Final Technical Report

Title of Project: High-throughput chemical imaging for optimizing biofuel synthesis using synthetic biology
Grant Number: DE-SC0019387
Institution: Boston University

Mary Dunlop, Ph.D., Boston University (Principal Investigator)
Ji-Xin Cheng, Ph.D., Boston University (Co-Investigator)
Wilson Wong, Ph.D., Boston University (Co-Investigator)

Description/Abstract
Fatty acids can be produced biosynthetically in microbes and these compounds can serve as precursors to biodiesels and other high value oleochemicals. However, progress on engineering fatty acid biosynthesis, and biofuel synthesis more generally, has been hindered by current quantification methods that are either indirect or not amenable to high-throughput or single-cell resolution screening. In this project, we assembled an interdisciplinary team with complimentary expertise in synthetic biology and microscopy, metabolic engineering, and chemical imaging to address these challenges. We used chemical imaging to directly measure lipid biosynthesis in *Escherichia coli* engineered to produce fatty acids, obtaining detailed single-cell resolution measurements. We deployed stimulated Raman scattering (SRS) microscopy in concert with multiplexed genome engineering and gene circuit design strategies from synthetic biology to optimize production of fatty acids. These results provided novel insight into cell-to-cell heterogeneity present in biofuel production strains. In addition, we introduced new chemical imaging methods which are label-free and do not require fluorescent reporters. These efforts were complemented by other studied developing foundational tools for regulation and control, which offer excellent potential for advancing researchers' ability to rapidly design, build, and test strains for enhanced biofuel synthesis.

Keywords
metabolic engineering, biofuel synthesis, stimulated Raman scattering, deep learning, synthetic biology, single-cell analysis

Summary of Main Results

1. Stimulated Raman scattering (SRS) imaging of biofuel synthesis in live cells

Longitudinal Single-Cell Imaging of Engineered Strains with Stimulated Raman Scattering to Characterize Heterogeneity in Fatty Acid Production
Tague, Lin, Lugagne, O'Connor, Burman, Wong, Cheng, Dunlop – Advance Science DOI: 10.1002/advs.202206519, 2023 (Reference [1])

Understanding metabolic heterogeneity is critical for optimizing microbial production of valuable chemicals, but requires tools that can quantify metabolites at the single-cell level over time. Here, longitudinal hyperspectral stimulated Raman scattering (SRS) chemical imaging is developed to

directly visualize free fatty acids in engineered *Escherichia coli* over many cell cycles. Compositional analysis is also developed to estimate the chain length and unsaturation of the fatty acids in living cells. This method reveals substantial heterogeneity in fatty acid production among and within colonies that emerges over the course of many generations. Interestingly, the strains display distinct types of production heterogeneity in an enzyme-dependent manner. By pairing time-lapse and SRS imaging, the relationship between growth and production at the single-cell level are examined. The results demonstrate that cell-to-cell production heterogeneity is pervasive and provides a means to link single-cell and population-level production.

