Synthetic biology tools for sensing and evolving microbes

CRISPR-diagnostics and transposon-mediated genome re-wiring

by

Max 'Atti' English

M.Sci., University of Cambridge (2017)
B.A., University of Cambridge (2016)

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

Doctor of Philosophy

at the

Massachusetts Institute of Technology

September 2022

Signature of Author _____

Department of Biological Engineering

Certified by _____

James J. Collins, Ph.D.
Termeer Professor of Medical Engineering and Science
Thesis supervisor

Accepted by _____

Katharina Ribbeck, Ph.D.
Andrew (1956) and Erna Viterbi Professor
Graduate Program Chair, Dept. Biological Engineering

Thank you to my thesis committee for your advice and support:

James J. Collins, Ph.D.
Termeer Professor of Medical Engineering and Science
Thesis Supervisor

Forest M. White, Ph.D.
Ned C. (1949) and Janet C. (Bemis) Rice Professor
Thesis Committee Chair

Kristala L. Jones Prather, Ph.D.
Arthur Dehon Little Professor, Department Executive Officer

Core publications and collaborations

This thesis draws directly from three core papers, each of which was a close collaboration with other members of the Collins Lab. Parts of the text are derived (partially or in whole) from sections of text originally written for these publications. Similarly, several of the figures presented here are derived from preliminary or published figures from these publications. Given the many contributors to these papers, I would like to start by thanking my good friends and co-authors Raph (RVG), Helena (HP) and Luis (LRS) for all of their incredible work on these publications. I would also like to thank Miguel (MAA), who has made invaluable contributions to the directed genome evolution project, and from whom I have learned the true power of dictionary comprehension.

At the end of each chapter, I highlight the important contributions of these authors to the data presented here.

Core publications used directly in this work:

English MA*, Gayet RV* and Collins JJ. Designing biological circuits: synthetic biology within the operon model and beyond. *Annual Review of Biochemistry* **90**: 221-244 (2021).

Gayet RV*, de Puig H*, **English MA***, Soenksen LR*, Nguyen PQ, Mao AS, Angenent-Mari NM and Collins JJ. Creating CRISPR-responsive smart materials for diagnostics and programmable cargo release. *Nature Protocols* **15**: 3030-3063 (2020).

English MA*, Soenksen LR*, Gayet RV*, de Puig H* *et al.* Programmable CRISPR-responsive smart materials. *Science* **785**: 780-785 (2019).

Abstract

Mobile genetic elements (MGEs) have played a fundamental role in the evolution of complex organisms, driving innovation through competition, collaboration and co-option. Transposons, in particular, are an ancient family of MGEs whose diverse functions have provided a rich source of DNA-binding and nuclease domains for their cellular hosts, and more recently for biological engineering technologies. In this thesis, I re-purpose transposons and their evolutionary descendants as synthetic biology tools for two distinct applications: new platforms for directed genome evolution, and CRISPR-based nucleic acid sensors. Prokaryotic CRISPR-Cas adaptive immune systems evolved in part from ancestral transposon domains, and the inherent programmability of their RNA-guided nucleases has underpinned their use as specific and sensitive *in vitro* diagnostics. Here, I present our efforts to expand the use of these CRISPR-based sensors to control the large-scale properties of smart biomaterial systems, and thereby enable programmed cargo release and the development of low-cost, paper-fluidic diagnostic devices. Transitioning to *in vivo* applications for MGE-derived tools, I describe the development of an engineered, autonomous transposon platform for continuous genome-wide mutagenesis and dynamic regulatory network re-wiring. I use this platform to study the impacts of transposon functionalization on the evolution of parallel *E. coli* populations towards diverse carbon source utilization and antibiotic resistance phenotypes. Through the implementation of barcode-based tracking and longitudinal next-generation sequencing, I then re-construct transposon lineages within the genomes of host cells and investigate the impact of environmental complexity and genetic contingency on host-transposon interactions. Moving forwards, we envision this directed genome evolution platform being used to discover and optimize strains for biopharmaceutical applications, and as a well-defined testbed to study the role of MGEs in the emergence and re-wiring of complex natural gene regulatory networks.

Keywords— CRISPR diagnostics - DNA hydrogels - CRISPR gene editing - genome engineering - directed evolution - NGS pipeline design - Tn-Seq - transposon mutagenesis

Table of Contents

List of Figures

List of Abbreviations

CRISPR ..... Clustered Regularly Interspaced Short Palindromic Repeats

Cas ..... CRISPR-associated proteins

CAST ..... CRISPR-associated transposon

MGE ..... Mobile genetic element

Tn-Seq ..... Transposon insertion Sequencing

RB-Tn-Seq ..... Randomly-Barcoded Transposon insertion Sequencing

PEG ..... Poly(ethylene glycol)

DNA ..... Deoxyribonucleic Acid

RNA ..... Ribonucleic Acid

GFP ..... Green Fluorescent Protein

RFP ..... Red Fluorescent Protein

PCR ..... Polymerase chain reaction

RT-qPCR ..... Reverse Transcription quantitative (real time) PCR

NGS ..... Next-Generation Sequencing

TEMED ..... Tetramethylethylenediamine

PA ..... Poly(acrylamide)

PA-DNA ..... Poly(acrylamide)-DNA

APS ..... Ammonium Persulfate

DMSO ..... Dimethyl Sulfoxide

LB ..... Lysogeny Broth

RBS ..... Ribosome Binding Site

PBS ..... Phosphate Buffered Saline

EDTA ..... Ethylenediaminetetraacetic acid

TAE ..... Tris Acetate EDTA

MM ..... Master Mix

MBN ..... Mung Bean Nuclease

FACS ..... Fluorescence activated cell sorting

KLD ..... Kinase, Ligase, DpnI

gRNA ..... Guide RNA

sgRNA ..... Synthetic guide RNA

crRNA ..... CRISPR RNA

1 | Introduction

Since the implementation of the first synthetic gene networks in living cells (1, 2), synthetic biology has emerged as a powerful framework for both forward-engineering and hypothesis testing in biological research (3). As an engineering discipline, synthetic biology has always been motivated by its applications to real-world challenges. These include the design of therapeutic devices and diagnostics, as well as the industrial or on-demand production of fuels, complex chemicals, medicines, and materials (4, 5). In parallel, minimal and well-controlled analogues of complex biological systems can be assembled and studied in isolation, and synthetic reporter circuits can be used to create human-interpretable readouts for natural cellular processes (6). In this context, synthetic biology has played an important role in investigations into the architectures of gene regulatory networks (GRNs) and the functional consequences of noise in cells (7, 8).

1.1 Synthetic biology across different scales

In this thesis, I present two separate avenues of investigation that span a broad spectrum of biological scale and complexity (Figure 1.1). The first is a series of efforts alongside my co-authors towards the development of CRISPR-responsive materials and their use in *in vitro* diagnostics and tissue engineering (Chapter 2)(9, 10). Importantly, this builds on previous work from the lab to incorporate electronic and wireless readouts into paper-based infectious disease diagnostics (11, 12), and expands the use of programmable CRISPR-based DNA sensors to the field of material science (13).

My second area of investigation reflects the use of synthetic circuits as tools to address fundamental questions in biology, and to develop technologies to re-wire complex, endogenous networks (6). Recent advances in genetics tools (14, 15), directed evolution (16, 17) and automated culture systems (18, 19) make synthetic biology well poised to study the processes underlying genome evolution and the emergence of complexity in living systems (20, 21). Of particular interest to me is the central role played by mobile genetic element (MGE) radiations in the evolution of regulatory networks (22–24), and the potential uses of the diverse biomolecular

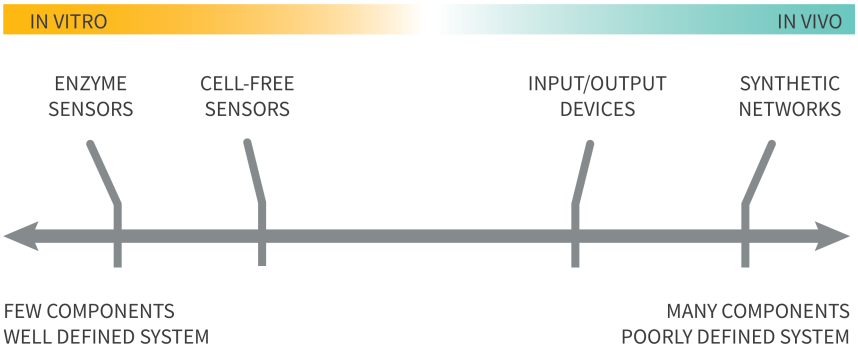

Figure 1.1: The engineering objectives of synthetic biology span systems with vastly different complexities. In Chapter 2, I present research at the far left hand side of this scale: enzyme-based diagnostics operating *in vitro*. In Chapter 3, I present research focused on perturbing and re-wiring complex cellular networks *in situ*

functions that they encode (25–28). Here, I present my efforts to re-purpose both CRISPR-associated (29–31) and randomly-integrating (15) transposon families as synthetic biology tools for dynamic network evolution (Chapter 3). This includes the development of a generalizable DNA-barcode based mechanism to track transposon movement through genomes over time using longitudinal next-generation sequencing (NGS). In the future, this platform could be readily employed to study genome evolution under complex environmental conditions, and to study the impact of predictability in the emergence of anticipatory gene regulation (32).

1.2 Summary of research objectives

Aim 1 - To develop a new class of CRISPR-responsive materials for applications in tissue engineering, therapeutics and diagnostics.

- Aim 1.1 - Demonstrate the programmable control of two established bio-compatible materials with Cas12a: poly(ethylene glycol) and polyacrylamide hydrogels.
- Aim 1.2 - Using DNA materials, develop electronic circuit elements whose properties can be modulated using Cas12a.
- Aim 1.3 - Prototype a low-cost, paper-based diagnostic for infectious diseases that implements electronic readouts of Cas12a reactions for the wireless transmission of results.

Aim 2 - To engineer diverse mobile genetic element families as synthetic biology tools for dynamic gene network re-wiring.

- Aim 2.1 - Functionalize CRISPR-associated transposon systems for RNA-guided gene activation and transcriptional reporting.
- Aim 2.2 - Develop an engineered, autonomous transposon platform for continuous *in vivo* mutagenesis.
- Aim 2.3 - Modify existing Tn-Seq pipelines to enable the tracking of barcoded transposon populations during parallel evolution.

2 | Nucleic acid detection with synthetic biology

Synthetic biology approaches to the development of disease diagnostics can be divided into two broad categories: living cell-based biosensors that rely on engineered gene circuits to detect and amplify signals (33, 34), and cell-free sensors that couple enzymatic reactions or biomolecular interactions to human-interpretable readouts (11). Perhaps the two most important analytes in the context of infectious diseases are pathogen-specific antigens (typically detected by lateral-flow immunoassays) and nucleic acids (typically detected by PCR in centralised laboratories) (35). Nucleic acids (NAs) are information-rich biological cues. Genomic sequencing is routinely used to identify different organisms and track clinically relevant markers such as antibiotic resistance genes (36). Similarly, knowledge of the abundances of RNA transcripts in a sample can aid clinical decision making through the prediction of treatment responsiveness (37) and microbial virulence states (38–40).

The development of effective point-of-care nucleic acid tests (NATs) for clinical or environmental samples requires the conversion of biomarker detection events into user-friendly readouts. Successful technologies also make use of sensing platforms that are easy to reprogram towards new targets (e.g. a novel virus or antibiotic resistance marker) with minimal re-design of the overall system (12, 38, 39). However, as the ongoing SARS-CoV-2 pandemic has emphasised, our ability to rapidly deploy point-of-care NATs at scale still lags behind that of antigen tests, with the majority of synthetic biology-based NATs in pre-clinical phases of development (41). Ultimately, the objective is to engineer re-programmable platforms for point-of-care sensors that fulfil the World Health Organisation's ASSURED (affordable, sensitive, specific, user-friendly, rapid, equipment-free, delivered) criteria for field deployment (42). In this chapter, I highlight promising recent advances in the development of programmable CRISPR-based diagnostics (13), and summarize our efforts to couple these reactions to the control of biomaterials and electronic circuits (9, 10).

2.1 CRISPR-based diagnostics

Research from the Zhang Lab, in collaboration with the Collins Lab, established the blueprint for a diagnostic platform that builds on the discovery of prokaryotic clustered regularly inter-spaced short palindromic repeat (CRISPR) and CRISPR-associated (Cas) nucleases (43). Microbial CRISPR-Cas systems have been successfully adapted into powerful genetic engineering technologies, owing to the inherent programmability of their effector proteins (14). Crucially, the target specificity of the Cas nuclease is defined at the nucleotide level through their associated guide RNA (gRNA): tight binding between the gRNA of the ribonucleoprotein complex and the target DNA strand can be readily designed *in silico* based on complementary Watson-Crick-Franklin base-pairing interactions.

Beyond the canonical double-stranded DNA (dsDNA) cleavage activity characteristic of the Cas9 effectors most commonly used in gene editing, several classes of Cas enzymes are able to hydrolyze single-stranded DNA (ssDNA) and RNA (ssRNA). The discovery of these alternative activities has fueled the rapid development of a new class of diagnostics, such as SHERLOCK (which detects RNA using Cas13a) (40, 44), HOLMES (45), and DETECTR (both of which detect dsDNA using Cas12a) (46). The latter example uses Cas12a, a protein that exhibits two distinct - but coupled - activities (Figure 2.1). When the Cas12a-gRNA complex encounters a dsDNA molecule matching the sequence of its gRNA (a 'trigger' DNA molecule), the enzyme cleaves the DNA in a sequence-specific way (termed 'targeted' or 'cis' cleavage) and remains bound to one of the resulting DNA fragments. This first cleavage event initiates a second, non-specific nuclease activity (termed 'collateral' or 'trans' cleavage); the activated Cas12a complex cleaves nearby ssDNA molecules with a very high turnover.

In the first generation of CRISPR diagnostics, DNA or RNA detection were coupled to fluorescence readouts of Cas12a/Cas13a activation (40, 46) (Figure 2.2). This typically requires costly instrumentation and infrastructure to measure, reducing their practicality as point-of-care or consumer diagnostics. More recent modifications to the SHERLOCK system have included lateral flow readouts, which improves usability and removes the need for fluorimeters (44, 47, 48).

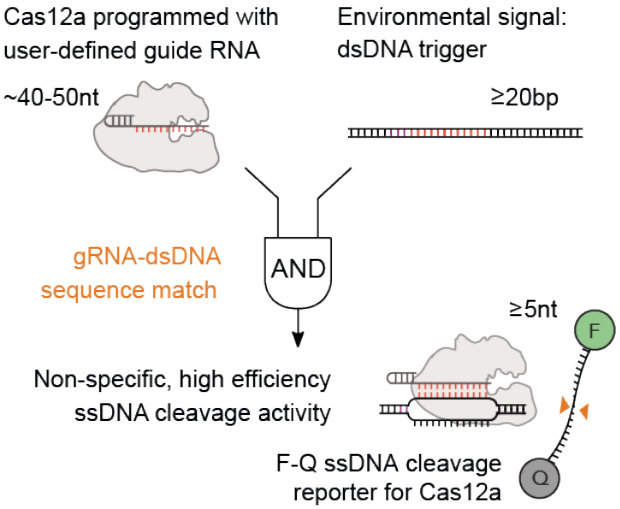

Figure 2.1: Cas12a forms a ribonucleoprotein complex with a guide RNA molecule. This complex is then able to bind double-stranded DNA (dsDNA) regions that fulfil two criteria: (i) the DNA sequence is complementary to the sequence of the variable region of the gRNA, and (ii) this target region has an appropriate flanking protospacer adjacent motif (PAM) for Cas12a. Once the Cas12a-gRNA complex binds the dsDNA target, the dsDNA is cut with approximately single-turnover kinetics. However, a secondary nuclease activity also becomes active, targeting single-stranded (ssDNA) with high turnover. This activity can be readily detected using oligonucleotides modified with a fluorophore (F) and a quencher (Q). Adapted from English *et al.*(9). Copyright AAAS.

However, the underlying reaction remains the same, and applications have yet to emerge outside of disease diagnostics. Furthermore, the ability to couple CRISPR-Cas enzyme activities to electronic signals would expand their uses towards wireless and embedded sensors, and allow for the coupling of the test readouts to existing data transmission and processing pipelines (9, 49).

2.2 Programmable control of DNA-materials with CRISPR-Cas12a

Smart materials that respond to biologically-relevant signals now play an important role in emerging technologies for drug delivery, diagnostics, tissue engineering and medical devices (50, 51). These autonomous systems can be designed to integrate signals ranging from enzymatic activity (52, 53) and the presence of metabolites (54, 55), to changes in pH (52, 56) and temperature (57, 58). For the reasons discussed in Section 2.1, materials that detect and respond to DNA and RNA are of particular interest in the development of diagnostics (59), or for applications that require the context-dependent release of therapeutics (50, 60).

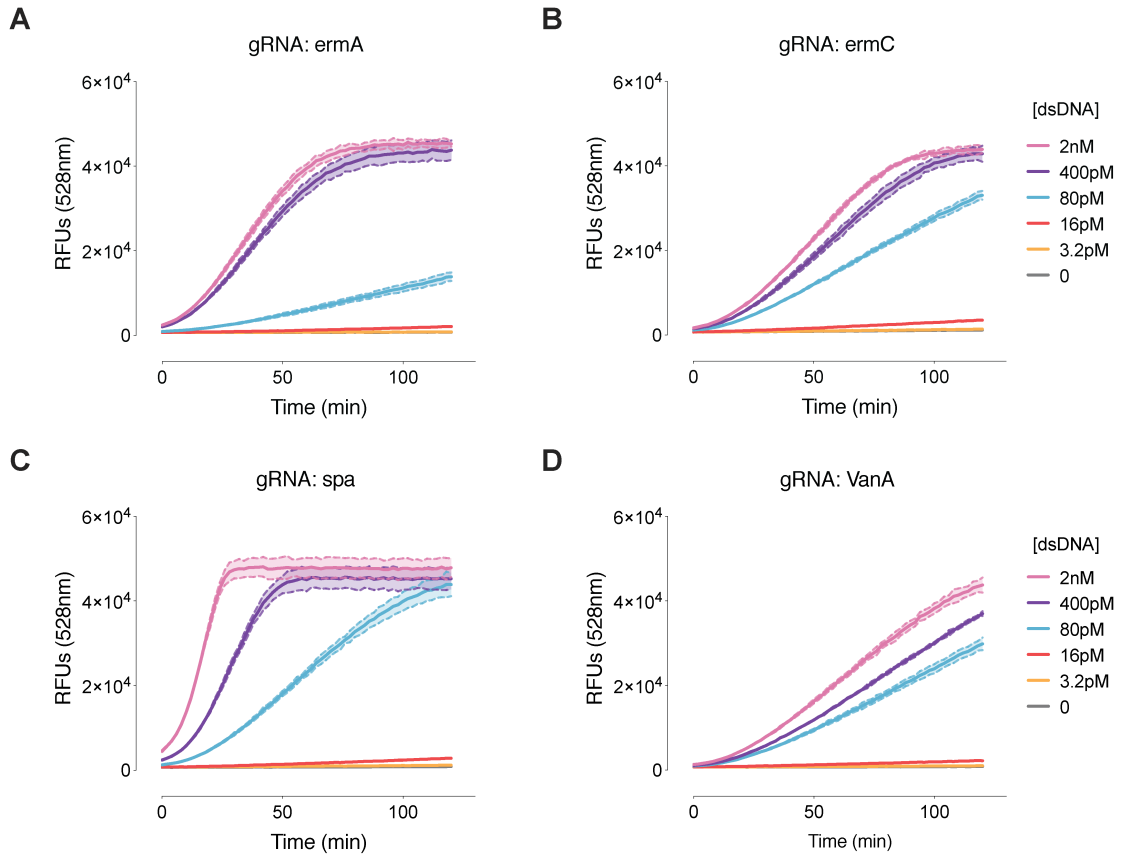

Figure 2.2: A single Cas12a enzyme can be rapidly re-programmed to detect multiple pathogen-associated DNA marker sequences. Here, we designed gRNAs complementary to four genes in *S. aureus* antibiotic-resistance mechanisms: the antibiotic-resistance genes *ermA* (A) and *ermC* (B), the virulence factor gene *spa* (C), and the vancomycin-resistance gene *vanA* (D). As shown, the specific gRNA-target combination does affect the timecourse and the overall sensitivity of the system.Adapted from English *et al.*(9). Copyright AAAS.

Incorporating DNA into materials as both a structural and information-encoding element is emerging as a promising approach to address this need (61). This has been facilitated by the decreasing cost of DNA-synthesis (62) and the relative ease with which sequence complementarity can be predicted and programmed at the sequence-level (63–65). Moreover, DNA is generally biocompatible and biodegradable, attractive properties for materials in biomedical applications.

Current methods to develop DNA-responsive materials typically rely on the strand displacement of structural DNA elements. However, the one-to-one stoichiometry of the hybridization reaction can limit the sensitivity of these systems. Furthermore, the fact that the same material elements are simultaneously responsible for structure and detection means that extensive redesign is often required for each new target. We reasoned that the natural signal amplification inherent to the mechanism of NA detection with Cas12a, combined with its robustness *in vitro*, would enable the sensitive and programmable actuation of DNA materials in response to biological cues (9).

Recognizing the strengths of CRISPR-based nucleic acid detection platforms, we set out to (i) expand their application space to the programmable control of materials, and (ii) build on the available readout formats for diagnostic tests. To demonstrate the ability of Cas12a to actuate the properties of materials in response to biological cues at different scales, we selected three polymer chemistries: poly(ethylene glycol) hydrogels for the controlled release of pendant small-molecules and enzymes anchored through ssDNA linkers, polyacrylamide-DNA (PA-DNA) hydrogels for the entrapment of larger cargos including nanoparticles and human cells, and carbon black-DNA composites that act as CRISPR-responsive electronic fuses (Fig. 2.3). We then employed the PA-DNA hydrogels in a low-cost, paper-based device (μPAD) for the diagnostic detection of pathogen-associated dsDNA and ssRNA. The use of a CRISPR-responsive material to control the flow of buffer through the device allowed us measure the readout electronically, and couple the sensitive system to an RFID tag for the wireless transmission of the test result. The culmination of this work is published in two accompanying manuscript (9, 10), which provide extensive detail on the methods and results.

