AFRL-AFOSR-VA-TR-2022-0402

Theory-based design of synthetic genetic circuits incorporating biophysical models, stochastic dynamics, and evolutionary robustness

Alper, Hal
UNIVERSITY OF TEXAS AT AUSTIN
110 INNER CAMPUS DR
AUSTIN, TX, 78712
USA

07/19/2022
Final Technical Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**	
		START DATE	**END DATE**
20220719	Final	20140815	20220214

4. TITLE AND SUBTITLE
Theory-based design of synthetic genetic circuits incorporating biophysical models, stochastic dynamics, and evolutionary robustness

5a. CONTRACT NUMBER	**5b. GRANT NUMBER**	**5c. PROGRAM ELEMENT NUMBER**
	FA9550-14-1-0089	61102F
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Hal Alper

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSITY OF TEXAS AT AUSTIN
110 INNER CAMPUS DR
AUSTIN, TX 78712
USA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
Air Force Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203	AFRL/AFOSR RTB2	AFRL-AFOSR-VA-TR-2022-0402

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This BRI grant focuses on developing a theoretical framework for synthetic biology design with a specific emphasis on robustness. Specifically, current “parts-off-the-shelf” approaches are unable to deliver parts and circuits with adequate functional robustness to genetic, environmental, and evolutionary pressures. This work uses a DNA-design centric approach to create new and robust parts and circuits guided by biophysical, thermodynamic, statistical, and stochastic models in addition to evolution. Such a theoretical framework can: (1) predict a genetic part’s function according to its DNA sequence, (2) develop optimal sequences for performance, (3) specifically program robustness, and (4) predict response and fault modes of a system. This project will focus on the establishment of robust, fault-tolerant cis-acting sequences and the development of fundamental principles and models for synthetic circuit design. To demonstrate the power and generality of this approach, we will investigate two model organisms, E. coli and yeast.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	22
U	U	U		

19a. NAME OF RESPONSIBLE PERSON	**19b. PHONE NUMBER** *(Include area code)*
BENNETT IBEY	000-0000

Standard Form 298 (Rev.5/2020)
Prescribed by ANSI Std. Z39.18

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED START DATE	3. DATES COVERED END DATE
20220515	Final Report	20140815	20220214

4. TITLE AND SUBTITLE
Theory-based design of synthetic genetic circuits incorporating biophysical models, stochastic dynamics, and evolutionary robustness

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
	FA9550-14-1-0089	
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Alper, Hal, S.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Texas at Austin
3925 West Braker Lane, Stop A9000
Austin, TX 78759-5316

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
875 N Randolph St Room 3112
Arlington VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution A: Unlimited Distribution

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This BRI grant focuses on developing a theoretical framework for synthetic biology design with a specific emphasis on robustness. Specifically, current "parts-off-the-shelf" approaches are unable to deliver parts and circuits with adequate functional robustness to genetic, environmental, and evolutionary pressures. This work uses a DNA-design centric approach to create new and robust parts and circuits guided by biophysical, thermodynamic, statistical, and stochastic models in addition to evolution. Such a theoretical framework can: (1) predict a genetic part's function according to its DNA sequence, (2) develop optimal sequences for performance, (3) specifically program robustness, and (4) predict response and fault modes of a system. This project will focus on the establishment of robust, fault-tolerant cis-acting sequences and the development of fundamental principles and models for synthetic circuit design. To demonstrate the power and generality of this approach, we will investigate two model organisms, E. coli and yeast.

15. SUBJECT TERMS
Synthetic biology; circuits; model-guided approaches; yeast; bacteria

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	C. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
UU	UU	UU	UU	20

PREVIOUS EDITION IS OBSOLETE.

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Hal Alper	512-471-4417

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE.
Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE.
State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED.
Indicate the time during which the work was performed and the report was written.

4. TITLE.
Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER.
Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER.
Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER.
Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER.
Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER.
Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

PREVIOUS EDITION IS OBSOLETE.

Final Report
Lead PI: Hal Alper
Program Officer: Bennett Ibey

I. Heading

Lead PI Name:	Hal Alper
Organization:	The University of Texas at Austin
Air Force Award #:	FA9550-14-1-0089
Award Title:	Theory-based design of synthetic genetic circuits incorporating biophysical models, stochastic dynamics, and evolutionary robustness

II. Scientific and Technical Objectives

This BRI grant focuses on developing a theoretical framework for synthetic biology design with a specific emphasis on robustness. Specifically, current "parts-off-the-shelf" approaches are unable to deliver parts and circuits with adequate functional robustness to genetic, environmental, and evolutionary pressures. This work uses a DNA-design centric approach to create new and robust parts and circuits guided by biophysical, thermodynamic, statistical, and stochastic models in addition to evolution. Such a theoretical framework can: (1) predict a genetic part's function according to its DNA sequence, (2) develop optimal sequences for performance, (3) specifically program robustness, and (4) predict response and fault modes of a system. This project will focus on the establishment of robust, fault-tolerant cis-acting sequences and the development of fundamental principles and models for synthetic circuit design. To demonstrate the power and generality of this approach, we will investigate two model organisms, *E. coli* and yeast.

III. Approach

The specific objectives of this research will focus on the establishment of robust, fault-tolerant cis-acting sequences and the development of fundamental principles and models for synthetic circuit design. As part of the first phase of this work, we proposed to establish a DNA-compiler approach (Task 1) and use the embodied models to engineer three different sequence-based parts, and to build these into robust, perturbation-resistant parts: promoters (Task 2), translation components (Task 3), and nucleic acid-based switches (Task 4). In each instance we will develop underlying sequence-based models for specifying how the different parts and circuits function, and apply these models to ensure error-resistant function. We will test each of these components under a catalogue of robustness metrics. Finally, we will expand these models and concepts to provide systems-level models that can be stochastically modeled to probe evolutionary space and predict robust circuit design (Task 5) across two model organisms, *E. coli* and yeast. In each of these cases, an emphasis will be placed on robustness of parts and circuits though designed features such as fault resistance, redundancy, and evolutionary tolerance. The deliverables for this work include a novel DNA compiler capable for specifying sequence for designing robust parts and circuits, in addition to robust promoters, translation elements, riboswitches, and riboregulators designed for bacteria and yeast. In the second phase of this project, work included enabling transitions of technologies and synthetic parts developed during the course of the BRI to Air Force Labs (Task 6), demonstrate the robustness/stability of synthetic circuits in both *E. coli* and *S. cerevisiae* (Task 7), catalogue

failure modes of synthetic circuits and provide insulation against these modes (Task 8), and extend the various techniques and paradigms of theory-based synthetic biology to an alternative organism (Task 9).

IV. Concise Accomplishments

Major Accomplishments in this entire project:

Theory based approaches

- Established a riboswitch calculator, a predictive biophysical model that designs RNA-based biosensors
- Expanded a Non-Repetitive Parts calculator, an algorithm to design large toolboxes of highly non-repetitive parts
- Created an Extra Long sgRNA Arrays calculator to design constructs with up to 20 sgRNAs expressed.
- Developed models to *de novo* design yeast terminator activity
- Established model based design of RNA aptamer/sensor function
- Created T7-based circuit design strategies
- Developed theoretical insights to enable new, complex circuits for multicellular patterning

Experimental Advances

- Established novel halo-tryptophan repressors and created orthogonal components for repressor-based logic circuits.
- Reduced DNA size of yeast regulatory elements to 1/10.
- Established strategies for sgRNA-based gene regulation and multiplex rewiring.
- Established T7-based circuits in yeast for sgRNA expression.
- Demonstrated the ability to get contrasting patterns with multicellular communities.
- Established homeostasis and addiction circuitry to enable biosafety modality and fine control of gene expression.
- Created a method for retron-mediated genome editing and evolution.
- Validated various complex circuits including homeostasis circuits, cell patterning toggle switches, universal plasmid systems, and logic gates that work across host organisms.
- Demonstrated a microdroplet picoinjection process for high-throughput analysis and use of biosensors

V. Expanded Accomplishments

In this final project report, we highlight some of the major accomplishments achieved throughout this multi-investigator, multi-University BRI program. This section is not meant to be comprehensive, but rather provide a broad overview of the accomplishments described in section IV above.

Development of various computational calculator tools for predicting and designing DNA

Throughout this project, our team developed a series of calculators for the design and prediction of functional DNA elements in cells. we developed the Operon Calculator, a DNA Compiler for bacterial operons that designs synthetic operon sequences for maximum tunable expression control and maximum evolutionary robustness (**Figure 1**). The Operon Calculator

applies over 10 predictive biophysical models and design rules to create operons that express desired proteins with the following objectives: design ribosome binding sites with targeted translation rates, optimize synonymous codon usage, eliminate internal promoters, eliminate internal terminators, minimize RNAse E/III binding sites, eliminate ribosomal pause sequences, eliminate highly translated internal start codons, remove repetitive DNA sequences, remove transposon insertion sites, eliminate recombinase sites, and remove excluded restriction sites. Users select their promoter, CDS, and terminator sequences, and the algorithm will design the entire operon sequence, using multi-objective optimization, producing several Pareto-optimal solutions. A web interface to the Operon Calculator algorithm is freely available and was published (**Figure 1**).

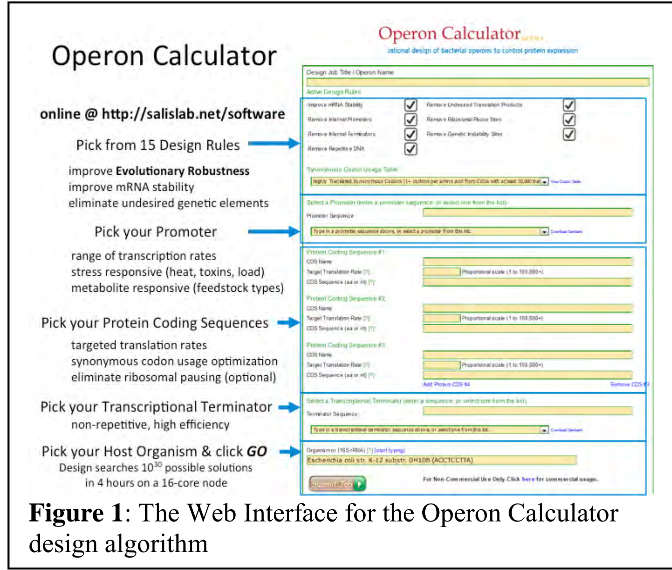


Figure 1: The Web Interface for the Operon Calculator design algorithm

The next challenge that is pervasive throughout this work was the selection of non-repetitive DNA elements and constructs. This issue became more prominent as CRISPR/sgRNA design increased throughout the time of this BRI. To solve this challenge, we developed a new algorithm, called the Non-Repetitive Genetic Parts Calculator, that designs the largest possible toolbox of highly non-repetitive genetic parts that all satisfy a user-specified set of design constraints. A design constraint could be a degenerate DNA sequences, a target RNA structure, and/or an arbitrary model function that evaluates the genetic part's function and returns true or false. The Non-Repetitive Genetic Parts Calculator then uses graph theory and an advanced path generation algorithm to identify the mathematically largest possible number of constraint-satisfying genetic parts that do not share more than L base pairs (nucleotides) of repetitive DNA. The maximum repeat size L is user-specified; we typically set L = 10 for maximum stringency.