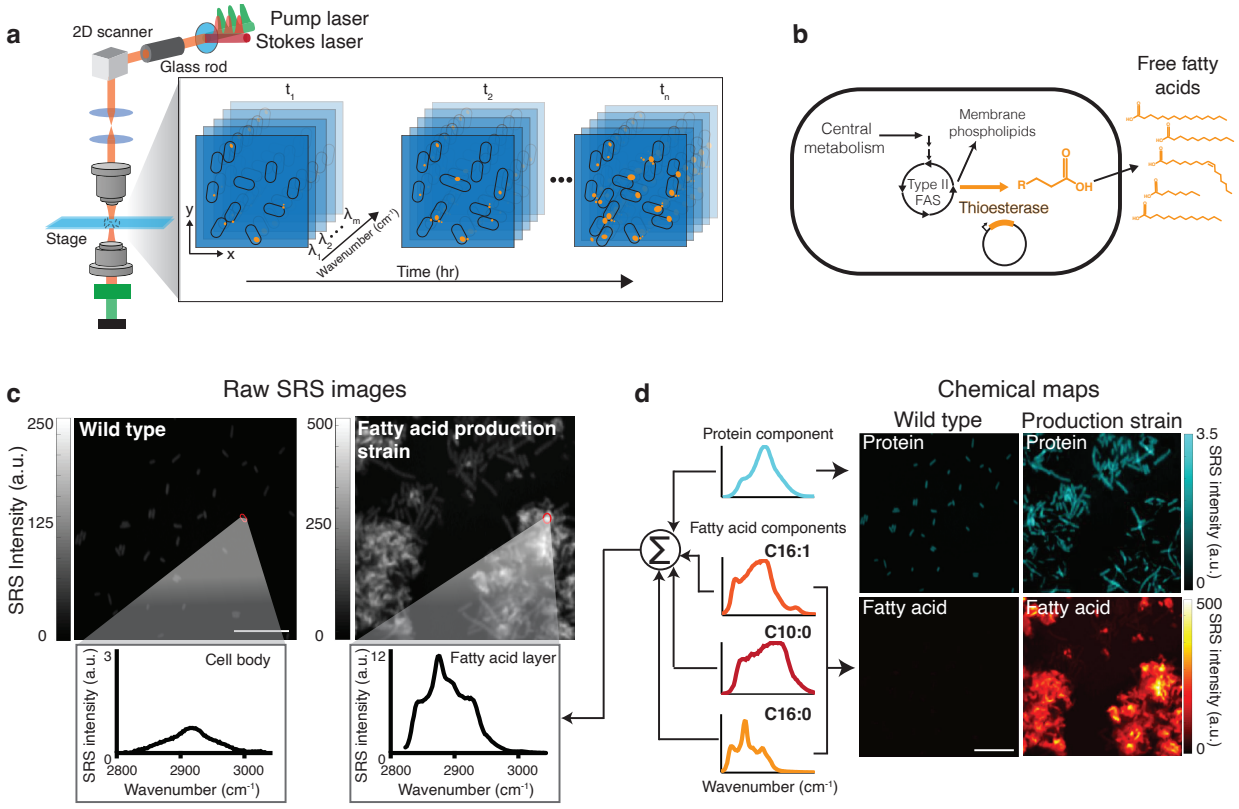

Figure 1. SRS imaging of *E. coli* production strains shows single-cell free fatty acid levels. (a) Schematic of the optical setup for SRS imaging to produce hyperspectral images using a Stokes and pump laser focused on a live sample. Hyperspectral SRS images contain three-dimensional data: x and y coordinates and wavenumber, which provides spectral information. Longitudinal SRS imaging adds a fourth dimension, time. (b) Schematic of free fatty acid production in *E. coli*. Expression of a cytosolic thioesterase results in free fatty acid accumulation through the type II fatty acid synthesis (FAS) pathway. Free fatty acids can vary in chain length and unsaturation, largely dictated by thioesterase specificity. (c) Representative raw SRS data from wild type *E. coli* and a strain overexpressing a cytosolic thioesterase (*Ab*TE*). The summation of Raman spectra at each pixel is shown. Representative regions are outlined in red with the corresponding Raman spectra shown below the image. Fatty acids and proteins emit strong Raman signals in the C–H region (~2900 cm^{-1}). Scale bar, 10 µm. (d) Spectra at each pixel of the SRS image can be decomposed to generate chemical maps. Protein and fatty acid components are decomposed using spectral standards to produce chemical maps. Spectral standards shown in schematic are Bovine serum albumin (cyan), palmitoleic acid (C16:1, orange), capric acid (C10:0, red), and palmitic acid

(C16:0, yellow). Protein and fatty acid chemical maps for both strains are shown. Scale bar, 10 μm.

Visualization of a Limonene Synthesis Metabolon Inside Living Bacteria by Hyperspectral SRS Microscopy
Zhang, Shin, Tague, Lin, Zhang, Ge, Wong, Dunlop, Cheng – Advanced Science DOI: 10.1002/advs.202203887, 2022 (Reference [2])

Monitoring biosynthesis activity at single-cell level is key to metabolic engineering but is still difficult to achieve in a label-free manner. Using hyperspectral stimulated Raman scattering imaging in the 670–900 cm^{-1} region, localized limonene synthesis are visualized inside engineered *Escherichia coli*. The colocalization of limonene and GFP-fused limonene synthase is confirmed by co-registered stimulated Raman scattering and two-photon fluorescence images. The finding suggests a limonene synthesis metabolon with a polar distribution inside the cells. This finding expands the knowledge of de novo limonene biosynthesis in engineered bacteria and highlights the potential of SRS chemical imaging in metabolic engineering research.

