked into 200 μL of the appropriate media in 96-deep well blocks (USA Scientific 1896-2000) and incubated for 3 h at 25°C, 30°C, or 37°C and 250 r.p.m. in an INFORS-HT Multitron Pro. After 3 h, the cells are diluted 1:500 into 150 μL of fresh media in black 96-well optical bottom plates (Thermo Scientific 165305) and incubated at the same growth temperature with 1 mm orbital shaking in a BioTek Synergy H1 plate reader. Optical density measurements at 600-nm wavelength (OD_{600}) were made every 20 min for 12 h.

To calculate growth rates, a standard curve was made to correlate plate reader measurements with OD_{600} measurements from the BioRad XX spectrophotometer used for all other assays. Briefly, OD_{600} of *B. subtilis* PY79 cultures (in the same media as the test strains) were measured using the BioRad XX spectrophotometer. These cultures were serially diluted (OD_{600} = 1.6, 0.8, 0.4, 0.2, 0.1, 0.05, 0.025, 0) into fresh media with 50μg/mL chloramphenicol to prohibit growth. 150μL of each were added to the 96-well optical bottom plates and measured concurrently with test strains. Growth rates (min/doubling) were calculated using the slope of the linear portion of the exponential agrowth curves.

4-4-5. Conjugation assays.

Conjugation assays were performed on filters as previously described. Briefly, starter cultures are diluted to OD_{600} 0.025 in 20mL of medium in 250 mL Erlenmeyer flasks (Corning 4450-250) and grown at the same temperature and speed as starter cultures. After 1 h, 2% xylose (wt/vol) is added to the donor culture to induce expression of RapI and the cells are grown for an additional hour. Next, 2.5mL of donors and recipients at OD_{600} 0.9 are mixed, collected on a mating filter (Thermo Scientific 145-0020) by vacuum, and placed on a solid mating support consisting of 1.5% agar with Spiz+$MgCl_2$ (0.2 g NH4SO4, 1.4 g K2HPO4, 37.5 mg KH2PO4, 0.1 g Na3citrate-2H2O, 20 mg MgSO4-7H2O per liter + 125mM MgCl2) for 3 h at 37°C. Cells were then rinsed off the filter, diluted, and spread on plates with selective antibiotics and/or D-alanine to determine the numbers of transconjugates, donors, and recipients. For all conjugation assays, donor starter cultures and O/N plates contain antibiotics to prevent ICE*Bs1* excision/loss.

Slight modifications to the filter mating protocol were made for spot and liquid matings. For spot and liquid matings, starter cultures are diluted into 2 mL of medium in 15 mL culture tubes. 150 μL of donor and recipients at OD_{600} 0.9 are mixed and either incubated at 37°C for 3 h (liquid mating) or centrifuged at 12,000 r.p.m. for 1 min, resuspended in 15 μL of media, and spotted onto agar plates.

4-4-6. Nitrogenase activity assay.

Nitrogenase activity was determined by acetylene reduction as previously described. Starter cultures are diluted to OD600 0.025 in 2mL of medium in 24-deep well blocks and grown for 2.5 h at 250 r.p.m in an INFORS-HT Multitron Pro. The cells were then collected by centrifugation (1 min, 12000 r.p.m.) and resuspended in 2mL BB minimal media in 10mL glass vials with PTFE-silicone septa screw caps. Headspace in the bottles was repeatedly evacuated and flushed with argon gas using a vacuum manifold equipped with a copper catalyst O2 trap. Acetylene gas was freshly generated from CaC2 in a Burris bottle and 1mL was injected into each bottle to start the reaction. Cultures were incubated at 30°C, 250 r.p.m. for 22 h. Ethylene production was analyzed by gas chromatography on an Agilent 7890A GC system (Agilent Technologies, Inc.) equipped with a PAL headspace autosampler and flame ionization detector as follows. 250mL

headspace preincubated to 35°C was sampled and separated on a GS-CarbonPLOT column (0.32 mm x 30 m, 3 micron; Agilent) at 60°C and a He flow rate of 1.8 mL/min. Detection occurred in a FID head heated to 300°C with a gas flow of 35 mL/min H2 and 400 mL/min air. Under these conditions, acetylene eluted at 3.0 min after injection and ethylene at 3.65 min. Ethylene production was quantified by integrating the 3.65 min peak using Agilent GC/MSD ChemStation Software.

4-4-7. Fluorescence measurement.

Starter cultures are diluted to OD600 0.025 in 2mL of medium in 15 mL culture tubes and grown for 2.5 h at the same temperature and speed. After 2.5 h, the cells are diluted 1:200 into 200 μL media with different concentrations of IPTG (0, 1, 5, 10, 20, 50, 70, 100, 200, 500, 1000 μM). These cultures are grown for 3 h, then diluted 1:15 into phosphate buffered saline and measured using a BD Biosciences Fortessa flow cytometer with blue (488 nm) and red (640 nm) lasers. An injection volume of 10 μL and flow rate of 0.5 μL/s were used. Cytometry data was analyzed using FlowJo (TreeStar Inc., Ashland, OR) and populations were gated on forward and side scatter heights. The gated populations consisted of at least 30,000 cells. The median fluorescence of the gated populations was used calculated using FlowJo and used for all reporting.

Chapter 5: Conclusions

5-1: Utility of subtle regulatory methods for controlling gene expression

In natural systems, gene expression is regulated by an astounding number of variables, many of which are ignored by synthetic biologists. This is because a few specific regulatory sequences, *e.g.*, RBSs, promoters, terminators, can drastically change gene expression and synthetic biologists have primarily focused on maximizing changes in gene expression. However, as synthetic genetic systems become more intricate, genetic engineers will need to make more nuanced adjustments to transcription/translation. To achieve a finer level of control, additional regulatory mechanisms must be considered. Just as natural genetic systems use multiple mechanisms to precisely control gene expression, genetic engineers could layer regulatory methods to build synthetic genetic programs. Layering subtle regulatory mechanisms may allow genetic engineers to move away from extreme overexpression of protein and/or RNA regulators, *e.g.*, repressor proteins, activator proteins, RNA IN/OUT, CRISPR/dCas9, to create more robust genetic circuits. If genetically engineered bacteria are ever approved for commercial release, they need to be able to survive in the target environment and closely mimicking natural regulation may be the best way to reduce the metabolic burden placed on cells by synthetic programs. This should help genetically modified bacteria compete with wild type organisms. Thus, it will become important to characterize more natural forms of gene expression control so that they can be incorporated into synthetic designs.

One example of a biological variable that affects gene expression is DNA topology[466]. DNA can change from negatively to positively supercoiled or flip from B-form to Z-form[467]. Each form of DNA differentially affects transcription initiation and protein binding to DNA[466,468,469]. However, topology is largely ignored when genetic engineers design DNA. This is surprising because DNA topology can be changed using transcription, protein binding, and DNA looping[470], which are all processes dictated by DNA sequence. Therefore, genetic engineers could design DNA sequences that convert between topologies and use DNA structure to tune gene expression. Another biological variable that is largely ignored by genetic engineers is tRNA modification. tRNA modifications can affect the rate and fidelity of translation by altering ribosome-binding affinity and reducing misreads/frame shifts. tRNA modifications can be

induced by cell stress or growth rate and can change the translation rate of specific codons[471]. Therefore, when synthetic biologists 'recode' genes, they may be able to use codons recognized by differentially modified tRNAs to promote translation under specific conditions. Antisense transcription (discussed in Chapter 3) is another form of regulation that could be layered on top of existing modes of regulation to fine tune gene expression. Now that we have started to quantify the affect of antisense promoters on gene expression, we expect that antisense promoters can be used in creative ways to tune gene expression. For example, operon arrangement could now be used as a design parameter. Engineers could design constructs with convergent operons to create antisense transcription and coordinate expression of the two operons[300]. Alternatively, inducible antisense promoters could be embedded in the coding regions of genes to predictably differentially regulate expression. The advantage of these alternative methods for tuning gene expression is the ability to layer regulation. Antisense transcription, DNA topology, and codon usage can all be considered when designing constructs that use characterized promoters, RBSs, and terminators. As the desire to precisely tune gene expression increases, I believe that features such as DNA topology and tRNA modification rates will become more important.

5-2: ADDITIONAL FACTORS THAT AFFECT ANTISENSE TRANSCRIPTION

As discussed in Chapter 3, promoter strength can affect the amount of repression generated by antisense transcription. However, additional factors, such as RBS strength and promoter firing kinetics, may also affect antisense transcription-mediated repression. The ODE model developed in Chapter 3 predicts that polymerases fired from the forward promoter, *i.e.*, the promoter driving expression of the protein of interest, are more likely to survive collisions than those fired from the antisense promoter. We interpret this result to mean that ribosomes translating the forward mRNA prevent RNAPs fired from the forward promoter from backsliding/stalling and increase the number of RNAPs that survive collision. Therefore, we hypothesize that changes in the RBS alter the amount of repression. Weaker RBSs should increase repression by increasing susceptibility to collisions. Weaker RBSs could also increase fold repression by reducing the number of ribosomes present on the mRNA, thereby leaving more mRNA available for binding by the antisense RNA (asRNA). Preliminary experiments

with different RBSs show that RBS strength is inversely correlated with fold repression (Fig 5-1). However, additional experiments will be needed to identify the mechanism behind RBS-mediated changes in repression. These experiments may include single molecule techniques to measure the frequency of RNAP ejection following collision or additional gene expression experiments in which both the forward and antisense promoters drive expression of translated mRNAs.

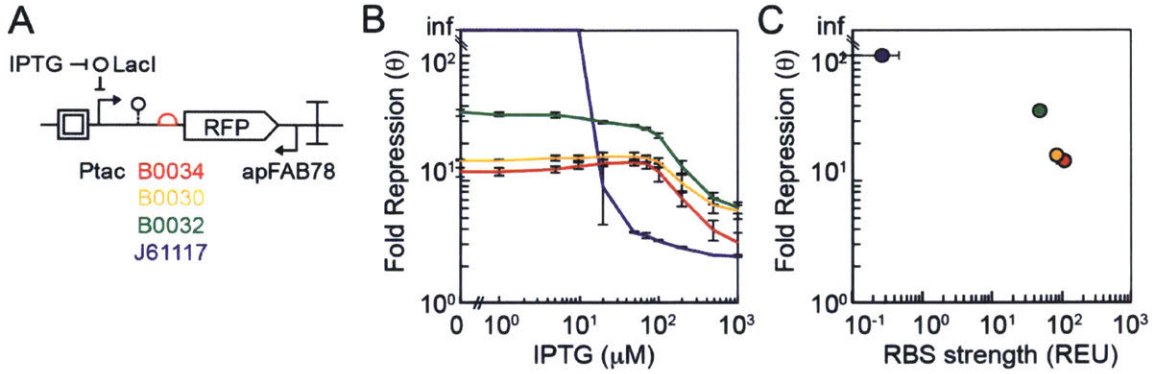

Figure 5-1: Impact of RBS sense on antisense transcription
(A) Antisense transcription reporter plasmids used to quantify repression. Plasmids are the same as in Fig 3-2, except RBS is varied. (B) Fold repression (Equation 3-1) is shown as a function of the induction of the forward promoter. The colors correspond to RBSs of different strength. (C) Maximum fold repression as a function of RBS strength. In all panels, data represents the mean of three experiments performed on different days and the error bars are the standard deviation of these replicates.

Promoter kinetics may also influence repression mediated by antisense transcription. Promoter kinetics can vary widely based on the ability of RNAP to bind, isomerize, and initiate productive transcription at a given DNA sequence[316]. Promoters that fire RNAPs in bursts may be less susceptible to antisense transcription than those that fire deterministically, *i.e.*, at regular intervals. To visualize why this it true, consider two promoters that both produce the same amount of mRNA in a given period of time (as measured by the amount of fluorescent protein produced in that period of time). One promoter is deterministic and one fires RNAPs in bursts. At the bursty promoter, several RNAPs will initiate transcription in quick succession[472]. This group of polymerases will transcribe the gene of interest and then fall off the DNA. Eventually the bursty promoter will fire again and another group of RNAPs will traverse the DNA. When compared to a deterministic promoter, the bursty promoter will create longer stretches of time during which no RNAPs are fired and the template strand of DNA is free of RNAP. During this

time, RNAPs fired from the antisense promoter will not cause collisions. Thus, for a given antisense promoter, genes expressed with bursty promoters will be less susceptible to repression via antisense transcription. Additionally, trailing RNAPs, *i.e.*, those closely following another actively transcribing RNAP, may also help prevent backsliding/stalling of the leading RNAP. Thus, promoter dynamics may affect the quantitative effect of antisense transcription.