We applied the Non-repetitive Genetic Parts Calculator (**Figure 2**) to design 3500 highly non-repetitive E. coli promoters (σ^{70}) with a maximum shared repeat length of 10 base pairs (L = 10). Our designs varied promoter consensus sequences to

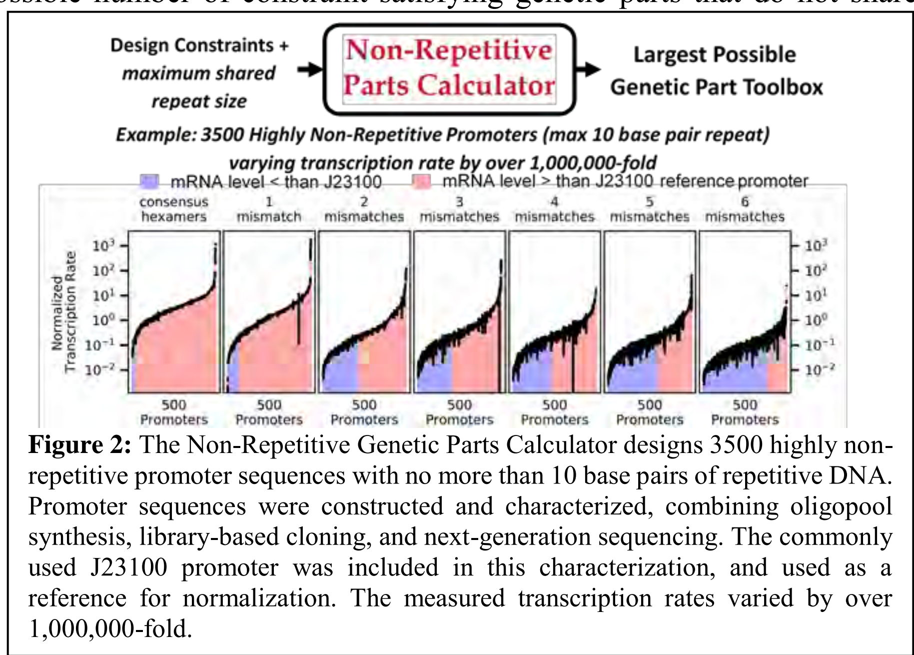


Figure 2: The Non-Repetitive Genetic Parts Calculator designs 3500 highly non-repetitive promoter sequences with no more than 10 base pairs of repetitive DNA. Promoter sequences were constructed and characterized, combining oligopool synthesis, library-based cloning, and next-generation sequencing. The commonly used J23100 promoter was included in this characterization, and used as a reference for normalization. The measured transcription rates varied by over 1,000,000-fold.

create a toolbox of promoters that maximally varied transcription initiation rates. Specifically, we designed 500 of these promoters to have consensus -35 and -10 hexamer sequences controlling RNAP/σ^{70} recruitment. We then designed sets of 500 promoters that each contained between 1 to 6 mismatches to the consensus -35/-10 hexamer sequences. We then combined oligopool synthesis, library-based cloning, and next-generation sequencing (DNA-Seq, RNA-Seq) to

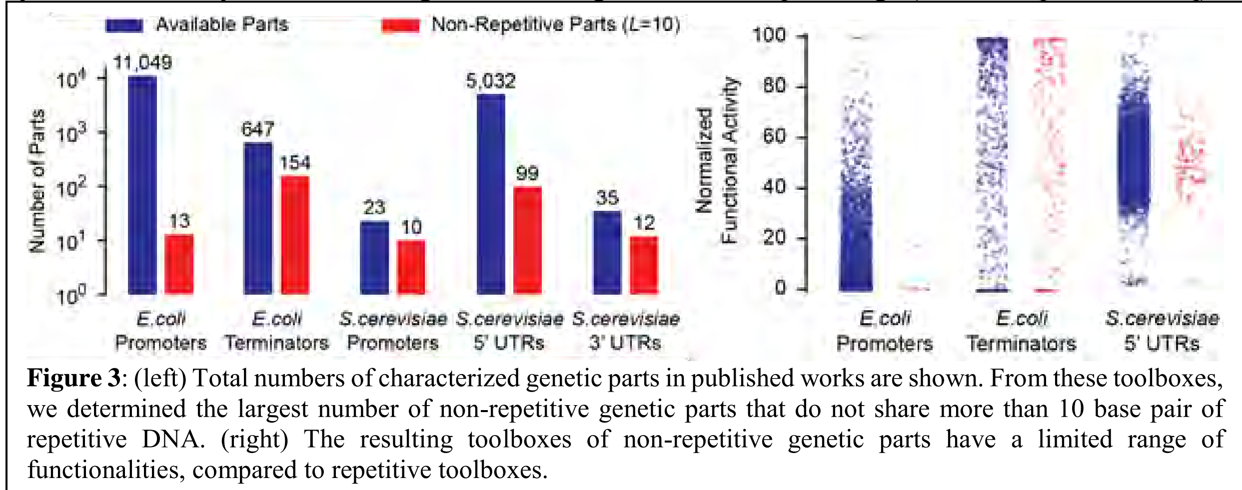


Figure 3: (left) Total numbers of characterized genetic parts in published works are shown. From these toolboxes, we determined the largest number of non-repetitive genetic parts that do not share more than 10 base pair of repetitive DNA. (right) The resulting toolboxes of non-repetitive genetic parts have a limited range of functionalities, compared to repetitive toolboxes.

construct and characterize all 3500 promoters simultaneously in *E. coli* cells cultured in exponential growth conditions (**Figure 3**).

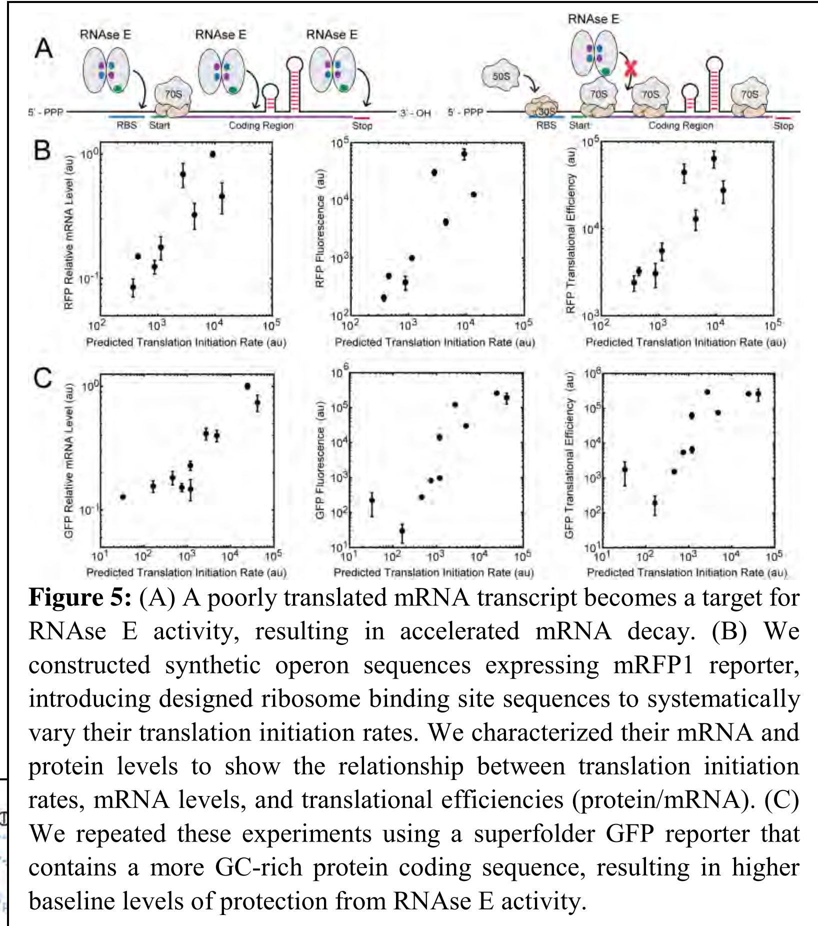


Figure 5: (A) A poorly translated mRNA transcript becomes a target for RNAse E activity, resulting in accelerated mRNA decay. (B) We constructed synthetic operon sequences expressing mRFP1 reporter, introducing designed ribosome binding site sequences to systematically vary their translation initiation rates. We characterized their mRNA and protein levels to show the relationship between translation initiation rates, mRNA levels, and translational efficiencies (protein/mRNA). (C) We repeated these experiments using a superfolder GFP reporter that contains a more GC-rich protein coding sequence, resulting in higher baseline levels of protection from RNAse E activity.

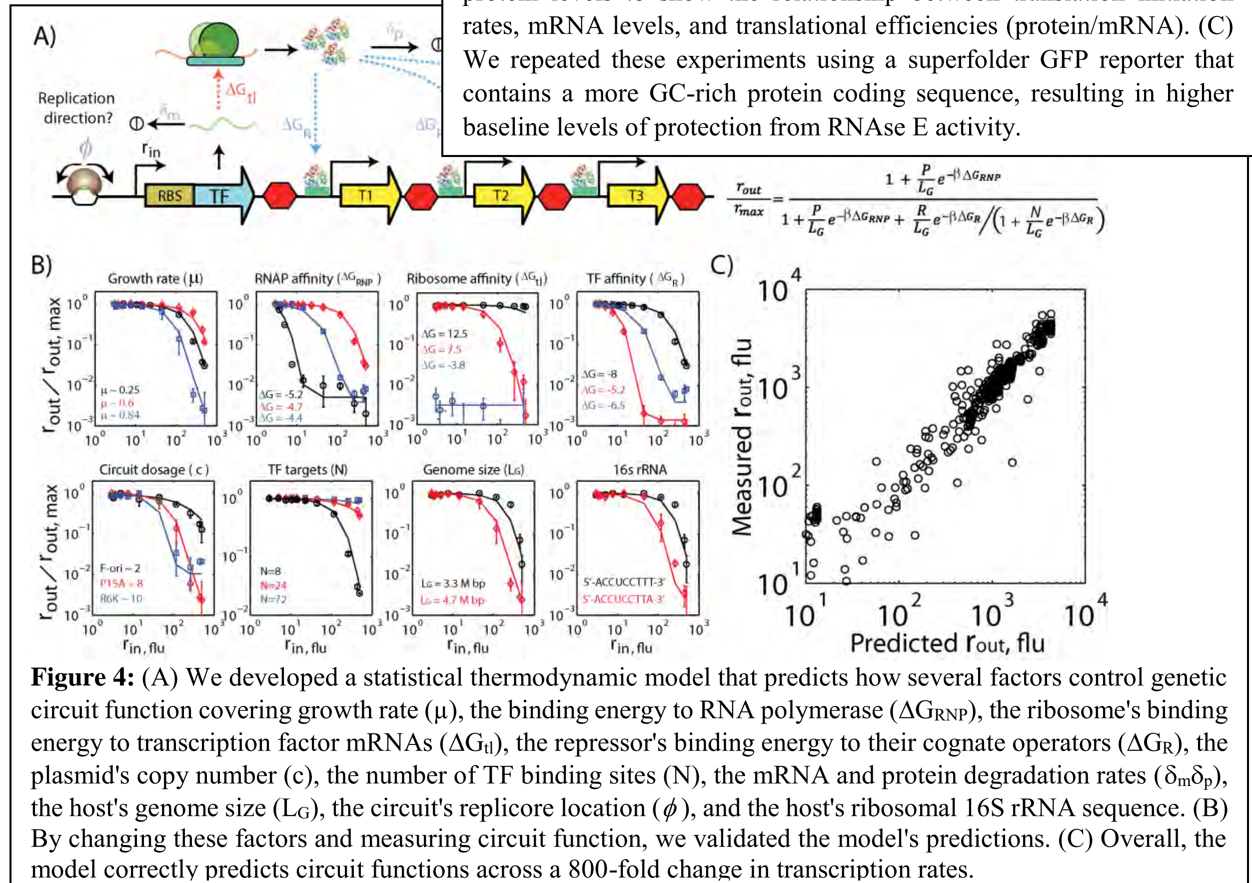


Figure 4: (A) We developed a statistical thermodynamic model that predicts how several factors control genetic circuit function covering growth rate (μ), the binding energy to RNA polymerase (ΔG_{RNP}), the ribosome's binding energy to transcription factor mRNAs (ΔG_{tl}), the repressor's binding energy to their cognate operators (ΔG_R), the plasmid's copy number (c), the number of TF binding sites (N), the mRNA and protein degradation rates ($\delta_m \delta_p$), the host's genome size (L_G), the circuit's replicore location (ϕ), and the host's ribosomal 16S rRNA sequence. (B) By changing these factors and measuring circuit function, we validated the model's predictions. (C) Overall, the model correctly predicts circuit functions across a 800-fold change in transcription rates.