Microsecond Fingerprint Stimulated Raman Spectroscopic Imaging by Ultrafast Tuning and Spatial-Spectral Learning
Lin, Lee, Tague, Lugagne, Zong, Deng, Shin, Tian, Wong, Dunlop, Cheng – Nature Communications DOI: 10.1038/s41467-021-23202-z, 2021 (Reference [3])

Label-free vibrational imaging by stimulated Raman scattering (SRS) provides unprecedented insight into real-time chemical distributions. Specifically, SRS in the fingerprint region (400–1800 cm^{-1}) can resolve multiple chemicals in a complex bio-environment. However, due to the intrinsic weak Raman cross-sections and the lack of ultrafast spectral acquisition schemes with high spectral fidelity, SRS in the fingerprint region is not viable for studying living cells or large-scale tissue samples. Here, we report a fingerprint spectroscopic SRS platform that acquires a distortion-free SRS spectrum at 10 cm^{-1} spectral resolution within 20μs using a polygon scanner. Meanwhile, we significantly improve the signal-to-noise ratio by employing a spatial-spectral residual learning network, reaching a level comparable to that with 100 times integration. Collectively, our system enables high-speed vibrational spectroscopic imaging of multiple biomolecules in samples ranging from a single live microbe to a tissue slice.

2. Metabolic engineering to enhance fatty acid biofuel synthesis

Transcriptional Tuning of Mevalonate Pathway Enzymes to Identify Impact on Limonene Production in *E. coli*
Shin, South, Dunlop – ACS Omega DOI: 10.1021/acsomega.2c00483, 2022 (Reference [4])

Heterologous production of limonene in micro-organisms through the mevalonate (MVA) pathway has traditionally imposed metabolic burden and reduced cell fitness, where imbalanced stoichiometries among sequential enzymes result in the accumulation of toxic intermediates. Although prior studies have shown that changes to mRNA stability, RBS strength, and protein homology can be effective strategies for balancing enzyme levels in the MVA pathway, testing different variations of these parameters often requires distinct genetic constructs, which can

exponentially increase assembly costs as pathways increase in size. Here, we developed a multi-input transcriptional circuit to regulate the MVA pathway, where four chemical inducers, L-arabinose (Ara), choline chloride (Cho), cuminic acid (Cuma), and isopropylβ-D-1-thiogalactopyranoside (IPTG), each regulate one of four orthogonal promoters. We tested modular transcriptional regulation of the MVA pathway by placing this circuit in an engineered *Escherichia coli* "marionette" strain, which enabled systematic and independent tuning of the first three enzymes (AtoB, HMGS, and HMGR) in the MVA pathway. By systematically testing combinations of chemical inducers as inputs, we investigated relationships between the expressions of different MVA pathway submodules, finding that limonene yields are sensitive to the coordinated transcriptional regulation of HMGS and HMGR.