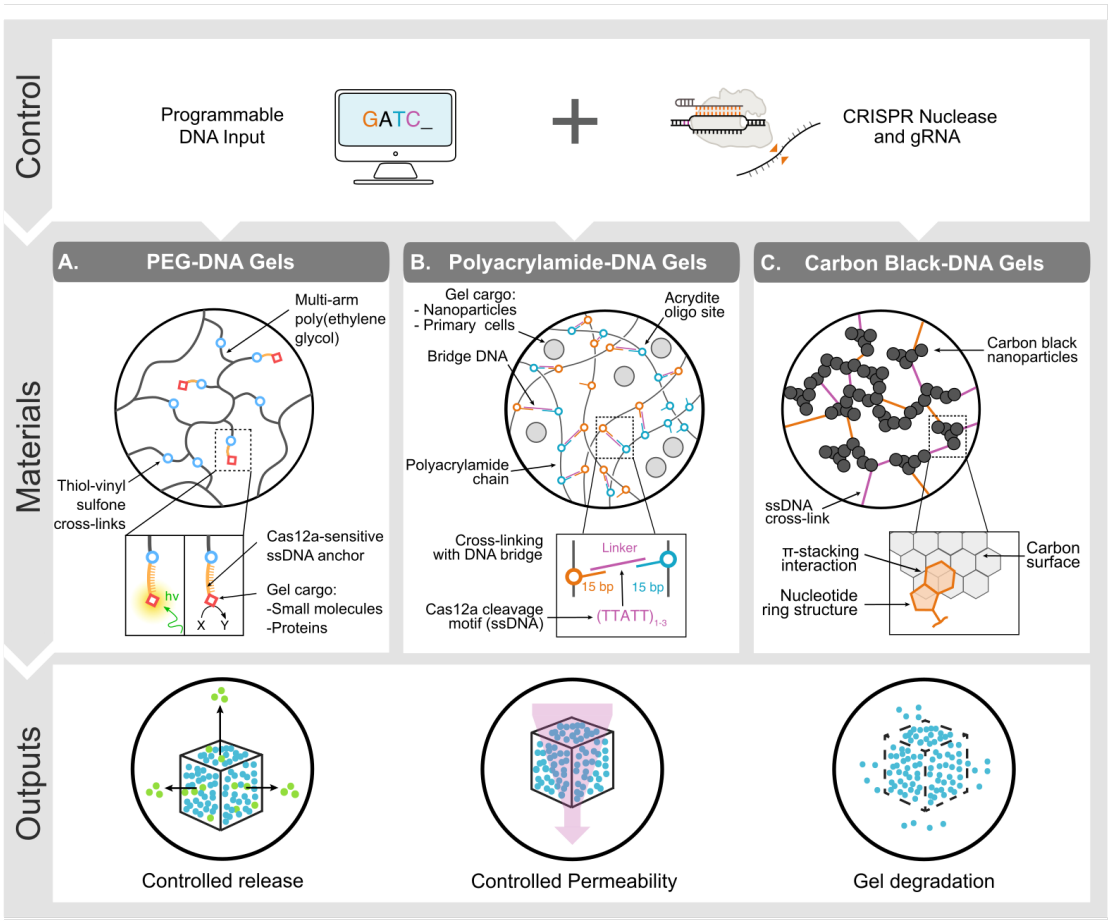

Figure 2.3: We synthesized three different materials that contained DNA as a functional or structural element, and demonstrated their programmable actuation using Cas12a-gRNA for the controlled release of small-molecules, enzymes, nanoparticles, and living cells. We then constructed a paper-based diagnostic device using polyacrylamide-DNA hydrogels and Cas12a for the detection of pathogen DNA and RNA. Adapted from English *et al.*(9). Copyright AAAS.

2.2.1 Releasing small, tethered cargos from PEG-DNA hydrogels

The targeted dsDNA cleavage activity of Cas12a can be used to preferentially release anchored cargos with near-single turnover; however, we focused on the collateral ssDNase activity of the enzyme, as it allows for the efficient transduction of external stimuli into changes in material properties through catalytic signal amplification. To illustrate the programmable actuation of materials using Cas12a, we covalently tethered a fluorophore (Cy3) into PEG hydrogels through an ssDNA linker (Fig. 2.4A) and monitored its release into solution upon Cas12a-induced cleavage. The Cas12a-gRNA complex was insufficient to catalyze cargo release; however, introduction of the mecA dsDNA trigger initiated the hydrolysis of ssDNA anchors (Fig 2.4B). By contrast, a randomly permutated version of the mecA dsDNA (scrambled control) failed to do so. We then used horseradish peroxidase (HRP) enzyme as a model for larger biomolecule payloads, demonstrating that biological function was preserved after immobilization in hydrogels using ssDNA anchors and subsequent release by Cas12a (Fig. 2.4C). Within 10 min of exposure to a low-concentration (10 nM) dsDNA stimulus, we detected sufficient HRP activity in the supernatant for a visual readout (Fig. 2.4C, inset). Further incubation allowed us to routinely discriminate trigger and scrambled dsDNA down to 100 pM. These experiments are consistent with the efficient ssDNase activity of activated Cas12a (46).

To demonstrate that changes to the gRNA were sufficient to entirely reprogram the target responsiveness of the material, we designed gRNAs to target a panel of genes involved in *S. aureus* antibiotic-resistance mechanisms (Fig. 2.2). These include the antibiotic-resistance genes *ermA* and *ermC* (66, 67), the virulence factor gene *spa* (68), and the vancomycin-resistance gene *vanA* (69). Out of 25 combinations of gRNAs and dsDNA, those in which the sequence of the trigger matched the gRNA resulted in substantially higher fluorophore payload release from the hydrogel matrix (Fig. 2.4D). These results correlated to similar observations of the reactions performed in solution and suggest that different gRNA-trigger pairs activate Cas12a to different extents (46).

For many applications, the rate at which a molecule is delivered from a carrier

conveys important biological information (51). The speed of CRISPR-mediated hydrogel actuation corresponds to the amount of input dsDNA; conversely, for a given level of input, the response dynamics can be hard-coded into the system by altering the properties of the starting material. For example, pore size is expected to alter the mobility of macromolecules in polymer networks (70). On the basis of our macroscopic observations of programmed anchor hydrolysis, we hypothesized that this could be used to further tune the relationship between dsDNA input and Cas12a-mediated response. By modulating the cross-linking density of a PEG-DNA hydrogel and measuring the rate of fluorophore release by Cas12a-gRNA, we established another strategy by which the behavior of the CRISPR-responsive material could be controlled (Fig. 2.4E)(45, 46).

In addition to controlling global dynamics of ssDNA cleavage through bulk material properties, we capitalized on the sequence-defined, addressable nature of the ssDNA linkers and the selectivity of the collateral cleavage activity of Cas12a for ssDNA over dsDNA (Fig. 2.1).We attached two different fluorophores (Cy3 and 6-FAM) into PEG hydrogels with distinct ssDNA linkers and preprogrammed the differential sensitivity of one linker over the other to Cas12a collateral degradation by hybridizing it with a complementary blocking strand *in situ*. Whereas the release of the unprotected fluorophore was unaffected, the speed of release of the hybridized reporter was markedly reduced (Fig. 2.4F).

2.2.2 Releasing large, enmeshed cargos from PA-DNA hydrogels

The high catalytic efficiency of dsDNA-activated Cas12a-gRNA makes it well suited for converting dsDNA signals into bulk material changes. To demonstrate this, we designed DNA cross-linked polyacrylamide (PA) hydrogels (71, 72) by separately incorporating two noncomplementary oligonucleotides into PA chains. We then cross-linked the PA-DNA precursors using an oligonucleotide strand that forms bridges between the PA-DNA chains. These cross-links contained single-stranded, AT-rich Cas12a collateral cleavage sites (Fig. 2.5A). In these hydrogels, degradation of DNA crosslinks physically disrupts the polymer networks (70, 73).

The Cas12a-induced degradation of PA-based CRISPR gels was initially evalu-

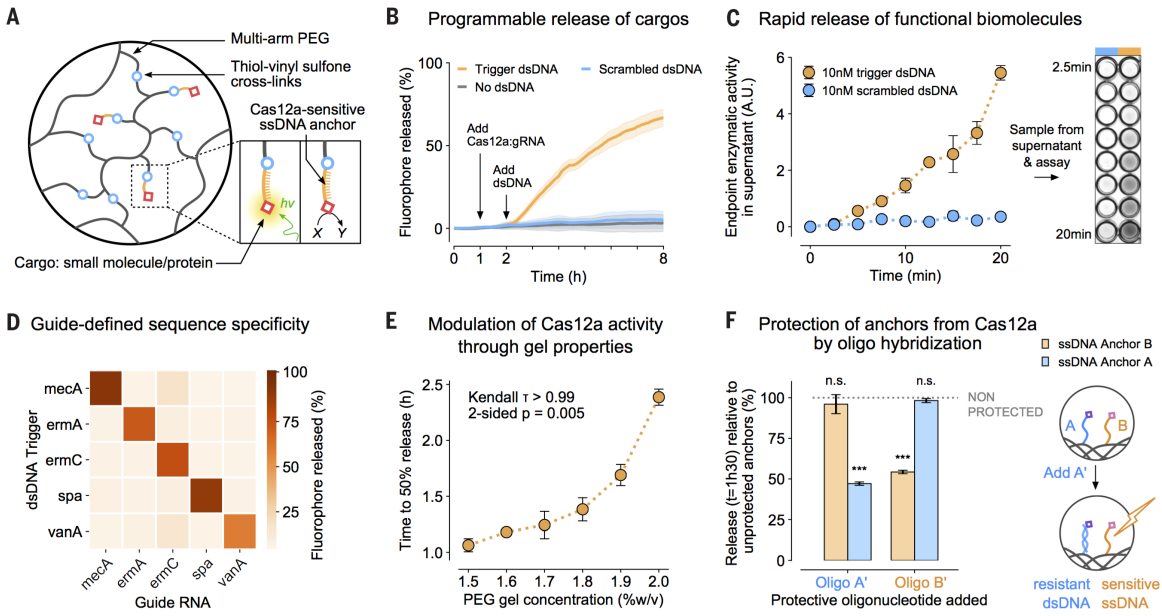

Figure 2.4: Cas12a-mediated release of small molecules and enzymes from PEG hydrogels. (A) ssDNA acts as a cleavable linker for attaching payloads to an inert PEG matrix. hn, light energy. (B) Release of a tethered fluorophore by Cas12a is initiated only upon introduction of a specific dsDNA trigger and not a scrambled dsDNA control sequence. (C) Functional enzymes can be anchored into the hydrogel and released by Cas12a in sufficient quantities for visual detection in an HRP activity assay within minutes. A.U., arbitrary units. (D) Activation of Cas12a and fluorophore release (t = 8 hours) is defined by the complementarity between a dsDNA sequence and the gRNA of Cas12a. (E) Cross-linking density of the PEG hydrogels modulates the release rate of the cargo by Cas12a. The correlation was analyzed using a Kendall rank test. (F) Prehybridization of the ssDNA linkers with a matching oligonucleotide selectively reduces the release rate of molecules anchored in the gel (observed at t = 1.5 hours). The means were compared with independent samples that were not preprotected with oligonucleotides (gray). Differences in the means of the test conditions and the unprotected controls were analyzed using a t test [Bonferroni-adjusted a = 0.0125, P values: not significant (n.s.) $P > 0.05$, ***$P < 0.0001$]. All plots show mean ± SD for n ≥ 3 replicates. Adapted from English *et al.* (9). Copyright AAAS.

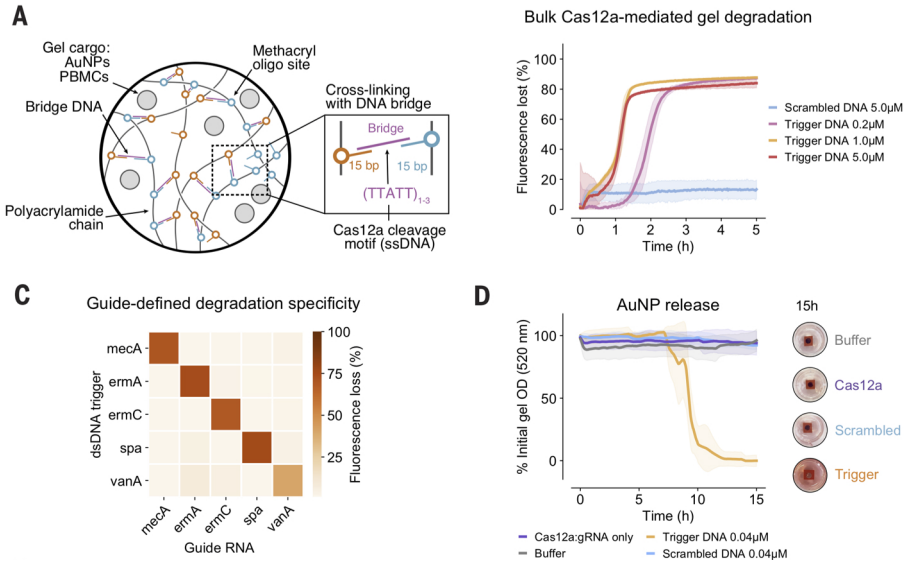

Figure 2.5: Programmable release of NPs and live cells from PA-DNA hydrogels. (A) ssDNA bridges lock DNA-functionalized PA chains into a 3D network. (B) Cas12a- mediated degradation of PA-DNA gels stained with EvaGreen intercalating DNA dye. (C) Degradation of gel with 25 combinations of gRNAs and dsDNA triggers and comparison of signals after 12 hours. (D) Release of AuNPs from 7% (w/v) PA-DNA gels using Cas12a collateral cleavage, tracked by measuring gel optical density. The Cas12a-gRNA and dsDNA trigger were encapsulated in the gel with the AuNPs (concentrations shown include supernatant volume). Adapted from English *et al.* (9). Copyright AAAS.

ated with a DNA-intercalating dye to label bridge sequences in PA-DNA gels and track gel integrity. The bridges were degraded upon exposure to gRNA-Cas12a and trigger dsDNA, as revealed by the dissipation of gel fluorescence at rates dependent on trigger concentration (Fig. 2.5B). Compared with experiments performed in solution, gel degradation appeared more robust to the introduction of sequence mismatches between the gRNA and dsDNA trigger. Using fluorescein isothiocyanate-dextran particles physically entrapped in the hydrogel, we also visualized the degradation of millimeter-scale PA-DNA hydrogels.

Programmable degradation of PA-DNA hydrogels was assessed by testing 25 combinations of different gRNAs and dsDNA triggers. Consistent with the nondestructive cargo-release experiments (Fig. 2.4D), PA-DNA hydrogel degradation occurred only when the gRNA and dsDNA sequenceswere complementary (Fig. 2.5C), demonstrating Cas12a-gRNA's ability to discriminate between inputs.

Though biomolecules can be tethered to materials through well-defined, single linkers, physical entrapment in a polymer matrix represents a more general strategy

to control the release of larger payloads. We tested the release of NPs by encapsulating PEG-coated gold NPs (AuNPs) in PA-DNA hydrogels. Loading gels with both Cas12a-gRNA and a dsDNA trigger led to total NP release through Cas12a activation and gel degradation, whereas gels loaded with a scrambled dsDNA trigger showed no significant release of AuNPs relative to a buffer-only background (Fig. 2.5D). This was consistent with the disruption of the percolated network upon cross-link cleavage (74, 75).

2.2.3 Paper-based diagnostics from CRISPR-responsive materials

We used a tunable PA-DNA hydrogel to control the permeability and electrical readout of a paper-based microfluidic device (Fig. 2.6). Paper-based technologies have shown promise for point-of-care diagnostics, as they are low cost, equipment-free, and easy to use (12, 76). Our device (Fig. 2.6A) expands on the concept of microfluidic paper-based analytical devices (μPADs) that rely on the capacity of hydrogels to obstruct flow through porous channels (77).

The layers of the device were folded to create a multilayered structure in which the hydrophilic regions are topologically aligned. Capillary-driven flow through the device terminated in a fifth layer where the output was measured (Fig. 2.6A). In this system, an intermediary layer contains PA-DNA gel precursors (Ps-X and Ps-Y) that, when mixed with ssDNA crosslinker, form a hydrogel in the paper channels (55, 78). The extent of gel formation, and therefore the rate of buffer flow, is dependent on the extent of degradation of the ssDNA gel cross-linker during a preincubation step. The activation of Cas12a can be confirmed by adding a fluorescent ssDNA reporter during this step. By degrading the cross-linker by using Cas12a, we were able to couple the level of buffer flow to the concentration of dsDNA trigger added to a Cas12a reaction incubated for 4 hours.

When nonspecific dsDNA trigger is present during preincubation, ssDNA cross-linkers are not cleaved, allowing for hydrogel assembly in the microchannel. Conversely, in the presence of a specific dsDNA trigger, unimpeded flow can be visually detected by adding dyes to the μPAD channel. We found the rate of buffer flow through a μPAD to be inversely related to the concentration of an MRSA dsDNA

trigger. Using this visual output, we were able to detect dsDNA concentrations down to 400 pM.

To optimize our CRISPR-μPAD for field diagnostic applications, we used reverse transcription (RT) to expand the range of detectable biomarkers to RNA and coupled the RT to an isothermal amplification [recombinase polymerase amplification (RPA)] step to improve the limit of detection. We used RT-RPA followed by a μPAD readout to detect synthetic Ebola genomic RNA (79) down to 11 aM, a sensitivity matching other state-of-the-art CRISPR based diagnostics (40, 44, 46). This approach is promising for point-of-care diagnostics and has overall better performances in terms of sensitivity, portability, and cost than other molecular diagnostics.

Visual readouts of buffer flow are commonly used, yet they are difficult to couple to downstream hardware for data processing. To overcome this limitation, we modified the CRISPR-actuated fluidic system to read buffer flow as an electric signal: the microfluidic channel in the final layer was sandwiched between two electrodes and connected to an ohmmeter (Fig. 2.6A). Electrical conductivity between the electrodes relied on electrolytes provided by the flowing buffer and was directly correlated to the buffer penetration length in the μPAD channel (80). Using this approach, sub-nanomolar concentrations of dsDNA trigger were successfully detected at a 5-min end point, without DNA amplification, demonstrating the potential of the CRISPR μPAD for embedded sensor applications (Fig. 2.6B). Nonspecific dsDNA trigger did not activate Cas12a, thus leaving the electrical circuit open (Fig. 2.6B). We were able to reduce the preincubation time required to observe a signal to 1 hour by tuning the properties of the acrylamide precursors.

The wireless, decentralized logging of individual clinical tests during infectious disease outbreaks could address challenges with record keeping and logistics. To integrate CRISPR-Cas reactions with electronic monitoring systems through hydrogel actuation, we incorporated a wireless radio-frequency identification (RFID) module into the μPAD. The original design was modified such that buffer flow would shortcircuit an interdigitated silver electrode, thereby modulating the efficiency of signal transmission by a flexible RFID tag (Fig. 2.6C). We then conducted

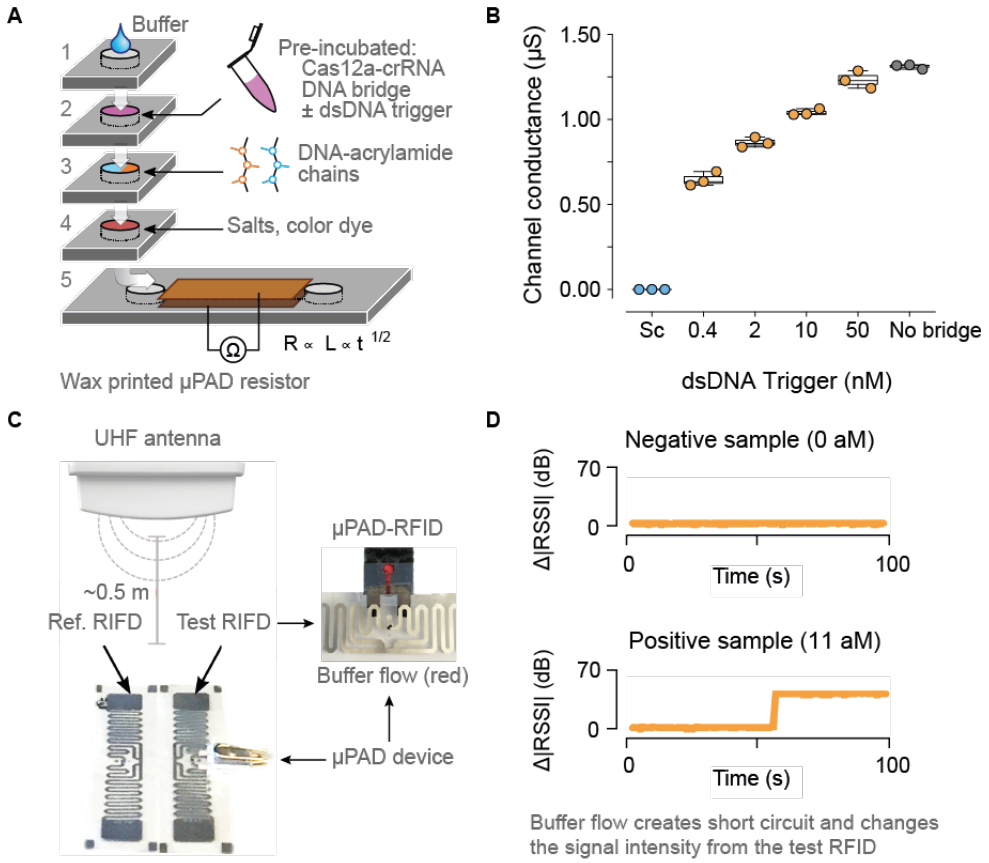

Figure 2.6: Cas12a digestion of hydrogel precursors modulates permeability of a paper-based microfluidic device (μPAD) with dual visual and electronic readouts for diagnostic applications.. (A) Schematic of the stackable mPAD design (77) modified for operation with CRISPR gels and electrical readout. Layers 1 to 4 contain hydrophilic regions that form a continuous channel on folding and feed into a lateral flow channel in layer 5. The channel in layer 5 was covered with conductive tape to measure conductivity as a function of buffer wicking. In the presence of target trigger, Cas12a cleaves the DNA linker, preventing hydrogel cross-linking in the channel and enabling flow. (B) End point measurements (5 min) of resistance facross the channel for different concentrations of dsDNA MRSA trigger input after a 4-hour predigestion step. Sc = 50 nM scrambled dsDNA. (C) Schematic illustrating the integration of the paper-fluidic device with an RFID flexible tag. Cas12a activation in the preincubation step results in the short-circuiting of an interdigitated electrode arrangement in the loop RFID tag, thereby altering the received signal strength indicator (RSSI) compared with a reference tag (Ref.). (D) Representative signal traces for positive and negative results in the experimenter-blinded trial of the RFID μPAD device. Samples containing either 0 aM (negative) or 11 aM (positive) EBOV ssRNA trigger were amplified by RT-RPA, incubated with the ssDNA gel bridging strand and Cas12a-gRNA for 4 hours, and assayed on a μPAD-RFID device. Adapted from English *et al.* (9). Copyright AAAS.

an experimenter-blinded trial consisting of 12 samples (containing either 11 or 0 aM Ebola ssRNA amplified by RT-RPA) divided across three geographic locations. The experimenter preincubated the samples with Cas12a and Ebola-specific gRNA for 4 hours and then recorded the RFID-μPAD signals over the course of 2 min. Buffer flow through the μPAD in Ebola-positive samples caused shortcircuiting of the RFID tag antenna, which was detected in real time as a change in the signal strength compared with an unmodified reference RFID tag (Fig. 2.6D). All positive and negative samples were correctly assigned using the RFID-μPAD.