Fully descriptive, thermodynamic models for yeast transcription were developed throughout the course of this project. This work enabled a compiling of data from this project to both develop and understand new experimentally derived promoter elements.

Beyond these models, several biophysical models were created for engineered components. Included in this are predictions for translation rate from a transcript. Prior to this work, it was unclear how an operon's sequence controls the half-life (stability) of the resulting mRNA transcript, which determines mRNA and protein levels. To determine the quantitative physical rules governing these interactions, we

constructed over 120 synthetic bi-cistronic operons and characterized their mRNA & protein levels, utilizing RT-qPCR and flow cytometry. A major finding of these experiments is the quantification of strong coupling between translation and mRNA decay (**Figure 5**). Specifically, a 100-fold change in translation initiation rate yields about a 10-fold change in mRNA level. The mechanism for this interaction is called ribosome protection, whereby a highly translated mRNA is bound by several ribosomes engaged in translation elongation. These bound ribosomes protect the mRNA from RNAse activity, particularly RNAse E. When the mRNA's translation initiation rate is lowered, the mRNA is no longer protected by elongating ribosomes, and thereby can be cut by RNAse E. We performed additional experiments to quantify the sequence determinants of RNAse E binding across 5' UTRs and protein coding sequences, including polarity effects within multi-cistronic operons. We also performed similar experiments to determine the sequence determinants of RNAse III activity. These systematic measurements enable us to determine the physical rules governing the several mRNA decay pathways, and thereby predict mRNA's half-life from its sequence.

Development of minimally-sized eukaryotic regulatory factors

Efforts within this BRI have focused on minimizing the size of regulatory DNA in yeast systems through efforts to establish minimal promoter and terminators. Prior to this work, fungal promoters often span hundreds of base pairs, nearly ten times the amount of bacterial counterparts. This size limits large-scale synthetic biology efforts in yeasts.

Through a series of library-based synthesis, analysis, and robustness tests, we have created a set of nine non-homologous, purely synthetic, robust, minimal core elements for yeast. 15 million random oligonucleotides comprised of two necessary promoter parts, a TATA box and a transcription start site spaced 30 random nucleotides apart, and activated by a native UAS were used to drive the expression of a fluorescent reporter gene (**Figure 6**). These candidate sequences were screened *in vivo* via flow cytometry to yield 18 unique, completely synthetic sequence elements (**Figure 6**). These elements were subjected to a series of robustness tests in an effort to compile a set of promoter core elements which ideally function modularly and robustly with any UAS. We ensured each promoter core element could (1) be activated by two native constitutive UAS elements, a 240 bp sequence isolated from the mitotic cyclin *CLB2* promoter (UAS_{CLB}) and a 275 bp sequence isolated from the mitochondrial citrate synthase *CIT1* promoter (UAS_{CIT}), (2) be

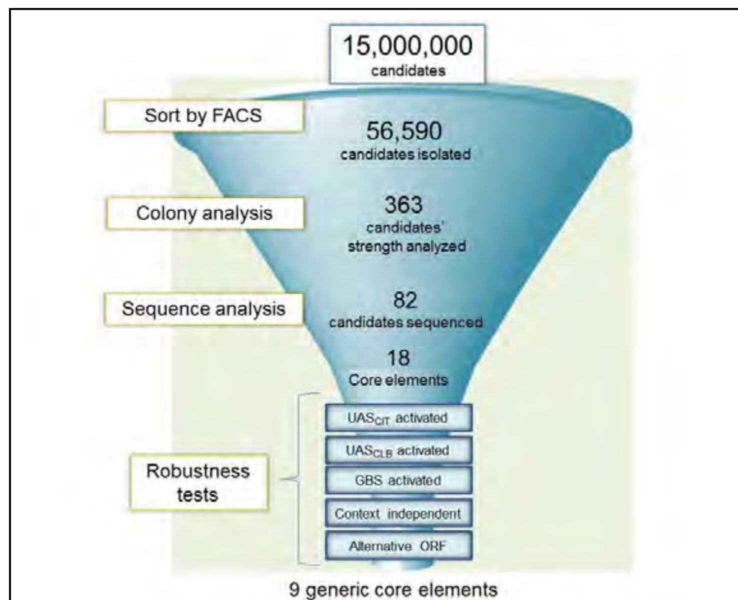


Figure 6: *Schematic of robust, minimal promoter core element isolation.* 27 libraries of 15 million candidates were created. 0.15% of the most promising libraries were sorted by fluorescence activated cell sorting (FACS). After sequence and strength confirmation, 363 candidates were narrowed to just 18 core elements. These elements were subjected to a battery of tests to create a final set of nine robust, minimal and modular core elements.

activated by the native galactose-inducible transcription factor binding site for Gal4 (GBS), (3) maintain expression strength with an alternative upstream context indicating context independency, and (4) drive expression of two alternative open reading frames (ORF), GFP-derived yECitrine and LacZ gene, β-galactosidase. With these tests, we narrowed our pool of promoter core elements to just nine (**Figure 6**). By measuring mRNA levels of fluorescent reporter gene with qPCR, we demonstrated that these core elements function at the transcriptional level. Moreover, we show that promoters constructed with these core elements recruit an assorted arrangement of pre-initiation complexes depending on UAS present, illustrating the set's mechanistic diversity. The final set bears no resemblance to genomic DNA sequences indicated by the high minimum E_{value} obtained from BLAST, safeguarding against homologous recombination events, and little sequence similarity to each other, allowing for multi-pathway constructs to be easily assembled without concern for recombination within a cassette. Lastly, the GC content of the promoter core elements spans a wide range from as low as 47% to as high as 73%.

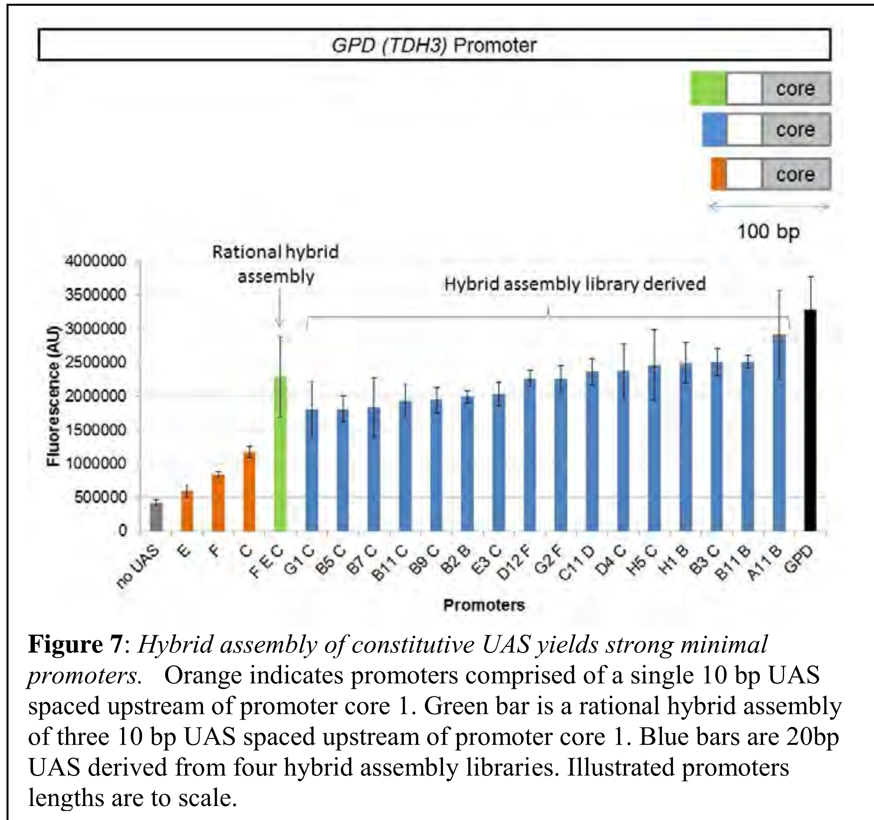
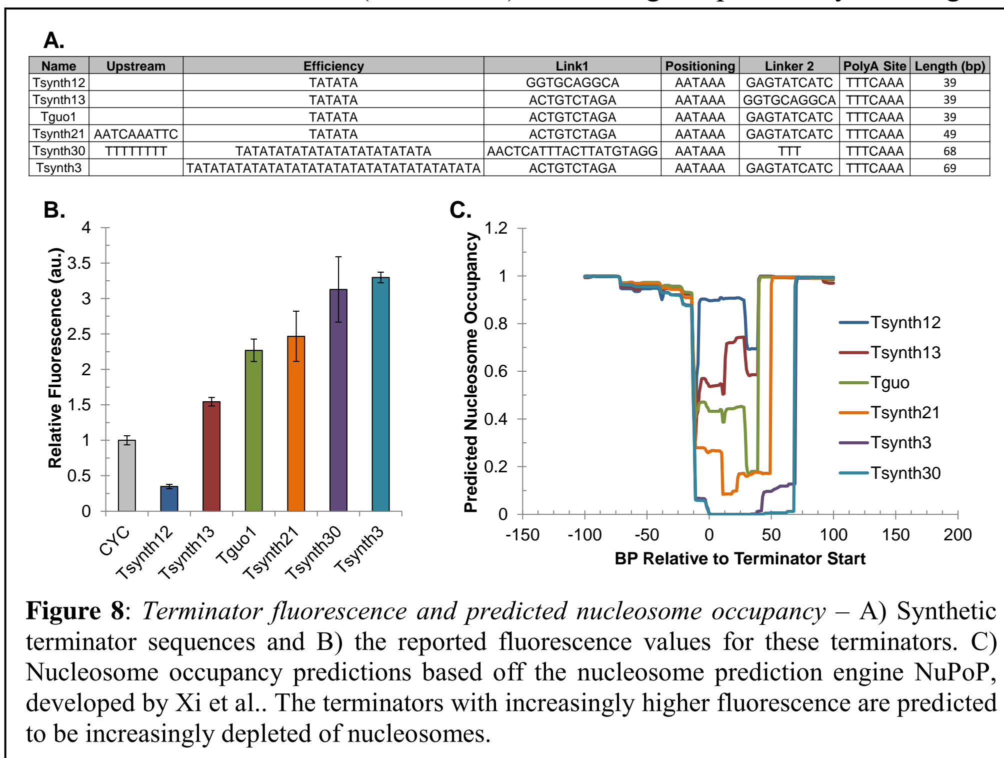


Figure 7: *Hybrid assembly of constitutive UAS yields strong minimal promoters.* Orange indicates promoters comprised of a single 10 bp UAS spaced upstream of promoter core 1. Green bar is a rational hybrid assembly of three 10 bp UAS spaced upstream of promoter core 1. Blue bars are 20bp UAS derived from four hybrid assembly libraries. Illustrated promoters lengths are to scale.