5-3: MODIFICATIONS THAT MAY EXPAND THE HOST RANGE & CONJUGATION FREQUENCY OF ICE*Bs1*

There may be several ways to expand the host range of ICE*Bs1*. We have been able to conjugate ICE*Bs1* into half of the bacterial recipients tested so far (see Chapter 4). However, it would be better if we could use ICE*Bs1* to engineer more of the recipient strains. Ideally, we would be able to conjugate ICE*Bs1* into any Gram-positive bacteria. We believe that the host range of ICE*Bs1* is limited, in some cases, by mechanisms that can be circumvented with some simple engineering. For example, restriction modification systems may be preventing ICE*Bs1* DNA from persisting in recipient cells. Restriction modification systems degrade foreign DNA and bacteria these systems to prevent 'infection' by selfish DNA elements[473,474]. Conjugative plasmids and other broad host range DNA elements often carry use antirestriction modification proteins to avoid degradation by host enzymes[475]. Wild type ICE*Bs1* does not encode any antirestriction modification proteins, however one could easily add antirestriction modification proteins, such as Orf18 from Tn*916*, to the element[476]. This may increase the recipient host range by preventing DNA degradation.

Insufficient expression of the integrase Int, may be preventing integration of the ICE*Bs1* element into some of the recipient bacterias' genomes. If ICE*Bs1* does not integrate, it must rely on its ability to replicate like a plasmid to persist in the recipient bacteria. Unfortunately, the plasmid form of ICE*Bs1* is not as stable as the integrated form and it may be lost during recipient cell division. To increase expression of Int in the recipient, the integrase gene *int* could be placed under a broad host range promoter, such as Pint from Tn*916*. A synthetic system designed to

kick start transcription in any bacteria could also be used to express *int*[461]. Low integration efficiency may also be due to a dearth of integration sites in the recipient. The ICE*Bs1* integrase prefers to integrate at a specific DNA sequence (attB). When this sequence is missing in the recipient, ICE*Bs1* will integrate randomly in the bacteria's chromosome[455]. ICE*Bs1*'s random integration is much less efficient than site specific integration. Thus, poor conjugation efficiency may be due to poor integration efficiency caused by a lack of good integration sites in the recipient strain. To circumvent this, we could replace the integrase in ICE*Bs1* with a more promiscuous integrase or with a protein that targets a different sequence.

Second strand synthesis may also prevent successful expression of the ICE*Bs1* integrase. When ICE*Bs1* conjugates, it enters the recipient cell as single stranded DNA. The second strand of DNA is made by the host DNA polymerase, which recognizes a second strand origin (sso) on ICE*Bs1*. Since double stranded DNA is required for transcription and recombination with the host genome, it is essential for integration. Unfortunately, ssos can have a very narrow host range and we do not know how well the ICE*Bs1* sso will function in all recipient bacteria[462]. Additional ssos can be added to ICE*Bs1* to increase second strand synthesis in diverse bacterial species.

Finally, the antibiotic resistance cassettes that we are using to isolate transconjugates could be altered to potentially increase the known host range of ICE*Bs1*. As discussed in Chapter 4, we selected *tet(M)* and *cat* as the antibiotic resistance markers to use for our ICE*Bs1* conjugation experiments because they are from broad host range elements Tn*916* and pC194. However, it is possible that these resistance cassettes are not functional in the some of the recipient bacteria. If the resistance proteins are poorly expressed or nonfunctional in a recipient bacteria, we would not be able to isolate transconjugates, even if ICE*Bs1* was successfully conjugated and integrated, giving us a false positive result. Other antibiotic resistance markers or selection methods could be used to test ICE*Bs1* conjugation to potentially broaden it's known host range.

Before modifying ICE*Bs1* to increase host range, it is necessary to identify the mechanism(s) preventing ICE*Bs1* acquisition by potential recipient strains. If ICE*Bs1* is getting digested by a recipient's restriction modification system, then a new sso will not improve conjugation efficiency. It will be straightforward to test several hypotheses for inefficient ICE*Bs1* conjugation and these experiments should be done before building new ICE varients so that one does not waste time building solutions to the wrong problems.

Bibliography

1. Evans, A. S. & Brachman, P. S. *Bacterial Infections of Humans: Epidemiology and Control.* (Springer, 2013).
2. Cunha, B. A. The cause of the plague of Athens: plague, typhoid, typhus, smallpox, or measles? *Infect. Dis. Clin. North Am.* **18,** 29–43 (2004).
3. Bos, K. I. *et al.* A draft genome of Yersinia pestis from victims of the Black Death. *Nature* **478,** 506–510 (2011).
4. Cherry, J. D. Historical review of pertussis and the classical vaccine. *J. Infect. Dis.* **174 Suppl 3,** S259–263 (1996).
5. Halliday, S. Death and miasma in Victorian London: an obstinate belief. *BMJ* **323,** 1469–1471 (2001).
6. Crookshank, E. The history and present position of the germ theory of disease. *Public Health* **1,** 16–19 (1888).
7. Chung, K.-T. & Ferris, D. H. Marinus Willem Beijerinck (1851-1931): Pioneer of general microbiology. *ASM News* **62,** 539–543 (1996).
8. Burns, R. C. & Hardy, R. W. F. *Nitrogen Fixation in Bacteria and Higher Plants.* (Springer Science & Business Media, 2012).
9. Roh, J. Y., Choi, J. Y., Li, M. S., Jin, B. R. & Je, Y. H. Bacillus thuringiensis as a specific, safe, and effective tool for insect pest control. *J. Microbiol. Biotechnol.* **17,** 547–559 (2007).
10. Fujimura, K. E., Slusher, N. A., Cabana, M. D. & Lynch, S. V. Role of the gut microbiota in defining human health. *Expert Rev. Anti Infect. Ther.* **8,** 435–454 (2010).
11. O'Hara, A. M. & Shanahan, F. The gut flora as a forgotten organ. *EMBO Rep.* **7,** 688–693 (2006).
12. Cogen, A. L. *et al.* Selective Antimicrobial Action Is Provided by Phenol-Soluble Modulins Derived from Staphylococcus epidermidis, a Normal Resident of the Skin. *J. Invest. Dermatol.* **130,** 192–200 (2010).
13. Witkin, S. The vaginal microbiome, vaginal anti-microbial defence mechanisms and the clinical challenge of reducing infection-related preterm birth. *BJOG Int. J. Obstet. Gynaecol.* **122,** 213–218 (2015).
14. Ahrné *et al.* The normal Lactobacillus flora of healthy human rectal and oral mucosa. *J. Appl. Microbiol.* **85,** 88–94 (1998).
15. Farnworth, E. R. (Ted). *Handbook of Fermented Functional Foods, Second Edition.* (Taylor & Francis, 2008).
16. Chance, R. E. & Frank, B. H. Research, Development, Production, and Safety of Biosynthetic Human Insulin. *Diabetes Care* **16,** 133–142 (1993).
17. Paddon, C. J. & Keasling, J. D. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.* **12,** 355–367 (2014).
18. Ajikumar, P. K. *et al.* Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia coli. *Science* **330,** 70–74 (2010).
19. Truffer, F. *et al.* Compact portable biosensor for arsenic detection in aqueous samples with Escherichia coli bioreporter cells. *Rev. Sci. Instrum.* **85,** 015120 (2014).
20. Gupta, S., Bram, E. E. & Weiss, R. Genetically Programmable Pathogen Sense and Destroy. *ACS Synth. Biol.* **2,** 715–723 (2013).
21. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 6578–6583 (1998).
22. 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).

23. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335–338 (2000).
24. Marcobal, A., De las Rivas, B., Landete, J. M., Tabera, L. & Muñoz, R. Tyramine and phenylethylamine biosynthesis by food bacteria. *Crit. Rev. Food Sci. Nutr.* **52,** 448–467 (2012).
25. Kurland, C. G. & Dong, H. Bacterial growth inhibition by overproduction of protein. *Mol. Microbiol.* **21,** 1–4 (1996).
26. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10,** 354–360 (2013).
27. Chen, Y.-J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* **10,** 659–664 (2013).
28. Solomon, J. M. & Grossman, A. D. Who's competent and when: regulation of natural genetic competence in bacteria. *Trends Genet.* **12,** 150–155 (1996).
29. Chassy, B. M., Mercenier, A. & Flickinger, J. Transformation of bacteria by electroporation. *Trends Biotechnol.* **6,** 303–309 (1988).
30. Dunny, G. M., Lee, L. N. & LeBlanc, D. J. Improved electroporation and cloning vector system for gram-positive bacteria. *Appl. Environ. Microbiol.* **57,** 1194–1201 (1991).
31. Ando, H., Lemire, S., Pires, D. P. & Lu, T. K. Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing. *Cell Syst.* **1,** 187–196 (2015).
32. Errington, J. & Pughe, N. Upper limit for DNA packaging by Bacillus subtilis bacteriophage phi 105: isolation of phage deletion mutants by induction of oversized prophages. *Mol. Gen. Genet. MGG* **210,** 347–351 (1987).
33. Griffiths, A. J., Gelbart, W. M., Miller, J. H. & Lewontin, R. C. Bacterial Conjugation. (1999).
34. Brucker, R. M. *et al.* Amphibian Chemical Defense: Antifungal Metabolites of the Microsymbiont Janthinobacterium lividum on the Salamander Plethodon cinereus. *J. Chem. Ecol.* **34,** 1422–1429 (2008).
35. Ding, Y., Wang, J., Liu, Y. & Chen, S. Isolation and identification of nitrogen-fixing bacilli from plant rhizospheres in Beijing region. *J. Appl. Microbiol.* **99,** 1271–1281 (2005).
36. Clarke, G. *et al.* Minireview: Gut Microbiota: The Neglected Endocrine Organ. *Mol. Endocrinol.* **28,** 1221–1238 (2014).
37. Lyte, M. Microbial Endocrinology in the Microbiome-Gut-Brain Axis: How Bacterial Production and Utilization of Neurochemicals Influence Behavior. *PLoS Pathog.* **9,** (2013).
38. Sukumar, P. *et al.* Involvement of auxin pathways in modulating root architecture during beneficial plant–microorganism interactions. *Plant Cell Environ.* **36,** 909–919 (2013).
39. Dahl, R. H. *et al.* Engineering dynamic pathway regulation using stress-response promoters. *Nat. Biotechnol.* **31,** 1039–1046 (2013).
40. Moser, F. *et al.* Genetic Circuit Performance under Conditions Relevant for Industrial Bioreactors. *ACS Synth. Biol.* **1,** 555–564 (2012).
41. Holtz, W. J. & Keasling, J. D. Engineering Static and Dynamic Control of Synthetic Pathways. *Cell* **140,** 19–23 (2010).
42. Anesiadis, N., Kobayashi, H., Cluett, W. R. & Mahadevan, R. Analysis and design of a genetic circuit for dynamic metabolic engineering. *ACS Synth. Biol.* **2,** 442–452 (2013).
43. Zhang, F. & Keasling, J. Biosensors and their applications in microbial metabolic engineering. *Trends Microbiol.* **19,** 323–329 (2011).
44. Dietrich, J. A., Shis, D. L., Alikhani, A. & Keasling, J. D. Transcription Factor-Based Screens and Synthetic Selections for Microbial Small-Molecule Biosynthesis. *ACS Synth. Biol.* **2,** 47–58 (2013).
45. Schendzielorz, G. *et al.* Taking Control over Control: Use of Product Sensing in Single Cells to Remove Flux Control at Key Enzymes in Biosynthesis Pathways. *ACS Synth. Biol.* **3,** 21–29 (2014).