2.2.4 Conclusions

We have demonstrated several strategies to interface biological signals with materials that combine the inherent programmability of Cas enzymes with hydrogel systems. These strategies offer control over a variety of complex behaviors and properties, including the release of molecules, NPs, and live cells, as well as bulk hydrogel degradation, electronic signal transduction, and microfluidic valve actuation. By exploiting the enzymatic properties of Cas12a, we have designed a platform that improves on hydrogel programmability and versatility, as only the gRNA molecule needs to be changed to allow hydrogel response to a user-defined DNA sequence. The catalytic activity of Cas12a improves sensitivity compared with DNA-responsive hydrogels requiring stochiometric amounts of DNA triggers for material activation. Finally, we demonstrate various forms of output that expand the capabilities of CRISPR-responsive materials and enhance existing biomaterial-based approaches for tissue engineering, molecular diagnostics, and bioelectronic interfaces with programmable readouts.

First author contributions

MAE, RVG, LRS, HP and JJC conceived and planned the experiments. HP screened a panel of guide RNA sequences for use in the different material systems and performed qPCR benchmarking. RVG developed the PEG-DNA hydrogels and performed experiments with MAE. MAE developed the PA-DNA hydrogels and performed experiments with HP. LRS and HP performed cell release experiments with the PA-DNA hydrogels (not described here). LRS developed the paper-fluidic device with contributions from MAE and HP. All authors contributed equally to analyzing the data and writing the associated manuscripts.

3 | Engineering and evolving natural gene regulatory networks

Since the field's inception, synthetic biologists have sought to apply engineered tools to the study of natural gene networks to answer questions in fundamental biology (7). The mechanisms underlying the emergence and expansion of complex gene regulatory networks (GRNs) across the kingdoms of life, and the relationship between genotype and phenotype, have been a major area of research since the first models of transcriptional control were proposed by Jacob and Monod (81, 82). The overwhelming majority of this work is based on the observation and comparison of natural systems. Progress in phylogenomics has led to the establishment of powerful models for the processes of gene duplication, transfer, and the emergence of new architectures and functions (see BOX)(83–86). In parallel, there has been a longstanding application of directed evolution to the study of evolutionary trajectories for proteins (16, 87), gene circuits (88), and whole organisms (89, 90). However, experiments designed to explicitly validate proposed evolutionary models of gene network evolution through real-time observation remain rare (91–93).

I argue that the most efficient way not only to understand, but also to generate complex behaviors in living systems, is by making targeted alterations in natural regulatory networks (94). New tools that enable in-context perturbations to gene regulation are an important addition to the synthetic biology toolbox, as they facilitate the design of circuits regulating native genetic elements. In the future, the applications of such systems could include engineering platforms for cell therapies, as the desired cellular behaviors often involve a complete remodelling of the cellular phenotype (95), including the activation of transcriptional programs that would be far too complex to recreate artificially. Here we propose several approaches that could help build technologies for advanced synthetic biology applications.

Sources of genetic novelty in cellular populations.

- **Duplication**. The duplication of operators (96), regulators (97), structural genes (92, 98), operons (99, 100), and even whole genomes (101) is a major source of novel genetic material for genome expansion (102, 103)
- **Functionalization**. In the most general model, these duplication processes are often followed by sub-functionalisation and/or neo-functionalization of the paralogs (104). In a process termed 'paralog interference', the evolutionary trajectories of the duplicate genes are proposed to impede on one-another due to extensive overlaps in their bioactivity and interaction networks (105).
- **Structural rearrangement**. The recombination, inversion and translocation of existing (duplicated) genes (106–108) generates structural variation.
- **De novo emergence**. The evolution of non-coding regions into functional genes is possible, but is more common in eukaryotes than in prokaryotes due to differences in the amount of non-coding DNA (109).
- **Mobile genetic elements (e.g. transposable elements)**. Comparative genomics indicates roughly 10,000 transposable element-derived fragments are under purifying selection in the human genome, accounting for 48% of the total sequence. 25% of characterised human promoters contain sequences from transposable elements (22, 110).
- **Horizontal gene transfer**. This is well documented for both genes and operators (111, 112). In *E. coli*, 98% of metabolic phenotype specialisations can be accessed via a single DNA transfer event (113).

3.1 A focus on mobile genetic elements

In order to study and forward-engineer complex natural GRNs, new tools are needed to rewire and evolve such systems in a way that scales to whole genomes rather than small, well-defined sequences. If we look to nature, mobile genetic elements (MGEs) would appear to be a good starting point. In particular, transposons are an ancient and diverse form of MGE that have helped drive major evolutionary transitions (24), and have provided much of the functional raw material for extant gene regulatory networks (22, 25). Indeed, the proliferation of transposase genes across kingdoms has provided a rich source of DNA-binding and nuclease domains (114, 115), while the transposons themselves can disperse novel or existing cis-regulatory sequences throughout genomes to establish co-regulated groups of genes (23, 116). This has led to extensive debate as to whether transposons are entirely parasitic (117, 118), or are in fact more generally of adaptive benefit to their hosts as has been observed in some cases (119). To

date however, there have been few systematic, forward-engineering studies on their potentiating role in the evolution of complexity and adaptability in gene regulatory networks (120–122).

Since their discovery (123), several natural transposon families have been re-purposed as powerful molecular tools for both loss- and gain-of-function genetic screens (15). Typical experimental pipelines couple genome-wide transposon insertion mutagenesis with next-generation sequencing to measure differential fitness in pooled populations. As examples, these screens have quantified gene essentiality across genomes and environments (124), and have helped to elucidate mechanisms of antimicrobial resistance (125, 126) and the genetic determinants of cancer (127). However, the majority of these studies use simple transposon constructs (128, 129) that fail to either capture gene-gene interactions through multiplexed insertion, or investigate the effects of dynamic gene-network re-wiring. Furthermore, the lack of an option for precise genomic targeting makes them inefficient as tools for genome editing and gene delivery.

3.2 CRISPR-associated transposon systems

The persistent and rapid growth of public genome sequence databases has given researchers the unprecedented ability to explore the diversity of MGE-encoded functionality, and understand their evolution. This is a particularly rich area for exploration, as the constant arms race between these genetic 'parasites' and their hosts fuels evolutionary innovation that can be harnessed by engineers (130). As a particularly pertinent example, Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) enzymes have been identified and characterized in a range of bacterial species (131). The resultant Cas nuclease toolbox has been harnessed for use in genome engineering (132, 133), transcriptional recording (134), diagnostics (13), and smart materials (9). Interestingly, these systems are in fact directly descended from the appropriation of mobile genetic element systems by their microbial hosts (25).

The value of these tools for synthetic biologists lies in their inherent programmability: a simple change in the sequence of a guide RNA is sufficient to modify the target of a Cas nuclease, or their catalytically inactive variants (e.g. dCas9). These Cas derivatives can then act as platforms for the targeting of genetic regulators to user-defined loci, thereby modulating endogenous gene expression (14). Transcriptional repression using dCas proteins (CRISPRi) is relatively species-agnostic, and is commonly used in bacteria (135) and eukaryotes (136). Conversely, transcriptional activators (CRISPRa) have been engineered in eukaryotes through the fusion of dCas targeting domains to well-established activation domains (137). Programmable CRISPRa tools have found numerous applications in eukaryotic hosts, yet analogous approaches have proved more challenging to interface with endogenous transcriptional machineries in bacteria (138). Moreover, even the most recent bacterial CRISPRa technologies are still limited to enterobacteria (139). This represents a bottleneck in the engineering of naturally isolated bacteria with poorly understood gene regulation mechanisms, and in which the manipulation of endogenous

genes can facilitate the study of original biological features (140).

More recent exploration of this evolutionary linkage between CRISPR systems and mobile genetic elements has yielded novel RNA-guided nucleases (26, 27), and led to the discovery of CRISPR-associated transposons (CASTs) (29, 31, 141). We envision the development of Cas-guided transposons as a potential solution to the long-standing issue of a generalizable gene activation platform in bacteria (29, 30). By adapting this approach it may be possible to remodel native promoters of bacterial genes in situ by insertion of engineered DNA sequences (142). Such sequences could harbor features designed to control gene expression like binding sites for sequence-targeted RNA polymerases (143), thus providing new genetic actuation handles in genetically intractable hosts. In this section, I present preliminary results demonstrating the use of CRISPR-guided transposon systems (specifically shCAST – a CAST from *Scytonema hofmanni*) for the targeted regulation of synthetic and endogenous genetic systems and discuss potential future applications in the creation of reconfigurable genetic circuits.

3.2.1 Targeted transcriptional control with shCAST

While existing CRISPR-associated transposon (CAST) platforms present the exciting possibility for generalized, *in situ* transcriptional regulation (29–31), this capability has yet to be experimentally validated. As a proof-of concept demonstration, we implemented a simple workflow to test the impact of the insertion of engineered CAST systems on downstream gene expression by modifying the three-plasmid protocol described by Strecker et al. (Fig. 3.1A) (29). We reasoned that the gRNA-defined insertion of a mini-Tn7-like transposon upstream of a fluorescent reporter (*gfp*) would modulate gene expression in a manner dependent on synthetic transcription factor and promoter sequences inserted into the transposon itself, and allow for simple plate-based readouts of transposition activity (Fig. 3.1B). Importantly, while the TypeV-K shCAST system offers the benefit of unidirectional insertion when compared to Type I systems, the shCAST system can mediate cointegrate formation due to the absence of an associated TnsA (Fig. 3.1C)(144). This could have implications for the heterogeneity of the resulting target expression profile.

First, we focused on introducing constitutive "ON" regulation at the promoter-less target reporter (Fig. 3.2A). We compared the baseline effect of upstream transposon insertion to that of a transposon harboring a well-characterized, synthetic σ-70 promoter (pJ23119) embedded in the right inverted terminal repeat (ITR) sequence (right end, RE)(Fig. 3.2B). Interestingly, we observed a significant ($\sim$9x) activation of *gfp* expression with just the unmodified transposon, which increased slightly upon the incorporation of a promoter. To establish a likely maximum threshold for gene activation, we encoded a full-length T7 RNA polymerase (T7 RNAP) gene within the transposon itself to drive high levels of transcription from a well-define promoter (145). In two different variants of this transposon (pT7_v1 and pT7_v2), we introduced the same orthogonal promoter sequence at two distinct sites within the RE ITR. As expected, the

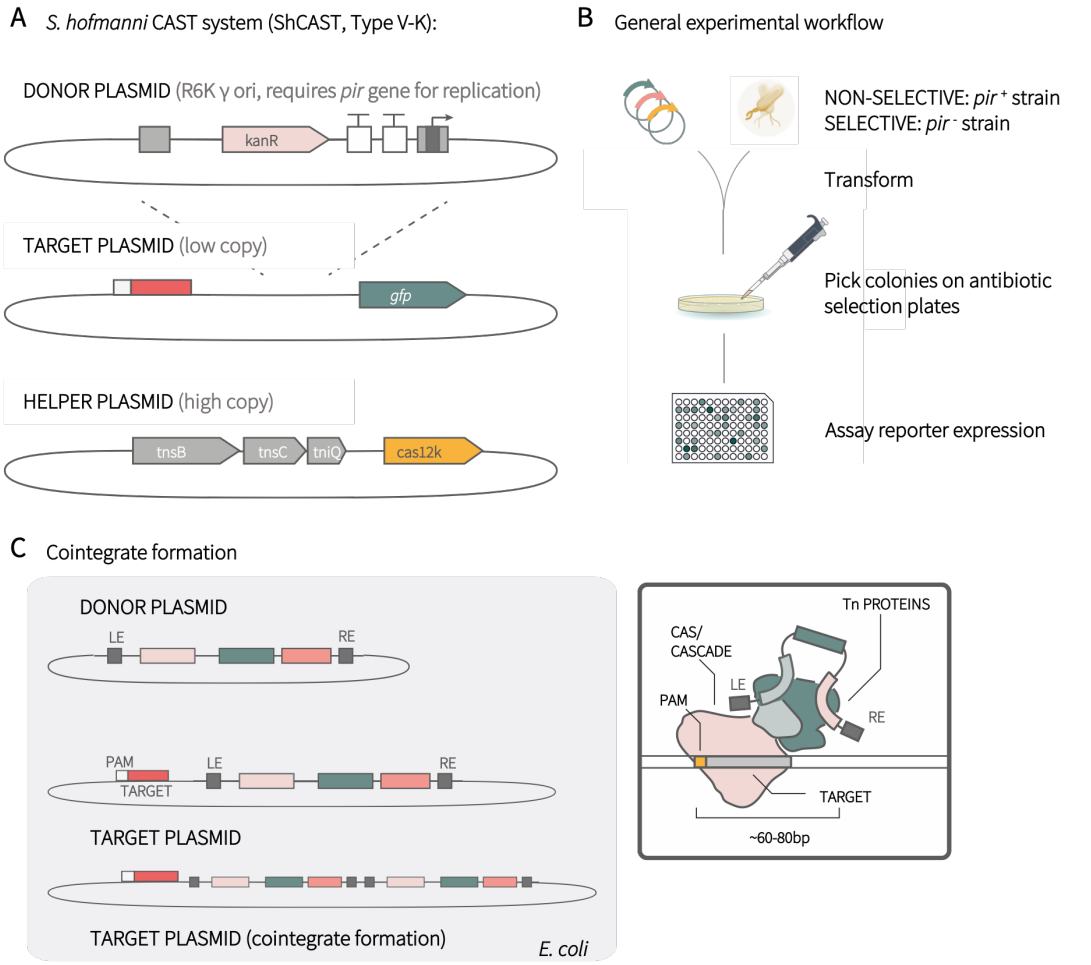

Figure 3.1: An experimental workflow for CAST-controlled circuits. (A) The *S. hofmanni* CRISPR-associated transposon (shCAST) system can be divided into sub-components across two plasmids to allow for independent modification of each element. The helper plasmid encodes the transposase genes, and the Type V-K CRISPR system consisting of *cas12k* and its associated gRNA. The donor plasmid harbors the mini-Tn7-like transposon with a selectable marker. Finally, a target plasmid is used to measure insertions through changes in the expression of a GFP reporter. (B) All three plasmids (pDonor, pHelper and pTarget) are co-transformed into *E. coli* cells, with the ability to stably maintain the R6K origin of replication on pDonor due the presence of a genomically integrated *pir* gene. After selecting for successful transformation (and in pir^{-} strains, strictly transposition as pDonor becomes a suicide vector), individual colonies are picked and reporter expression levels measured using a plate reader. *gfp*, green fluorescent protein; PAM, protospacer adjacent motif; LE, left end; RE, right end.

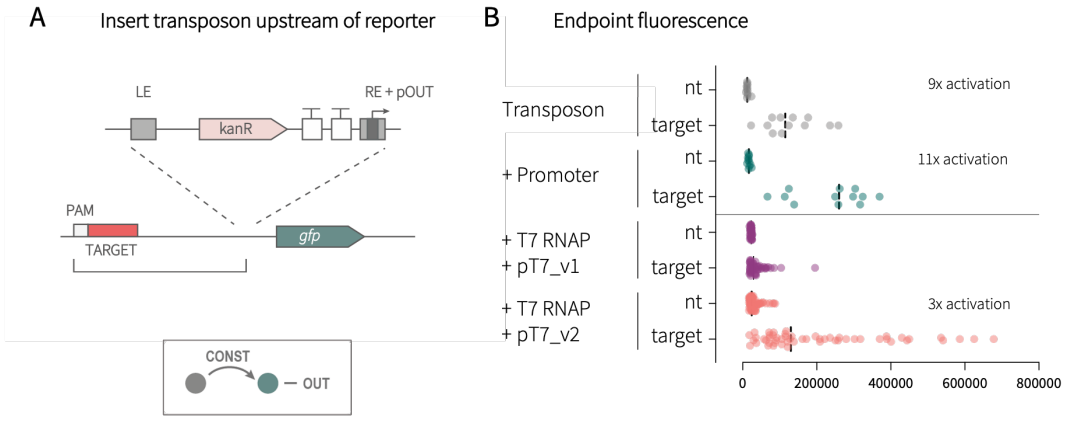

Figure 3.2: Introducing constitutive reporter expression with CAST systems. (A) We inserted promoters into the right end (RE) of the transposon, replacing sections of the wild-type sequence that had not been annotated as long (terminal) repeats (29). For the delivery of the T7 RNAP, we also included a cassette designed for low-level expression within the transposon (145). In this assay, successful transposition results in the insertion of the promoter upstream of a promoter-less fluorescent reporter and the establishment of constitutive expression. (B) Insertion of the unmodified transposon results in a significant 9-fold median increase in fluorescence compared to a non-targeting gRNA. This increases to an 11-fold activation with the addition of a pJ23119 promoter. Variants of a T7 RNAP-carrying transposon with two different T7 promoter insertions (pT7_v1 and pT7_v2) resulted in the heterogeneous activation of the target. In this assay, a pir^{+} strain was used, so there is no explicit selection for successful transposition. *gfp*, green fluorescent protein; PAM, protospacer adjacent motif; LE, left end; RE, right end; nt, non-targeting gRNA; target, targeting gRNA.

T7 RNAP did drive the maximum observed levels of gene activation, but with a significant level of heterogeneity: only one of the two promoter variants appeared functional (pT7_v2), and the spread amongst individual colonies was much higher than the previous two systems. This heterogeneity resulted in an overall lower median activation (∼3x), and is likely caused by a combination of the significant increase in size of the transposon (the T7 RNAP expression cassette is ∼4 kb) and the poorly-characterized impact of changes to the RE sequence on transposition efficiency.

Given the significant level of background activation observed from the 'wild-type' shCAST transposon, we anticipated that the introduction of externally inducible systems would be hampered by high levels of baseline expression in the "OFF" state. Indeed, an initial attempt to deliver an arabinose-regulated pBAD cassette within the transposon failed to introduce useful, titratable control over downstream *gfp* expression (Fig. 3.3). Importantly, in this example the promoter was contained within the transposon proper rather than implanted within the end sequence, demonstrating the potential pitfalls of controlling target expression across a ∼200 bp section of poorly characterized DNA sequence.

To address this challenge, we reasoned that active repression within the RE ITR itself would serve to dampen the baseline activation and then provide a mechanism to titrate activation after insertion (Fig. 3.4). We inserted two variants of the ATc-inducible pTet promoter (BBa_R0040 and PLtetO-1(146)) into the RE ITR sequence between the annotated long and short repeat

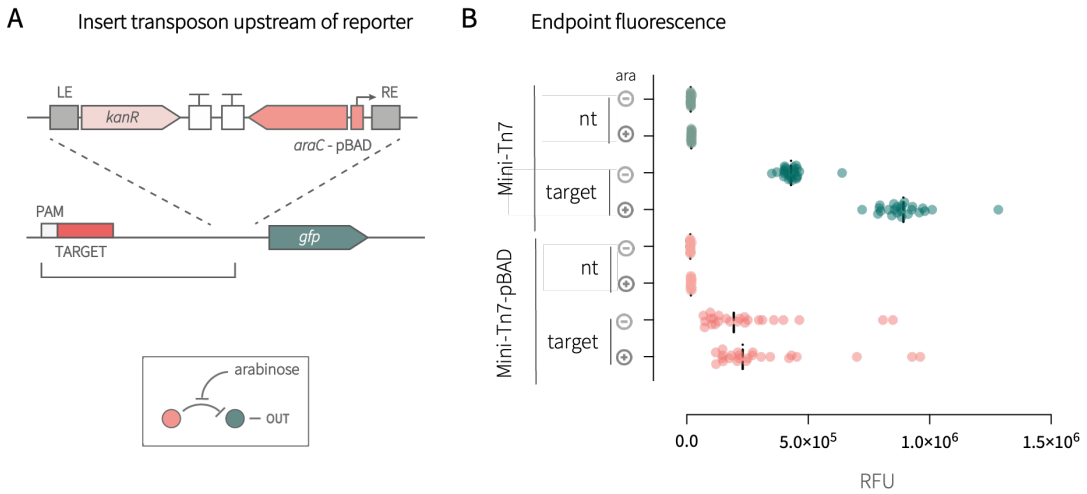

Figure 3.3: Unmodifed RE sequences do not support inducible expression. (A) We modified the transposon donor to encode an *araC*-pBAD promoter cassette oriented outwards from the interior, across the right end. Targeting the transposon upstream of a promoter-less reporter should introduce arabinose-inducible expression. (B) An unmodified version of the transposon shows significant baseline levels of gene activation, in agreement with Fig. 3.2. The increase in fluorescence upon the introduction of 0.2% arabinose is likely due to the increase in carbon source availability. The addition of the pBAD promoter resulted in a decrease in reporter expression, and this lower signal did not increase significantly upon induction with arabinose. In this assay, a pir^{+} strain was used, so there is no explicit selection for successful transposition. *gfp*, green fluorescent protein; PAM, protospacer adjacent motif; LE, left end; RE, right end; nt, non-targeting gRNA; target, targeting gRNA; ara, arabinose.