Operable with these nine core elements, we have identified generic minimal constitutive UAS elements: six 10 bp UAS elements of low strength, and sixteen 20 bp UAS elements of moderate to high strength (**Figure 6**). 10 bp UAS elements were screened *in vivo* via flow cytometry from a single library of 1.3 million sequences comprised ten random oligonucleotides efficiently spaced upstream of a promoter core element. Through rational hybrid-assembly of these 10 bp UAS, we are able to create promoters in a fraction of the sequence space of a native promoter and of substantial expressions strengths. The sixteen 20 bp UAS elements were screened *in vivo* via flow cytometry from multiple hybrid assembly libraries totaling more than 5 million sequences. These hybrid assembly libraries consist of ten random oligonucleotides positioned directly upstream of the previously isolated 10 bp UAS. Thus, with these completely synthetic UAS, we are able to craft fully operating promoters reaching 88% the strength of one of the strongest promoters used in the synthetic biology today, GPD, in just 16% of the sequence space (**Figure 7**). Furthermore, at just ~100 bp, these synthetic constitutive yeast promoters are roughly the size of bacterial promoters (**Figure 7**).

A similar strategy was taken for the development of minimal terminator elements achieving highly effective terminators with lengths between 30 – 85 bp. A heuristic model was created to understand and model these elements and it was found that predicted nucleosome depletion positively correlated with terminator function ($R^2 = 0.848$), indicating the possibility of using this for terminator design (**Figure 8**). It was therefore uncovered that a correlation between nucleosome depletion and termination efficiency could successfully improve protein levels within the cell. This work helped establish a biophysical-based forward design method for these elements that was demonstrated through this BRI.


A.

Name	Upstream	Efficiency	Link1	Positioning	Linker 2	PolyA Site	Length (bp)
Tsynth12		TATATA	GGTGCAGGCA	AATAAA	GAGTATCATC	TTTCAAA	39
Tsynth13		TATATA	ACTGTCTAGA	AATAAA	GGTGCAGGCA	TTTCAAA	39
Tguo1		TATATA	ACTGTCTAGA	AATAAA	GAGTATCATC	TTTCAAA	39
Tsynth21	AATCAAATTC	TATATA	ACTGTCTAGA	AATAAA	GAGTATCATC	TTTCAAA	49
Tsynth30	TTTTTTTT	TATATATATATATATATATATATA	AACTCATTTACTTATGTAGG	AATAAA	TTT	TTTCAAA	68
Tsynth3		TATATATATATATATATATATATATATATATATA	ACTGTCTAGA	AATAAA	GAGTATCATC	TTTCAAA	69

Figure 8: *Terminator fluorescence and predicted nucleosome occupancy* – A) Synthetic terminator sequences and B) the reported fluorescence values for these terminators. C) Nucleosome occupancy predictions based off the nucleosome prediction engine NuPoP, developed by Xi et al.. The terminators with increasingly higher fluorescence are predicted to be increasingly depleted of nucleosomes.

<u>Development of CRISPR-based approaches for bi-direction gene expression modulation</u>

Using dCas9-reprorgramming in *S. cerevisiae*, we have coupled work here (along with DNA synthesis by the JGI Synthesis group) to develop a plasmid-based library methodology that affords bi-directional, graded modulation of gene expression enabled by tiling the promoter regions of all 969 genes that comprise the ito977 model of *Saccharomyces cerevisiae*'s metabolic network. In many respects, this work served as a culmination of the activities on this BRI as it utilized the guiding principles that were discovered for positional effects of sgRNAs, synthetic elements (including synthetic, hybrid promoters and subsequent T7-RNA polymerase-based methods), and microdroplet-enabled sorting. When coupled with a CRISPR-dCas9 based modulation and next-generation sequencing, this method affords a library based, bi-direction titration of gene expression across all major metabolic genes. We utilized this approach in two case studies: growth enrichment on alternative sugars, glycerol and galactose, and chemical overproduction of betaxanthins, leading to the identification of unique gene targets. In particular, we identify essential genes and other targets that were missed by classic genetic approaches.

To do so, four sgRNA libraries were designed in a pooled format to target broad classes of metabolic function including carbohydrate and lipid metabolism (Pool 1), energy and cofactor metabolism (Pool 2), amino acid and nucleotide metabolism (Pool 3), and housekeeping/other (Pool 4). This library was used in conjunction with growth selections and with product selections. Given the high-resolution aspect of this dataset (*i.e.,* having both target identifications along with

their optimal expression levels), multiple modes of analysis are possible and we used this data for analysis of the genotype-phenotype relationship (**Figure 9**).

In this project, we demonstrated that many cases whereby moderate knockdown greatly enhances growth relative to the complete knockout. Likewise, we identified how tuning the expression of an essential gene can create a new phenotype that is otherwise inaccessible. More specifically, these libraries contained 86 essential genes. Within this set when selected for glycerol and galactose, 8 of these targets are identified as downregulation targets in galactose and 10 of these genes we identified as downregulation targets in glycerol (with an additional 10 of these genes enriched as overexpression targets). These examples in particular are poignant demonstrations of the benefit of an expression titration library as these targets would not be identified in a full knockout collection. As examples, knockdowns were identified for both *DIM1* and *GPI18*, both of which are essential genes. The knockdown condition substantially improved growth rates in their respective alternative carbon sources whereas the complete deletion of these genes is lethal. Thus, this approach identified a unique set of targets for which galactose and glycerol-improvement phenotypes were not previously ascribed. These approaches were also followed for improving product secretion of betaxanthins in a *S. cerevisiae* strain and a similar result was found here with intermediate expression (especially of essential genes) identified as newfound targets that would have been otherwise missed by standard, classical genetic approaches.

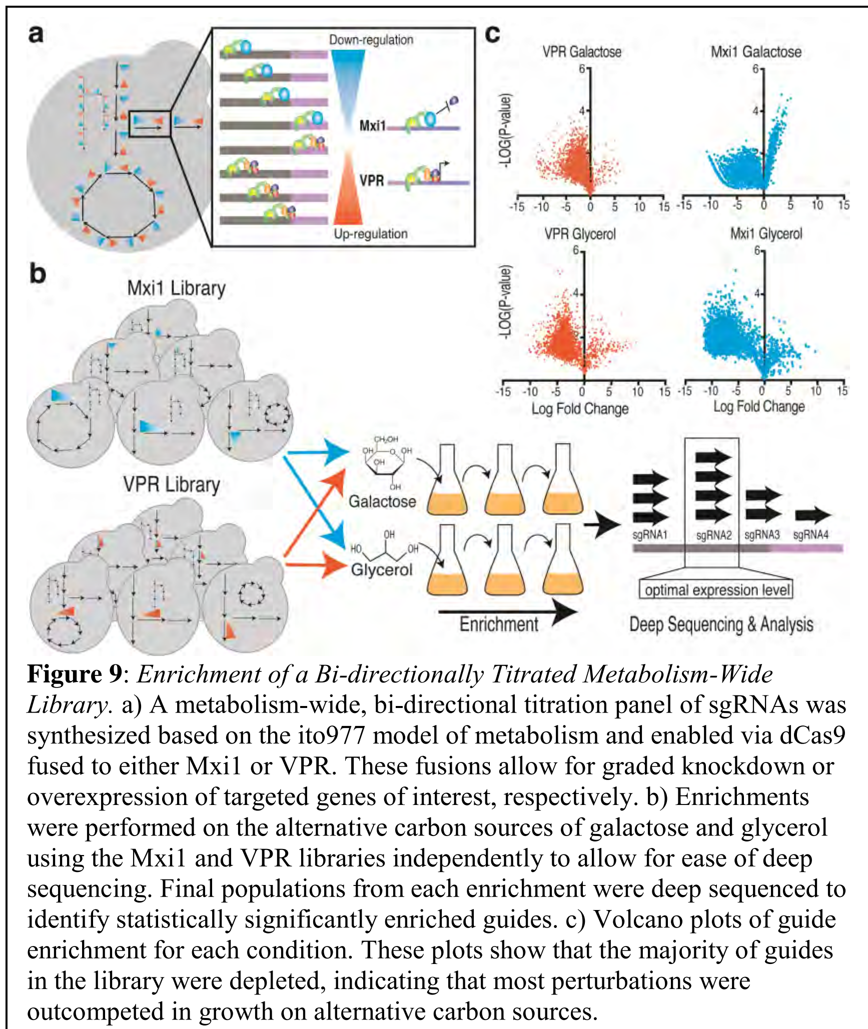


Figure 9: *Enrichment of a Bi-directionally Titrated Metabolism-Wide Library.* a) A metabolism-wide, bi-directional titration panel of sgRNAs was synthesized based on the ito977 model of metabolism and enabled via dCas9 fused to either Mxi1 or VPR. These fusions allow for graded knockdown or overexpression of targeted genes of interest, respectively. b) Enrichments were performed on the alternative carbon sources of galactose and glycerol using the Mxi1 and VPR libraries independently to allow for ease of deep sequencing. Final populations from each enrichment were deep sequenced to identify statistically significantly enriched guides. c) Volcano plots of guide enrichment for each condition. These plots show that the majority of guides in the library were depleted, indicating that most perturbations were outcompeted in growth on alternative carbon sources.

Evolution of natural product biosensors

Genetic biosensors are gaining traction as high-throughput tools for otherwise difficult chemical analyses, such as distinguishing between compounds with identical exact masses and/or lacking a chromophore. However, biosensors often require custom engineering for target analytes, and even with great improvements in bioinformatic mining it is difficult to find corresponding

biosensors. Therefore, we have developed a method (SELIS; Seamless Enrichment of Ligand-inducible Sensors) that can be broadly used to identify biosensors (**Figure 10**), and herein we develop a genetic biosensor that is responsive to a range of bicyclic monoterpenes used in the flavor, fragrance, cosmetic, and pharmaceutical industries, which are traditionally limited to analysis using low-throughput gas chromatography-mass spectrometry.

Briefly, SELIS relies on a positive screen for ligand-dependent regulator function via GFP production, and a conjoined negative selection in which transcriptional repression in the absence of ligand yields antibiotic resistance (**Figure 10**). Between the positive screen and negative selection, SELIS can quickly deconvolute libraries with over 10^5 members in under a week, and provides substantial genotype and phenotype data that allows optimally performing biosensors to be chosen by the researcher from amongst a range of successful candidates. It should be noted that since the fluorescence of clonal isolates is measured both with and without the effector during screening (**Figure 10c**) virtually all false positives are eliminated.