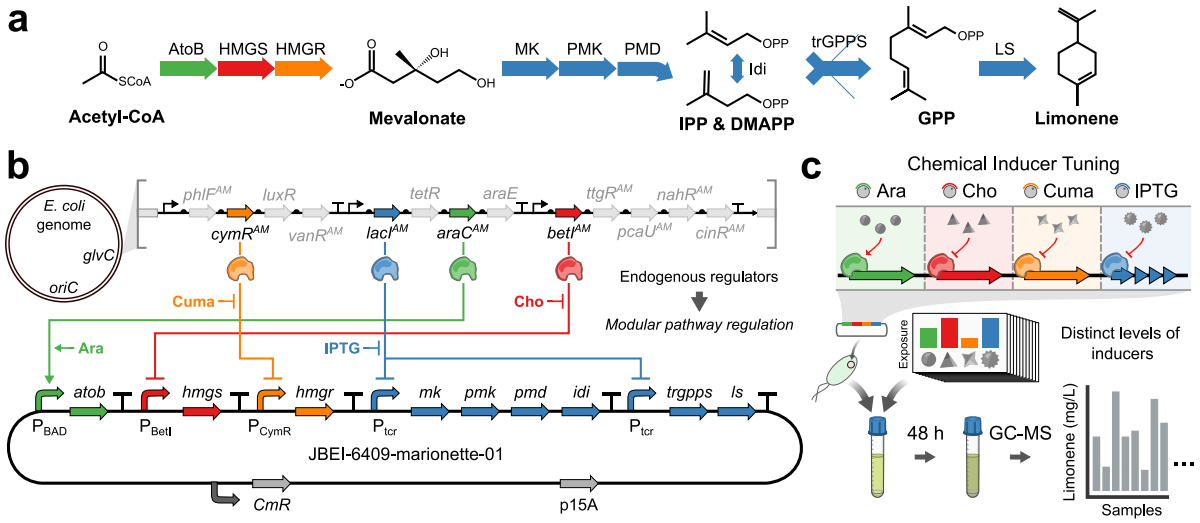

Figure 2. Modular transcriptional regulation of the mevalonate pathway to tune limonene production in *E. coli*. (a) Limonene production pathway, where acetyl-CoA serves as the starting substrate in the mevalonate (MVA) pathway, which then over multiple steps is consumed alongside NADPH and ATP to produce IPP or DMAPP and then limonene. Abbreviations: AtoB, acetoacetyl-CoA thiolase; HMGS, 3-hydroxy-3-methylglutaryl-CoA synthase; HMGR, 3-hydroxy-3-methylglutaryl-CoA reductase; MK, mevalonate kinase; PMK, phosphomevalonate kinase; PMD, mevalonate pyrophosphate decarboxylase; IPP, isopentenyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; Idi, isopentenyl diphosphate isomerase; trGPPS, truncated geranyl diphosphate synthase; GPP, geranyl diphosphate; LS, limonene synthase. (b) Marionette *E. coli* strain MG1655, where heterologous repressors are integrated into the genome at the *glvC* locus. These endogenous repressors are constitutively expressed and will bind to their cognate promoters on a plasmid to regulate downstream gene expression. Four inducible promoters (P_{BAD}, P_{BetI}, P_{CymR}, and P_{tcr}; induced by l-arabinose (Ara), choline chloride (Cho), cuminic acid (Cuma), and isopropyl β-d-1-thiogalactopyranoside (IPTG), respectively) are inserted to modularize the limonene synthesis pathway. (c) A marionette transcriptional circuit enables both modular and multivariate application of chemical inducers to tune the expression of enzymes. Limonene production after 48 hours of fermentation is measured by gas chromatography–mass spectrometry (GC-MS).