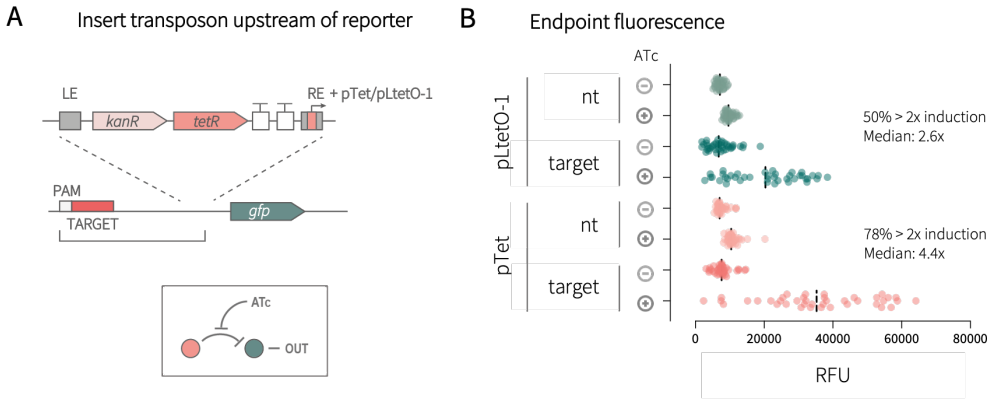

Figure 3.4: Repressor binding sites in the RE enable inducible target control. (A) We incorporated two *tetR* regulated promoter variants, and their cognate repressor, into the transposon end and its interior, respectively. (B) The repressor binding sites reduced the background levels of reporter activation in the absence of the inducer ATc. Across the two promoter variants, pTet BBa_R0040 (54 bp) and PLtetO-1 (74 bp), the background was reduced to the same level as the non-targeting gRNA. Addition of the inducer, ATc (100 ng/mL), resulted in ∼2.6 or ∼4.4-fold median reporter activation. In this assay, a pir^{+} strain was used, so there is no explicit selection for successful transposition. *gfp*, green fluorescent protein; PAM, protospacer adjacent motif; LE, left end; RE, right end; nt, non-targeting gRNA; target, targeting gRNA; ATc, anhydrotetracycline.

sections (29), and then assayed the impact of ATc induction on reporter expression after transposition (Fig. 3.4A). To supply the TetR repressor protein, we also included the *tetR* gene on the transposon under the control of the constitutive promoter pN25. Importantly, the presence of the *tetO* repressor binding site appeared to suppress the level of baseline activation in the absence of ATc, while the addition of ATc then alleviated this activity and lead to ∼2.5-4.5-fold median reporter activation depending on the pTet variant used (Fig. 3.4B). The apparent requirement for operator integration into the ITR itself could impose a size restriction on the promoter that can be used - pTet, PLtetO-1, pT7 and pJ23119 are all <75 bp in length, which is comparable to the longest continuous gap between annotated repeats in the RE (57 bp). Significantly larger promoters could disrupt transposition entirely.

Building on this model for active repression in the transposon end, we sought to increase the complexity of the signal integration occurring upstream of the reporter while maintaining a minimal promoter length. We modified the construct in Fig. 3.4, replacing the pTet promoter variants with a pT7-tetO promoter (Fig. 3.5A). This provides two layers of possible regulation: the ATc-mediated de-repression of the *tetO* element, and inducible control over the expression of the T7 RNAP itself. To achieve this, we performed this experiment in the BL21(DE3) strain, which harbors a LacI regulated T7 RNAP gene. Importantly however, this strain is pir^{-}, so unlike the experiments described in Fig. 3.2-3.4, the R6K replication origin of the pDonor plasmid is unsupported. Hence, following transformation of the pDonor, only cells with a successful transposition event will grow under kanamycin selection. This is in contrast to previous experiments in which transposition is not explicitly selected for, and therefore we expect to see 100%

of the transformed BL21(DE3) colonies showing reporter responsiveness. Indeed, all three of the colonies assayed showed high levels of reporter induction upon the addition of both IPTG and ATc (Fig. 3.5B), with titration curves showing the anticipated AND-like logic behaviour (Fig. 3.5C). The discrepancy between the fold-changes of the individual transformants in Fig. 3.5B warranted further investigation, and diagnostic PCR reactions confirmed the presence of two distinct types of transposition product: simple integrations (colonies 1 and 3), and a cointegrate (colony 2, Fig. 3.5D). This highlights an important consideration for synthetic circuit design using CAST systems: heterogeneity in the transposition process can lead to heterogeneity in the circuit response.

As a final proof-of-concept demonstration for the use of CAST systems in synthetic circuit design, we investigated their capacity to support the *in situ* generation of circuit diversity and memory in an analogous menchanism to recombinase-based logic circuits (147, 148). We focused on mimicking the behaviour of natural composite transposons (149), designing a test system in which the transposase can select from two possible transposon options defined by identical, layered LE ITR sequences (Fig. 3.6A). In this system, transposition of the longer sequence (Tn-1) delivers a promoter-less *gfp* reporter downstream of the target, while transposition of the shorter sequence (Tn-2) delivers just the kanamycin selection cassette. Given the inverse relationship between cargo size and transposition efficiency, we expect *a priori* that the smaller section will be preferentially mobilized. Importantly, we used a pir^{-} strain and blue-white screening to stringently select for colonies derived from successful transposition events.

To measure reporter insertion (and by proxy the transposition of Tn-1), we designed two gRNAs targeting the endogenous *lacZ* gene. Insertion of the *gfp* ORF downstream of the *lac* promoter should therefore create a reporter system responsive to IPTG-mediated LacI de-repression. Across both target sequences, we observed a bifurcation in the fold-activation of GFP expression upon IPTG induction: the majority of colonies did not respond to IPTG, whereas a small sub-population showed an equivalent level of activation to a control transposon that lacked the internal LE2 sequence (Fig. 3.6B). Across both target sequences, the non-responsive colonies corresponded to those with low absolute GFP levels, indicating that there was indeed a preference for the smaller Tn-2 and that the insertion of a reporter gene could be used to quantify the activity of an endogenous promoter (Fig. 3.6C).

3.2.2 Reconfigurable gene circuits based on CAST systems

With the proof-of-concept experiments described here, we have shown that engineered CAST systems can be used to activate target genes and detect the activities of endogenous promoters. We have also demonstrated a repression-based strategy to introduce inducible control over the downstream reporter, highlighting the significant impact that the ITR sequences can have on the bioactivity of the transposon and suggesting a blueprint for future engineering of these sequences.

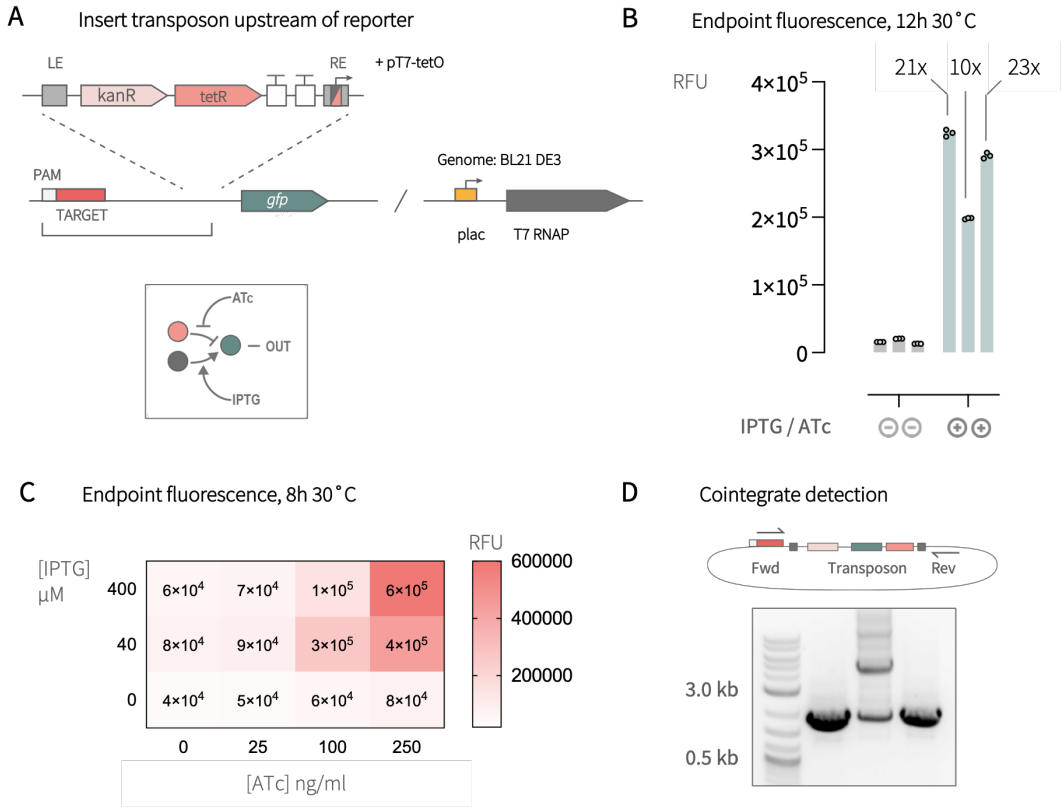

Figure 3.5: Signal integration at a dual-input promoter inserted with a CAST system. (A) In a similar approach to Fig. 3.4, we introduced a hybrid pT7-tetO promoter into the RE of the transposon, alongside the cognate repressor *tetR* carried in the transposon interior. By performing this experiment in a BL21(DE3) strain background, the T7 RNAP could be controlled via IPTG-induction. (B) Endpoint fluorescence measurements show a high fold-change increase in GFP levels upon the addition of both inducers (ATc, 250 ng/mL, IPTG, 400μM), with some heterogeneity across replicates. (C) The independent titration of both the inducers shows a characteristic AND-like logic behaviour. (D) Diagnostic PCR reactions for the three replicates shown in (B) (in corresponding order) indicate that the significantly lower reporter induction levels could be associated with the cointegration event. In this assay, a *pir*$^-$ strain was used, which explicitly selects for successful transposition. *gfp*, green fluorescent protein; PAM, protospacer adjacent motif; LE, left end; RE, right end; nt, non-targeting gRNA; target, targeting gRNA; ATc, anhydrotetracycline; IPTG, Isopropyl-β-D-1-thiogalactopyranoside

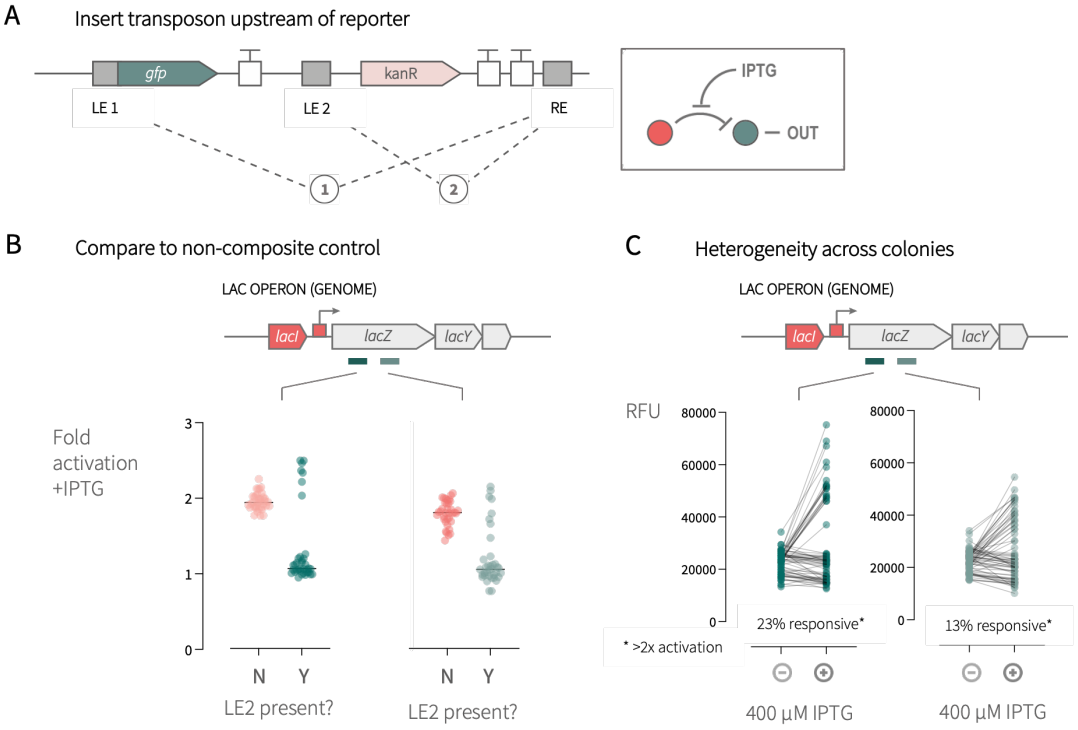

Figure 3.6: Composite transposons introduced a pre-defined degree of heterogeneity in circuit response. (A) Introducing two identical LE sections in sequence along with a single distal RE sequence creates two possible transposition donors: Tn-1 (outer LE1), and Tn-2 (inner LE2). To distinguish the outer from the inner product, we incorporated a promoter-less fluorescent reporter immediately downstream of LE1 but upstream of LE2. Finally, we used two separate *lacZ)*-targeting gRNAs to insert the composite transposon products into the endogenous *lac* operon, downstream of the IPTG-inducible promoter. (B) With a control donor transposon that lacks LE2, all of the insertion products show an ∼2-fold increase in GFP expression upon the addition of IPTG. In contrast, when two transposon donors are possible and one lacks the *gfp* reporter gene, there is a bifurcation in the response to IPTG. The overall fold change in signal for the responders decreases for the target site further from the *lac* promoter. (C) The proportion of replicates that show a ∼1-fold change in fluorescence upon IPTG addition also have low absolute fluorescence levels across both conditions. In contrast, a smaller sub-population show an IPTG-dependent increase in fluorescence indicating the presence of the Tn-2 product and its GFP cargo. This suggests that the longer Tn-1 donor is disfavoured, likely due to its larger size. In this assay, a pir^{-} strain was used, which explicitly selects for successful transposition. *gfp*, green fluorescent protein; PAM, protospacer adjacent motif; LE, left end; RE, right end; nt, non-targeting gRNA; target, targeting gRNA; IPTG, Isopropyl-β-D-1-thiogalactopyranoside.

Finally, we have implemented a composite transposon design that can be used to generate defined heterogeneity in a cell population akin to recombinase based memory circuits. In the future, this type of transposase-mediated structural diversification could be used to implement logic circuits, or enable lineage tracking given a sufficient number of changeable bits.

Moving forward, these CAST systems could be incorporated into reconfigurable synthetic gene circuits, in which the transposition events (reversibly) alter the underlying network structure based on inducible gRNA expression, and hence the overall behaviour of the system (150). To achieve this, steps must be taken to reduce the heterogeneity in the transposition process across a population. This could include maximising transposition efficiency by modifying ITR sequences, cargo size, gRNA sequence, and subcomponent stoichiometry (31). To this end, an engineered solution to the problem of cointegrate formation in Type V-K systems (Fig. 3.5) has recently been developed that uses a nicking endonuclease to replace the absent TnsA functionality (144). Ultimately, defined design rules for the incorporation of ORFs and promoters in the transposon will need to be established, particularly in the case where promoters must be embedded in the ITR sequences themselves.

3.3 Engineering autonomous transposons for directed genome evolution

For very complex systems, our ability to design optimal solutions from sequence alone can be a major challenge, primarily because of the size of the design space to be explored (151, 152). Directed evolution can be used to optimize and re-purpose existing tools or strains towards specific tasks without the need for a prior understanding of the underlying sequence-function relationships. Generally, this approach uses iterative rounds of diversification and selection to explore the sequence-activity landscapes of single molecules or circuits (16). This idea was explored early on using targeted mutagenesis experiments in small directional synthetic gene networks to generate pre-defined logic behaviors (153). Directed evolution is now regularly used to expand the repertoire of canonical synthetic biology tools (154–156), for example by creating transcription factors with new target sites or responding to synthetic molecules (157) (Fig. 3.7A). However, sequential rounds of *in vitro* mutagenesis and *in vivo* selection are not suitable for the untargeted evolution of large-scale, highly connected networks, as they require a detailed knowledge of critical actuation points within the circuit.

To improve the throughput and biological relevance of these approaches, recent efforts have focused on *in vivo* continuous directed evolution (158), in which synthetic circuits couple a complex biological property such as protein solubility to a fitness-conferring gene (159). Continuous, *in vivo* evolution could be a powerful approach to re-wire and optimize endogenous gene networks, for reasons deeply rooted in evolutionary theory: highly connected networks with redundant or promiscuous connectivity (e.g. many overlapping transcription factor-promoter interactions) are very robust to evolutionary change, meaning that seemingly important network alterations often have a minor impact of the phenotype (94). While this may, at first sight, seem to work against the emergence of new functionalities, robustness actually enables the accumulation of no-cost genotype changes that could eventually generate entirely new phenotypes (160, 161). However, considering the size of the evolutionary landscape of even a small gene network, enabling technologies are needed to explore the vast sequence space on practical timescales. Rather than focusing on fine-grained nucleotide-level diversification, we can look to natural systems to identify coarse-grained, structural changes that drive genome-scale network evolution (Fig. 3.7B). In this section, I present the design and implementation of a transposon-mediated *in vivo* mutagenesis platform to complement the growing number of directed genome evolution platforms (162)*.

*The manuscript associated with this work is still in preparation, and the data presented here and their interpretation are accurate at the time of writing. However, they may be subject to change pending more detailed analysis. Supplementary data will be included in the final paper but are as yet unprepared.

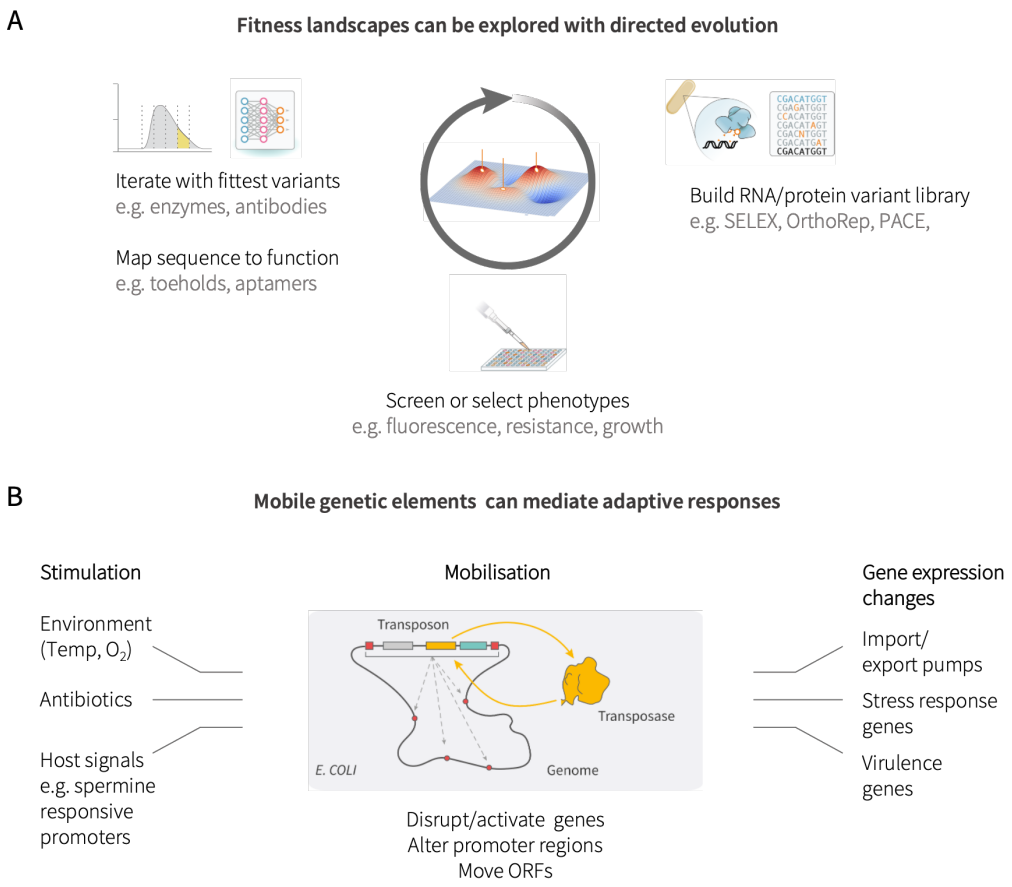

Figure 3.7: Directed evolution enables the non-rational generation of new behaviours, but few approaches incorporate genome-wide structural changes. (A) A typical directed evolution cycle will focus on generating nucleotide-level diversity for a defined module of interest (e.g. a protein or RNA). This can be done *in vitro* or *in vivo*. An appropriate selection or screening criteria must be carefully designed to identify variants with a desired property but not 'cheaters'. Following phenotyping, the proportion of the population exceeding some selection criteria can be carried forward for a next cycle of diversification and enrichment. Alternatively, structure-function models can be used to design or identify candidates for further iterations. (B) Nucleotide-level mutation does not scale well for genome-wide diversification. However, in microbes, transposons or insertion sequences can mediate adaptive responses and drive population-level diversity by moving within genomes. This can be initiated by defined external stimuli, and can lead to changes in the structure and connectivity of gene regulatory networks.

3.3.1 An engineered platform for directed genome evolution

Designing well-defined platforms to study and recreate complex natural systems in the laboratory is a major area of focus for synthetic biology (6), beginning with early efforts to study gene expression noise (163, 164). As a field, it has already begun to provide tools to introduce structural (165) or network connectivity (94) changes using pre-defined, functional components. Harnessing these coarse-grained perturbations for laboratory evolution could accelerate strain engineering through network re-wiring (166), and build rich data for network-level models of cell phenotypes (152). In turn, this would provide a platform to study the impact of analogous processes in the evolution of natural systems.