46. Zhang, F., Carothers, J. M. & Keasling, J. D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* **30,** 354–359 (2012).
47. Yi, T.-M., Huang, Y., Simon, M. I. & Doyle, J. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl. Acad. Sci.* **97,** 4649–4653 (2000).
48. Krishnanathan, K., Anderson, S. R., Billings, S. A. & Kadirkamanathan, V. A Data-Driven Framework for Identifying Nonlinear Dynamic Models of Genetic Parts. *ACS Synth. Biol.* **1,** 375–384 (2012).
49. Carbonell, P., Parutto, P., Baudier, C., Junot, C. & Faulon, J.-L. Retropath: Automated Pipeline for Embedded Metabolic Circuits. *ACS Synth. Biol.* (2013). doi:10.1021/sb4001273
50. Adams, B. L. *et al.* Evolved Quorum Sensing Regulator, LsrR, for Altered Switching Functions. *ACS Synth. Biol.* (2013). doi:10.1021/sb400068z
51. Umeyama, T., Okada, S. & Ito, T. Synthetic Gene Circuit-Mediated Monitoring of Endogenous Metabolites: Identification of GAL11 as a Novel Multicopy Enhancer of S-Adenosylmethionine Level in Yeast. *ACS Synth. Biol.* **2,** 425–430 (2013).
52. Stapleton, J. A. *et al.* Feedback Control of Protein Expression in Mammalian Cells by Tunable Synthetic Translational Inhibition. *ACS Synth. Biol.* **1,** 83–88 (2012).
53. Liu, D., Xiao, Y., Evans, B. & Zhang, F. Negative Feedback Regulation of Fatty Acid Production Based on a Malonyl-CoA Sensor–Actuator. *ACS Synth. Biol.* (2013). doi:10.1021/sb400158w
54. Siedler, S. *et al.* SoxR as a Single-Cell Biosensor for NADPH-Consuming Enzymes in Escherichia coli. *ACS Synth. Biol.* **3,** 41–47 (2014).
55. Medema, M. H., Breitling, R., Bovenberg, R. & Takano, E. Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. *Nat. Rev. Microbiol.* **9,** 131–137 (2011).
56. Fischbach, M. & Voigt, C. A. Prokaryotic gene clusters: A rich toolbox for synthetic biology. *Biotechnol. J.* **5,** 1277–1296 (2010).
57. Frasch, H.-J., Medema, M. H., Takano, E. & Breitling, R. Design-based re-engineering of biosynthetic gene clusters: plug-and-play in practice. *Curr. Opin. Biotechnol.* **24,** 1144–1150 (2013).
58. Temme, K., Zhao, D. & Voigt, C. A. Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proc. Natl. Acad. Sci.* **109,** 7085–7090 (2012).
59. Shao, Z. *et al.* Refactoring the Silent Spectinabilin Gene Cluster Using a Plug-and-Play Scaffold. *ACS Synth. Biol.* **2,** 662–669 (2013).
60. Oßwald, C. *et al.* Modular Construction of a Functional Artificial Epothilone Polyketide Pathway. *ACS Synth. Biol.* (2012). doi:10.1021/sb300080t
61. Steidler, L. *et al.* Treatment of Murine Colitis by Lactococcus lactis Secreting Interleukin-10. *Science* **289,** 1352–1355 (2000).
62. Anderson, J. C., Clarke, E. J., Arkin, A. P. & Voigt, C. A. Environmentally Controlled Invasion of Cancer Cells by Engineered Bacteria. *J. Mol. Biol.* **355,** 619–627 (2006).
63. Ruder, W. C., Lu, T. & Collins, J. J. Synthetic Biology Moving into the Clinic. *Science* **333,** 1248–1252 (2011).
64. Motta, J.-P. *et al.* Food-Grade Bacteria Expressing Elafin Protect Against Inflammation and Restore Colon Homeostasis. *Sci. Transl. Med.* **4,** 158ra144–158ra144 (2012).
65. Wang, S., Kong, Q. & Curtiss III, R. New technologies in developing recombinant attenuated Salmonella vaccine vectors. *Microb. Pathog.* **58,** 17–28 (2013).
66. Huh, J. H., Kittleson, J. T., Arkin, A. P. & Anderson, J. C. Modular Design of a Synthetic Payload Delivery Device. *ACS Synth. Biol.* **2,** 418–424 (2013).
67. Gupta, S., Bram, E. E. & Weiss, R. Genetically Programmable Pathogen Sense and Destroy. *ACS Synth. Biol.* **2,** 715–723 (2013).

68. Hwang, I. Y. *et al.* Reprogramming Microbes to Be Pathogen-Seeking Killers. *ACS Synth. Biol.* (2013). doi:10.1021/sb400077j
69. Prindle, A. *et al.* Genetic Circuits in Salmonella typhimurium. *ACS Synth. Biol.* **1,** 458–464 (2012).
70. Volzing, K., Borrero, J., Sadowsky, M. J. & Kaznessis, Y. N. Antimicrobial Peptides Targeting Gram-negative Pathogens, Produced and Delivered by Lactic Acid Bacteria. *ACS Synth. Biol.* **2,** 643–650 (2013).
71. Hasty, J. Engineered Microbes for Therapeutic Applications. *ACS Synth. Biol.* **1,** 438–439 (2012).
72. Danino, T., Lo, J., Prindle, A., Hasty, J. & Bhatia, S. N. In Vivo Gene Expression Dynamics of Tumor-Targeted Bacteria. *ACS Synth. Biol.* **1,** 465–470 (2012).
73. Archer, E. J., Robinson, A. B. & Süel, G. M. Engineered E. coli That Detect and Respond to Gut Inflammation through Nitric Oxide Sensing. *ACS Synth. Biol.* **1,** 451–457 (2012).
74. Antunes, M. S. *et al.* Programmable Ligand Detection System in Plants through a Synthetic Signal Transduction Pathway. *PLoS ONE* **6,** e16292 (2011).
75. Widmaier, D. M. *et al.* Engineering the Salmonella type III secretion system to export spider silk monomers. *Mol. Syst. Biol.* **5,** (2009).
76. Bernhardt, K. *et al.* New tools for self-organized pattern formation. *BMC Syst. Biol.* **1,** S10 (2007).
77. Xia, X.-X. *et al.* Native-sized recombinant spider silk protein produced in metabolically engineered Escherichia coli results in a strong fiber. *Proc. Natl. Acad. Sci.* (2010). doi:10.1073/pnas.1003366107
78. Widmaier, D. M. & Voigt, C. A. Quantification of the physiochemical constraints on the export of spider silk proteins by Salmonella type III secretion. *Microb. Cell Factories* **9,** 78 (2010).
79. Aquea, F. *et al.* A molecular framework for the inhibition of Arabidopsis root growth in response to boron toxicity. *Plant Cell Environ.* **35,** 719–734 (2012).
80. Antunes, M. S. *et al.* A synthetic de-greening gene circuit provides a reporting system that is remotely detectable and has a re-set capacity. *Plant Biotechnol. J.* **4,** 605–622 (2006).
81. Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **10,** 410–422 (2009).
82. Khoury, G. A., Smadbeck, J., Kieslich, C. A. & Floudas, C. A. Protein folding and de novo protein design for biotechnological applications. *Trends Biotechnol.* doi:10.1016/j.tibtech.2013.10.008
83. Lewis, N. E., Nagarajan, H. & Palsson, B. O. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* **10,** 291–305 (2012).
84. Peralta-Yahya, P. P. *et al.* Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.* **2,** 483 (2011).
85. Jonnalagadda, S. B., Becker, J. U., Sel'kov, E. E. & Betz, A. Flux regulation in glycogen-induced oscillatory glycolysis in cell-free extracts of Saccharomyces carlsbergensis. *Biosystems* **15,** 49–58 (1982).
86. Loo, L. W. M. *et al.* cis-Expression QTL Analysis of Established Colorectal Cancer Risk Variants in Colon Tumors and Adjacent Normal Tissue. *PLoS ONE* **7,** e30477 (2012).
87. Klavins, E. Proportional-integral control of stochastic gene regulatory networks. in *2010 49th IEEE Conference on Decision and Control (CDC)* 2547–2553 (2010). doi:10.1109/CDC.2010.5717525
88. Bernard, P. & Couturier, M. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J. Mol. Biol.* **226,** 735–745 (1992).
89. Weiss, R. Cellular computation and communications using engineered genetic regulatory networks. (Massachusetts Institute of Technology, 2001).
90. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403,** 339–342 (2000).
91. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27,** 946–950 (2009).

92. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10,** 354–360 (2013).
93. Cambray, G. *et al.* Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res.* **41,** 5139–5148 (2013).
94. Rodrigo, G. & Jaramillo, A. AutoBioCAD: Full Biodesign Automation of Genetic Circuits. *ACS Synth. Biol.* **2,** 230–236 (2013).
95. Voigt, C. A. Genetic parts to program bacteria. *Curr. Opin. Biotechnol.* **17,** 548–557 (2006).
96. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci.* **99,** 16587–16591 (2002).
97. Ellefson, J. W. *et al.* Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat. Biotechnol.* **advance online publication,** (2013).
98. 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).
99. Haseltine, E. L. & Arnold, F. H. Synthetic Gene Circuits: Design with Directed Evolution. *Annu. Rev. Biophys. Biomol. Struct.* **36,** 1–19 (2007).
100. Collins, C. H., Arnold, F. H. & Leadbetter, J. R. Directed evolution of Vibrio fischeri LuxR for increased sensitivity to a broad spectrum of acyl-homoserine lactones. *Mol. Microbiol.* **55,** 712–723 (2005).
101. Sleight, S. C. & Sauro, H. M. Randomized BioBrick Assembly: A Novel DNA Assembly Method for Randomizing and Optimizing Genetic Circuits and Metabolic Pathways. *ACS Synth. Biol.* **2,** 506–518 (2013).
102. Shong, J. & Collins, C. H. Engineering the esaR Promoter for Tunable Quorum Sensing-Dependent Gene Expression. *ACS Synth. Biol.* **2,** 568–575 (2013).
103. Balagaddé, F. K., You, L., Hansen, C. L., Arnold, F. H. & Quake, S. R. Long-Term Monitoring of Bacteria Undergoing Programmed Population Control in a Microchemostat. *Science* **309,** 137–140 (2005).
104. Cardinale, S. & Arkin, A. P. Contextualizing context for synthetic biology – identifying causes of failure of synthetic biological systems. *Biotechnol. J.* **7,** 856–866 (2012).
105. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS ONE* **4,** e5553 (2009).
106. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
107. Hillson, N. J., Rosengarten, R. D. & Keasling, J. D. j5 DNA Assembly Design Automation Software. *ACS Synth. Biol.* **1,** 14–21 (2012).
108. Leguia, M., Brophy, J. A., Densmore, D., Asante, A. & Anderson, J. C. 2ab assembly: a methodology for automatable, high-throughput assembly of standard biological parts. *J. Biol. Eng.* **7,** 2 (2013).
109. Kok, S. de *et al.* Rapid and Reliable DNA Assembly via Ligase Cycling Reaction. *ACS Synth. Biol.* (2014). doi:10.1021/sb4001992
110. Paetzold, B., Carolis, C., Ferrar, T., Serrano, L. & Lluch-Senar, M. In Situ Overlap and Sequence Synthesis During DNA Assembly. *ACS Synth. Biol.* **2,** 750–755 (2013).
111. Clancy, K. & Voigt, C. A. Programming cells: towards an automated 'Genetic Compiler'. *Curr. Opin. Biotechnol.* **21,** 572–581 (2010).
112. Friedland, A. E. *et al.* Synthetic Gene Networks That Count. *Science* **324,** 1199–1202 (2009).
113. Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456,** 516–519 (2008).
114. Stanton, B. C. *et al.* Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.* **advance online publication,** (2013).
115. Endy, D. Foundations for engineering biology. *Nature* **438,** 449–453 (2005).

116. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nat. Rev. Genet.* **11,** 367–379 (2010).
117. Liang, J. C., Bloom, R. J. & Smolke, C. D. Engineering Biological Systems with Synthetic RNA Molecules. *Mol. Cell* **43,** 915–926 (2011).
118. Lim, W. A. Designing customized cell signalling circuits. *Nat. Rev. Mol. Cell Biol.* **11,** 393–403 (2010).
119. Weber, W. & Fussenegger, M. Synthetic gene networks in mammalian cells. *Curr. Opin. Biotechnol.* **21,** 690–696 (2010).
120. Liu, W., Yuan, J. S. & Stewart Jr, C. N. Advanced genetic tools for plant biotechnology. *Nat. Rev. Genet.* **14,** 781–793 (2013).
121. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* gkt520 (2013). doi:10.1093/nar/gkt520
122. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152,** 1173–1183 (2013).
123. Liu, C. C. *et al.* An adaptor from translational to transcriptional control enables predictable assembly of complex regulation. *Nat. Methods* **9,** 1088–1094 (2012).
124. Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B. & Arkin, A. P. Rationally designed families of orthogonal RNA regulators of translation. *Nat. Chem. Biol.* **8,** 447–454 (2012).
125. Beerli, R. R. & Barbas, C. F. Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol.* **20,** 135–141 (2002).
126. Garg, A., Lohmueller, J. J., Silver, P. A. & Armel, T. Z. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res.* **40,** 7584–7595 (2012).
127. Moscou, M. J. & Bogdanove, A. J. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science* **326,** 1501–1501 (2009).
128. Takeda, Y., Folkmanis, A. & Echols, H. Cro regulatory protein specified by bacteriophage lambda. Structure, DNA-binding, and repression of RNA synthesis. *J. Biol. Chem.* **252,** 6177–6183 (1977).
129. Ptashne, M. & Hopkins, N. The operators controlled by the lambda phage repressor. *Proc. Natl. Acad. Sci. U. S. A.* **60,** 1282–1287 (1968).
130. Zhan, J. *et al.* Develop reusable and combinable designs for transcriptional logic gates. *Mol. Syst. Biol.* **6,** (2010).
131. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335–338 (2000).
132. Guet, C. C., Elowitz, M. B., Hsing, W. & Leibler, S. Combinatorial Synthesis of Genetic Networks. *Science* **296,** 1466–1470 (2002).
133. Hasty, J., Dolnik, M., Rottschäfer, V. & Collins, J. J. Synthetic Gene Network for Entraining and Amplifying Cellular Oscillations. *Phys. Rev. Lett.* **88,** 148101 (2002).
134. Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 3581–3586 (2005).
135. Gaber, R. *et al.* Designable DNA-binding domains enable construction of logic circuits in mammalian cells. *Nat. Chem. Biol.* **advance online publication,** (2014).
136. Lohmueller, J. J., Armel, T. Z. & Silver, P. A. A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. *Nucleic Acids Res.* **40,** 5180–5187 (2012).
137. Peacock, R. W. S., Sullivan, K. A. & Wang, C. L. Tetracycline-Regulated Expression Implemented through Transcriptional Activation Combined with Proximal and Distal Repression. *ACS Synth. Biol.* **1,** 156–162 (2012).
138. Mercer, A. C., Gaj, T., Sirk, S. J., Lamb, B. M. & Barbas, C. F. Regulation of Endogenous Human Gene Expression by Ligand-Inducible TALE Transcription Factors. *ACS Synth. Biol.* (2013). doi:10.1021/sb400114p