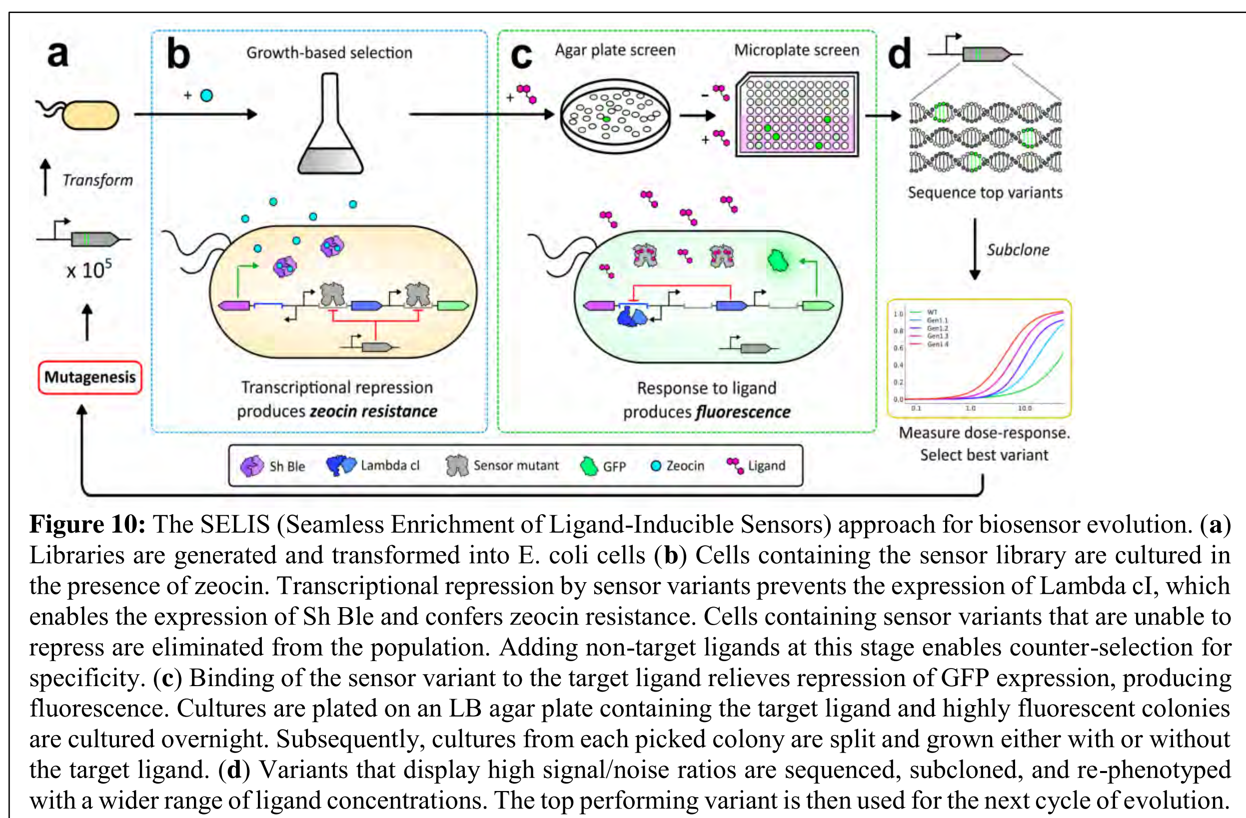


Figure 10: The SELIS (Seamless Enrichment of Ligand-Inducible Sensors) approach for biosensor evolution. (**a**) Libraries are generated and transformed into E. coli cells (**b**) Cells containing the sensor library are cultured in the presence of zeocin. Transcriptional repression by sensor variants prevents the expression of Lambda cI, which enables the expression of Sh Ble and confers zeocin resistance. Cells containing sensor variants that are unable to repress are eliminated from the population. Adding non-target ligands at this stage enables counter-selection for specificity. (**c**) Binding of the sensor variant to the target ligand relieves repression of GFP expression, producing fluorescence. Cultures are plated on an LB agar plate containing the target ligand and highly fluorescent colonies are cultured overnight. Subsequently, cultures from each picked colony are split and grown either with or without the target ligand. (**d**) Variants that display high signal/noise ratios are sequenced, subcloned, and re-phenotyped with a wider range of ligand concentrations. The top performing variant is then used for the next cycle of evolution.

To evolve a broadly useful monoterpene biosensor, we used the camphor-responsive CamR repressor from *Pseudomonas putida*, as a starting point for evolution. The Pcamr promoter was constructed by extracting a CamR-protected region from the natural P. putida CamR-bearing plasmid (GenBank: D14680.1), and placing it upstream from the self-cleaving ribozyme PlmJ. CamR was introduced into the SELIS pipeline by placing this Pcamr promoter construct upstream of a sfGFP gene for the positive screen, and having a separate Pcamr promoter drive the expression of the Lambda cI repressor for the negative selection, thereby creating pSELIScamr.

Using SELIS, CamR was evolved for increased responsiveness to the minor effectors borneol, fenchol, eucalyptol and camphene. CamR libraries, generated via random mutagenesis, were co-transformed into *E. coli* with the pSELIScamr plasmid, which were then grown in the

presence of zeocin (negative selection) and subsequently plated on solid media containing 1% DMSO and either borneol, fenchol, eucalyptol, or camphene (positive screen; **Figure 11a**). Highly fluorescent clones were isolated, grown in the presence and absence of the target monoterpene, and the ten clones with the high signal-to-noise ratios for each different effector were sequenced, a total of 40 CamR variants. Given that a majority (22/40) of the clones had only single substitutions, and that 34/40 clones had either the A139T or Y58F substitution, we hypothesized that we could best understand the major contributions to responsivity and specificity on a substitution-by-substitution basis. We therefore cloned all 30 unique single amino acid substitutions that were recovered from CamR evolution and screened them for activity against camphor, borneol, fenchol, eucalyptol, and camphene (**Figure 11b**). Notably, three substitutions Y58F, H102N, and A139T were found to increase the response to at least one terpene by 6- to 7.5-fold. Some 10 other substitutions increased responsivity to at least one terpene by 2- to 3-fold, five increased the response by 20-50%, and 12 variants were found to not significantly increase the response to any terpenes. Mapping all 11 of the most productive substitutions onto a homology model of CamR revealed that 9/11 substitutions were localized in the ligand binding pocket (**Figure 11c**).

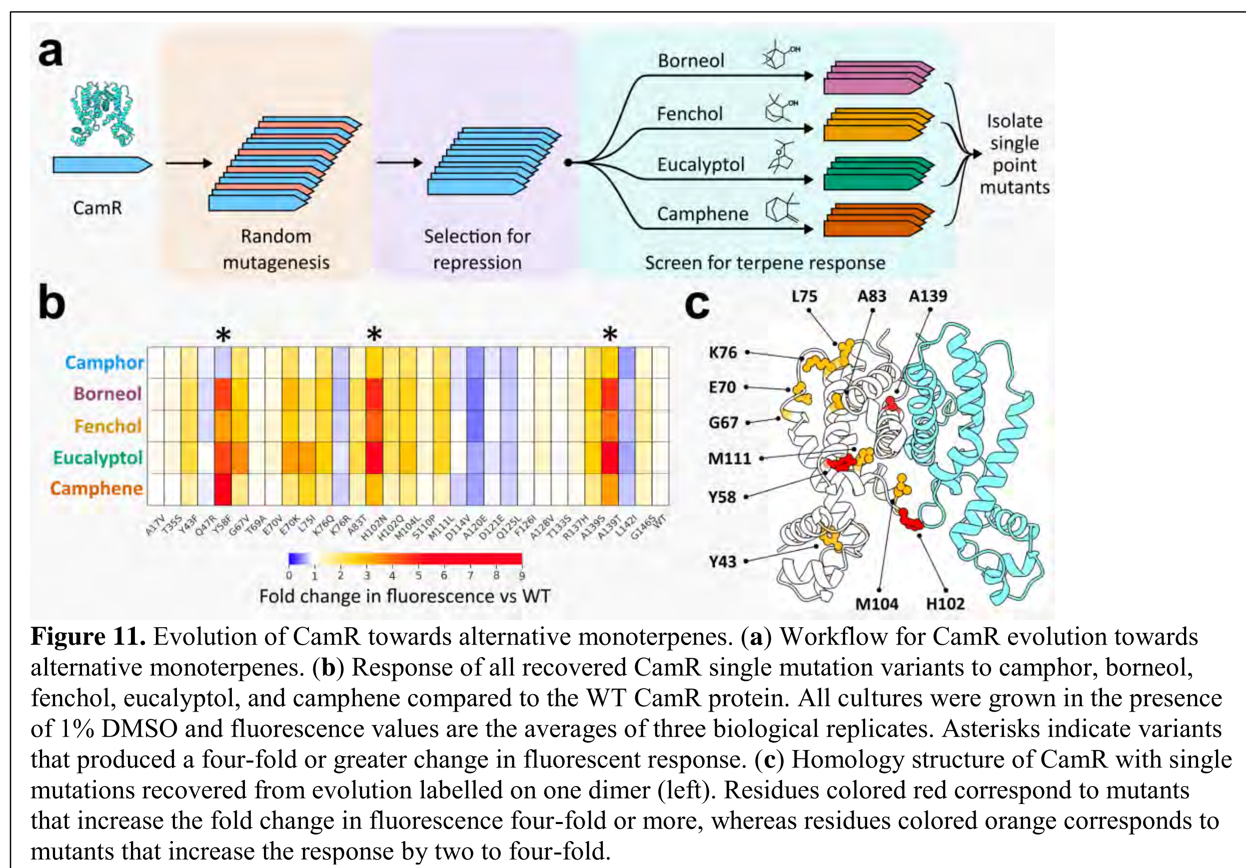


Figure 11. Evolution of CamR towards alternative monoterpenes. (**a**) Workflow for CamR evolution towards alternative monoterpenes. (**b**) Response of all recovered CamR single mutation variants to camphor, borneol, fenchol, eucalyptol, and camphene compared to the WT CamR protein. All cultures were grown in the presence of 1% DMSO and fluorescence values are the averages of three biological replicates. Asterisks indicate variants that produced a four-fold or greater change in fluorescent response. (**c**) Homology structure of CamR with single mutations recovered from evolution labelled on one dimer (left). Residues colored red correspond to mutants that increase the fold change in fluorescence four-fold or more, whereas residues colored orange corresponds to mutants that increase the response by two to four-fold.

Beyond monoterpenes, we have also applied the SELIS approach to develop highly specific and sensitive biosensors for therapeutic plant alkaloid intermediates. In particular, we focused on the Amaryllidaceae alkaloids, which include the FDA-approved anti-Alzheimer's drug galantamine that is currently sourced from daffodil plants due to challenges in its chemical synthesis. 4-Omethyl-norbelladine (4NB) is the key branchpoint intermediate to all known Amaryllidaceae

alkaloids, and thus was targeted for biosensor generation. The generalist transcription factor RamR from *Salmonella enterica* was found to be slightly responsive to both 4NB and its immediate precursor norbelladine (NB) and therefore was chosen as the starting point for evolution. RamR was introduced into the SELIS pipeline in a similar manner as for CamR.

Leveraging our newly evolved alkaloid biosensors, we developed circuits to report on the *in vivo* activity of an Amaryllidaceae alkaloid plant enzyme in *E. coli*. The enzyme Norbelladine 4-Omethyltransferase (Nb4OMT) from *Narcissus pseudonarcissus* (wild daffodil) was expressed on a plasmid that was co-transformed with our 4NB reporter construct. Since our RamR-4NB2 biosensor does not respond to the substrate NB, we were able to grow our engineered *E. coli* strains in the presence of NB and monitor its conversion to 4NB via the Nb4OMT enzyme (**Figure 12a**). The fluorescent response from our circuit was dependent on both the level of Nb4OMT expression as well as the amount of NB supplemented in the media, producing up to a 100-fold when compared to similar conditions using an inactive enzyme (**Figure 12b**).