Despite their established role in microbial genome dynamics and evolution (119), there have been few systematic, forward-engineering studies into the role of transposons in the adaptability and evolution of complexity in gene regulatory networks (120–122). Current continuous directed evolution technologies are primarily focused on accelerating nucleotide-level diversification on orthogonal plasmids or defined genetic loci (158). By contrast, we lack inducible control over larger structural changes in laboratory evolution experiments, severely limiting our understanding of these evolutionary processes. Many of the larger-scale processes generating genome-wide structural heterogeneity in cells – for example duplication, translocation, or recombination of DNA segments (108) – promise to accelerate fitness landscape exploration through mechanisms distinct from nucleotide mutagenesis (162). Currently, these approaches are limited by our ability to track the underlying changes across a genome and their reliance on artificially introduced recombination sites. While transposon-based tools have been developed for high-throughput loss- and gain-of-function genetic screens (e.g. transposon insertion sequencing, Tn-seq)(15), these pipelines employ a single, static round of mutagenesis and have not been optimized for continuous evolution. These tools are used to quantify gene essentiality across genomes and environments (124) or specific phenotypes of interest (125, 167), but cannot reveal evolutionary dynamics between host and parasite(117, 118).

To this end, we set out to develop an engineered, self-propagating transposon platform to study the role of MGEs in the evolution of complex natural gene regulatory networks, and to harness their capacity to accelerate genome diversification through continuous, genome-wide mutagenesis *in vivo*. Existing transposon tools for genetic screens have been developed on an ad-hoc basis, and lack the ability to continuously re-wire endogenous networks to probe their evolvability. Building on the hyperactive mariner transposase derivative Himar1C9 from the horn fly *Haematobia irritans* (168), we adopt a synthetic biology framework to transposon platform design that focuses on three areas: (i) establishing a systematized, modular assembly workflow to deliver functionalized transposon variants for gene-network re-wiring (169, 170), (ii) comparing the impact of engineered transposon variants on host phenotypes in parallel evolving populations, and (iii) using unique transposon barcodes to track individual lineages within host populations via longitudinal next-generation sequencing (NGS) readouts (171).

We demonstrate that this autonomous transposon platform can re-capitulate natural examples of insertion-mediated host adaptation to both nutrients and antibiotics. We then compare the impacts of transposon functionalization on the evolution of *E. coli* populations towards diverse carbon source utilization. Finally, as a proof-of-concept experiment with more complex, dynamic environments, we compare parallel evolutions across alternating carbon sources and investigate the impact of contingency on host adaptation. Moving forward, we anticipate this directed genome evolution platform playing a central role in longer-term evolution experiments, focusing on the dynamics of host-parasite interactions and in particular how this depends on contingency (172) and environmental complexity (e.g. temporal structure) (173–175).

3.3.2 Results

Construction and characterization

Our platform for continuous, transposon-mediated genome evolution can be broken down into three core functionalities: a titratable mechanism for transposase expression (in *cis* or in *trans*) (Fig. 3.8Ai), a DNA assembly strategy to functionalize transposons with natural and synthetic cargos (e.g. reporters, promoters and/or transcription factors) and thereby enable network rewiring (Fig. 3.8Aii) (94), and a high-resolution, genome-wide readout of insertion location based on next-generation sequencing (Fig. 3.8Aiii). To build this self-propagating transposon mutagenesis platform, we started from the well-established mariner transposon *himar1C9*. This transposon - derived from the horn fly *Haematobia irritans* - has been optimized for high insertion frequencies and is orthogonal to natural microbial mobile genetic elements (168). Furthermore, it yields a random distribution of insertions specifically at TA sites throughout the genome, and is compatible with a range of next-generation sequencing pipelines for insertion localization (15). To enable autonomous self-propagation of the transposon within the genome, we explored two strategies for the expression of the transposase: the first resembles a natural transposon, with the transposases acting in *cis* from within the region flanked by the inverted repeat (IR) sequences (see publication, in prep.). The second uses a medium copy helper plasmid (pHelper) to provide transposase acting in trans (Fig. 3.8Ai). In this work, we primarily focus on the plasmid-based transposase expression strategy, as it supports the rapid prototyping of transposon variants in a single common pHelper strain and their sequential introduction into a single genome (Fig. 3.8B).

To enable the external titration of transposase expression, we compared two pHelper designs: an arabinose-inducible araC-pBAD expression vector (177), and an ATc-inducible tetR-pTet expression vector. First, we compared the number of colonies formed following transformation of a pir^{-} pHelper strain with two transposon donor variants, representing a simple proxy for the frequency of initial transposition from the non-replicating donor plasmid (Fig. 3.8C). Active re-

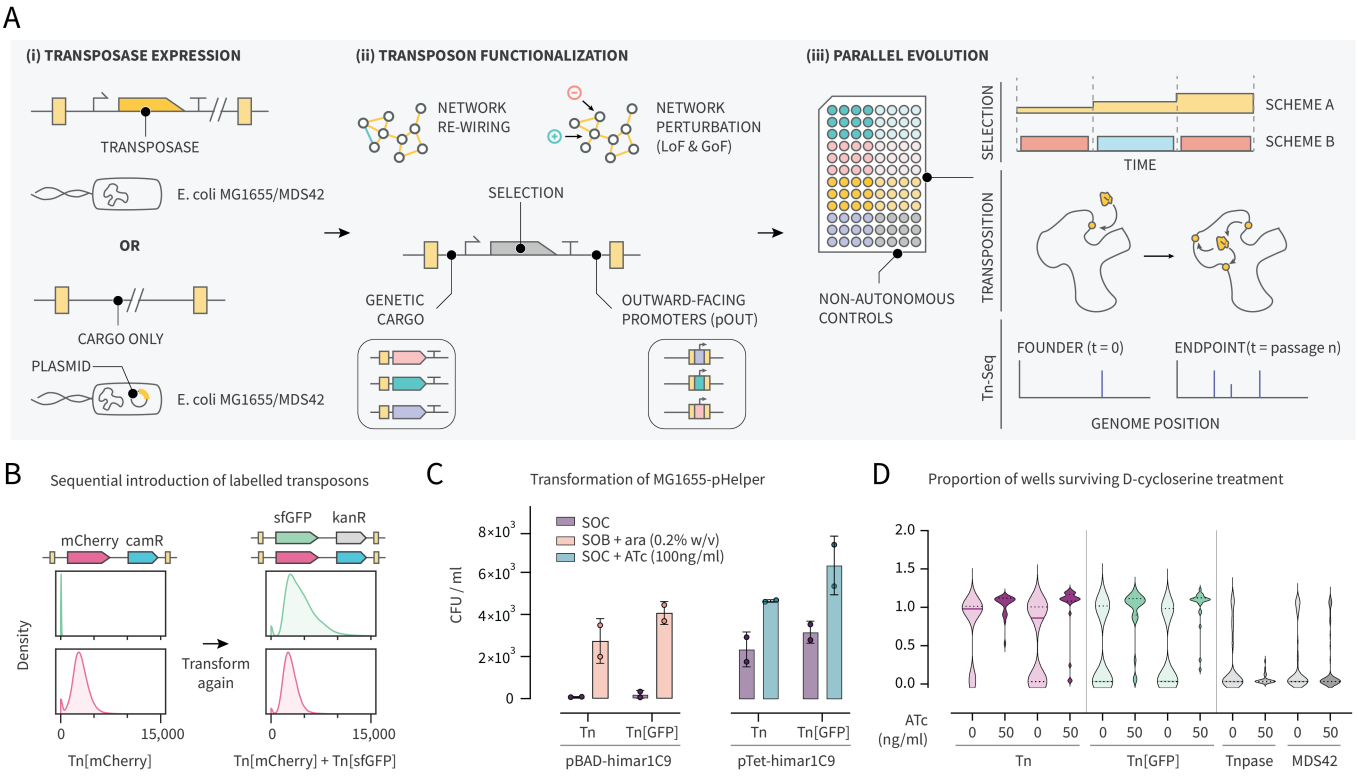

Figure 3.8: An engineered platform for continuous transposon-mediated mutagenesis and gene regulatory network re-wiring. (A) Our approach to directed genome evolution can be broken down into three core elements: (i) a mechanism to supply titratable levels of transposase in cells either from within the transposon (in cis) or from an independent pHelper plasmid (in trans); (ii) a DNA assembly platform to functionalize and barcode transposon variants, allowing in situ network re-wiring and external tracking, respectively; and (iii) a parallelized approach to evolution focusing on comparing replicate linages across different transposon variants and environmental conditions, with a modified Tn-Seq pipeline for genome-wide insertion site identification. (B) Transposase expression in trans enables the sequential transformation and insertion of transposons functionalized with orthogonal fluorescent reporter cassettes into the genome of *E. coli* MDS42 cells, as determined by flow cytometry (FITC, 488nm; mCherry, 561nm). (C) A comparison of two strategies to titrate transposase expression levels from the pHelper plasmid using either pBAD or pTet. We used relative CFU counts on selective plates as a proxy for insertion frequencies following transformation of chemically competent MDS42-pHelper cells with the transposon donor suicide vector ($3x10^9$ plasmid molecules, R6K origin of replication). The inducers (arabinose or ATc) were added after the heat shock, during the rescue in SOC/SOB (1 hr, 37°C). (D) The introduction of transposase and transposon elements into MDS42 cells lacking endogenous MGEs increases the rate of *in vivo* mutagenesis. As a proxy, we measured the proportion of replicate, bottlenecked cultures that develop resistance to the antibiotic D-cycloserine (20uM) via the spontaneous inactivation of the *cycA* gene (e.g. through transposon insertion) (120, 121, 176). Endpoint ODs (40 hrs, 37°C) were measured for n=24 colonies from duplicate transformations, n=24 colonies for the parental MDS42-pHelper strain, and n=48 for 'wild-type' MDS42. Tn, mariner transposon (kanR); Tn[GFP], mariner transposon harboring a green fluorescent protein gene; Tnpase, *himar1c9* transposase; ATc, anhydrotetracycline; ara, arabinose; Tn-Seq, transposon insertion sequencing.

pression of the pBAD promoter in the presence of glucose resulted in a relatively low background insertion frequency. Conversely, while the pTet helper resulted in a higher absolute number of insertion mutants, the OFF state also retained a relatively high level of activity. A similar pattern was observed in an established method for comparing *in vivo* mutation rates: resistance to the antibiotic D-cycloserine (D-cyc) occurs following inactivation of the *cycA* gene (176, 178). By measuring the relative frequency of the emergence of spontaneous D-cyc resistance (D-cyc^R) mutants in a non-selected population, a composite metric for the various sources of genomic mutation rates can be derived that includes MGE-mediated gene disruptions (121). Again, we found that the pBAD system offered tighter control over the genomic mutation rate while the pTet system offered a higher maximum and an intermediate uninduced state (see publication, in prep.). Furthermore, we demonstrated similar results in a simplified, plate-format version of this assay based on the proportion of growth-positive wells inoculated with a severely bottlenecked population of cells (Fig. 3.8D). By reducing the need for plating and colony counting, this approach would be compatible with higher-throughput, liquid-handling based pipelines for benchmarking transposon variants (17, 19). While the pBAD-regulated transposase expression platform offers tighter control, our need for an inducer that was orthogonal to the nutrient content of any future growth conditions meant that we employed the *tetR*-pTet pHelper expression vector in subsequent experiments.

Benchmarking using a known transposon-mediated adaptation

MGEs can play an important role as genetic switches, remodeling the host genome in response to changes in environment in ways that can sometimes be adaptive for the host. One well-documented example of this is the activation of the *bgl* operon to facilitate growth on the glycoside arbutin (Fig. 3.9A): frequent insertion of IS1 or IS5 into the operator of the regulatory gene *bglG* disrupts a H-NS repressor binding site, leading to an increase in *bglG* expression and the upregulation of the structural genes *bglF*, *bglB* and *bglH* (179, 180). Guided by previous work developing this phenomenon into an assay for MGE activity (121), we began by benchmarking the continuous *himar1C9* mutagenesis platform in this context. First, we modified this protocol by switching from solid to liquid media, using longitudinal OD measurements rather than CFU counts (Fig. 3.9C). This enabled a dramatic increase in throughput: we compared 48 replicates across four strains (MDS42, MDS42 pHelper, MDS42::Tn pHelper, and MDS42::Tn-pOUT pHelper) and two inducer conditions (no ATc, or 50ng/ml ATc) for a total of 576 parallel evolving populations (Fig. 3.9B).

These populations were passaged in 96-well plate format through minimal media with an increasing ratio of arbutin to glucose as a carbon source. High-OD endpoint phenotypes (11/48 replicates) were only observed in the Tn-pOUT strain exposed to the ATc inducer (Fig. 3.9B-D).

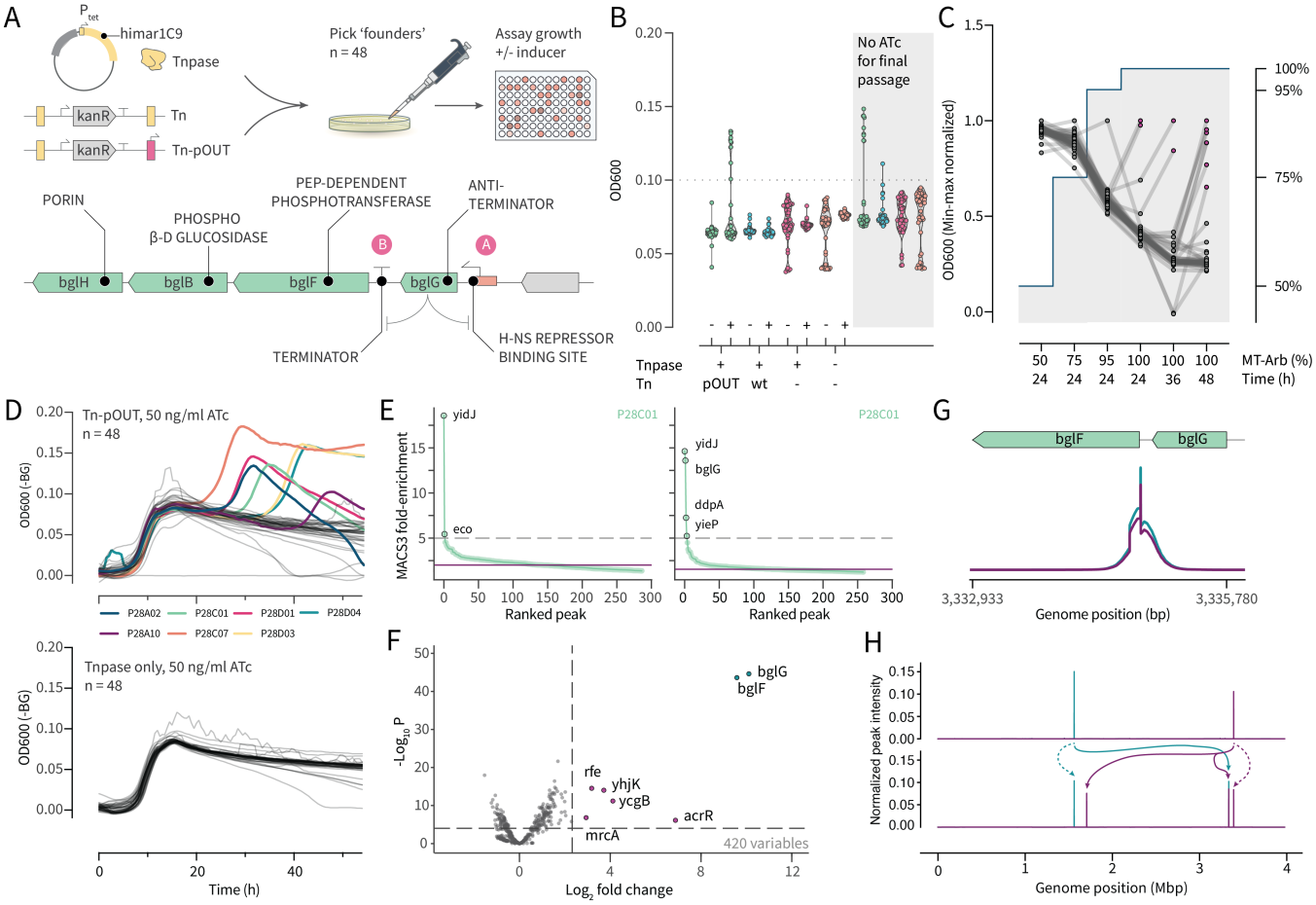

Figure 3.9: Validating transposon-dependent gene activation with a model, cryptic metabolic operon. (A) The adaptation of wild *E. coli* isolates to growth on the carbon source arbutin involves the disruption of an H-NS repressor binding site upstream of the positive regulator bglG by IS1 or IS5, and the subsequent activation of the structural genes bglF/B/H (121, 179). We introduced two transposon variants, Tn and Tn-pOUT (pJ23104), into MDS42-pHelper cells and cultured 576 parallel populations derived from unique, single-colony founders. (B) Endpoint OD600 measurements (48 hrs, 1.0g/L arbutin) for n=48 replicates per condition. Data in the grey box are cultures that were supplemented with ATc (50ng/mL) until the final passage. (C) Min-max normalized OD measurements for n=48 replicates in a single condition (Tn-pOUT + ATc throughout, Fig. 3.9B green dots) showing the emergence of high-growth variants in 100% MT-arbutin media (1.0g/L). (D) Growth curves for two conditions (Tn-pOUT and Tnpase only, both with ATc) in MT-arbutin media (1.0g/L), inoculated from the 24 hr timepoint in 95% MT-arbutin media from Fig. 3.9C. High-growth replicates are colored. (E) Fold-enrichment values for peaks identified using MACS3, based on aligned Tn-Seq reads from the start-point (TB, left) and endpoint (MT-arbutin, right) cultures for a single founder colony (P28C01). The initial insertion (yidJ) is the dominant peak before selection, and is retained alongside new high-intensity peaks post-selection including bglG. (F) Differentially-enriched insertion loci (labelled points) across n=11 paired lineages identified using Bio-Tradis (181). (G) All (11/11) of these replicates exhibited a sharp insertion peak at a TA site upstream of bglF (two representative traces shown). (H) Longitudinal sequencing enables the genome-wide tracking of transposon movement from their initial insertion location to their convergence on the *bgl* operon, including intermediate insertions (purple).

Based on this evidence for a transposon variant-specific phenotype, we modified an established Tn-Seq pipeline to compare the initial and endpoint insertion locations across all eleven replicates (Fig. 3.9E). As expected, for the pre-selection populations each of the 11 replicates was identified with a unique, highly enriched 'founder' insertion (except in one case, where the reads from the initial insertion aligned to a set of ribosome genes with high sequence similarity) (Fig. 3.9C). In every corresponding endpoint population that evolved a high-growth phenotype on arbutin, a second highly enriched peak was identified in the terminator region downstream of *bglG* and upstream of *bglF* (Fig. 3.9F). Furthermore, we consistently observed the preservation of the donor site during transposon movement, in keeping with other reports on transposon remobilization (31, 182): this is evident from the retention of the signature 'founder' insertion in the endpoint samples.

By comparing parallel replicates that represent distinct founder insertions, these results suggest that evolution with an autonomous Tn-pOUT reproducibly achieves an equivalent growth outcome to observations of natural IS1/IS5 insertions, but via a slightly distinct mechanism: the direct upregulation of *bglF/B/H* expression rather than disruption of H-NS mediated *bglG* repression (Fig. 3.9G) (179, 180). Interestingly, in some of the 11 replicate populations, we detected additional highly enriched peaks that we hypothesize represent intermediate insertion events preceding the high fitness insertion at *bglF* (Fig. 3.9F,H). Through genetic linkage, these intermediate insertions would be carried to fixation in the population. Taken together, our results confirm the capacity of our engineered, autonomous transposon system to replicate natural adaptive processes and highlight the important influence that specific functional differences (i.e. the presence of an outward-facing promoter) can have on the overall contribution of MGEs to the evolvability of their host populations.

Screening for transposon-dependent phenotypes

To test the power of our engineered, autonomous transposon system to identify insertion-dependent phenotypes, we designed a rapid screen of 31 carbon sources in a standard Biolog EcoPlate assay (183): after transforming the pHelper MDS42 strain with two transposon variants, we picked three founder colonies for each and included three control colonies from the parental strain. After several passages through fresh plates, we identified three carbon sources in which the transposon-harboring strain appeared to exhibit a partial or complete growth advantage: L-serine, glycyl-L-glutamic acid, and b-methyl-D-glucoside (Fig. 3.10A). Both glycyl-L-glutamic acid (184) and β-methyl-D-glucoside (185, 186) have previously been described as examples of growth conditions that differentiate laboratory strains of *E. coli* and environmental strains. The isolate capable of L-serine utilization showed a high overall tolerance compared to MD42 and MG1655, a phenotype that remained stable in the absence of antibiotic selection for the transposon (Fig. 3.10B). In the presence of glucose as an alternative carbon source, we observed detectable growth in up to 100 g/L of L-serine (Fig. 3.10C-D).

To verify how reproducibly this strain evolved and to identify the underlying genetic changes, we augmented the initial screen with a second experiment comparing 16 unique founder colonies for each of the same three strains across three passages in 12 g/L-serine in the presence of ATc. We observed growth in 14/16 colonies when the transposon was functionalized with a constitutive promoter, but no growth in cultures with either an unfunctionalized promotor or no transposon (Fig. 3.10E). Comparison of Tn-seq data for four of the 16 colonies before and after selection identified a reproducible insertion upstream of the *sdaA* gene (Fig. 3.10F-G), which encodes the 4Fe-4S enzyme L-serine deaminase I and catalyses the conversion of L-serine to pyruvate and ammonium (187). In *E. coli*, *sdaA* is typically regulated by a σ32 promoter as part of the heat shock response, and its activity is strongly dependent on the oxidative state of the iron-sulfur complex - oxidative inactivation occurs in aerobic conditions (188). The low enzyme activity observed in cell extracts under aerobic conditions might therefore explain the strong selection for transposon-mediated gene activation, and a re-wiring from the natural σ32-mediated expression to the synthetic, constitutive σ70 promoter harbored on the transposon. This experiment highlights the rapid and stable generation of useful mutant phenotypes that can be achieved with this platform (Fig. 3.10A), and the ease with which the transposons can be tracked to identify the underlying genotypes (Fig. 3.10G).