139. Purcell, O., Peccoud, J. & Lu, T. K. Rule-Based Design of Synthetic Transcription Factors in Eukaryotes. *ACS Synth. Biol.* (2013). doi:10.1021/sb400134k
140. Lienert, F. *et al.* Two- and three-input TALE-based AND logic computation in embryonic stem cells. *Nucleic Acids Res.* gkt758 (2013). doi:10.1093/nar/gkt758
141. 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 Res.* gks597 (2012). doi:10.1093/nar/gks597
142. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A System for the Continuous Directed Evolution of Biomolecules. *Nature* **472,** 499–503 (2011).
143. Rhodius, V. A. *et al.* Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. *Mol. Syst. Biol.* **9,** (2013).
144. Wang, B., Kitney, R. I., Joly, N. & Buck, M. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nat. Commun.* **2,** 508 (2011).
145. Anderson, J. C., Voigt, C. A. & Arkin, A. P. Environmental signal integration by a modular AND gate. *Mol. Syst. Biol.* **3,** (2007).
146. Daniel, R., Rubens, J. R., Sarpeshkar, R. & Lu, T. K. Synthetic analog computation in living cells. *Nature* **497,** 619–623 (2013).
147. Buchler, N. E., Gerland, U. & Hwa, T. On schemes of combinatorial transcription logic. *Proc. Natl. Acad. Sci.* **100,** 5136–5141 (2003).
148. Calles, B. & Lorenzo, V. de. Expanding the Boolean Logic of the Prokaryotic Transcription Factor XylR by Functionalization of Permissive Sites with a Protease-Target Sequence. *ACS Synth. Biol.* **2,** 594–603 (2013).
149. Ramalingam, K. I., Tomshine, J. R., Maynard, J. A. & Kaznessis, Y. N. Forward engineering of synthetic bio-logical AND gates. *Biochem. Eng. J.* **47,** 38–47 (2009).
150. Lou, C. *et al.* Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch. *Mol. Syst. Biol.* **6,** (2010).
151. Regot, S. *et al.* Distributed biological computation with multicellular engineered networks. *Nature* **469,** 207–211 (2011).
152. Ausländer, S., Ausländer, D., Müller, M., Wieland, M. & Fussenegger, M. Programmable single-cell mammalian biocomputers. *Nature* **487,** 123–127 (2012).
153. Tamsir, A., Tabor, J. J. & Voigt, C. A. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* **469,** 212–215 (2011).
154. Shis, D. L. & Bennett, M. R. Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proc. Natl. Acad. Sci.* **110,** 5028–5033 (2013).
155. Basu, S., Mehreja, R., Thiberge, S., Chen, M.-T. & Weiss, R. Spatiotemporal control of gene expression with pulse-generating networks. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 6355–6360 (2004).
156. Chen, D. & Arkin, A. P. Sequestration-based bistability enables tuning of the switching boundaries and design of a latch. *Mol. Syst. Biol.* **8,** (2012).
157. Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. Development of Genetic Circuitry Exhibiting Toggle Switch or Oscillatory Behavior in Escherichia coli. *Cell* **113,** 597–607 (2003).
158. Fung, E. *et al.* A synthetic gene–metabolic oscillator. *Nature* **435,** 118–122 (2005).
159. Tigges, M., Dénervaud, N., Greber, D., Stelling, J. & Fussenegger, M. A synthetic low-frequency mammalian oscillator. *Nucleic Acids Res.* **38,** 2702–2711 (2010).
160. Argos, P. *et al.* The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J.* **5,** 433–440 (1986).
161. Gopaul, D. N. & Van Duyne, G. D. Structure and mechanism in site-specific recombination. *Curr. Opin. Struct. Biol.* **9,** 14–20 (1999).

162. Ham, T. S., Lee, S. K., Keasling, J. D. & Arkin, A. P. A tightly regulated inducible expression system utilizing the fim inversion recombination switch. *Biotechnol. Bioeng.* **94,** 1–4 (2006).
163. Ham, T. S., Lee, S. K., Keasling, J. D. & Arkin, A. P. Design and Construction of a Double Inversion Recombination Switch for Heritable Sequential Genetic Memory. *PLoS ONE* **3,** e2815 (2008).
164. Moon, T. S. *et al.* Construction of a Genetic Multiplexer to Toggle between Chemosensory Pathways in Escherichia coli. *J. Mol. Biol.* **406,** 215–227 (2011).
165. Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P. & Endy, D. Amplifying Genetic Logic Gates. *Science* **340,** 599–603 (2013).
166. Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31,** 448–452 (2013).
167. Bonnet, J., Subsoontorn, P. & Endy, D. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci.* (2012). doi:10.1073/pnas.1202344109
168. Sorek, R., Lawrence, C. M. & Wiedenheft, B. CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea. *Annu. Rev. Biochem.* **82,** 237–266 (2013).
169. Sashital, D. G., Wiedenheft, B. & Doudna, J. A. Mechanism of Foreign DNA Selection in a Bacterial Adaptive Immune System. *Mol. Cell* **46,** 606–615 (2012).
170. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31,** 833–838 (2013).
171. Farzadfard, F., Perli, S. D. & Lu, T. K. Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas. *ACS Synth. Biol.* **2,** 604–613 (2013).
172. Gilbert, L. A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154,** 442–451 (2013).
173. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10,** 977–979 (2013).
174. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10,** 973–976 (2013).
175. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8,** 2180–2196 (2013).
176. Hussein, R. & Lim, H. N. Direct comparison of small RNA and transcription factor signaling. *Nucleic Acids Res.* **40,** 7269–7279 (2012).
177. Levine, E., Zhang, Z., Kuhlman, T. & Hwa, T. Quantitative Characteristics of Gene Regulation by Small RNA. *PLoS Biol* **5,** e229 (2007).
178. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507,** 62–67 (2014).
179. Nishimasu, H. *et al.* Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156,** 935–949 (2014).
180. Del Vecchio, D., Ninfa, A. J. & Sontag, E. D. Modular cell biology: retroactivity and insulation. *Mol. Syst. Biol.* **4,** (2008).
181. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10,** 1116–1121 (2013).
182. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).
183. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–832 (2013).
184. Gossen, M., Bonin, A. L. & Bujard, H. Control of gene activity in higher eukaryotic cells by prokaryotic regulatory elements. *Trends Biochem. Sci.* **18,** 471–475 (1993).

185. Simons, R. W. & Kleckner, N. Translational control of IS10 transposition. *Cell* **34,** 683–691 (1983).
186. Kittle, J. D., Simons, R. W., Lee, J. & Kleckner, N. Insertion sequence IS10 anti-sense pairing initiates by an interaction between the 5′ end of the target RNA and a loop in the anti-sense RNA. *J. Mol. Biol.* **210,** 561–572 (1989).
187. Ma, C. & Simons, R. W. The IS10 antisense RNA blocks ribosome binding at the transposase translation initiation site. *EMBO J.* **9,** 1267–1274 (1990).
188. Qi, L., Lucks, J. B., Liu, C. C., Mutalik, V. K. & Arkin, A. P. Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. *Nucleic Acids Res.* gks168 (2012). doi:10.1093/nar/gks168
189. Liu, C. C., Qi, L., Yanofsky, C. & Arkin, A. P. Regulation of transcription by unnatural amino acids. *Nat. Biotechnol.* **29,** 164–168 (2011).
190. Callura, J. M., Dwyer, D. J., Isaacs, F. J., Cantor, C. R. & Collins, J. J. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proc. Natl. Acad. Sci.* 201009747 (2010). doi:10.1073/pnas.1009747107
191. Nielsen, A. A., Segall-Shapiro, T. H. & Voigt, C. A. Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression. *Curr. Opin. Chem. Biol.* doi:10.1016/j.cbpa.2013.10.003
192. Bintu, L. *et al.* Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.* **15,** 116–124 (2005).
193. Bintu, L. *et al.* Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.* **15,** 125–135 (2005).
194. Voigt, C. A., Wolf, D. M. & Arkin, A. P. The Bacillus subtilis sin Operon An Evolvable Network Motif. *Genetics* **169,** 1187–1202 (2005).
195. Strogatz, S. H. *Nonlinear dynamics and chaos: with application to physics, biology, chemistry, and engineering.* (Westview press, 2000).
196. Ang, J., Harris, E., Hussey, B. J., Kil, R. & McMillen, D. R. Tuning Response Curves for Synthetic Biology. *ACS Synth. Biol.* **2,** 547–567 (2013).
197. Gottesman, S. Proteases and Their Targets in Escherichia Coli1. *Annu. Rev. Genet.* **30,** 465–506 (1996).
198. Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V. & Ebright, R. H. Structural Organization of the RNA Polymerase-Promoter Open Complex. *Cell* **101,** 601–611 (2000).
199. Davis, J. H., Rubin, A. J. & Sauer, R. T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res.* **39,** 1131–1141 (2011).
200. Kittleson, J. T., Cheung, S. & Anderson, J. C. Rapid optimization of gene dosage in E. coli using DIAL strains. *J. Biol. Eng.* **5,** 10 (2011).
201. Gottesman, S., Roche, E., Zhou, Y. & Sauer, R. T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* **12,** 1338–1347 (1998).
202. Cox, R. S., Surette, M. G. & Elowitz, M. B. Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.* **3,** (2007).
203. Chen, S. *et al.* Automated Design of Genetic Toggle Switches with Predetermined Bistability. *ACS Synth. Biol.* **1,** 284–290 (2012).
204. Koshland, D. E., Goldbeter, A. & Stock, J. B. Amplification and adaptation in regulatory and sensory systems. *Science* **217,** 220–225 (1982).
205. Vilar, J. M. & Saiz, L. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr. Opin. Genet. Dev.* **15,** 136–144 (2005).
206. Johnson, S., Lindén, M. & Phillips, R. Sequence dependence of transcription factor-mediated DNA looping. *Nucleic Acids Res.* gks473 (2012). doi:10.1093/nar/gks473