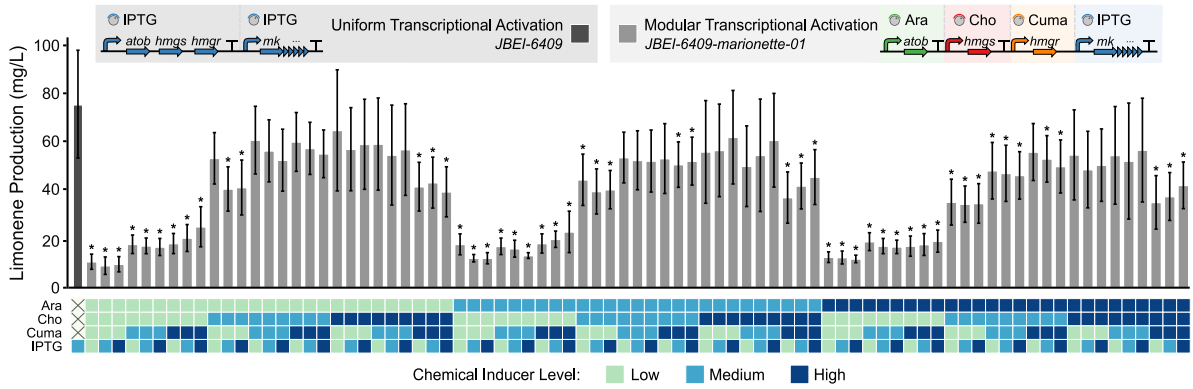

Figure 3. Limonene production in marionette *E. coli* strains across 81 chemical inducer combinations. Transcriptional regulation of different submodules of the MVA pathway significantly impacts limonene titers in batch culture after 48 hours of fermentation. Error bars represent standard deviation among replicates for each combination of chemical inducers (n = 6-8 biological replicates). Chemical inducer levels corresponding to low, medium, and high: arabinose (2, 10, 50 µM), choline chloride (50, 200, 500 µM), cuminic acid (2, 5, 20 µM), and IPTG (25, 100, 200 µM). Strains harboring the JBEI-6409 plasmid serve as a positive control. This uniform transcriptional activation is achieved using a single inducer, where P_{lacUV5} and P_{trc} promoters were induced with a medium concentration (100 µM) of IPTG. This JBEI-6409 strain produced a limonene titer of 76 mg/L. Modular multivariate transcriptional regulation was implemented with strains harboring the JBEI-6409-marionette-01 plasmid. Asterisks mark statistically significant differences in limonene titers between a given modular transcriptional activation sample (cells with JBEI-6409-marionette-01) and the uniform transcriptional activation control (cells with JBEI-6409). Statistical significance for normalized data was determined with Welch's unequal variances t-tests (α=0.05), followed by p-value correction for multiple testing using the Benjamini-Hochberg procedure (FDR = 0.10).

<u>Programmable Gene Regulation for Metabolic Engineering Using Decoy Transcription Factor Binding Sites</u>
Wang, Tague, Whelan, Dunlop – Nucleic Acids Research DOI: 10.1093/nar/gkaa1234 (Reference [5])

Transcription factor decoy binding sites are short DNA sequences that can titrate a transcription factor away from its natural binding site, therefore regulating gene expression. In this study, we harness synthetic transcription factor decoy systems to regulate gene expression for metabolic pathways in *Escherichia coli*. We show that transcription factor decoys can effectively regulate expression of native and heterologous genes. Tunability of the decoy can be engineered via changes in copy number or modifications to the DNA decoy site sequence. Using arginine biosynthesis as a showcase, we observed a 16-fold increase in arginine production when we introduced the decoy system to steer metabolic flux towards increased arginine biosynthesis, with negligible growth differences compared to the wild type strain. The decoy-based production strain retains high genetic integrity; in contrast to a gene knock-out approach where mutations were common, we detected no mutations in the production system using the decoy-based strain. We further show that transcription factor decoys are amenable to multiplexed library screening by demonstrating enhanced tolerance to pinene with a combinatorial decoy library. Our study shows

that transcription factor decoy binding sites are a powerful and compact tool for metabolic engineering.

3. Synthetic biology tool development for improving fatty acid synthesis

An Optogenetic Toolkit for Light-Inducible Antibiotic Resistance
Sheets, Tague, Dunlop – Nature Communications DOI: 10.1038/s41467-023-36670-2, 2023
(Reference [6])

Antibiotics are a key control mechanism for synthetic biology and microbiology. Resistance genes are used to select desired cells and regulate bacterial populations, however their use to-date has been largely static. Precise spatiotemporal control of antibiotic resistance could enable a wide variety of applications that require dynamic control of susceptibility and survival. Here, we use light-inducible Cre recombinase to activate expression of drug resistance genes in *Escherichia coli*. We demonstrate light-activated resistance to four antibiotics: carbenicillin, kanamycin, chloramphenicol, and tetracycline. Cells exposed to blue light survive in the presence of lethal antibiotic concentrations, while those kept in the dark do not. To optimize resistance induction, we vary promoter, ribosome binding site, and enzyme variant strength using chromosome and plasmid-based constructs. We then link inducible resistance to expression of a heterologous fatty acid enzyme to increase production of octanoic acid. These optogenetic resistance tools pave the way for spatiotemporal control of cell survival.