A modular assembly pipeline facilitates rapid prototyping and barcode tracking

The results presented in this work so far, as well as those from diverse insertion mutagenesis screens (15), demonstrate that the functionalization of the transposon and the design of the donor plasmid can impact the outcomes of a given screen. Modular, combinatorial assembly pipelines facilitate the rapid prototyping of genetic constructs from lists of defined parts, and allow for pooled screens (169, 170). We modified an existing Golden Gate assembly platform for transposon mutagenesis constructs ('Magic Pools') to allow for the incorporation of libraries of outward-facing promoters and additional genetic cargos (Fig. 3.11A). Importantly, this platform centers on the introduction of barcodes into individual transposons such that their unique molecular identity can be cross-referenced with their genomic location via random barcode Tn-Seq (RB-Tn-Seq) (Fig. 3.11, inset) (171). While this is typically implemented to allow for standardized Bar-Seq workflows when comparing insertion mutant libraries across several conditions, in the context of our autonomous mutagenesis platform it performs a distinct function – allowing for the longitudinal tracking of barcoded transposons as they propagate through genomes via computational demultiplexing.

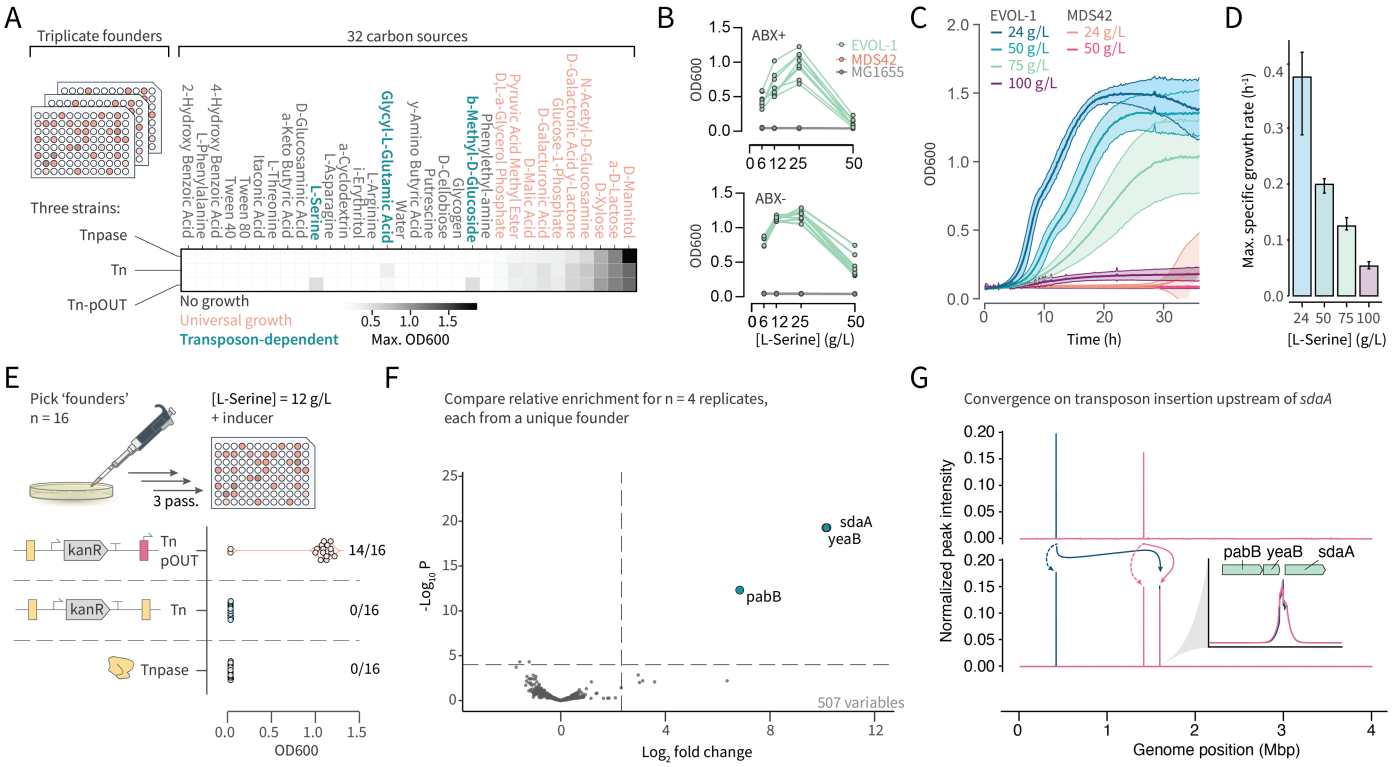

Figure 3.10: Screening for transposon-dependent carbon source utilization phenotypes. (A) For a preliminary screen of 31 carbon sources (Biolog Ecoplate format), we used triplicate founder colonies for two transposon variants (Tn, Tn-pOUT with pJ23104) and the MDS42-pHelper parental strain as a control. The heatmap shows the maximum OD across the three replicates for each condition. We identified three distinct groups: no growth (18), universal growth (10), and transposon-dependent growth (3). (B) L-serine utilization was observed in 1/3 of the Tn-pOUT replicates (EVOL-1). Further comparison of this isolate to MG1655 and MDS42 demonstrated its high tolerance to L-serine levels, both with (upper) and without (lower) antibiotic selection for the transposon and transposase. (C) Detectable growth at even higher concentrations of L-serine (up to 100g/L) was observed in the presence of glucose (4.0g/L). (D) Growth rate estimates for EVOL-1 from the data in (C) calculated using the curveball python package [73]. (E) A repeated screen focusing on L-serine increased the number of unique founder replicates to 16. Reproducible evolution of L-serine tolerance was only observed with the Tn-pOUT transposon (14/16 replicates). (F) With Tn-Seq data from 4 of these samples at both the start- and endpoints, we identified three genes from a single region showing high differential enrichment: *pabB*, *yeaB* and *sdaA*. (G) Both *pabB* and *yeaB* are upstream of *sdaA* (L-serine deaminase I), with genome-wide maps showing convergence from distinct founder peaks and high-resolution maps showing convergent transposon insertion in the promoter region (inset).

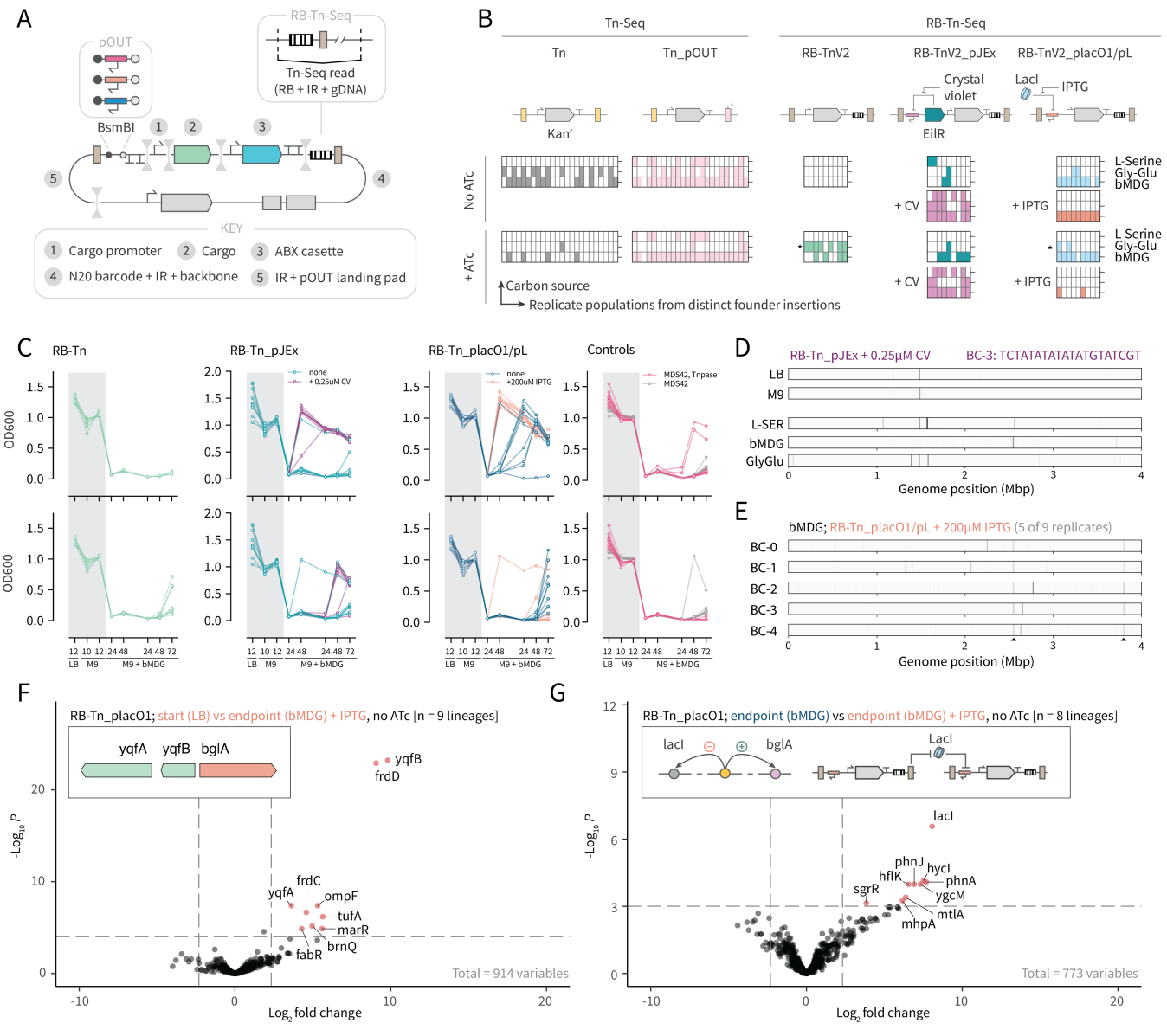
A
RB-Tn-Seq
pOUT
Tn-Seq read
(RB + IR + gDNA)
BsmBI
1
2
3
4
5
KEY
1 Cargo promoter
2 Cargo
3 ABX casette
4 N20 barcode + IR + backbone
5 IR + pOUT landing pad
B
Tn-Seq
RB-Tn-Seq
Tn
Tn_pOUT
RB-TnV2
RB-TnV2_pJEx
RB-TnV2_placO1/pL
Crystal violet
LacI
IPTG
Kanr
EilR
No ATc
+ ATc
+ CV
+ IPTG
L-Serine
Gly-Glu
bMDG
Carbon source
Replicate populations from distinct founder insertions
C
RB-Tn
RB-Tn_pJEx
RB-Tn_placO1/pL
Controls
OD600
none
+ 0.25uM CV
+200uM IPTG
MDS42, Tnpase
MDS42
LB
M9
M9 + bMDG
D
RB-Tn_pJEx + 0.25μM CV
BC-3: TCTATATATATATGTATCGT
LB
M9
L-SER
bMDG
GlyGlu
Genome position (Mbp)
E
bMDG; RB-Tn_placO1/pL + 200μM IPTG (5 of 9 replicates)
BC-0
BC-1
BC-2
BC-3
BC-4
Genome position (Mbp)
F
RB-Tn_placO1; start (LB) vs endpoint (bMDG) + IPTG, no ATc [n = 9 lineages]
yqfA
yqfB
bglA
yqfB
frdD
frdC
yqfA
ompF
tufA
marR
brnQ
fabR
Total = 914 variables
-Log10 P
Log2 fold change
G
RB-Tn_placO1; endpoint (bMDG) vs endpoint (bMDG) + IPTG, no ATc [n = 8 lineages]
lacI
bglA
LacI
lacI
phnJ
hycI
hflK
phnA
sgrR
ygcM
mtlA
mhpA
Total = 773 variables
-Log10 P
Log2 fold change

Figure 3.11: (Previous page.) **A modular assembly platform for transposon functionalization and barcode-based lineage tracking.** (A) A modified two-layer Golden Gate assembly pipeline focusing on modular plasmid assembly (BbsI) and promoter library insertion (BsmBI) within the transposon enables rapid prototyping and transposon barcoding. (B) Comparison of the evolutionary impacts of the two original transposons (Tn and Tn-pOUT, n=24 per condition) with three second generation transposons based on RB-TnV2 with additional outward promoters (RB-TnV2_pJEx, RB-TnV2_pLacO1/pL, n=9 per condition) across the three carbon sources identified in Fig. 3A). Each cell represents a replicate culture, with each column derived from the same initial founder colony and each row corresponding to a unique carbon source. Colored cells indicate growth (endpoint OD600>0.2). Rows marked with a * indicate contamination from RB-TnV2_pJEx cultures as determined based on the unique sequencing barcode of each founder. (C) OD600 measurements for each carbon source were used to track the emergence of growth phenotypes in cultures with (top row) and without (bottom row) ATc and the pOUT-specific inducers CV or IPTG. We included two controls: MDS42-pHelper expression the transposase only, and the parental MDS42 strain (n=12 per condition). (D) Computational barcode demultiplexing from pooled sequencing runs for each unique variant-environment combination enabled the re-construction of insertion mutant lineages. For a single lineage, condition-specific insertion spectra evolve from the initial founder insertion(s). (E) By comparing independent lineages for a single carbon source-inducer combination, convergent insertion loci emerge (black arrows). (F). Differential enrichment of RB-Tn-Seq reads between paired start- and endpoint samples for independent lineages further confirms reproducible insertion sites. For MDS42 cells harboring transposons with a pLacO1 promoter in bMDG + IPTG, the two common peaks in Fig. 3.11E correspond to frdD and yqfB (likely activating bglA, inset). (G) For RB-TnV2_pLacO1/pL, growth phenotypes in bMDG emerge later for this transposon variant in the absence of IPTG (Fig. 3.11C). Differential enrichment implicates a secondary insertion in lacI in the absence of IPTG, likely rescuing the activation potential of the pLacO1 transposons (inset). RB, random barcode; IR, inverted repeat; CV, crystal violet; IPTG, Isopropyl β-D-1-thiogalactopyranoside; bMDG, β-methyl-D-glucoside;

To systematically investigate the impact of transposon functionalization on network re-wiring, we evolved parallel replicates of five different transposon systems across three different carbon sources, comparing the two original transposon variants (Tn, Tn-pOUT) used in this work with three additional barcoded variants generated using the modular, combinatorial assembly pipeline (RB-TnV2, RB-TnV2_pJEx, and RB-TnV2_placO1/pL) (Fig. 3.11A). Importantly, we studied replicate populations initiated from a single-insert founder (n=24 or n=9). We also included two controls; the transposase-expressing pHelper MDS42, and unmodified MDS42. Barcoded transposon variants RB-TnV2_pJEx and RB-TnV2_placO1/pL are functionalized with inducible promoters: pJEx is based on the Jungle Express platform and harbors the engineered crystal violet-responsive repressor *eliR* (189), while placO1/pL is regulated by the endogenous LacI protein and its substrate analogue IPTG (146). We sought to compare the impact of these transposons on the adaptation of MDS42 populations to growth on the three carbon sources identified in the initial screen (L-serine, glycyl-L-glutamic acid, and β-methyl-D-glucoside, Fig. 3.10A), focusing on any differences observed between the inducible promoters with and without their respective ligands (Fig. 3.11B).

For each carbon source, we used longitudinal OD measurements across three passages to identify the emergence of growth phenotypes (Fig 3.11C). To track the location of the transposons within the genomes of the evolving cells, we performed RB-Tn-seq on pooled samples at three timepoints across the experiment: LB (24h), M9 (24h), and at the end of the third passage on the selective carbon source (48h). Importantly, the unique barcode defining each founder colony supports both the computational demultiplexing of the underlying lineages of each pooled sample (Fig

3.11D), and the identification of cross-contamination between wells (a common problem in long-term plate-based assays) (Fig 3.11B) (17). For each unique combination of carbon source and inducer(s) with n > 1 growth phenotypes, we compared demultiplexed, replicate lineages to identify reproducible insertion sites associated with carbon source utilization (Fig. 3.11E). As an example, in the case of the TnV2_placO1/pL transposon induced with IPTG, two conserved insertions emerge: one in *yqfB* (adjacent to *bglA*), and one in *frdD* (Fig. 3.11F). The *bglA* gene encodes 6-phospho-β-glucosidase A, the upregulation of which has previously been associated with mutants able to utlize β-methyl-D-glucoside as a carbon source (185, 190). The potential role of *frdD* in this phenotype is yet to be determined, but it appears to be specific to the TnV2_placO1/pL lineage.

The OD traces for the IPTG- and IPTG+ replicates of the TnV2_placO1/pL evolution (Fig. 3.11C, third column) show a delayed but reproducible emergence of growth phenotype in the absence of IPTG. Given the requirement for the upregulation of *bglA* (Fig. 3.11F), we hypothesized that transposon insertion into the endogenous *lacI* gene would relieve repression on the outward facing placO1/pL promoter and establish quasi-constitutive gene activation at other insertion sites. As evidence for this multi-site insertion mechanism, comparing the differential enrichment of reads across the n=8 paired IPTG- and IPTG+ conditions identified a common, additional *lacI* insertion in the IPTG- context (Fig 3.11G). The artificial regulatory interaction between *lacI* and *bglA* introduced by the insertion of the placO1/pL promoter is therefore disrupted through a secondary knock-out of *lacI*. This example highlights two important advantages of our transposon-mediated evolution platform moving forwards: first, the possibility for multi-loci mutagenesis enables the investigation of gene-gene interactions (191), and second, the ability to incorporate endogenous regulatory sequences into the transposon establishes a mechanism to create new network connectivities *in situ* (94).

3.3.3 Discussion

In this work, we develop an engineered, autonomous transposon platform for continuous genome-wide mutagenesis and dynamic regulatory network re-wiring. To validate its generalizability and robustness, we then use this platform to study the impacts of transposon functionalization on the evolution of parallel *E. coli* populations towards diverse carbon source utilization and antibiotic resistance phenotypes. Through the implementation of barcode-based tracking and longitudinal next-generation sequencing, we are able to re-construct transposon lineages within the genomes of pooled host cells and investigate the impact of environmental complexity and genetic contingency on host-transposon interactions. Moving forwards, we envision this directed genome evolution platform being used to discover and optimize strains for biopharmaceutical applications, and as a well-defined testbed to study the role of MGEs in the emergence and re-wiring of complex natural gene regulatory networks. Throughout, we have ensured the compatibility of the NGS readout for this platform with current and future (near)-continuous, automated culture systems.

Expanding the throughput and timescale for the individual evolution experiments will enable a more statistical approach to measurements of evolvability, and facilitate the discovery of rare genetic innovations. Furthermore, long-term host-parasite dynamics could be readily tracked using fluorescence or sequencing-based readouts to improve our understanding of host adaptations to parasite burden and study intra-genomic competition between transposons.

3.3.4 Future directions

In this section, I highlight some exciting applications for the engineered, continuous transposon mobilisation platform. Importantly, these experiments could be readily performed without significant modifications to the existing system. As such, they represent potential immediate next steps.

Studying intra-genomic parasite dynamics

Computational models for intra-genomic competition between mobile genetic elements predicts the emergence of 'super-parasites' that compete for transposase resources in *trans* (173). These non-autonomous, minimal elements lack their own transposase genes, and their resulting compact size can allow them to propagate at the expense of the larger autonomous copies. However, this competition can lead to parasite population collapse as the copy number of the transposase genes decreases and becomes susceptible to random, mutation-driven inactivation.

The continuous transposon mobilisation platform presented here could be used to track the population-wide dynamics of intra-genomic compeition between labelled transposon populations (Fig. 3.12). For example, two transposon variants could be introduced into the same genome: one encoding a transposase gene in *cis*, the other acting as a super-parasite. These could be labelled with different fluorescent reporters (for cytometry-based readouts) and/or DNA barcodes (for NGS readouts), to allow for longitudinal tracking of the relative abundance of each variant. Cytometry-based fluorescence readouts in particular could enable assaying of transposon copy numbers at the single-cell level with high temporal frequency. Alternatively, two orthogonal transposon systems (e.g. Tn5 and *mariner*) could be initiated and inter-variant competition assays used to study how these parasites compete with one another, and the impact of this competition on their host. One final possible area of investigation, likely requiring longer-term laboratory evolution experiments, could be focused on observing the *de novo* emergence of super-parasitism itself in genomes harboring multiple copies of an autonomous transposon variant.

To achieve direct intra-genomic mapping of multi-insertion mutants, the implementation of single-cell workflows to identify the physical linkage between transposon copies would be necessary. One limitation of the current sequencing-based approach is our inability to distinguish whether insertion events are in the same genome, or in different positions within the genomes of two or more divergent but co-resident host lineages. Technological solutions to this problem

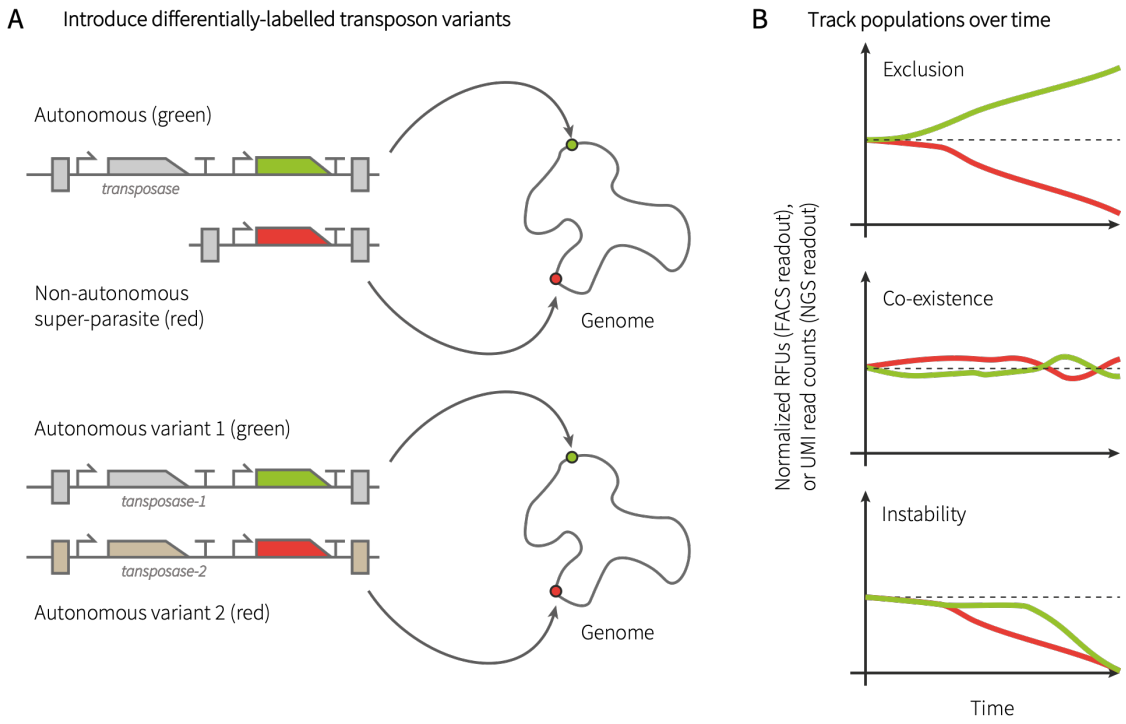

Figure 3.12: Differential transposon labelling would enable sub-population tracking in intra-genomic parasite competition assays. (A) We envisage studying two broad, hypothetical scenarios involving the parallel or sequential delivery of labelled transposons into the same genome. The first focuses on the concept of super-parasitism, in which a non-autonomous transposon population competes for the same transposase resource as the autonomous variants. Alternatively, two orthogonal transposon systems (e.g. Tn5 and *mariner* could be introduced, and compared to controls with two differentially-labelled copies of the same transposon. This would help identify factors that contribute to the relative long-term stability of mobile genetic elements.(B) Flow cytometry analysis could provide longitudinal single-cell measurements of relative fluorescence across the two channels, as a proxy for relative transposon copy numbers. This data could then be aggregated to provide population-averaged fluorescence measurements. Alternatively, relative proportions of the two transposon variants could be estimated from NGS data using unique molecular identifiers as proxies for raw counts. In this context, digital droplet PCR would also be applicable.

could stem from the recently-described droplet Tn-Seq (dTn-Seq) approach (192), but single-cell bacterial sequencing remains challenging (193, 194). Alternatively, dynamic, self-modifying barcode sequences would be necessary to introduce a time-variant record of the underlying parasite lineage as it propagates within a population (195–197).