207. Legewie, S., Dienst, D., Wilde, A., Herzel, H. & Axmann, I. M. Small RNAs Establish Delays and Temporal Thresholds in Gene Expression. *Biophys. J.* **95,** 3232–3238 (2008).
208. Buchler, N. E. & Cross, F. R. Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol. Syst. Biol.* **5,** (2009).
209. Lu, M. S., Mauser, J. F. & Prehoda, K. E. Ultrasensitive Synthetic Protein Regulatory Networks Using Mixed Decoys. *ACS Synth. Biol.* **1,** 65–72 (2012).
210. Lee, T.-H. & Maheshri, N. A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Mol. Syst. Biol.* **8,** (2012).
211. Cookson, N. A. *et al.* Queueing up for enzymatic processing: correlated signaling through coupled degradation. *Mol. Syst. Biol.* **7,** (2011).
212. Shen, J., Liu, Z., Zheng, W., Xu, F. & Chen, L. Oscillatory dynamics in a simple gene regulatory network mediated by small RNAs. *Phys. Stat. Mech. Its Appl.* **388,** 2995–3000 (2009).
213. Borujeni, A. E., Channarasappa, A. S. & Salis, H. M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* gkt1139 (2013). doi:10.1093/nar/gkt1139
214. Geyer, P. K. The role of insulator elements in defining domains of gene expression. *Curr. Opin. Genet. Dev.* **7,** 242–248 (1997).
215. Lou, C., Stanton, B., Chen, Y.-J., Munsky, B. & Voigt, C. A. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol.* **30,** 1137–1142 (2012).
216. Jeong, W. & Kang, C. Start site selection at lacUV5 promoter affected by the sequence context around the initiation sites. *Nucleic Acids Res.* **22,** 4667–4672 (1994).
217. Walker, K. A. & Osuna, R. Factors Affecting Start Site Selection at the Escherichia coli fis Promoter. *J. Bacteriol.* **184,** 4783–4791 (2002).
218. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-Sequence Determinants of Gene Expression in Escherichia coli. *Science* **324,** 255–258 (2009).
219. Goodman, D. B., Church, G. M. & Kosuri, S. Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science* **342,** 475–479 (2013).
220. Kosuri, S. *et al.* Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proc. Natl. Acad. Sci.* **110,** 14024–14029 (2013).
221. Qi, L., Haurwitz, R. E., Shao, W., Doudna, J. A. & Arkin, A. P. RNA processing enables predictable programming of gene expression. *Nat. Biotechnol.* **30,** 1002–1006 (2012).
222. Chen, Y.-J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* **10,** 659–664 (2013).
223. Yao, A. I. *et al.* Promoter Element Arising from the Fusion of Standard BioBrick Parts. *ACS Synth. Biol.* **2,** 111–120 (2013).
224. Villalobos, A., Ness, J. E., Gustafsson, C., Minshull, J. & Govindarajan, S. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* **7,** 1–8 (2006).
225. Rhodius, V. A., Mutalik, V. K. & Gross, C. A. Predicting the strength of UP-elements and full-length E. coli σE promoters. *Nucleic Acids Res.* **40,** 2907–2924 (2012).
226. Brewster, R. C., Jones, D. L. & Phillips, R. Tuning Promoter Strength through RNA Polymerase Binding Site Design in Escherichia coli. *PLoS Comput Biol* **8,** e1002811 (2012).
227. Weller, K. & Recknagel, R.-D. Promoter Strength Prediction Based on Occurrence Frequencies of Consensus Patterns. *J. Theor. Biol.* **171,** 355–359 (1994).
228. Seo, S. W. *et al.* Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metab. Eng.* **15,** 67–74 (2013).
229. de Hoon, M. J. L., Makita, Y., Nakai, K. & Miyano, S. Prediction of Transcriptional Terminators in Bacillus subtilis and Related Species. *PLoS Comput Biol* **1,** e25 (2005).

230. Lesnik, E. A. *et al.* Prediction of rho-independent transcriptional terminators in Escherichia coli. *Nucleic Acids Res.* **29,** 3583–3594 (2001).
231. Rackham, O. & Chin, J. W. A network of orthogonal ribosome·mRNA pairs. *Nat. Chem. Biol.* **1,** 159–166 (2005).
232. 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).
233. Lovett, S. T., Hurley, R. L., Sutera, V. A., Aubuchon, R. H. & Lebedeva, M. A. Crossing Over Between Regions of Limited Homology in Escherichia coli: RecA-Dependent and RecA-Independent Pathways. *Genetics* **160,** 851–859 (2002).
234. Sleight, S. C., Bartley, B. A., Lieviant, J. A. & Sauro, H. M. Designing and engineering evolutionary robust genetic circuits. *J. Biol. Eng.* **4,** 12 (2010).
235. Arkin, A. P. & Fletcher, D. A. Fast, cheap and somewhat in control. *Genome Biol.* **7,** 114 (2006).
236. Dong, H., Nilsson, L. & Kurland, C. G. Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. *J. Bacteriol.* **177,** 1497–1504 (1995).
237. Mather, W. H., Hasty, J., Tsimring, L. S. & Williams, R. J. Translational Cross Talk in Gene Networks. *Biophys. J.* **104,** 2564–2572 (2013).
238. Grigorova, I. L., Phleger, N. J., Mutalik, V. K. & Gross, C. A. Insights into transcriptional regulation and σ competition from an equilibrium model of RNA polymerase binding to DNA. *Proc. Natl. Acad. Sci.* **103,** 5332–5337 (2006).
239. Levine, J. H., Lin, Y. & Elowitz, M. B. Functional Roles of Pulsing in Genetic Circuits. *Science* **342,** 1193–1200 (2013).
240. 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).
241. Jayanthi, S., Nilgiriwala, K. S. & Del Vecchio, D. Retroactivity Controls the Temporal Dynamics of Gene Transcription. *ACS Synth. Biol.* **2,** 431–441 (2013).
242. Cardinale, S., Joachimiak, M. P. & Arkin, A. P. Effects of Genetic Variation on the E. coli Host-Circuit Interface. *Cell Rep.* **4,** 231–237 (2013).
243. Canton, B., Labno, A. & Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* **26,** 787–793 (2008).
244. Tagami, H., Inada, T., Kunimura, T. & Aiba, H. Glucose lowers CRP* levels resulting in repression of the lac operon in cells lacking cAMP. *Mol. Microbiol.* **17,** 251–258 (1995).
245. Purcell, O., Grierson, C. S., Bernardo, M. di & Savery, N. J. Temperature dependence of ssrA-tag mediated protein degradation. *J. Biol. Eng.* **6,** 10 (2012).
246. Kelly, J. R. *et al.* Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng.* **3,** 4 (2009).
247. Cho, B.-K., Charusanti, P., Herrgård, M. J. & Palsson, B. Ø. Microbial regulatory and metabolic networks. *Curr. Opin. Biotechnol.* **18,** 360–364 (2007).
248. Yaman, F., Bhatia, S., Adler, A., Densmore, D. & Beal, J. Automated Selection of Synthetic Biology Parts for Genetic Regulatory Networks. *ACS Synth. Biol.* **1,** 332–344 (2012).
249. Bhatia, S. & Densmore, D. Pigeon: A Design Visualizer for Synthetic Biology. *ACS Synth. Biol.* **2,** 348–350 (2013).
250. Huynh, L., Tsoukalas, A., Köppe, M. & Tagkopoulos, I. SBROME: A Scalable Optimization and Module Matching Framework for Automated Biosystems Design. *ACS Synth. Biol.* **2,** 263–273 (2013).
251. Roehner, N. & Myers, C. J. A Methodology to Annotate Systems Biology Markup Language Models with the Synthetic Biology Open Language. *ACS Synth. Biol.* (2013). doi:10.1021/sb400066m

252. Arkin, A. P. A wise consistency: engineering biology for conformity, reliability, predictability. *Curr. Opin. Chem. Biol.* **17,** 893–901 (2013).
253. Ceroni, F., Furini, S., Stefan, A., Hochkoeppler, A. & Giordano, E. A Synthetic Post-transcriptional Controller To Explore the Modular Design of Gene Circuits. *ACS Synth. Biol.* **1,** 163–171 (2012).
254. Seo, J.-H. *et al.* Multiple-omic data analysis of Klebsiella pneumoniae MGH 78578 reveals its transcriptional architecture and regulatory features. *BMC Genomics* **13,** 679 (2012).
255. Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* **7,** 1534–1550 (2012).
256. Becker, A. H., Oh, E., Weissman, J. S., Kramer, G. & Bukau, B. Selective ribosome profiling as a tool for studying the interaction of chaperones and targeting factors with nascent polypeptide chains and ribosomes. *Nat. Protoc.* **8,** 2212–2239 (2013).
257. Shin, J. & Noireaux, V. An E. coli Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. *ACS Synth. Biol.* **1,** 29–41 (2012).
258. Siegal-Gaskins, D., Noireaux, V. & Murray, R. M. Biomolecular resource utilization in elementary cell-free gene circuits. *ArXiv13070178 Q-Bio* (2013).
259. Karzbrun, E., Shin, J., Bar-Ziv, R. H. & Noireaux, V. Coarse-Grained Dynamics of Protein Synthesis in a Cell-Free System. *Phys. Rev. Lett.* **106,** 048104 (2011).
260. Noireaux, V., Bar-Ziv, R. & Libchaber, A. Principles of cell-free genetic circuit assembly. *Proc. Natl. Acad. Sci.* **100,** 12672–12677 (2003).
261. Karig, D. K., Jung, S.-Y., Srijanto, B., Collier, C. P. & Simpson, M. L. Probing Cell-Free Gene Expression Noise in Femtoliter Volumes. *ACS Synth. Biol.* **2,** 497–505 (2013).
262. Lentini, R. *et al.* Fluorescent Proteins and in Vitro Genetic Organization for Cell-Free Synthetic Biology. *ACS Synth. Biol.* **2,** 482–489 (2013).
263. Davidson, E. A., Meyer, A. J., Ellefson, J. W., Levy, M. & Ellington, A. D. An in vitro Autogene. *ACS Synth. Biol.* **1,** 190–196 (2012).
264. Niederholtmeyer, H., Xu, L. & Maerkl, S. J. Real-Time mRNA Measurement during an in Vitro Transcription and Translation Reaction Using Binary Probes. *ACS Synth. Biol.* **2,** 411–417 (2013).
265. Chizzolini, F., Forlin, M., Cecchi, D. & Mansy, S. S. Gene Position More Strongly Influences Cell-Free Protein Expression from Operons than T7 Transcriptional Promoter Strength. *ACS Synth. Biol.* (2013). doi:10.1021/sb4000977
266. Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V. & Murray, R. M. Linear DNA for Rapid Prototyping of Synthetic Biological Circuits in an Escherichia coli Based TX-TL Cell-Free System. *ACS Synth. Biol.* (2013). doi:10.1021/sb400131a
267. Shearwin, K. E., Callen, B. P. & Egan, J. B. Transcriptional interference – a crash course. *Trends Genet. TIG* **21,** 339–345 (2005).
268. Pruss, G. J. & Drlica, K. DNA supercoiling and prokaryotic transcription. *Cell* **56,** 521–523 (1989).
269. Chen, C.-C. & Wu, H.-Y. in *Gene Expression and Regulation* (ed. Ma, J.) 469–480 (Springer New York, 2006).
270. Georg, J. & Hess, W. R. cis-Antisense RNA, Another Level of Gene Regulation in Bacteria. *Microbiol. Mol. Biol. Rev.* **75,** 286–300 (2011).
271. Helmrich, A., Ballarino, M., Nudler, E. & Tora, L. Transcription-replication encounters, consequences and genomic instability. *Nat. Struct. Mol. Biol.* **20,** 412–418 (2013).
272. Vilette, D., Ehrlich, S. D. & Michel, B. Transcription-induced deletions in Escherichia coli plasmids. *Mol. Microbiol.* **17,** 493–504 (1995).
273. French, S. Consequences of replication fork movement through transcription units in vivo. *Science* **258,** 1362–1365 (1992).