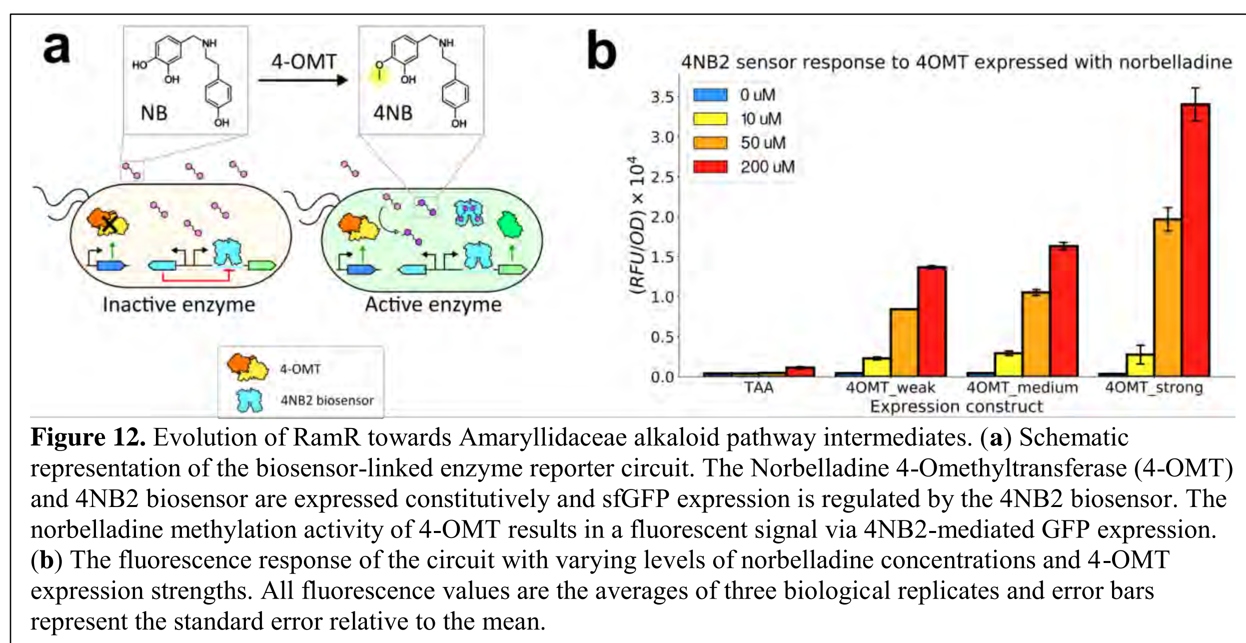


Figure 12. Evolution of RamR towards Amaryllidaceae alkaloid pathway intermediates. (**a**) Schematic representation of the biosensor-linked enzyme reporter circuit. The Norbelladine 4-Omethyltransferase (4-OMT) and 4NB2 biosensor are expressed constitutively and sfGFP expression is regulated by the 4NB2 biosensor. The norbelladine methylation activity of 4-OMT results in a fluorescent signal via 4NB2-mediated GFP expression. (**b**) The fluorescence response of the circuit with varying levels of norbelladine concentrations and 4-OMT expression strengths. All fluorescence values are the averages of three biological replicates and error bars represent the standard error relative to the mean.

Complex circuitry and patterning by cells

Through this effort, we focused on a mechanism based on cell-to-cell contact inhibition and developed an efficient mathematical approach to predict and design patterns. This approach blends graph-theoretic ideas with dynamical systems techniques to uncover bifurcation mechanisms for patterning, and is applicable to broad classes of signaling models. We represent the network as a graph where each vertex represents the dynamics of an identical individual cell and where graph edges represent cell-to-cell signaling. To predict steady-state patterns we find equitable partitions of the graph vertices and assign them into disjoint classes. We then use results from monotone systems theory to prove the existence of patterns that are structured in such a way that all the cells in the same class have the same final fate. As application examples, we considered neural networks, reaction-diffusion systems, and random graphs to illustrate the practical implications of our results. Importantly, our results do not restrict the number of cells or reactants, allow multiple communicating species with possibly different interconnection structure, and do not assume symmetric connections between two connected cells.

Our next task was the experimental validation of the theory with a synthetic gene network. The broader objective is to showcase a theory-based design of a synthetic system that enables spontaneous pattern formation. We developed what we call a "compartmental lateral inhibition" (CLI) system using diffusible molecules and engineered communication channels to mimic contact-dependent inhibition (CDI) that is common in developmental biology; see **Figure 13** for a comparison of CDI and CLI systems.

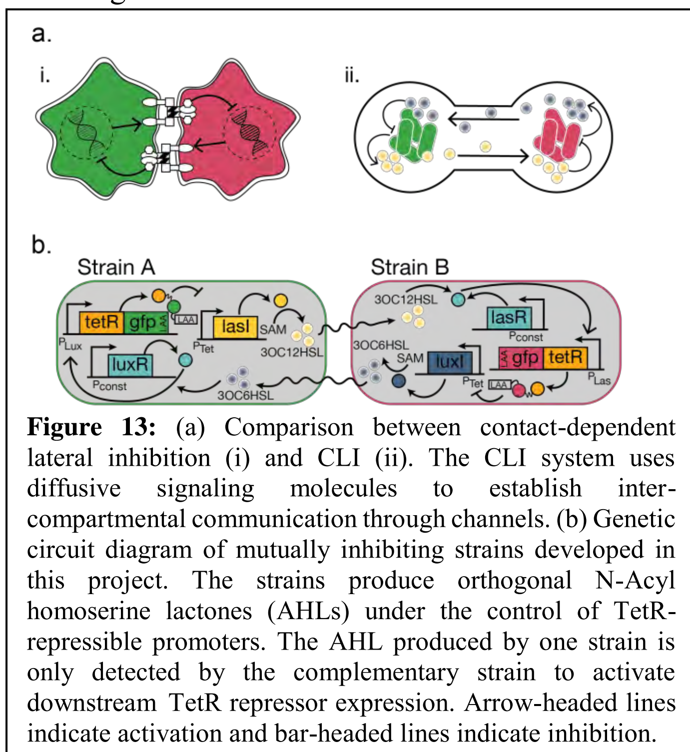


Figure 13: (a) Comparison between contact-dependent lateral inhibition (i) and CLI (ii). The CLI system uses diffusive signaling molecules to establish inter-compartmental communication through channels. (b) Genetic circuit diagram of mutually inhibiting strains developed in this project. The strains produce orthogonal N-Acyl homoserine lactones (AHLs) under the control of TetR-repressible promoters. The AHL produced by one strain is only detected by the complementary strain to activate downstream TetR repressor expression. Arrow-headed lines indicate activation and bar-headed lines indicate inhibition.

The CLI system comprises two bacterial colonies, Strain A and Strain B (Figure 1b), seeded in compartments connected by channels through which signaling molecules diffuse. Connected compartments communicate with each other via orthogonal QS systems, LuxR/LuxI from V. *fischeri* and LasR/LasI from P. *aeruginosa*. The use of orthogonal QS systems prevents self-communication within the same colonies. Strain A senses N-(3-oxo-hexanoyl)-L-homoserine lactone (3OC6HSL) through LuxR transcriptional activator binding to 3OC6HSL to activate 3OC6HSL-LuxR-responsive promoter *pLux76*. Strain B senses the orthogonal AHL signal, N-(3-oxo-dodecanoyl)-L-homoserine lactone (3OC12HSL), with LasR activator binding to 3OC12HSL-LasR-responsive promoter *pLas81*. To send AHLs that are recognized by the opposite strain, Strain A produces 3OC12HSL into medium with *lasI* AHL synthase, and Strain B produces 3OC6HSL with *luxI* AHL synthase. In each strain, corresponding AHL activates transcription of *tetR* transcriptional repressor fused to a reporter fluorescent protein (sfGFP). TetR in one strain represses transcription of the orthogonal AHL synthase that activates the other strain. Thus, the overall circuit establishes mutual inhibition between Strain A and Strain B.

We adapted the aforementioned theory to the CLI system and characterized the parameter regions that lead to contrasting patterns of gene expression, where either Strain A or Strain B expresses TetR, and disables the other strain from doing so. This study revealed the importance of geometry as a controllable factor for the designer, as the spatial arrangement of compartments and channels is crucial for determining patterning. We experimentally implemented mutually inhibiting systems by sequentially adjusting individual components to achieve the overarching objective. We constructed the geometrical culturing platform from agarose gel and polydimethylsiloxane mold, and simulated its design validity using partial differential equation modeling. Experiments with the engineered strains on the agarose gel platform showed agreement with the theoretical work, exhibiting contrasting gene expression patterns tunable by spatial arrangements of the inter-compartmental network and initial conditions (**Figure 14**).

This work successfully demonstrated control over a multi-strain community in theoretical and experimental frameworks, and identified design principles required for differentiation and pattern formation during developmental processes in natural biological systems.

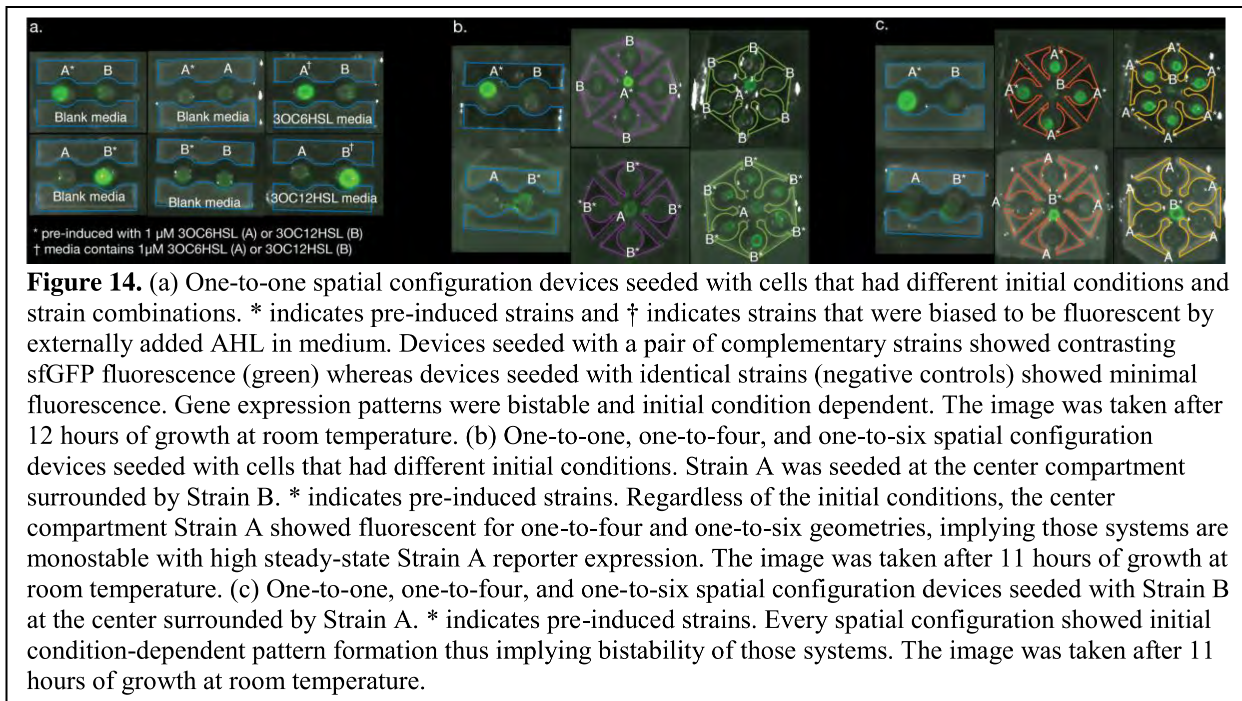


Figure 14. (a) One-to-one spatial configuration devices seeded with cells that had different initial conditions and strain combinations. * indicates pre-induced strains and † indicates strains that were biased to be fluorescent by externally added AHL in medium. Devices seeded with a pair of complementary strains showed contrasting sfGFP fluorescence (green) whereas devices seeded with identical strains (negative controls) showed minimal fluorescence. Gene expression patterns were bistable and initial condition dependent. The image was taken after 12 hours of growth at room temperature. (b) One-to-one, one-to-four, and one-to-six spatial configuration devices seeded with cells that had different initial conditions. Strain A was seeded at the center compartment surrounded by Strain B. * indicates pre-induced strains. Regardless of the initial conditions, the center compartment Strain A showed fluorescent for one-to-four and one-to-six geometries, implying those systems are monostable with high steady-state Strain A reporter expression. The image was taken after 11 hours of growth at room temperature. (c) One-to-one, one-to-four, and one-to-six spatial configuration devices seeded with Strain B at the center surrounded by Strain A. * indicates pre-induced strains. Every spatial configuration showed initial condition-dependent pattern formation thus implying bistability of those systems. The image was taken after 11 hours of growth at room temperature.