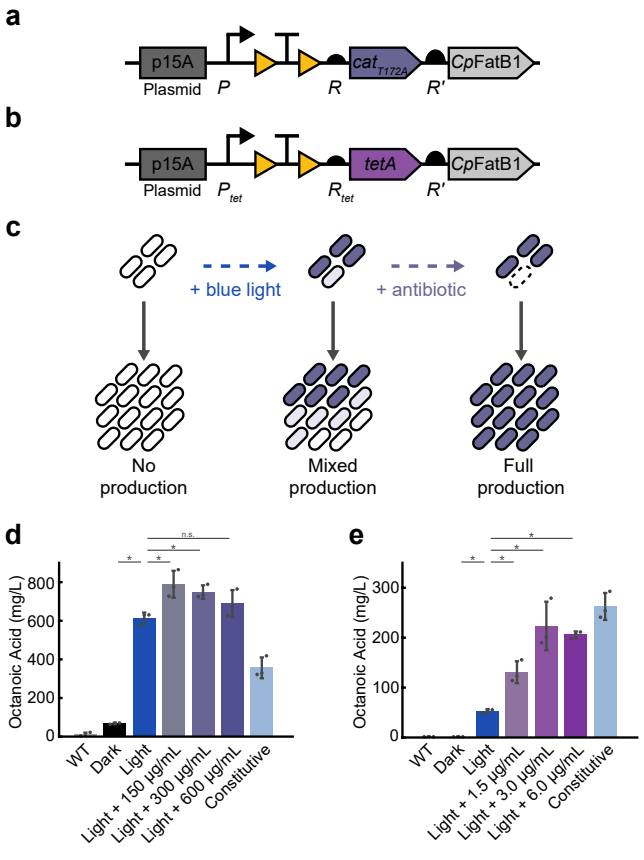

Figure 4. Light-inducible production of octanoic acid. (a) Activation of OptoCre-cat using cat_{T172A} or (b) OptoCre-tetA is coupled with CpFatB1 by introducing the gene downstream of the drug resistance marker under the strong RBS R'. (c) Blue light induces expression of the resistance gene and CpFatB1, but does not guarantee all individuals within a community are expressing the genes. Non-producers have a growth advantage due to the burden of CpFatB1. Addition of chloramphenicol prevents growth of any individuals not producing the resistance gene and consequently CpFatB1. (d) Octanoic acid production coupled with OptoCre-cat measured by GC-MS. (e) Octanoic acid production coupled with OptoCre-tetA measured by GC-MS. Significance was determined using a two-tailed Welch's t-test: *P < 0.05; n.s. not significant. Error bars show standard deviation around the mean (n = 3 biological replicates).

Light inducible protein degradation in *E. coli* with the LOVdeg tag
Tague, Coriano-Ortiz, Sheets, Dunlop – eLife (Accepted, to appear) 2024 (Reference [7])

Molecular tools for optogenetic control allow for spatial and temporal regulation of cell behavior. In particular, light controlled protein degradation is a valuable mechanism of regulation because it can be highly modular, used in tandem with other control mechanisms, and maintain functionality throughout growth phases. Here, we engineered LOVdeg, a tag that can be appended to a protein of interest for inducible degradation in *Escherichia coli* using blue light. We demonstrate the modularity of LOVdeg by using it to tag a range of proteins, including the LacI repressor, CRISPRa activator, and the AcrB efflux pump. Additionally, we demonstrate the utility of pairing the LOVdeg tag with existing optogenetic tools to enhance performance by developing a combined EL222 and LOVdeg system. Finally, we use the LOVdeg tag in a metabolic engineering application to demonstrate post-translational control of metabolism. Together, our results highlight the modularity and functionality of the LOVdeg tag system, and introduce a powerful new tool for bacterial optogenetics.

4. Foundational tools for analysis and control of gene regulation

DeLTA: Automated Cell Segmentation, Tracking, and Lineage Reconstruction using Deep Learning
Lugagne, Lin, Dunlop – PLOS Computation Biology DOI: 10.1371/journal.pcbi.1007673, 2020 (Reference [8])

The study introduces a deep learning-based pipeline for automated segmentation, tracking, and lineage reconstruction from single-cell microscopy data.

Cell-Machine Interfaces for Characterizing Gene Regulatory Network Dynamics
Lugagne, Dunlop – Current Opinion in Systems Biology DOI: 10.1016/j.coisb.2019.01.001, 2019 (Reference [9])

This review article describes new tools and approaches for interfacing cells and computers to improve data analysis and control.

Controlled Protein Activities with Viral Proteases, Antiviral Peptides, and Antiviral Drugs
Tague, McMahan, Tague, Dunlop, Ngo – ACS Chemical Biology DOI: 10.1021/acschembio.3c00138, 2023 (Reference [10])

The study introduces a new tool for metabolic engineering and synthetic biology studies based on a catalytically inactive NS3 protease which works in diverse organisms.