274. Sharma, C. M. *et al.* The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464,** 250–255 (2010).
275. Singh, S. S. *et al.* Widespread suppression of intragenic transcription initiation by H-NS. *Genes Dev.* (2014). doi:10.1101/gad.234336.113
276. Saida, F., Uzan, M., Odaert, B. & Bontems, F. Expression of Highly Toxic Genes in E. coli: Special Strategies and Genetic Tools. *Curr. Protein Pept. Sci.* **7,** 47–56 (2006).
277. O'Connor, C. D. & Timmis, K. N. Highly repressible expression system for cloning genes that specify potentially toxic proteins. *J. Bacteriol.* **169,** 4457–4462 (1987).
278. Worrall, A. F. & Connolly, B. A. The chemical synthesis of a gene coding for bovine pancreatic DNase I and its cloning and expression in Escherichia coli. *J. Biol. Chem.* **265,** 21889–21895 (1990).
279. Fozo, E. M., Hemm, M. R. & Storz, G. Small Toxic Proteins and the Antisense RNAs That Repress Them. *Microbiol. Mol. Biol. Rev.* **72,** 579–589 (2008).
280. Chatterjee, A. *et al.* Convergent transcription confers a bistable switch in Enterococcus faecalis conjugation. *Proc. Natl. Acad. Sci.* **108,** 9721–9726 (2011).
281. Chatterjee, A. *et al.* Convergent Transcription in the Butyrolactone Regulon in Streptomyces coelicolor Confers a Bistable Genetic Switch for Antibiotic Biosynthesis. *PLoS ONE* **6,** e21974 (2011).
282. Hongay, C. F., Grisafi, P. L., Galitski, T. & Fink, G. R. Antisense Transcription Controls Cell Fate in Saccharomyces cerevisiae. *Cell* **127,** 735–745 (2006).
283. Wurtzel, O. *et al.* A single-base resolution map of an archaeal transcriptome. *Genome Res.* **20,** 133–141 (2010).
284. Wade, J. T. & Grainger, D. C. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat. Rev. Microbiol.* **12,** 647–653 (2014).
285. Selinger, D. W. *et al.* RNA expression analysis using a 30 base pair resolution Escherichia coli genome array. *Nat. Biotechnol.* **18,** 1262–1268 (2000).
286. Filiatrault, M. J. *et al.* Transcriptome Analysis of Pseudomonas syringae Identifies New Genes, Noncoding RNAs, and Antisense Activity. *J. Bacteriol.* **192,** 2359–2372 (2010).
287. Georg, J. *et al.* Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Mol. Syst. Biol.* **5,** 305 (2009).
288. Güell, M. *et al.* Transcriptome Complexity in a Genome-Reduced Bacterium. *Science* **326,** 1268–1271 (2009).
289. Dujon, B. The yeast genome project: what did we learn? *Trends Genet.* **12,** 263–270 (1996).
290. Yelin, R. *et al.* Widespread occurrence of antisense transcription in the human genome. *Nat. Biotechnol.* **21,** 379–386 (2003).
291. He, Y., Vogelstein, B., Velculescu, V. E., Papadopoulos, N. & Kinzler, K. W. The Antisense Transcriptomes of Human Cells. *Science* **322,** 1855–1857 (2008).
292. Thomason, M. K. *et al.* Global Transcriptional Start Site Mapping Using Differential RNA Sequencing Reveals Novel Antisense RNAs in Escherichia coli. *J. Bacteriol.* **197,** 18–28 (2015).
293. Tutukina, M. N., Shavkunov, K. S., Masulis, I. S. & Ozoline, O. N. Intragenic promotor-like sites in the genome of escherichia coli discovery and functional implication. *J. Bioinform. Comput. Biol.* **05,** 549–560 (2007).
294. Dornenburg, J. E., DeVita, A. M., Palumbo, M. J. & Wade, J. T. Widespread Antisense Transcription in Escherichia coli. *mBio* **1,** e00024–10 (2010).
295. Peters, J. M. *et al.* Rho and NusG suppress pervasive antisense transcription in Escherichia coli. *Genes Dev.* **26,** 2621–2633 (2012).
296. Raghavan, R., Sloan, D. B. & Ochman, H. Antisense Transcription Is Pervasive but Rarely Conserved in Enteric Bacteria. *mBio* **3,** e00156–12 (2012).

297. Nicolas, P. *et al.* Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in Bacillus subtilis. *Science* **335,** 1103–1106 (2012).
298. Beaume, M. *et al.* Cartography of Methicillin-Resistant S. aureus Transcripts: Detection, Orientation and Temporal Expression during Growth Phase and Stress Conditions. *PLoS ONE* **5,** e10725 (2010).
299. Thomason, M. K. & Storz, G. Bacterial antisense RNAs: How many are there and what are they doing? *Annu. Rev. Genet.* **44,** 167–188 (2010).
300. Sesto, N., Wurtzel, O., Archambaud, C., Sorek, R. & Cossart, P. The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. *Nat. Rev. Microbiol.* **11,** 75–82 (2013).
301. Mason, E., Henderson, M. W., Scheller, E. V., Byrd, M. S. & Cotter, P. A. Evidence for phenotypic bistability resulting from transcriptional interference of bvgAS in Bordetella bronchiseptica. *Mol. Microbiol.* **90,** 716–733 (2013).
302. Kawano, M., Aravind, L. & Storz, G. An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Mol. Microbiol.* **64,** 738–754 (2007).
303. Giangrossi, M. *et al.* A novel antisense RNA regulates at transcriptional level the virulence gene icsA of Shigella flexneri. *Nucleic Acids Res.* **38,** 3362–3375 (2010).
304. Lee, E.-J. & Groisman, E. A. An antisense RNA that governs the expression kinetics of a multifunctional virulence gene. *Mol. Microbiol.* **76,** 1020–1033 (2010).
305. Liu, Y. & Kobayashi, I. Negative Regulation of the EcoRI Restriction Enzyme Gene Is Associated with Intragenic Reverse Promoters. *J. Bacteriol.* **189,** 6928–6935 (2007).
306. Lee, T.-H. & Maheshri, N. A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Mol. Syst. Biol.* **8,** 576 (2012).
307. Rhodius, V. A. *et al.* Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. *Mol. Syst. Biol.* **9,** 702 (2013).
308. Buchler, N. E. & Cross, F. R. Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol. Syst. Biol.* **5,** 272 (2009).
309. Chen, D. & Arkin, A. P. Sequestration-based bistability enables tuning of the switching boundaries and design of a latch. *Mol. Syst. Biol.* **8,** 620 (2012).
310. Hirakawa, H., Harwood, C. S., Pechter, K. B., Schaefer, A. L. & Greenberg, E. P. Antisense RNA that affects Rhodopseudomonas palustris quorum-sensing signal receptor expression. *Proc. Natl. Acad. Sci.* **109,** 12141–12146 (2012).
311. Tutukina, M. N., Shavkunov, K. S., Masulis, I. S. & Ozoline, O. N. Antisense transcription within the hns locus of Escherichia coli. *Mol. Biol.* **44,** 439–447 (2010).
312. Brantl, S. & Wagner, E. G. H. An Antisense RNA-Mediated Transcriptional Attenuation Mechanism Functions in Escherichia coli. *J. Bacteriol.* **184,** 2740–2747 (2002).
313. Brantl, S. Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr. Opin. Microbiol.* **10,** 102–109 (2007).
314. Callen, B. P., Shearwin, K. E. & Egan, J. B. Transcriptional Interference between Convergent Promoters Caused by Elongation over the Promoter. *Mol. Cell* **14,** 647–656 (2004).
315. Palmer, A. C., Ahlgren-Berg, A., Egan, J. B., Dodd, I. B. & Shearwin, K. E. Potent Transcriptional Interference by Pausing of RNA Polymerases over a Downstream Promoter. *Mol. Cell* **34,** 545–555 (2009).
316. Sneppen, K. *et al.* A Mathematical Model for Transcriptional Interference by RNA Polymerase Traffic in Escherichia coli. *J. Mol. Biol.* **346,** 399–409 (2005).
317. André, G. *et al.* S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of Clostridium acetobutylicum. *Nucleic Acids Res.* **36,** 5955–5969 (2008).
318. Tian, J. *et al.* Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432,** 1050–1054 (2004).

319. Sharon, E. *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* **30,** 521–530 (2012).
320. Kosuri, S. *et al.* Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proc. Natl. Acad. Sci.* **110,** 14024–14029 (2013).
321. Goodman, D. B., Church, G. M. & Kosuri, S. Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science* **342,** 475–479 (2013).
322. Lou, C., Stanton, B., Chen, Y.-J., Munsky, B. & Voigt, C. A. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol.* **30,** 1137–1142 (2012).
323. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nat. Methods* **11,** 508–520 (2014).
324. Liang, S.-T. *et al.* Activities of constitutive promoters in Escherichia coli1. *J. Mol. Biol.* **292,** 19–37 (1999).
325. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci.* **99,** 16587–16591 (2002).
326. Eiamphungporn, W. & Helmann, J. D. Extracytoplasmic Function σ Factors Regulate Expression of the Bacillus subtilis yabE Gene via a cis-Acting Antisense RNA. *J. Bacteriol.* **191,** 1101–1105 (2009).
327. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11,** 499–507 (2014).
328. Stanton, B. C. *et al.* Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.* **10,** 99–105 (2014).
329. Raveh-Sadka, T. *et al.* Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nat. Genet.* **44,** 743–750 (2012).
330. 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).
331. Deininger, P. Molecular cloning: A laboratory manual: 2nd ed. Edited by J. Sambrook, E. F. Fritsch, and T. Maniatis. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989 (in 3 volumes). *Anal. Biochem.* **186,** 182–183 (1990).
332. Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B. & Arkin, A. P. Rationally designed families of orthogonal RNA regulators of translation. *Nat. Chem. Biol.* **8,** 447–454 (2012).
333. Lucks, J. B., Qi, L., Mutalik, V. K., Wang, D. & Arkin, A. P. Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 8617–8622 (2011).
334. Brennan, R. G. & Link, T. M. Hfq structure, function and ligand binding. *Curr. Opin. Microbiol.* **10,** 125–133 (2007).
335. Stazic, D., Lindell, D. & Steglich, C. Antisense RNA protects mRNA from RNase E degradation by RNA–RNA duplex formation during phage infection. *Nucleic Acids Res.* **39,** 4890–4899 (2011).
336. Crampton, N., Bonass, W. A., Kirkham, J., Rivetti, C. & Thomson, N. H. Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucleic Acids Res.* **34,** 5416–5425 (2006).
337. Roberts, J. & Park, J.-S. Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination. *Curr. Opin. Microbiol.* **7,** 120–125 (2004).
338. Nudler, E. RNA Polymerase Backtracking in Gene Regulation and Genome Instability. *Cell* **149,** (2012).
339. Proshkin, S., Rahmouni, A. R., Mironov, A. & Nudler, E. Cooperation Between Translating Ribosomes and RNA Polymerase in Transcription Elongation. *Science* **328,** 504–508 (2010).
340. Epshtein, V. & Nudler, E. Cooperation Between RNA Polymerase Molecules in Transcription Elongation. *Science* **300,** 801–805 (2003).

341. Epshtein, V., Toulmé, F., Rahmouni, A. R., Borukhov, S. & Nudler, E. Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO J.* **22,** 4719–4727 (2003).
342. Kaplan, D. L. & O'Donnell, M. Rho Factor: Transcription Termination in Four Steps. *Curr. Biol.* **13,** R714–R716 (2003).
343. Vogel, U. & Jensen, K. F. The RNA chain elongation rate in Escherichia coli depends on the growth rate. *J. Bacteriol.* **176,** 2807–2813 (1994).
344. Ma, N. & McAllister, W. T. In a Head-on Collision, Two RNA Polymerases Approaching One Another on the Same DNA May Pass by One Another. *J. Mol. Biol.* **391,** 808–812 (2009).
345. Hobson, D. J., Wei, W., Steinmetz, L. M. & Svejstrup, J. Q. RNA Polymerase II Collision Interrupts Convergent Transcription. *Mol. Cell* **48,** 365–374 (2012).
346. Nielsen, A. A., Segall-Shapiro, T. H. & Voigt, C. A. Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression. *Curr. Opin. Chem. Biol.* **17,** 878–892 (2013).
347. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27,** 946–950 (2009).
348. Na, D. *et al.* Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nat. Biotechnol.* **31,** 170–174 (2013).
349. Carrier, T. A. & Keasling, J. D. Engineering mRNA stability in E. coli by the addition of synthetic hairpins using a 5′ cassette system. *Biotechnol. Bioeng.* **55,** 577–580 (1997).
350. Carothers, J. M., Goler, J. A., Juminaga, D. & Keasling, J. D. Model-Driven Engineering of RNA Devices to Quantitatively Program Gene Expression. *Science* **334,** 1716–1719 (2011).
351. Cox, R. S., Surette, M. G. & Elowitz, M. B. Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.* **3,** 145 (2007).
352. Stevens, J. T. & Carothers, J. M. Designing RNA-Based Genetic Control Systems for Efficient Production from Engineered Metabolic Pathways. *ACS Synth. Biol.* **4,** 107–115 (2015).
353. Gowrishankar, J., Leela, J. K. & Anupama, K. R-loops in bacterial transcription. *Transcription* **4,** 153–157 (2013).
354. Llopis, P. M. *et al.* Spatial organization of the flow of genetic information in bacteria. *Nature* **466,** 77–81 (2010).
355. Kawano, M., Storz, G., Rao, B. S., Rosner, J. L. & Martin, R. G. Detection of low-level promoter activity within open reading frame sequences of Escherichia coli. *Nucleic Acids Res.* **33,** 6268–6276 (2005).
356. Price, M. N., Huang, K. H., Arkin, A. P. & Alm, E. J. Operon formation is driven by co-regulation and not by horizontal gene transfer. *Genome Res.* **15,** 809–819 (2005).
357. Nevozhay, D., Zal, T. & Balázsi, G. Transferring a synthetic gene circuit from yeast to mammalian cells. *Nat. Commun.* **4,** 1451 (2013).
358. Murphy, K. F., Adams, R. M., Wang, X., Balázsi, G. & Collins, J. J. Tuning and controlling gene expression noise in synthetic gene networks. *Nucleic Acids Res.* **38,** 2712–2726 (2010).
359. Lasa, I. *et al.* Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc. Natl. Acad. Sci.* **108,** 20172–20177 (2011).
360. Kuczynski, J. *et al.* Experimental and analytical tools for studying the human microbiome. *Nat. Rev. Genet.* **13,** 47–58 (2012).
361. Round, J. L. & Mazmanian, S. K. The gut microbiome shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* **9,** 313–323 (2009).
362. Keijser, B. J. F. *et al.* Pyrosequencing Analysis of the Oral Microflora of Healthy Adults. *J. Dent. Res.* **87,** 1016–1020 (2008).
363. Torsvik, V. & Øvreås, L. Microbial diversity and function in soil: from genes to ecosystems. *Curr. Opin. Microbiol.* **5,** 240–245 (2002).