Increasing throughput and complexity with automated culture

Natural evolutionary trajectories are influenced by the many dynamic environmental variables that define a specific niche. During laboratory evolution, the number of selection pressures is often limited to one or two variables, effectively reducing the phenotype space that can be explored (18, 198–200). This could also drive a reduction in gene network complexity in laboratory evolution (174). In contrast, even the simplest of natural niches experience fluctuating conditions across a range of variables. Observations from extant genomes suggest that environmental complexity could be a primary driving force for gene regulatory complexity (32, 198, 201). To address this need, automated culture systems can be designed with continuous evolution in mind. This gives experimenters the power to control multiple environmental conditions in parallel, and to adjust the stringency of selection in real time in response to near-continuous fitness measurements (17, 19). Furthermore, robotic liquid handling systems can maintain hundreds of cultures in 96-well plate format, enabling a more statistical approach to investigations into evolvability that focuses on the repeatability of individual trajectories. We propose adopting an experimental workflow that uses parallelized, feedback-controlled culture systems to perform transposon-mediated directed genome evolution, as this would allow for real-time measurements of growth rates and a provide a more tailored selection regimen.

Experimentally testing the predictive-dynamic framework

The predictive-dynamic framework proposes that in niches where two or more environmental variables or cues are statistically coupled through time, gene regulatory networks will evolve to encode this information in order to better exploit resources or prepare for stresses (Fig. 3.13)(198, 200). As an example, the up-regulation of genes required for respiration in the low oxygen environment of the human GI tract is observed in *E. coli* following a temperature increase: this has been proposed as an evolved response to the environmental correlation structure experienced when the bacteria enter a mammalian host (32).

While a series of anecdotal examples lend support to this theory, the spontaneous emergence of novel network linkages under analogous conditions has not been reproduced in the laboratory. A continuous, transposon-mediated genome re-wiring platform could support directed evolution experiments with dynamic environmental conditions to assess whether temporal correlations in these variables drive network re-wiring. These experiments would likely require the adoption of the robotic automated culture platforms described in a previous section, as it would rely on

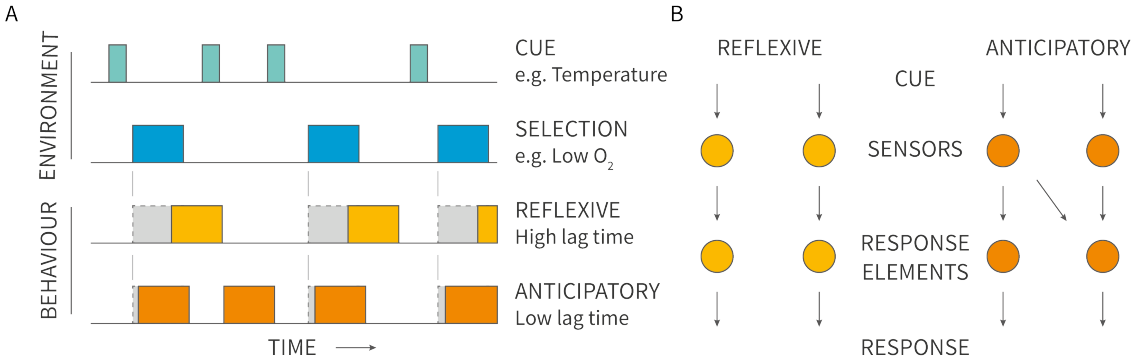

Figure 3.13: A predictive-dynamic framework for gene network evolution. (A) In natural environments, temporal correlations may emerge between different variables. If one such variable serves as a reliable cue for another, anticipatory regulation can evolve as an alternative to reflexive responses. (B) At the network level, this corresponds to the emergence of novel connections between pathways. Adapted from Freddolino *et al.*(201)

alternating growth conditions with well-defined, pre-programmed timings. To define the different parameter regimes under which this might occur, the following outstanding questions about the evolution of genetic co-regulation must be answered:

- Are coupled environmental selection pressures sufficient to drive the evolution of anticipatory gene regulation patterns? If so, under what range of inter-variable correlations does this robustly occur?
- Across selection pressures, can the opportunity cost of a reflexive regulatory strategy by quantified? How does this vary with selection stringency?
- What effect does varying the environmental dynamics (e.g. no selection pressure, or uncorrelated selection pressures) have on the observed evolutionary trajectories?
- Are anticipatory regulatory strategies generally lost as strains adapt to constant, non-selective conditions?

Author contributions

MAE and JJC conceived and planned the experiments. MAE performed all experiments related to the CRISPR-guided transposon system, and analyzed the associated data. MAE developed the self-propagating *mariner* transposon platform and performed all of the experiments with contributions from MAA. MAE analyzed the associated experimental data, except for the Tn-Seq and RB-Tn-Seq data which was analyzed by MAA. MAA developed the bioinformatics pipeline for the Tn-Seq and RB-Tn-Seq data data analysis with input from MAE. MAE wrote the manuscript (unpublished at the time of writing) with critical feedback from MAA and JJC.

4 | Discussion

In this thesis, I have presented two distinct areas of research: the development of smart material systems that respond to nucleic acid cues in their environment (Chapter 2), and the engineering of transposon-based tools for synthetic biology and directed evolution (Chapter 3). Fundamentally, these two separate aims are linked through their re-purposing of the natural diversity and functionality of mobile genetic elements to create tools for biological engineering. The programmability and sensitivity of the CRISPR-Cas enzymes we harnessed to control large-scale material properties for cargo release and diagnostics highlights the impressive biochemical capabilities that can be discovered within MGE lineages (25–27). Conversely, we have shown that synthetic biology approaches can be used to 'domesticate' natural transposon systems, and with that their intrinsic capability for dynamic network re-wiring and genome engineering. This represents an important advance in ongoing efforts to better integrate synthetic genetic elements into complex natural systems (6, 202). The focus here is on perturbing the endogenous functions of the host to steer them towards a useful phenotype using smaller synthetic controllers, adapters and reporters to modulate, re-wire, and measure natural networks (203, 204). Importantly, the regulatory and biochemical heavy-lifting are performed by the host, rather than large, heterologously-expressed gene circuits. These network-level engineering approaches mediated by mobile genetic elements are likely to become more important as we engineer complex living systems for bioproduction (4, 5) and therapeutics (34, 205).

As we develop new MGE-derived tools to engineer and evolve cellular networks, we also create a laboratory platform to study their analogous natural processes (Fig. 4.1). As such, we not only want to obtain a desired cellular phenotype, but also track the underlying genomic and population-level changes that establish this phenotype and compare these to existing evolutionary models (152). To this end, I have presented our efforts to develop next-generation sequencing assays to track barcoded populations of transposons as they propagate and move through the genomes of an evolving cellular population. By expanding both the number of parallel replicates and the complexity of the environmental space that we explore through automation, our hope is that our transposon-mediated directed genome evolution platform could be used to generate 'proofs-by-construction' for proposed mathematical models of host-parasite interactions and anecdotal observations of predictive behaviours in gene regulation (32, 200, 206). Ultimately, through long-term evolution experiments across unpredictable environments, this platform could be used to observe and study the expansion of gene regulatory networks and answer fundamental questions on how this increase in complexity occurs (104).

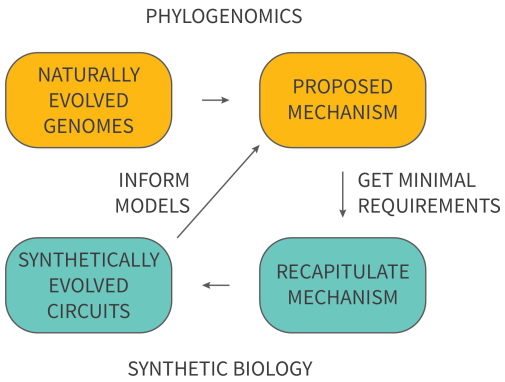

Figure 4.1: A general approach to studying complex evolutionary processes with synthetic systems. An ever-increasing abundance of metagenomic and genomic sequencing data across diverse organisms has empowered well-defined evolutionary models. For example, by reconstructing phylogenies for families of proteins or protein sub-domains, it is possible to identify duplication events that support sub- or neo-functionalization. To study these processes over reasonable timescales in the laboratory and with sufficient experimenter control, semi-synthetic analogues could be established as a model system.

5 | Methods

5.1 CRISPR-responsive materials

For a detailed explanation of the methods used to create CRISPR-responsive materials, please see the full protocol in Gayet et al. (10) and the methods of the original publication (9).

5.2 CRISPR-associated transposon systems

Strains and media

All cloning protocols and reporter expression experiments were performed using Miller LB broth (Fisher), with 15 g/L Bacto agar (BD) added for solid media. Unless otherwise stated, we used kanamycin (50 μg/ml), carbenicillin (100 μg/ml), and chloramphenicol (50 μg/ml) for antibiotic selections. Cloning was performed in using NEB Stable competent cells (pir^-), and One Shot PIR1 Chemically Competent *E. coli* (pir^+). The inducers IPTG, Isopropyl-β-D-1-thiogalactopyranoside (IPTG) and anydrotetracycline (ATc) were added to the growth media from stocks stored at -20°C, at the concentrations described in the text. Arabinose was supplemented into the growth media at a final concentration of 0.2% w/v. Note that on the pHelper shCAST plasmid, expression of the transposase and Cas12k is driven by a *lac* promoter (plac). However, the One Shot PIR1 Chemically Competent cells in which the majority of the experiments are performed have the genomic Δlac169 mutation, which means they lack a functioning *lacI* gene and so expression from plac is effectively constitutive in this background.

Plasmids and cloning

The shCAST plasmids from Strecker et al. (29), corresponding to Addgene plasmids #127924, #127921, and #127926 were a gift from Feng Zhang. The T7 RNAP casette was derived from pCS6, which was a gift from Matthew Bennett (Addgene plasmid #55752). The modified pDonor and pTarget plasmids used in this work were created using a combination of HiFI DNA assembly, KLD, and restriction-ligation cloning. All pTarget and pHelper plasmids were cloned in NEB Stable Competent *E. coli*, whereas the pDonor plasmuds with an R6K origin of replication were cloned in One Shot PIR1 Chemically Competent *E. coli* (ThermoFisher). Correct assemblies were verified using colony PCR and Sanger sequencing (Quintara Biosciences, Cambridge, MA).

The gRNA sequences used in this work (pTarget PSP1, randomized PSP1, *lacZ* sgRNA-3, and *lacZ* sgRNA-4) are described in (29), and were inserted into the pHelper backbones by Golden Gate assembly. Briefly, the two oligos encoding the spacer were first phosphorylated and annealed by combining 1μL Oligo 1 (100μM), 1μL Oligo 2 (100μM), 1μL 10X T4 ligase buffer (NEB),

0.5μL T4 PNK (NEB), and nuclease free water to a total volume of 10μL in a PCR tube. The components were mixed and annealing was performed in a thermal cycler with the following conditions: 37°C for 1hr, 95°C for 5 min, ramp to 25°C at 5°C/min. Following annealing, the 10μL reaction was diluted by adding 90μL nuclease free water. We then set up the golden gate assembly reaction by combining 1μL annealed oligo, 1μL pHelper plasmid (25ng/μl), 1μL T4 DNA Ligase buffer (NEB), 0.5μL T4 DNA Ligase (NEB), 0.5μL BspQI (10,000 U/ml), and then nuclease free water to 10μL. The reaction mix was then incubated according to the following program on a thermal cycler: 30 cycles of digestion and ligation (37°C for 5 min, 16°C for 5 min), followed by a final digestion step (60°C for 10 min). For each assembly, 3μL was transformed into NEB Stable competent cells, and purified plasmid was obtained from 6mL of overnight culture using a QIAprep Spin Miniprep Kit (Qiagen).

Transformations and fluorescent reporter assays

For the experiments performed in Fig. 3.2 - 3.4, purified plasmid stocks were diluted to a working concentration of 5 ng/μL. Then, 1μL each of the pTarget, pDonor and pHelper plasmids was co-transformed into chemically competent One Shot PIR1 cells. Following heat shock, the transformations were rescued in SOC at 30°C for 2 hrs, before plating on triple-antibiotic selection media and growing overnight at 30°C. Then, individual colonies were inoculated into the wells of a 96-well plate conatianing 300μL LB and antibiotics. These cultures were grown to saturation overnight at 30°C, diluted 1:100 into fresh media, and then grown in the presence of the inducer(s) for 8-12h at 37°C. Fluorescence and OD measurements of 100μL of culture were taken using a CLARIOstar microplate reader (BMG Labtech).

For the experiment described in Fig. 3.5, the pTarget and pHelper plasmids were first transformed into BL21(DE3) competent cells (NEB). From these successful transformants, we then made fresh chemically competent BL21(DE3) cells. Finally, we transformed 1μL of the pDonor plasmid (100ng/μL) into these cells, and plated on LB-agar with 1mM IPTG, and triple antibiotic selection. IPTG induction of the CAST components on the pHelper plasmid is necessary because BL21(DE3) cells have an intact copy of the *lacI* gene. Furthermore, because BL21(DE3) is a *pir*$^-$ strain, there is stringent selection for successful transposition events. Three colonies from the plate were picked, and inolculated into 6mL LB media (Fig. 3.5B), or 300μL LB media in a 96-well plate (Fig. 3.5C). These cultures were grown overnight at 30°C, and then transferred them to fresh media at a dilution of 1:100 with the corresponding concentration of inducers. After 8-12h growth at 30°C, fluorescence and OD measurements of 100μL culture were taken using a CLARIOstar microplate reader (BMG Labtech).

For the experiment described in Fig. 3.6, pDonor plasmids (100ng) were transformed into BL21-AI One Shot Chemically Competent *E. coli* (ThermoFisher) harboring either the pre-transformed pHelper plasmids with sgRNA-3, or sgRNA-4. As described in (29), these gRNA target the

endogenous *lacZ* gene. IPTG induction of the CAST components on the pHelper plasmid is necessary because BL21-AI cells have an intact copy of the *lacI* gene. Furthermore, because BL21-AI is a *pir*$^{-}$ strain, there is stringent selection for successful transposition events. After rescuing the transformations at 30°C for 2 hrs, we plated the cells on LB-agar with dual antibiotic selection and X-gal/IPTG for blue-white colony screening and pHelper induction. We then inoculated colonies into 300μL LB media with antibiotics in 96-well plates, and grew the cells overnight at 30°C. Once the cultures had reached saturation, we then transferred them to fresh media with 400μM IPTG, at a dilution of 1:100. After 8 hrs growth at 30°C, fluorescence and OD measurements of 100μL culture were taken using a CLARIOstar microplate reader (BMG Labtech).

5.3 Directed genome evolution with *mariner* transposons

Strains and routine plasmid construction

The underlying genetic part sequences used to construct the plasmids in this study were obtained from the following plasmids: the TetR-regulated himar1C9 was from pHdCas9, a gift from Harris Wang (Addgene plasmid # 137080); the AraC/pBAD-regulated himar1C9 was form pSAM_AraC, a gift from Harry Mobley (Addgene plasmid # 91569); the mariner transposon flanked with MmeI-modified inverted repeats was derived from pSAM_Ec, a gift from Matthew Mulvey (Addgene plasmid # 102939).

To construct the pHelper plasmids, we used NEB Turbo *E. coli* competent cells. pDonor plasmids harboring an R6K y origin of replication were assembled in Invitrogen OneShot PIR1 and PIR2 competent cells. All plasmids were constructed with a combination of NEB HiFi assembly protocols and restriction-ligation cloning, except those described in Fig. 3.11. Following transformation of the chemically competent cloning strains, clones were isolated on LB agar containing either 50ug/ml chloramphenicol, 50ug/ml kanamycin or 100 ug/ml carbenicillin. Diagnostic colony PCR reactions were performed in OneTaq HotStart 2x Master Mix (NEB) with sample loading dye, and 2-5 positive colonies were identified by resolving the amplicons on 1% agarose gels with SYBR Safe (brand). The clones were grown overnight in 6 ml LB with the corresponding antibiotic selection, and plasmid DNA was extracted with a GenCatchTM Plasmid DNA Mini-Prep Kit (Epoch Life Sciences). Finally, each plasmid prep was sequenced at ∼1x coverage to verify the insert and backbone sequences (Sanger Sequencing, Quintara Bio).

Transposon-mediated mutagenesis and evolution experiments were performed in two *E. coli* K-12 strains: MG1655, and its genome-reduced derivative MDS42 (121).

Growth media and inductions

Standard cloning procedures and non-selective outgrowths were performed using Miller LB broth (Fisher), with 15 g/L Bacto agar (BD) added for solid media. Unless otherwise stated, we used kanamycin (50 ug/ml), carbenicillin (100 ug/ml), and chloramphenicol (20 ug/ml) for antibiotic selections.

For the carbon source utilization experiments described in Fig. 3.10, we used M9 minimal media supplemented with different carbon sources. To make 1L of the base media, we combined 200 mL Difco 5x M9 minimal salts solution (BD), 1mL $MgSO_4$ (1M), 0.3mL $CaCl_2$ (1M), and 10mL of a trace elements stock solution (100x). To make the 100x trace elements stock solution, we combined the following (in order) to 800ml of milliQ water: 5g EDTA, 498mg $FeCl_3$ (anhydrous), 84mg ZnCl2, 765ul $CuCl_2$-$2H_2O$ (0.1M), 210ul $CoCl_2$-$6H_2O$ (0.2M), 1.6ml H_3BO_3 (0.1M), and 8.1ul $MnCl_2$-$4H_2O$ (1M) before bringing the total volume to 1L and sterilizing over a 0.22um filter. NaOH was used to adjust the pH to 7.5 immediately after the EDTA was added.

For the Biolog experiments in Fig. 3.10, we added milliQ water (EMD Millipore) to this solution and brought the volume to 1L before sterlizing over a 0.22um filter. For the downstream experiments with either L-serine (6-100g/L), glycyl-L-glutamic acid (12g/L), or beta-methyl-D-glucoside (4g/L), we incorporated the solid carbon source to the base media before bringing the volume to 1L and sterilizing over a 0.22um filter. To make M9-glucose, we added 1mL biotin (1mg/mL), 1mL thiamin (1mg/ml), and 20mL glucose (20% w/v stock) to the base media before bringing the volume to 1L and sterilizing over a 0.22um filter.

For the D-cycloserine and arbutin growth experiments in Fig. 3.8 & Fig. 3.9, we used the MT mineral salts medium described by Hall (179). This consists of 423mg sodium citrate, 100mg $MgSO_4 \cdot 7H_2O$, 1.0g $(NH_4)_2SO_4$, 540mg $FeCl_3$, 1.0mg thiamine, 3.0g KH_2PO_4, and 7.0g K_2HPO_4, with either glucose (2.0g/L) or arbutin (1.0g/L) as carbon sources. For the arabinose induction experiments, we combined arabinose (1.0g/l) and glycerol (2.0g/L). In the arbutin growth experiments, we used glucose (1.0g/L) and glycerol (2.0g/L) as the carbon source.

Unless otherwise stated, the transcriptional inducers used in this study were added to growth media at the following concentrations: anhydrotetracycline (ATc); 50ng/ml, arabinose; 0.1% w/v, isopropyl β-D-1-thiogalactopyranoside (IPTG); 200uM, and crystal violet (CV); 0.25uM. The ATc and IPTG stock solutions were stored frozen at -20°C.

Transformation and D-cycloserine assays

For the arabinose-inducible transposase system, *E. coli* strains harbouring an pBAD-controlled *himar1C9* transposase on a medium-copy p15A plasmid were transformed with $3x10^9$ pDonor plasmid molecules (∼10ng), and rescued for 1 hour at 37°C in either SOC (uninduced), or SOB+0.2% arabinose (induced). We compared to transposon constructs: an unmodified donor

containing a neoR/kanR cassette (labelled Tn in the text), and a larger variant containing kanR and mNeonGreen expression cassettes (labelled Tn[GFP] in the text). 450uL from each transformation was then plated on two replicate plates. After growth at 37°C overnight, colonies were counted. For the ATc-inducible system, E. coli strains harbouring a TetR/pTet controlled *himar1C9* transposase on a medium-copy p15A plasmid were transformed with $3x10^9$ plasmid molecules (∼10ng), and rescued for 1 hour at 37°C in either SOC, or SOC + 100ng/ml ATc. 450uL from each transformation was then plated on two replicate plates. After growth at 37°C overnight, colonies were counted.