Comprehensive Screening of a Light-Inducible Split Cre Recombinase with Domain Insertion Profiling
Tague, Andreani, Fan, Timp, Dunlop – ACS Synthetic Biology DOI: 10.1021/acssynbio.3c00328 (Reference [11])

We introduce a pooled library approach using optogenetic dimers to rapidly screen split protein constructs, offering a streamlined method for achieving precise post-translational control.

Products Delivered

Publications

[1] N. Tague *et al.*, "Longitudinal Single-Cell Imaging of Engineered Strains with Stimulated Raman Scattering to Characterize Heterogeneity in Fatty Acid Production," *Advanced Science*, vol. 10, no. 20, p. 2206519, 2023, doi: 10.1002/advs.202206519.

[2] J. Zhang *et al.*, "Visualization of a Limonene Synthesis Metabolon Inside Living Bacteria by Hyperspectral SRS Microscopy," *Advanced Science*, vol. 9, no. 32, p. 2203887, 2022, doi: 10.1002/advs.202203887.

[3] H. Lin *et al.*, "Microsecond fingerprint stimulated Raman spectroscopic imaging by ultrafast tuning and spatial-spectral learning," *Nature Communications*, vol. 12, no. 1, p. 3052, Dec. 2021, doi: 10.1038/s41467-021-23202-z.

[4] J. Shin, E. J. South, and M. J. Dunlop, "Transcriptional Tuning of Mevalonate Pathway Enzymes to Identify the Impact on Limonene Production in Escherichia coli," *ACS Omega*, May 2022, doi: 10.1021/ACSOMEGA.2C00483.

[5] T. Wang, N. Tague, S. A. Whelan, and M. J. Dunlop, "Programmable gene regulation for metabolic engineering using decoy transcription factor binding sites," *Nucleic acids research*, vol. 49, no. 2, 2021, doi: 10.1093/nar/gkaa1234.

[6] M. B. Sheets, N. Tague, and M. J. Dunlop, "An optogenetic toolkit for light-inducible antibiotic resistance," *Nat Commun*, vol. 14, no. 1, Art. no. 1, Feb. 2023, doi: 10.1038/s41467-023-36670-2.

[7] N. Tague, C. Coriano-Ortiz, M. B. Sheets, and M. J. Dunlop, "Light inducible protein degradation in E. coli with the LOVdeg tag." bioRxiv, p. 2023.02.25.530042, Oct. 26, 2023. doi: 10.1101/2023.02.25.530042. (Accepted: *eLife*, to appear)

[8] J.-B. Lugagne, H. Lin, and M. J. Dunlop, "DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning," *PLoS Computational Biology*, p. PMCID: PMC7153852, 2020.

[9] J.-B. Lugagne and M. J. Dunlop, "Cell-machine interfaces for characterizing gene regulatory network dynamics," *Current Opinion in Systems Biology*, vol. 14, p. PMCID: PMC6774389, Apr. 2019, doi: 10.1016/j.coisb.2019.01.001.

[10] E. P. Tague, J. B. McMahan, N. Tague, M. J. Dunlop, and J. T. Ngo, "Controlled Protein Activities with Viral Proteases, Antiviral Peptides, and Antiviral Drugs," *ACS Chem. Biol.*, vol. 18, no. 5, pp. 1228–1236, May 2023, doi: 10.1021/acschembio.3c00138.

[11] N. Tague, V. Andreani, Y. Fan, W. Timp, and M. J. Dunlop, "Comprehensive Screening of a Light-Inducible Split Cre Recombinase with Domain Insertion Profiling," *ACS Synth. Biol.*, vol. 12, no. 10, pp. 2834–2842, Oct. 2023, doi: 10.1021/acssynbio.3c00328.

Theses

Nathan Tague, Ph.D., Biomedical Engineering, 2023
Thesis: Utilizing light as an input and output for synthetic biology and metabolic engineering

Jing Zhang, Ph.D., Biomedical Engineering, 2023
Thesis: High-throughput single-cell imaging and sorting by stimulated Raman scattering microscopy and laser-induced ejection

Haonan Lin, Ph.D., Biomedical Engineering 2021
Thesis: Stimulated Raman spectroscopic imaging: Data science-driven innovations and application

Deeya Burman, M.S., Biomedical Engineering, 2020
Project: Design, build, and test of *Escherichia coli* strains engineered to produce free fatty acids