364. Costello, E. K. *et al.* Bacterial Community Variation in Human Body Habitats Across Space and Time. *Science* **326,** 1694–1697 (2009).
365. Spor, A., Koren, O. & Ley, R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat. Rev. Microbiol.* **9,** 279–290 (2011).
366. Behnsen, J., Deriu, E., Sassone-Corsi, M. & Raffatellu, M. Probiotics: Properties, Examples, and Specific Applications. *Cold Spring Harb. Perspect. Med.* **3,** (2013).
367. Steidler, L. *et al.* Treatment of Murine Colitis by Lactococcus lactis Secreting Interleukin-10. *Science* **289,** 1352–1355 (2000).
368. Mimee, M., Tucker, A. C., Voigt, C. A. & Lu, T. K. Programming a Human Commensal Bacterium, Bacteroides thetaiotaomicron, to Sense and Respond to Stimuli in the Murine Gut Microbiota. *Cell Syst.* **1,** 62–71 (2015).
369. Foo, J. L. *et al.* Improving Microbial Biogasoline Production in Escherichia coli Using Tolerance Engineering. *mBio* **5,** e01932–14 (2014).
370. Wery, J., Mendes da Silva, D. I. & de Bont, J. A. A genetically modified solvent-tolerant bacterium for optimized production of a toxic fine chemical. *Appl. Microbiol. Biotechnol.* **54,** 180–185 (2000).
371. Xue, G.-P., Johnson, J. S. & Dalrymple, B. P. High osmolarity improves the electro-transformation efficiency of the gram-positive bacteria Bacillus subtilis and Bacillus licheniformis. *J. Microbiol. Methods* **34,** 183–191 (1999).
372. Yeo, S.-K. & Liong, M.-T. Effect of electroporation on viability and bioconversion of isoflavones in mannitol-soymilk fermented by lactobacilli and bifidobacteria. *J. Sci. Food Agric.* **93,** 396–409 (2013).
373. NEB 10-beta Competent E. coli (High Efficiency) | NEB. Available at: https://www.neb.com/products/c3019-neb-10-beta-competent-e-coli-high-efficiency. (Accessed: 13th April 2016)
374. del Solar, G., Alonso, J. C., Espinosa, M. & Díaz-Orejas, R. Broad-host-range plasmid replication: an open question. *Mol. Microbiol.* **21,** 661–666 (1996).
375. Del Solar, G. & Espinosa, M. Plasmid copy number control: an ever-growing story. *Mol. Microbiol.* **37,** 492–500 (2000).
376. Smith, M. A. & Bidochka, M. J. Bacterial fitness and plasmid loss: the importance of culture conditions and plasmid size. *Can. J. Microbiol.* **44,** 351–355 (1998).
377. Sørensen, S. J., Bailey, M., Hansen, L. H., Kroer, N. & Wuertz, S. Studying plasmid horizontal transfer in situ: a critical review. *Nat. Rev. Microbiol.* **3,** 700–710 (2005).
378. Wozniak, R. A. F. & Waldor, M. K. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.* **8,** 552–563 (2010).
379. Roberts, A. P. & Mullany, P. A modular master on the move: the Tn916 family of mobile genetic elements. *Trends Microbiol.* **17,** 251–258 (2009).
380. Auchtung, J. M., Lee, C. A., Monson, R. E., Lehman, A. P. & Grossman, A. D. Regulation of a Bacillus subtilis mobile genetic element by intercellular signaling and the global DNA damage response. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 12554–12559 (2005).
381. Waldor, M. K., Tschäpe, H. & Mekalanos, J. J. A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in Vibrio cholerae O139. *J. Bacteriol.* **178,** 4157–4165 (1996).
382. Rodríguez-Blanco, A., Lemos, M. L. & Osorio, C. R. Integrating Conjugative Elements as Vectors of Antibiotic, Mercury, and Quaternary Ammonium Compound Resistance in Marine Aquaculture Environments. *Antimicrob. Agents Chemother.* **56,** 2619–2626 (2012).
383. Sullivan, J. T. & Ronson, C. W. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc. Natl. Acad. Sci.* **95,** 5145–5149 (1998).

384. Drenkard, E. & Ausubel, F. M. Pseudomonas biofilm formation and antibiotic resistance are linked to phenotypic variation. *Nature* **416,** 740–743 (2002).
385. Trieu-Cuot, P., Carlier, C., Martin, P. & Courvalin, P. Plasmid transfer by conjugation from Escherichia coli to Gram-positive bacteria. *FEMS Microbiol. Lett.* **48,** 289–294 (1987).
386. Auchtung, J. M., Lee, C. A., Garrison, K. L. & Grossman, A. D. Identification and characterization of the immunity repressor (ImmR) that controls the mobile genetic element ICEBs1 of Bacillus subtilis. *Mol. Microbiol.* **64,** 1515–1528 (2007).
387. Lee, C. A., Auchtung, J. M., Monson, R. E. & Grossman, A. D. Identification and characterization of int (integrase), xis (excisionase) and chromosomal attachment sites of the integrative and conjugative element ICEBs1 of Bacillus subtilis. *Mol. Microbiol.* **66,** 1356–1369 (2007).
388. Lee, C. A. & Grossman, A. D. Identification of the origin of transfer (oriT) and DNA relaxase required for conjugation of the integrative and conjugative element ICEBs1 of Bacillus subtilis. *J. Bacteriol.* **189,** 7254–7261 (2007).
389. Thomas, J., Lee, C. A. & Grossman, A. D. A Conserved Helicase Processivity Factor Is Needed for Conjugation and Replication of an Integrative and Conjugative Element. *PLoS Genet* **9,** e1003198 (2013).
390. Leonetti, C. T. *et al.* Critical Components of the Conjugation Machinery of the Integrative and Conjugative Element ICEBs1 of Bacillus subtilis. *J. Bacteriol.* **197,** 2558–2567 (2015).
391. Wright, L. D., Johnson, C. M. & Grossman, A. D. Identification of a Single Strand Origin of Replication in the Integrative and Conjugative Element ICEBs1 of Bacillus subtilis. *PLoS Genet* **11,** e1005556 (2015).
392. Himeno, T., Imanaka, T. & Aiba, S. Nucleotide sequence of the penicillinase repressor gene penI of Bacillus licheniformis and regulation of penP and penI by the repressor. *J. Bacteriol.* **168,** 1128–1132 (1986).
393. Bhavsar, A. P., Zhao, X. & Brown, E. D. Development and Characterization of a Xylose-Dependent System for Expression of Cloned Genes inBacillus subtilis: Conditional Complementation of a Teichoic Acid Mutant. *Appl. Environ. Microbiol.* **67,** 403–410 (2001).
394. Babic, A., Berkmen, M. B., Lee, C. A. & Grossman, A. D. Efficient Gene Transfer in Bacterial Cell Chains. *mBio* **2,** e00027–11 (2011).
395. Anagnostopoulos, C. & Spizizen, J. REQUIREMENTS FOR TRANSFORMATION IN BACILLUS SUBTILIS1. *J. Bacteriol.* **81,** 741–746 (1961).
396. Harwood, C. R. & Cutting, S. M. *Molecular biological methods for Bacillus.* (Wiley, 1990).
397. Johnson, C. M. & Grossman, A. D. The Composition of the Cell Envelope Affects Conjugation in Bacillus subtilis. *J. Bacteriol.* **198,** 1241–1249 (2016).
398. Dagert, M. & Ehrlich, S. D. Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. *Gene* **6,** 23–28 (1979).
399. Bhatty, M. *et al.* Enterococcus faecalis pCF10-encoded surface proteins PrgA, PrgB (aggregation substance) and PrgC contribute to plasmid transfer, biofilm formation and virulence. *Mol. Microbiol.* **95,** 660–677 (2015).
400. Binet, R. & Maurelli, A. T. Fitness Cost Due to Mutations in the 16S rRNA Associated with Spectinomycin Resistance in Chlamydia psittaci 6BC. *Antimicrob. Agents Chemother.* **49,** 4455–4464 (2005).
401. Fromm, H., Edelman, M., Aviv, D. & Galun, E. The molecular basis for rRNA-dependent spectinomycin resistance in Nicotiana chloroplasts. *EMBO J.* **6,** 3233–3237 (1987).
402. Walsh, C. T. Enzymes in the D-alanine branch of bacterial cell wall peptidoglycan assembly. *J. Biol. Chem.* **264,** 2393–2396 (1989).
403. Reichmann, N. T., Cassona, C. P. & Gründling, A. Revised mechanism of d-alanine incorporation into cell wall polymers in Gram-positive bacteria. *Microbiology* **159,** 1868–1877 (2013).

404. Ferrari, E., Henner, D. J. & Yang, M. Y. Isolation of an Alanine Racemase Gene from Bacillus subtilis and its Use for Plasmid Maintenance in B. subtilis. *Nat. Biotechnol.* **3,** 1003–1007 (1985).
405. Shukuo, K., Takashi, N. & Masanaru, M. Process for the manufacture of 5'-purine nucleotide by the fermentation method. (1966).
406. Kiyoshi, N., Zenroku, S. & Haruo, T. Process for the production of ribosylphosphates of 8-azapurine derivatives by fermentation. (1967).
407. Carvalho, M. da G. S. *et al.* Enterococcus caccae sp. nov., isolated from human stools. *Int. J. Syst. Evol. Microbiol.* **56,** 1505–1508 (2006).
408. Franz, C. M. A. P., Huch, M., Abriouel, H., Holzapfel, W. & Gálvez, A. Enterococci as probiotics and their implications in food safety. *Int. J. Food Microbiol.* **151,** 125–140 (2011).
409. Rönkä, E. *et al.* Probiotic and milk technological properties of Lactobacillus brevis. *Int. J. Food Microbiol.* **83,** 63–74 (2003).
410. Azcarate-Peril, M. A. *et al.* Analysis of the Genome Sequence of Lactobacillus gasseri ATCC 33323 Reveals the Molecular Basis of an Autochthonous Intestinal Organism. *Appl. Environ. Microbiol.* **74,** 4610–4625 (2008).
411. Kitazawa, H. *et al.* Expression of mRNA encoding IFNα in macrophages stimulated with Lactobacillus gasseri. *FEMS Microbiol. Lett.* **120,** 315–321 (1994).
412. Capra, M. l., Mercanti, D. j., Rossetti, L. c., Reinheimer, J. a. & Quiberoni, A. Isolation and phenotypic characterization of Lactobacillus casei and Lactobacillus paracasei bacteriophage-resistant mutants. *J. Appl. Microbiol.* **111,** 371–381 (2011).
413. Azevedo, M. S. P. *et al.* Lactobacillus acidophilus and L. reuteri modulate cytokine responses in gnotobiotic pigs infected with human rotavirus. *Benef. Microbes* **3,** 33–42 (2012).
414. Schlegel, L. *et al.* Streptococcus infantarius sp. nov., Streptococcus infantarius subsp. infantarius subsp. nov. and Streptococcus infantarius subsp. coli subsp. nov., isolated from humans and food. *Int. J. Syst. Evol. Microbiol.* **50,** 1425–1434 (2000).
415. Kleerebezem, M. *et al.* Complete genome sequence of Lactobacillus plantarum WCFS1. *Proc. Natl. Acad. Sci.* **100,** 1990–1995 (2003).
416. Berge, O., Heulin, T. & Balandreau, J. Diversity of diazotroph populations in the rhizosphere of maize (Zea mays L.) growing on different French soils. *Biol. Fertil. Soils* **11,** 210–215 (1991).
417. von der Weid, I., Paiva, E., Nóbrega, A., Dirk van Elsas, J. & Seldin, L. Diversity of Paenibacillus polymyxa strains isolated from the rhizosphere of maize planted in Cerrado soil. *Res. Microbiol.* **151,** 369–381 (2000).
418. Parke, J. L. Biological inoculant effective against aphanomyces. (1996).
419. Weid, I. von der, Duarte, G. F., Elsas, J. D. van & Seldin, L. Paenibacillus brasilensis sp. nov., a novel nitrogen-fixing species isolated from the maize rhizosphere in Brazil. *Int. J. Syst. Evol. Microbiol.* **52,** 2147–2153 (2002).
420. Berge, O., Guinebretière, M.-H., Achouak, W., Normand, P. & Heulin, T. Paenibacillus graminis sp. nov. and Paenibacillus odorifer sp. nov., isolated from plant roots, soil and food. *Int. J. Syst. Evol. Microbiol.* **52,** 607–616 (2002).
421. Kado, C. I., Schnathorst, W. C. & Azad, H. R. Method of using Bacillus polymyxa 9A to protect plants against verticillium wilt. (1987).
422. Stanton, C., Ross, R. P., Fitzgerald, G. F. & Sinderen, D. V. Fermented functional foods based on probiotics and their biogenic metabolites. *Curr. Opin. Biotechnol.* **16,** 198–203 (2005).
423. Kimoto, H., Kurisaki, J., Tsuji, N. M., Ohmomo, S. & Okamoto, T. Lactococci as probiotic strains: adhesion to human enterocyte-like Caco-2 cells and tolerance to low pH and bile. *Lett. Appl. Microbiol.* **29,** 313–316 (1999).
424. Seldin, L., Elsas, J. D. V. & Penido, E. G. C. Bacillus azotofixans sp. nov., a Nitrogen-Fixing Species from Brazilian Soils and Grass Roots. *Int. J. Syst. Bacteriol.* **34,** 451–456 (1984).