Organismal engineering with complex circuitry

To enable the facile construction and rapid characterization of genetic circuits especially those comprising of orthogonal transcriptional elements such as T7 RNA polymerase (T7 RNAP), we have designed a highly modular architecture for their assembly termed Universal Plasmid System (UPS). Briefly, in the UPS framework, each component of the plasmid is encoded as modular units /parts flanked by BsaI restriction sites followed by unique ligation sites (**Figure 15**). First, a plasmid backbone comprising of an origin of replication and selectable marker of choice is assembled along with specific landing pads for the insertion of defined transcriptional units (**Figure 15**). In parallel, gene circuits composed of modular elements such as promoters, ribosomal binding sites, coding

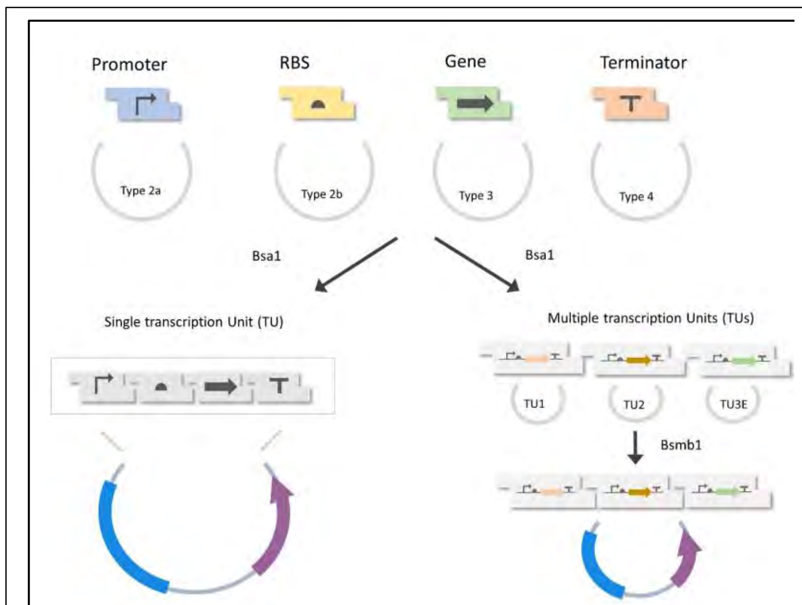


Figure 15: Modular architecture for the rapid assembly of genetic circuit in prokaryotes. Genetic circuits are comprised of modular units – promoters, RBSs, genes and terminators. Each element is maintained as a separate part flanked by Type IIS restriction enzyme recognition sites. By designing appropriate ‘sticky’ ends, these parts can be seamlessly assembled into genetic circuits into plasmid backbones of choice.

sequences and terminators are assembled to form transcriptional units (TUs). Each TU is flanked by appropriate 'connector' sequences that dictate the order of their assembly into the plasmid backbone (**Figure 15**).

After evaluating the activity of these T7 RNAP homeostasis circuits in various strains of *E. coli*, we sought to evaluate their portability across different bacteria. To do so, we characterized the performance of the T7 RNAP homeostasis circuits in different host chassis. To this end, we selected two species – *Salmonella typhimurium* and *Pseudomonas putida*. *S.typhimurium* strains has been demonstrated to selectively colonize the hypoxic conditions, such as those prevalent in

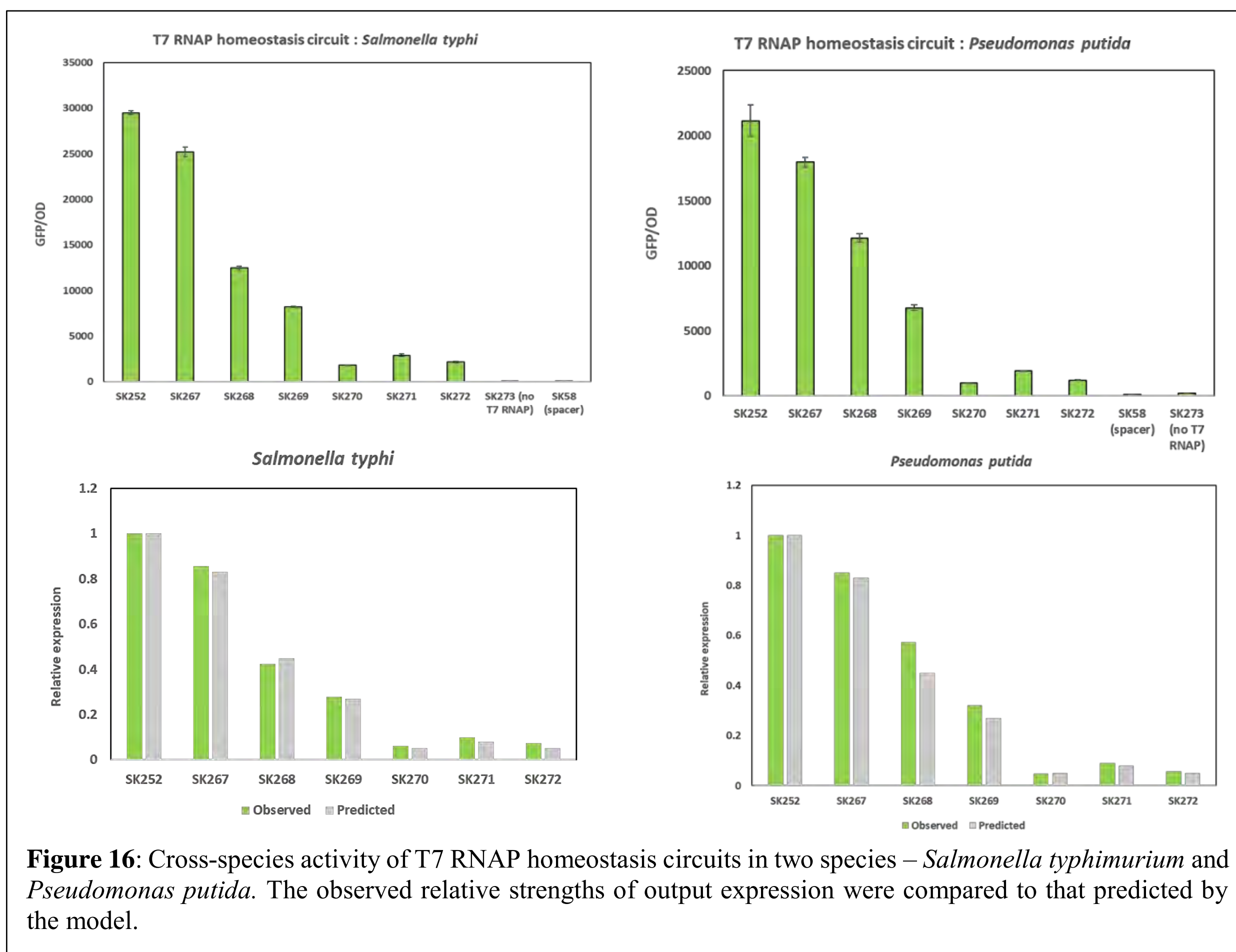


Figure 16: Cross-species activity of T7 RNAP homeostasis circuits in two species – *Salmonella typhimurium* and *Pseudomonas putida.* The observed relative strengths of output expression were compared to that predicted by the model.

the tumor microenvironments and thus serves as an ideal chassis for the selective delivery of therapeutic payloads to these environments. *P.putida* is a well characterized microbe which has been extensively used for industrial production of a wide variety of chemicals as well as being engineered for bioremediation applications. The panel of homeostasis circuits were transformed in both species, and the output (sfGFP) levels were characterized. In both species, the homeostasis circuits showed highly orthogonal expression of the target gene, without the need of any additional engineering of the plasmid (**Figure 16**). Furthermore, in both species, the relative output levels were nearly identical to that obtained in *E. coli* strains, suggesting that such T7 RNAP homeostasis circuits can be used for precise and predictable gene expression platform across different prokaryotes (**Figure 16**).

Construction and characterization of cellular circuits capable of exerting Boolean-style logic control over expression

Synthetic control of gene expression—whether from choice of promoter to higher order Boolean-style logic—is an important tool for metabolic engineering. However, the construction of these circuits is limited by lengthy design cycles and cross-talk among native cellular transcription factors. In this project, we aim to demonstrate that an orthogonal T7-based system linked with CRISPR dCas9 regulation can enable both simple regulation and complex circuitry in the yeast *Saccharomyces cerevisiae*. Specifically, through leveraging a suite of heterologous T7 polymerase variants that can be used to induce CRISPR dCas9 based gene expression, we can easily develop an approach to gene expression control that circumvents complex promoter engineering efforts. Thus far, our results show that, dependent on the combination of T7 polymerase 'switch' component and dCas9+sgRNA 'actuator' component, we can achieve AND/OR/NAND/NOR style control in response to external stimuli of combinations of galactose and/or copper. We characterize our AND/NAND-style circuitry further, using time-course induction studies, and have found these cell lines to have the greatest increase in reporter signal when the environmental stimulus is introduced during early exponential phase growth. Overall, we expect the technology developed in this work to greatly reduce the time and labor necessary for developing Boolean gene circuits for novel applications including metabolic pathway control in the future. These efforts led to the successful construction of model-guided logic gates that functioned in *S. cerevisiae* (**Figure 17**) taking advantage of the models and paradigms developed using this BRI.

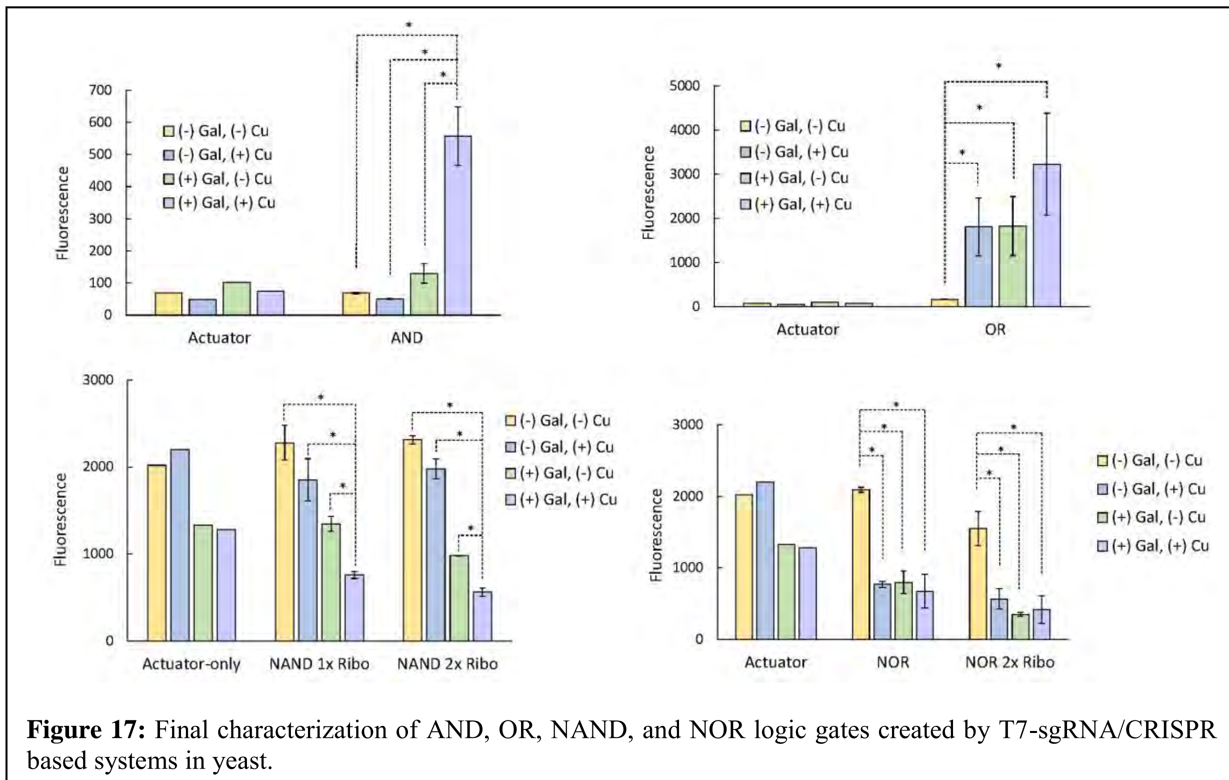


Figure 17: Final characterization of AND, OR, NAND, and NOR logic gates created by T7-sgRNA/CRISPR based systems in yeast.

VI. Major Problems/Issues

Beyond Covid-19 related delays that occurred during the project, this project was impacted by a co-PI, Dr. Howard Salis, improperly charging funds toward unrelated research (in this case, Covid-related explorations). Upon identifying these budgeting anomalies, the subcontract with this PI was terminated. Remaining funds and tasks were moved back to UT-Austin and project tasks have been re-allocated. This caused some setbacks in the work especially in Task 9 with efforts in extremophiles that has been mainly incomplete in this project. Work in Tasks 7-8 has enabled further collaborations with Dr. Soha Hassoun at Tufts University. Her work on machine learning has already started to develop new models for synthetic parts that will be published in future papers.

VII. Technology Transfer

Several patents have been generated from this work and have subsequently been licensed to companies. Many of the constructs from this work have been placed on Addgene and have been distributed across the World. The online computational genetic calculators have been accessed by research groups around the World. Direct technology transfer with AFRL labs has taken place during this period.

VIII. Productivity

The following published papers and have resulted from the funded project (note: additional papers are in review and under preparation):

Published Papers in this BRI Project

(Note: 4 additional papers under final preparation)

1. Heidi Redden and Hal Alper, 2015, *The development and characterization of synthetic minimal yeast promoters,* **Nature Communications**, 6:7810.
2. Ng C.Y., I. Farasat, C. Maranas, and H.M. Salis, 2015, *Rational Design of a Synthetic Entner-Doudoroff pathway for Improved and Controllable NADPH Regeneration*, **Metabolic Engineering**, v29.
3. Tian, Tian, and Howard M. Salis, 2015, *A predictive biophysical model of translational coupling to coordinate and control protein expression in bacterial operons,* ***Nucleic acids research*** 43.14: 7137-7151.
4. John Leavitt and Hal Alper, 2015, *Advances and Current Limitations in Transcript-Level Control of Gene Expression*, ***Current Opinion in Biotechnology***, 34, 98-104.
5. Espah, B. A., Mishler, D. M., Wang, J., Huso, W., & Salis, H. M, 2016. ***Breakthrough Article:*** *Automated physics-based design of synthetic riboswitches from diverse RNA aptamers*. ***Nucleic acids research***, *44*(1), 1-13.
6. Espah Borujeni, Amin, and Howard M. Salis, 2016. *Translation Initiation is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism*. ***Journal of the American Chemical Society***.
7. Farasat, Iman, and Howard M. Salis., 2016, *A Biophysical Model of CRISPR/Cas9 Activity for Rational Design of Genome Editing and Gene Regulation*. ***PLoS Computational Biology*** 12.1: e1004724.
8. Hsia J., Holtz W.J., Maharbiz M.M., Arcak M., & Keasling J.D., 2016. Modular Synthetic Inverters from Zinc Finger Proteins and Small RNAs. ***PLoS ONE*** 11(2): e0149483.

9. John Leavitt, Alice Tong, Joyce Tong, Jonathan Pattie, and Hal Alper, 2016, *Coordinated Transcription Factor and Promoter Engineering to Establish Strong Expression Elements in Saccharomyces cerevisiae*, ***Biotechnology Journal***, 11(7), 866–876.
10. Drew S Tack, Jared W Ellefson, Ross Thyer, Bo Wang, Jimmy Gollihar, Matthew T Forster, and Andrew D Ellington, 2016, *Addicting diverse bacteria to a noncanonical amino acid*. ***Nature Chemical Biology***, 12(3), 138–140.
11. Gyorgy, A. and Arcak, M., 2017, Pattern Formation over Multigraphs, **IEEE Transactions on Network Science and Engineering**, Issue 99, p. 1-1.
12. Nicholas Morse, Madan Gopal, James Wagner, and Hal Alper, 2017, *Yeast terminator function can be modulated and designed based on predictions of nucleosome occupancy*, **ACS Synthetic Biology**, 6, 11, 2086–2095.
13. Matthew Deaner, Julio Mejia, and Hal Alper, 2017, *Enabling graded and large-scale multiplex of desired genes using a dual-mode dCas9 activator in Saccharomyces cerevisiae*, **ACS Synthetic Biology**, 6(10):1931-194.
14. Matthew Deaner and Hal Alper, 2017, *Systematic Testing of Enzyme Perturbation Sensitivities via Graded dCas9 Modulation*, **Metabolic Engineering**, 40:14-22.
15. Halper, Sean M., Daniel P. Cetnar, and Howard M. Salis. 2017, *An Automated Pipeline for Engineering Many-Enzyme Pathways: Computational Sequence Design, Pathway Expression-Flux Mapping, and Scalable Pathway Optimization*. ***Synthetic Metabolic Pathways***. Humana Press, New York, NY, 2018. 39-61.
16. Espah Borujeni, A., Cetnar, D., Farasat, I., Smith, A., Lundgren, N. and Salis, H.M., 2017, *Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences*. ***Nucleic acids research***, **45**, 5437-5448.
17. Leistra, A. N., Gelderman, G., Sowa, S. W., Moon-walker, A., Salis, H. M., and Contreras, L. M., 2018, *Biophysical Model of CsrA-mRNA Interactions Expands Canonical Understanding of the CsrA Global Regulator Protein*, ***Scientific Reports***, v8, 9892.
18. Gyorgyi A., Arcak M, 2018, *Pattern formation over multigraphs*. ***IEEE Transactions on Network Science and Engineering***, vol. 5, no. 1, pp. 55-64.
19. Gomez M, Arcak M, 2018, *A tug-of-war mechanism for pattern formation in a genetic network*. ***ACS Synthetic Biology***, vol.6, no.11, pp. 55-64.
20. Thyer R., Shroff R., Klein DR., d'Oelsnitz S., Cotham VC., Byrom M., Brodbelt J.S.,and Ellington A.D., 2018. *Custom selenoprotein production enabled by laboratory evolution of recoded bacterial strains*. ***Nature Biotechnology*** 36(7):624-631.
21. Kar S, Ellington AD. 2018, *Transcription networks based on hairpin promoter switches*. ***ACS Synth Biol***., 10.1021/acssynbio.8b00172.
22. d'Oelsnitz S, Ellington A., 2018, Continuous directed evolution for strain and protein engineering. ***Curr Opin Biotechnol***. 2018 Feb 13;53:158-163.
23. Ellefson JW, Ledbetter MP, Ellington AD. 2018, Directed evolution of a synthetic phylogeny of programmable Trp repressors. ***Nature Chemical Biology***, Apr;14(4):361-367.
24. Kristin Presnell and Hal Alper, 2018, *Thermodynamic and first-principles biomolecular simulations applied to synthetic biology: Promoter and aptamer designs*, ***Molecular Systems Design & Engineering***, 3(1), 19-37.

25. Nicholas Morse, James Wagner, Kevin Reed, Madan Gopal, Lars Lauffer , and Hal Alper, 2018, *T7 Polymerase Expression of Guide RNAs in vivo Allows Exportable CRISPR-Cas9 Editing in Multiple Yeast Hosts*, ***ACS Synthetic Biology***, 7(4), 1075-1084.
26. Matthew Deaner, Allison Holzman, and Hal Alper, 2018, *Modular Ligation Extension of Guide RNA Operons (LEGO) for multiplexed dCas9 regulation of metabolic pathways in Saccharomyces cerevisiae*, ***Biotechnology Journal***, 13(9):e1700582.
27. Garry DJ, Meyer AJ, Ellefson JW, Bull JJ, Ellington AD. *Predicting Evolution of the Transcription Regulatory Network in a Bacteriophage*. 2018, ***Genome Biol Evol***. 2018;10(10):2614-28.
28. Simon AJ, Morrow BR, Ellington AD. 2018, *Retroelement-based genome editing and evolution*. ***ACS Synthetic Biology***, 7, 2600–2611.
29. Mika Tei, Melinda Perkins, Justin Hsia, Murat Arcak and Adam Arkin. 2019, *Designing spatially distributed gene regulatory networks to elicit contrasting patterns*. ***ACS Synthetic Biology***, vol.8, no.1, pages 119 – 126.
30. Simon AJ, d'Oelsnitz S, Ellington AD. 2019, *Synthetic evolution*. ***Nature Biotechnology***, 37(7):730-43.
31. Thyer R, Ellington AD. 2019, *The Role of tRNA in Establishing New Genetic Codes*. ***Biochemistry***. 2019;58(11):1460-3.
32. Holly Ekas, Matthew Deaner, and Hal Alper, 2019, *Recent advancements in fungal-derived fuel and chemical production and commercialization*, ***Current Opinion in Biotechnology***, 57, 1-9.
33. Kristin Presnell, Madeleine Flexer-Harrison, and Hal Alper, 2019, *Design and synthesis of synthetic UP elements for modulation of gene expression in Escherichia coli,* ***Synthetic and Systems Biotechnology***, 4(2): 99–106
34. Reis, A. C., Halper, S. M., Vezeau, G. E., Cetnar, D. P., Hossain, A., Clauer, P. R., & Salis, H. M., 2019, *Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays*. ***Nature biotechnology***, 37(11), 1294-1301.
35. Hossain, Ayaan, Eriberto Lopez, Sean M. Halper, Daniel P. Cetnar, Alexander C. Reis, Devin Strickland, Eric Klavins, and Howard M. Salis., 2020, *Automated design of thousands of nonrepetitive parts for engineering stable genetic systems*. ***Nature biotechnology,*** 1-10.
36. Halper, Sean M., Ayaan Hossain, and Howard M. Salis, 2020, *The Synthesis Success Calculator: Predicting the Rapid Synthesis of DNA Fragments with Machine Learning*. ***ACS Synthetic Biology*** 9, 7, 1563–1571.
37. Leiby, N., Hossain, A., & Salis, H. M., 2020. EPiC Series in Computing. In ***Proceedings of the 12th International Conference on Artificial Intelligence*** (Vol. 70, pp. 163-172).
38. Gardner, E. C., Watkins, E. J., Gollihar, J., & Ellington, A. D. 2020, *Emulsion-based directed evolution of enzymes and proteins in yeast*. ***In Methods in Enzymology***. Academic Press. https://doi.org/10.1016/bs.mie.2020.04.053
39. Shuo-Fu Yuan and Hal Alper, 2020, *Improving Spinach2- and Broccoli-based biosensors for single and double analytes*, ***Biotechnology Notes***, 1, 2-8.
40. Emily Bowman, Matthew Deaner, Jan-Fang Cheng, Robert Evans, Ernst Oberortner, Yasuo Yoshikuni, and Hal Alper, 2020. *Bi-directional Titration of Yeast Gene Expression using a Pooled CRISPR Guide RNA Approach* ***PNAS*** 117 (31) 18424-18430
41. Emily Bowman, James Wagner, Shuo-Fu Yuan, Matt Deaner, Claire Palmer, Simon D'Oelsnitz, Lauren Cordova, Xin Li, Frank F. Craig and Hal Alper, 2021, *Sorting for*

secreted molecule production using a biosensor-in-microdroplet approach, ***PNAS***, 118 (36) e2106818118.
42. Sierra Brooks and Hal Alper, 2021, Applications, challenges, and needs for employing synthetic biology beyond the lab, ***Nature Communications*** 12 (1390).
43. d'Oelsnitz S, Kim W, Burkholder N, Javanmardi K, Zhang J, Alper HS, Ellington AD, 2022. Directed evolution of a biosensor generalist accelerates alkaloid pathway engineering. *bioRxiv* https://www.biorxiv.org/content/10.1101/2021.06.07.447399v1