425. Jeong, H. *et al.* Draft Genome Sequence of the Paenibacillus polymyxa Type Strain (ATCC 842T), a Plant Growth-Promoting Bacterium. *J. Bacteriol.* **193,** 5026–5027 (2011).
426. Smith, N. R., Gibson, T., Gordon, R. E. & Sneath, P. H. TYPE CULTURES AND PROPOSED NEOTYPE CULTURES OF SOME SPECIES IN THE GENUS BACILLUS. *J. Gen. Microbiol.* **34,** 269–272 (1964).
427. Rosado, A. S., van Elsas, J. D. & Seldin, L. Reclassification of Paenibacillus durum (formerly Clostridium durum Smith and Cato 1974) Collins et al. 1994 as a member of the species P. azotofixans (formerly Bacillus azotofixans Seldin et al. 1984) Ash et al. 1994. *Int. J. Syst. Bacteriol.* **47,** 569–572 (1997).
428. Eppinger, M. *et al.* Genome Sequences of the Biotechnologically Important Bacillus megaterium Strains QM B1551 and DSM319 ▿. *J. Bacteriol.* **193,** 4199–4213 (2011).
429. Deng, Y. *et al.* Complete Genome Sequence of Bacillus subtilis BSn5, an Endophytic Bacterium of Amorphophallus konjac with Antimicrobial Activity for the Plant Pathogen Erwinia carotovora subsp. carotovora. *J. Bacteriol.* **193,** 2070–2071 (2011).
430. Rooney, A. P., Price, N. P. J., Ehrhardt, C., Swezey, J. L. & Bannan, J. D. Phylogeny and molecular taxonomy of the Bacillus subtilis species complex and description of Bacillus subtilis subsp. inaquosorum subsp. nov. *Int. J. Syst. Evol. Microbiol.* **59,** 2429–2436 (2009).
431. Zhang, J. *et al.* Isolation and characterization of plant growth-promoting rhizobacteria from wheat roots by wheat germ agglutinin labeled with fluorescein isothiocyanate. *J. Microbiol.* **50,** 191–198 (2012).
432. Ruiu, L. Brevibacillus laterosporus, a Pathogen of Invertebrates and a Broad-Spectrum Antimicrobial Species. *Insects* **4,** 476–492 (2013).
433. Borriss, R. *et al.* Extracellular phytase activity of Bacillus amyloliquefaciens FZB45 contributes to its plant-growth-promoting effect a. *Microbiology* **148,** 2097–2109 (2002).
434. Idris, E. E., Iglesias, D. J., Talon, M. & Borriss, R. Tryptophan-Dependent Production of Indole-3-Acetic Acid (IAA) Affects Level of Plant Growth Promotion by Bacillus amyloliquefaciens FZB42. *Mol. Plant. Microbe Interact.* **20,** 619–626 (2007).
435. Duc, L. H., Hong, H. A., Barbosa, T. M., Henriques, A. O. & Cutting, S. M. Characterization of Bacillus Probiotics Available for Human Use. *Appl. Environ. Microbiol.* **70,** 2161–2171 (2004).
436. Green, D. H. *et al.* Characterization of Two Bacillus Probiotics. *Appl. Environ. Microbiol.* **65,** 4288–4291 (1999).
437. Madhaiyan, M., Poonguzhali, S., Kwon, S.-W. & Sa, T.-M. Bacillus methylotrophicus sp. nov., a methanol-utilizing, plant-growth-promoting bacterium isolated from rice rhizosphere soil. *Int. J. Syst. Evol. Microbiol.* **60,** 2490–2495 (2010).
438. Zeigler, D. R. Genome Sequence of Bacillus glycinifermentans TH008, Isolated from Ohio Soil. *Genome Announc.* **4,** (2016).
439. Birrer, G. A., Cromwick, A.-M. & Gross, R. A. γ-Poly(glutamic acid) formation by Bacillus licheniformis 9945a: physiological and biochemical studies. *Int. J. Biol. Macromol.* **16,** 265–275 (1994).
440. Drahos, D. J. & West, L. Bacillus licheniformis biofungicide. (2004).
441. Holzinger, A., Nagendra-Prasad, D. & Huys, G. Plant protection potential and ultrastructure of Bacillus subtilis strain 3A25. *Crop Prot.* **30,** 739–744 (2011).
442. Guidi, V., Lehner, A., Lüthy, P. & Tonolla, M. Dynamics of Bacillus thuringiensis var. israelensis and Lysinibacillus sphaericus Spores in Urban Catch Basins after Simultaneous Application against Mosquito Larvae. *PLOS ONE* **8,** e55658 (2013).
443. Myers, P., Yousten, A. A. & Davidson, E. W. Comparative studies of the mosquito-larval toxin of Bacillus sphaericus SSII-1 and 1593. *Can. J. Microbiol.* **25,** 1227–1231 (1979).

444. Ingham, C. J. & Jacob, E. B. Swarming and complex pattern formation in Paenibacillus vortex studied by imaging and tracking cells. *BMC Microbiol.* **8,** 36 (2008).
445. Nakamura, L. K., Shida, O., Takagi, H. & Komagata, K. Bacillus pycnus sp. nov. and Bacillus neidei sp. nov., round-spored bacteria from soil. *Int. J. Syst. Evol. Microbiol.* **52,** 501–505 (2002).
446. Logan, N. A. *et al.* Bacillus shackletonii sp. nov., from volcanic soil on Candlemas Island, South Sandwich archipelago. *Int. J. Syst. Evol. Microbiol.* **54,** 373–376 (2004).
447. Salvetti, S., Celandroni, F., Ghelardi, E., Baggiani, A. & Senesi, S. Rapid determination of vitamin B2 secretion by bacteria growing on solid media. *J. Appl. Microbiol.* **95,** 1255–1260 (2003).
448. Koehler, T. M., Dai, Z. & Kaufman-Yarbray, M. Regulation of the Bacillus anthracis protective antigen gene: CO2 and a trans-acting element activate transcription from one of two promoters. *J. Bacteriol.* **176,** 586–595 (1994).
449. Smith, J. L., Goldberg, J. M. & Grossman, A. D. Complete Genome Sequences of Bacillus subtilis subsp. subtilis Laboratory Strains JH642 (AG174) and AG1839. *Genome Announc.* **2,** e00663–14 (2014).
450. Zeigler, D. R. *et al.* The Origins of 168, W23, and Other Bacillus subtilis Legacy Strains. *J. Bacteriol.* **190,** 6983–6995 (2008).
451. Fitzpatrick, L. R. *et al.* Bacillus coagulans GBI-30, 6086 limits the recurrence of Clostridium difficile-Induced colitis following vancomycin withdrawal in mice. *Gut Pathog.* **4,** 13 (2012).
452. Chopra, I. & Roberts, M. Tetracycline Antibiotics: Mode of Action, Applications, Molecular Biology, and Epidemiology of Bacterial Resistance. *Microbiol. Mol. Biol. Rev.* **65,** 232–260 (2001).
453. Luchansky, J. B., Muriana, P. M. & Klaenhammer, T. R. Application of electroporation for transfer of plasmid DNA to Lactobacillus, Lactococcus, Leuconostoc, Listeria, Pediococcus, Bacillus, Staphylococcus, Enterococcus and Propionibacterium. *Mol. Microbiol.* **2,** 637–646 (1988).
454. Naglich, J. G. & Andrews, R. E. Tn916-dependent conjugal transfer of PC194 and PUB110 from Bacillus subtilis into Bacillus thuringiensis subsp. israelensis. *Plasmid* **20,** 113–126 (1988).
455. Menard, K. L. & Grossman, A. D. Selective Pressures to Maintain Attachment Site Specificity of Integrative and Conjugative Elements. *PLoS Genet* **9,** e1003623 (2013).
456. Spaink, H. P. Root Nodulation and Infection Factors Produced by Rhizobial Bacteria. *Annu. Rev. Microbiol.* **54,** 257–288 (2000).
457. Kandel, S. L., Herschberger, N., Kim, S. H. & Doty, S. L. Diazotrophic Endophytes of Poplar and Willow for Growth Promotion of Rice Plants in Nitrogen-Limited Conditions. *Crop Sci.* **55,** 1765–1772 (2015).
458. Akshit Puri, K. P. P. Evidence of nitrogen fixation and growth promotion in canola (Brassica napus L.) by an endophytic diazotroph Paenibacillus polymyxa P2b-2R. *Biol. Fertil. Soils* **52,** 119–125 (2016).
459. Temme, K., Zhao, D. & Voigt, C. A. Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proc. Natl. Acad. Sci.* **109,** 7085–7090 (2012).
460. Wang, L. *et al.* A Minimal Nitrogen Fixation Gene Cluster from Paenibacillus sp. WLY78 Enables Expression of Active Nitrogenase in Escherichia coli. *PLoS Genet* **9,** e1003865 (2013).
461. Kushwaha, M. & Salis, H. M. A portable expression resource for engineering cross-species genetic circuits and pathways. *Nat. Commun.* **6,** 7832 (2015).
462. Kramer, M. G., Espinosa, M., Misra, T. K. & Khan, S. A. Characterization of a single-strand origin, ssoU, required for broad host range replication of rolling-circle plasmids. *Mol. Microbiol.* **33,** 466–475 (1999).
463. Yang, L. *et al.* Permanent genetic memory with >1-byte capacity. *Nat. Methods* **11,** 1261–1266 (2014).
464. Cobb, R. E., Wang, Y. & Zhao, H. High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. *ACS Synth. Biol.* **4,** 723–728 (2015).

465. Aminov, R. I. Horizontal Gene Exchange in Environmental Microbiota. *Front. Microbiol.* **2,** (2011).
466. Hatfield, G. W. & Benham, C. J. DNA Topology-Mediated Control of Global Gene Expression in Escherichia coli. *Annu. Rev. Genet.* **36,** 175–203 (2002).
467. Bates, A. D. & Maxwell, A. *DNA Topology*. (Oxford University Press, 2005).
468. Litwin, T. R., Solà, M., Holt, I. J. & Neuman, K. C. A robust assay to measure DNA topology-dependent protein binding affinity. *Nucleic Acids Res.* **43,** e43 (2015).
469. Pagel, J. M., Winkelman, J. W., Adams, C. W. & Hatfield, G. W. DNA topology-mediated regulation of transcription initiation from the tandem promoters of the ilvGMEDA operon of Escherichia coli. *J. Mol. Biol.* **224,** 919–935 (1992).
470. Leng, F. & McMacken, R. Potent stimulation of transcription-coupled DNA supercoiling by sequence-specific DNA-binding proteins. *Proc. Natl. Acad. Sci.* **99,** 9139–9144 (2002).
471. Chan, C. T. Y. *et al.* Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat. Commun.* **3,** 937 (2012).
472. Chong, S., Chen, C., Ge, H. & Xie, X. S. Mechanism of transcriptional bursting in bacteria. *Cell* **158,** 314–326 (2014).
473. Boyer, H. W. DNA Restriction and Modification Mechanisms in Bacteria. *Annu. Rev. Microbiol.* **25,** 153–176 (1971).
474. Jeltsch, A. & Pingoud, A. Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *J. Mol. Evol.* **42,** 91–96 (1996).
475. Tock, M. R. & Dryden, D. T. The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.* **8,** 466–472 (2005).
476. Serfiotis-Mitsa, D. *et al.* The Orf18 Gene Product from Conjugative Transposon Tn916 Is an ArdA Antirestriction Protein that Inhibits Type I DNA Restriction–Modification Systems. *J. Mol. Biol.* **383,** 970–981 (2008).