The D-cycloserine growth assay presented in Fig. 3.8 was performed by transforming MDS42 harboring a TetR/pTet controlled himar1C9 transposase on a pHelper plasmid with the two transposon variants described above: Tn and Tn[GFP]. Individual colonies were picked and inoculated into a deep-well 96-well plate with 300uL TB (plus antibiotics) per well (n=24 per variant, per duplicate transformations). We also included two controls: the parental MDS42 strain (n=48), and MDS42 (n=24) with the pHelper plasmid. After an overnight incubation at 37°C, the TB plate was passaged into two parallel plates containing 300uL MT minimal salts media with 0.1% glucose and 0.2% glycerol per well, at a dilution of 1:100. One MT-glucose/glycerol plate was supplemented with ATc (50ng/ml). The cultures were grown for 8 hours at 37°C, and then passaged into a final pair of selection plates containing 400uL MT-glucose/glycerol media and 20uM D-cycloserine, with or without ATc. Importantly, each culture was diluted 10,000-fold into the selection media at this step to create a population bottleneck, and selection was maintained with kanamycin (25ug/ml) and chloramphenicol (25ug/ml).

Arbutin growth experiments

In a preliminary experiment (see accompanying publication), we transformed both *E. coli* MG1655 and MDS42 cells harboring either a tetR or araC regulated *himar1C9* transposase helper plasmid (pHelper) with a donor plasmid harboring the Tn-pOUT construct. Individual colonies from the transformations were picked alongside the respective pHelper controls and the parental strain, and inoculated into 300uL of terrific broth (TB, Difco) in individual wells of a deep-well 96 well plate. Throughout this experiment, we used both chloramphenicol and kanamycin at 25ug/mL. After growth overnight at 37°C with shaking at 900 rpm, we passaged the wells into 300uL non-selective MT minimal salts media at a dilution of 1:100. For the variants with the tetR-regulated transposase, the non-selective MT media contained glucose (1.0 g/L) and glycerol (2.0 g/L) as the carbon source, and duplicate plates were incubated for ∼8 hours either with or without 50ng/uL ATc (i.e. induced or uninduced conditions). For the variants with the araC-regulated transposase, the duplicate plates contained either MT media with glucose (1.0g/L) and glycerol (2.0g/L) as the uninduced state, or arabinose (1.0g/l) and glycerol (2.0g/L) as the induced state. Similarly, these duplicate plates were incubated for ∼8 hours at 37°C with shaking at 900 rpm. After passaging, the wells of the TB overnight plate were supplemented with 300uL

of 50% glycerol and frozen at -80°C. Following treatment with or without inducer, the cultures were passaged at a ratio of 1:2000 into 400uL fresh MT media containing arbutin (1.0g/L) by first diluting each well 10-fold in PBS. These cultures were incubated for 5 days at 37°C, and high-growth phenotypes were documented based on the characteristic color change of the media. For the arabinose-inducible system, the MT-arbutin media was incompatible with sustained induction. However, for the ATc-inducible system we maintained 50ng/mL throughout.

Focusing specifically on the ATc-inducible tetR variant of the pHelper, we repeated this experiment by re-inoculating fresh TB plates from the original glycerol stock plates and grew them overnight at 37°C. We then passaged the wells into duplicate plates of 300uL MT media with glucose (1.0 g/L) and glycerol (2.0 g/L) at a ratio of 1:100, either with or without inducer, and incubated them at 37°C for ∼8 hours. We then passaged 3uL per well from both the induced and uninduced plates into two intermediate conditions: 300uL total of MT-arbutin and MT-glucose/glycerol mixed at ratios of either 3:1 or 19:1. After a 24 hour incubation, these four plates were then finally passaged into 100% MT-arbutin media at a ratio of 1:100, and incubated for 5 days at 37°C. On the second day, 200uL of fresh media was added to each well. On the fifth day, endpoint OD600 measurements were taken with a CLARIOstar Plus plate reader (BMG Labtech).

The experiment presented in Fig. 3.9 was performed in a similar way, starting with the inoculation of single colonies into the wells of a 96-well deep well plate and then passaging the cultures through the following conditions (dilution factors for each passage are provided in brackets): 8 hours in 400uL TB, 8 hours in 400uL MT-glucose/glycerol (1:100 dilution), 24 hours in MT-arbutin:MT-glucose/glycerol mixed at a 1:1 ratio (1:100 dilution), 24 hours in MT-arbutin:MT-glucose/glycerol mixed at a 3:1 ratio (1:100 dilution), 24 hours in MT-arbutin:MT-glucose/glycerol mixed at a 19:1 ratio (1:100 dilution), 48 hours in 100% MT-arbutin (1:100 dilution). For each of the 24 hour incubations, endpoint OD600 readings were taken using a SpectraMax M5 plate reader. During the final selection in MT-arbutin, intermediate OD600 readings were taken by removing 100uL of culture per well and replacing the volume with 100uL of appropriate fresh media. A final endpoint OD600 reading was taken at 48 hours.

The growth curves shown in Fig. 3.9 were obtained by inoculating 100uL of MT-arbutin media (containing 50ng/ml ATc) with 1uL of the appropriate ATc+ culture following the 24 hour incubation on MT-arbutin:MT-glucose/glycerol mixed at a 19:1 ratio. In all the experiments described in this section, parental MG1655 and MDS42 strains were maintained without antibiotic. pHelper strains expressing the transposase only were maintained with chloramphenicol. Transposon insertion strains were maintained with both chloramphenicol and kanamycin.

Biolog carbon source assays

Chemically competent MDS42 cells (50uL) containing a pHelper plasmid with a chloramphenicol resistance marker were transformed in triplicate with 20ng of the pDonor plasmids Tn and Tn-pOUT, and plated on LB agar with dual chloramphenicol (20ug/ml) and kanamycin (50 ug/ml) selection. One colony from each of the triplicate plates was inoculated into 6ml of M9-glucose minimal media with ATc (50ng/ml), chloramphenicol (20ug/ml) and kanamycin (50ug/ml), alongside three control cultures inoculated from untransformed MDS42-pHelper cells with ATc (50ng/ml) and chloramphenicol (20ug/ml). After 16 hours of growth, the nine cultures were passaged into identical, fresh media at a dilution of 1:100 and grown for a further 8 hours. The cultures were diluted to a concentration of 10^6 cfu/ml, and then 1uL was aliquoted into the wells of three Biolog EcoPlates such that each plate contained 32 wells corresponding to MDS42 pHelper, 32 wells corresponding to MDS42 pHelper Tn, and 32 wells corresponding to MDS42 pHelper Tn-pOUT spanning each unique carbon source respectively. The wells of the Biolog EcoPlates contained 100uL of M9 salts media (without carbon source), supplemented with 50 ng/ml ATc. To reduce the potential burden imposed by antibiotic selection in this preliminary screen, we used chloramphenicol (10ug/ml) and kanamycin (25 ug/ml) to maintain selection for the transposase and transposon, respectively.

The three replicate plates were incubated at 37°C with shaking at 900 rpm and 90% humidity. OD600 measurements were taken every 6-12 hours using a SpectraMax M5 plate reader. After 24 hours of growth, 1uL of culture was taken from each well and used to inoculate fresh triplicate plates at a dilution of 1:100. This process was repeated every 24 hours for a total of 4 passages, with OD600 measurements continuing in parallel for both freshly inoculated and ancestral plates. A glycerol stock of the single MDS42 pHelper Tn-pOUT replicate (EVOL-1) that showed detectable growth on L-serine was stored at -80°C.

L-serine growth experiments

A glycerol stock of EVOL-1 was used to streak out colonies on LB-agar with chloramphenicol (20ug/ml) and kanamycin (50 ug/ml) selection. For comparison, we also grew out stocks of MDS42 pHelper, MG1655 pHelper, MDS42 and MG1655. Eight colonies were picked from each of the five strains (EVOL-1 plus four controls), and grown up overnight in M9-glucose media a deep well 96-well plate at 37°C. These overnight cultures were then each inoculated 1:300 into two different parallel sets of conditions: 300uL M9-glucose with either 24, 50, 75 or 100 g/L L-serine Fig. 3.10C, and 300uL M9 salts only with 6, 12, 24 or 50 g/L L-serine Fig. 3.10B. These cultures were grown for 24 hours, with 100 uL aliquots removed for OD600 measurements at 13hrs and 23hrs using a CLARIOstar Plus plate reader (BMG Labtech).

To measure the growth rates of the EVOL-1 strain across different L-serine concentrations, we inoculated 1uL of the overnight cultures into 100uL of M9-glucose with either 24, 50, 75 or

100g/L L-serine, and compared this to control cultures of MDS42 pHelper (n=8). We incubated the plate at 37°C in a CLARIOstar Plus plate reader (BMG Labtech), with orbital shaking at 700 rpm and OD measurements every 10 minutes. The growth curve data for each of the eight replicates across the four concentrations was then analyzed using the Curveball python package (https://github.com/yoavram/curveball). The models.fit_model function fits six related logistic growth models to the data and then selects the best model by comparing the Bayesian Information Criteria for each (207). The confidence intervals for the growth rate parameters were obtained from 15 bootstrapped replicates of the best fit model.

To assess the reproducibility of this adaptation process, we repeated the transformation process for the original carbon source screen: chemically competent MDS42 cells (50uL) containing a pHelper plasmid with a chloramphenicol resistance marker were transformed with 20ng of the pDonor plasmids Tn and Tn-pOUT, and plated on LB agar with dual chloramphenicol (20ug/ml) and kanamycin (50ug/ml) selection. To control for any effects of the antibiotic selection, we performed the same transformation with chemically competent MDS42 cells (50uL) containing a pHelper plasmid with an ampicillin/carbenicillin resistance marker. Sixteen colonies from each of the six MDS42 strains – Tn and Tn-pOUT with either pHelper(chlor) or pHelper(carb), and the parental controls pHelper(chlor) and pHelper(carb) – were inoculated into 300 uL M9-glucose minimal media with 50ng/ml ATc in a deep-well 96 well plate and incubated at 37°C for 10 hours. These cultures were then passaged into 300 uL of identical fresh media at a dilution of 1:300 and grown overnight. After this second passage in M9-glucose minimal media, 3uL of the overnight cultures were inoculated into two parallel plates of M9 salts media with 300uL of 12g/L L-serine per well either with or without 50ng/ml ATc. Appropriate antibiotic selection was maintained throughout with chloramphenicol (20ug/ml) or carbenicillin (100ug/ml), and kanamycin (50ug/ml) for strains containing the transposon. At 32 hour intervals, each plate was passaged into an identical one containing 300uL per well of fresh media at a dilution of 1:100, with a final OD600 measurement taken after 50 hrs of growth in the third passage.

For each of the two strains with detectable growth at this endpoint [MDS42 Tn-pOUT pHelper(chlor) and MDS42 Tn-pOUT pHelper(carb)], we stored four separate replicates for Tn-Seq and paired these with the corresponding 'start point' cultures from the first round of non-selective growth. We then pooled all the replicates for each strain at the start point (six pools), and all the replicates in the two growth-positive strains at the endpoint (two pools). This total of 24 samples (8 independent populations at matched start and endpoints, plus 6 start point pools and 2 endpoint pools) were prepped for Tn-Seq.

Tn-Seq sample preparation and sequencing

To generate Tn-seq data from frozen cell pellets, we used a modified version of the Tn-seq protocol described by Palani (208). First, we purified genomic DNA (gDNA) from cell pellets (fresh, or

flash-frozen and stored at -80°C) using a Qiagen DNeasy Blood & Tissue Kit following the manufacturers instructions. We included an RNase A treatment, and eluted samples in 100uL AE buffer. We then fragmented 100ng gDNA using an NEBNext Ultra II FS DNA Library Prep with Sample Purification Beads, again manufacturers instructions. As described by Palani, at the ligation step we replaced the standard NEB adapter with 2.5uL of a custom, pre-annealed Nextera-compatibe adapter (sENG-020 and sENG-021, 1uM) containing an 8N unique molecular identifier (UMI). We then performed a 0.8X SPRI selection, eluting in 17uL of 0.1x TE buffer and transferring 15uL of this to a new tube. We set up 50uL enrichment PCR reactions with 15uL of adapter-ligated gDNA template, 5uL each of the forward and reverse primer (sENG-022 and sENG-023), and 25uL of NEBNext Ultra II Q5 Master Mix. Using a T100 Thermal Cycler (Bio-Rad), the samples were incubated as follows (heated lid on): 98°C 30s, (98°C 10s, 65°C 1 min 15s) x 12 cycles, 65°C 5 mins, 4°C hold. We then performed a 0.9X SPRI selection, eluting in 17uL of 0.1x TE buffer and transferring 15uL of this to a new tube. We measured the concentration of DNA in a 1uL aliquot of each sample using a Qubit3 fluorimeter (ThermoFisher), and set up indexing PCR reactions with 10ng of template (made up to 7.5uL), 2.5uL each of a unique pair of i5 and i7 primers from a Nextera XT Index Kit (Illumina), and 12.5uL NEBNext Ultra II Q5 Master Mix. The samples were incubated as follows (heated lid on): 98°C 30s, (98°C 10s, 65°C 1 min 30s) x 10 cycles, 65°C 5 mins, 4°C hold. We then performed a final 1.2X SPRI selection on each sample, eluting in 30uL 0.1x TE buffer. We used 2% and 4% TAE-agarose gels to check the size distributions of the amplified products relative to a low molecular weight DNA ladder (NEB), and verified the absence of primer-dimer bands (∼128bp). After measuring the concentrations of each sample, we then pooled them by equal mass and submitted them for sequencing by Quintara Biosciences (Cambridge, MA). Samples were run in paired-end mode on a MiSeq (2x150 cycles, 1M reads) or HiSeqX (2x150 cycles, 300M reads) with a 15% PhiX spike-in.

Tn-seq analysis

Raw paired-end sequencing reads were first trimmed, quality filtered, and tagged for UMIs using fastp (209) v0.23.2 with parameters "-U –umi_loc=read2 –umi_len=9 –trim_front1=27 –trim_front2=17". To locate transposon insertion sites, we mapped filtered reads with Bowtie2 (210) v2.4.4 against reference genomes for *E. coli* MG1655 (U00096.3) or MDS42 (AP012306.1). In brief, we created bowtie indices from each reference genome using the 'bowtie2-build" function with default parameters. We then mapped paired-end reads to a reference genome using the "bowtie2" function with parameters "–sensitive-local –maxins 1000 –no-mixed –no-discordant –no-unal". We chose the maximum insertion parameter based on the size distribution of sequences observed on a 2% agarose gel prior to sequencing. Additionally, only read pairs in which both reads concordantly aligned to the reference genome were considered in subsequent analyses. We deduplicated mapped read pairs based on both their UMI and mapping coordinate using the

UMI-tools (211) v1.1.0 "dedup" function with parameters "–umi-separator=: –paired".

To locate genomic sites enriched with transposon insertions, we used the Model-based Analysis of Chip-Seq (MACS) v3.0.0a7 peak calling algorithm (212). Specifically, we identified genomic sites with significantly higher read coverage compared to background, within samples, using the "macs3 callpeaks" function with parameters "-f BAMPE -g 5e6 -B -q 0.01". Peaks were associated with their nearest genes using the BEDTools (213) v2.30.0 "closest" function. We plotted peaks and their corresponding fold-change over background in Python v3.6.13 using Matplotlib v3.2. The most highly enriched peaks were labeled with their nearest gene. If two genes were equally proximal to a peak, only one gene name was used as the label. Lastly, we calculated the distribution of transposon insertion events by counting the number of times forward and reverse reads mapped to each position in the genome. Mapping counts were normalised by the total number of mapped reads.

Statistical comparisons of transposon enrichment pre- and post-evolutions were conducted using Bio-Tradis (181). As input to Bio-Tradis we used alignments produced by the SMALT v0.7.6 read mapper (ref:https://www.sanger.ac.uk/tool/smalt-0/) as this was the only short read mapper that did not result in substantial amounts of soft-clipping, which interferes with the Bio-Tradis pipeline. We performed alignments using the "smalt map" function with parameters "-x -y 0.96 -r -1" and deduplicated alignments as described above. Since the native Bio-Tradis pipeline only accepts FASTQ files as the initial input, we modified the existing scripts to accept deduplicated BAM files in order to mitigate statistical artefacts created by PCR amplification biases. The resultant insertion site files were then annotated with genomic features using the "tradis_gene_insert_sites" function with default settings. Transposon enrichment pre- and post-evolution was then compared using the "tradis_comparison.R" function with parameters -f -t 128, which produced fold-changes and adjusted p-values for each gene above the detection threshold. We produced volcano plots using the EnhancedVolcano (214) v.1.4.0 package in R v3.6.1.

Magic pools platform

To facilitate the rapid golden gate assembly of transposon donor plasmids, we modified the Magic Pools pipeline descrbed by Liu et al. (169). A schematic of the overall design is provided in Fig. 3.11. In short, the donor plasmid was split into the following functional components: Part 1, a promoter to be used for the expression of the cargo; Part 2, cargo gene(s) for expression from within the transposon; Part 3, a pre-assembled antibiotic resistance cassette (including promoter, ORF and terminator(s)); Part 4, one end of the transposon including the FseI-SbfI sites for barcode insertion and the terminal repeat, as well as the backbone components for plasmid cloning (R6K origin of replication, oriT sequence, and ampicillin resistance cassette); Part 5, one end of the transposon with BsmBI sites for pOUT promoter insertion, as well as

insulating terminator sequences. A portion of Part 4 (encoding the mariner transposon end) and all of Part 5 were first ordered as g-Blocks (IDT). The linear Part4 g-Block fragment was combined with a linear section of plasmid backbone components using NEBuilder HiFi DNA Assembly Master Mix (NEB). Flanking BbsI cut sites for each part were installed by PCR using Q5 Hot Start High-Fidelity 2X Master Mix (NEB), and the resulting products verified via agarose gel and column purified. Each part was stored as a linear DNA amplicon at -20°C.

We began by assembling pMP-001. Part 4 (50 fmol) and all other parts (100 fmol) were combined in a 20uL golden gate assembly reaction, containing 2uL T4 DNA ligase buffer (NEB), 1uL T4 DNA ligase (NEB), and 1uL BbsI-HF (NEB). The samples were incubated for 60 cycles x (37°C,5 mins -> 16°C, 5 mins), and then 3uL was transformed into One Shot PIR1 competent cells (Invitrogen). Correct assembly of the transposon donor plasmid was verified via colony PCR and sanger sequencing. To assembly the pOUT constructs 'pMP-001_x', distinct promoter regions were then inserted into the Part 5 segment of pMP-001. We first PCR-amplified the promoter regions of interest using Q5 Hot Start High-Fidelity 2X Master Mix (NEB) to insert orientation-specific BsmBI cut sites, and the resulting products were verified via agarose gel and column purified. Each part was stored as a linear DNA amplicon at -20°C. Golden gate assembly reactions for pOUT promoter insertion were prepared by combining the plasmid backbone (50fmol), the promoter amplicon (100fmol), water (to a total reaction volume of 10uL), T4 DNA ligase buffer (1uL, NEB), T4 DNA ligase (0.5uL , NEB), and (0.5uL BsmBI-v2-HF, NEB). The samples were incubated for 60 cycles x (42°C, 5 min -> 16°C, 5 min) -> 60°C for 5 min, and then 3uL from each reaction was transformed into One Shot PIR1 competent cells (Invitrogen). Correct promoter insertions were verified by sanger sequencing.

To create N20 barcodes for each transposon variant, we adapted the protocol described by Wetmore et al.(171). In short, the barcode section is ordered as a ssDNA oligo, and then two flanking PCR primers were used to create a double stranded barcode amplicon with flanking SbfI-FseI cut sites for insertion into the transposon backbone. To identify the optimal template concentration, we set the flanking primer concentration to 0.5uM and varied the barcode template input concentration (0.2, 0.1, 0.02, or 0.01uM). We then performed PCR reactions with Q5 Hot Start High-Fidelity 2X Master Mix (NEB), setting the annealing temperature to 64°C and elongation time to 60 seconds for a total of 6 cycles to preserve barcode diversity. We found that a starting concentration of 0.02uM template maximized the barcode input while minimizing the formation of undesired products (e.g. unamplified template, primer dimers, or amplicon oligomers) as determined by agarose gel. Following this optimization, we repeated three 50uL PCR reactions in parallel and gel purified the product bands using a Monarch DNA Gel Extraction Kit (NEB) to create a pool of barcode amplicons. Separately, we set up restriction digest reactions for the different transposon donor plasmids and the barcode amplicon in 50uL reactions as follows: 2ug DNA, 5uL CutSmart Buffer (NEB), 1uL SbfI-HF (NEB) and 1uL FseI (NEB). After incubating the samples for 2 hours at 37°C, we gel-purified the products using a Monarch DNA Gel Extrac-

tion Kit (NEB). Finally, we performed separate ligation reactions to insert barcodes from the same digested pool into the different donor backbones, combining template (0.1 pmol), barcode (0.6 pmol), 2uL T4 DNA ligase buffer (NEB) and 1uL T4 DNA ligase (NEB) in 20uL reactions that were then incubated at 16°C overnight.

To check that the barcoding process was successful, we transformed 5uL of the pMP-001-BC ligation into One Shot PIR1 competent cells (Invitrogen) and used Sanger sequencing to verify that 8 different colonies contained 8 unique barcodes. To prepare barcoded libraries of each transposon donor, we transformed 5uL of the ligation reaction into each of two aliquots of One Shot PIR1 (Invitrogen) competent cell tubes, plating 4 x 250ul per tube following rescue with 950uL of SOC. We then incubated the LB-agar plates overnight at 30°C, resuspended the colonies in 2.5 ml of LB-Kanamycin (50ug/ml) per plate, and pooled the resuspension for each donor variant (giving 20ml LB per ligation reaction). We added this to 250mL of LB with Kanamycin (50ug/ml) in 1L flasks and grew the cultures for 1 hour at 37°C. After dividing this culture into 50mL aliquots we then used Zympo-spin VI columns (C1013-10) to perform scaled-up minipreps on the cell pellet using buffers from a GenCatch Plasmid DNA Mini-Prep Kit (Epoch Life Sciences).

RB-Tn-Seq sample preparation and sequencing

We prepared gDNA samples from cell culture pellets in the same way we prepared samples for Tn-seq. To modify the NGS sample preparation workflow to include the transposon barcode, we modified the primers described by Wetmore et al. (171) and Palani (208) for enrichment of the transposon-genome junctions via PCR. Furthermore, we used an adapter and enrichment primer sets designed for compatibility with the NEBNext Multiplex Oligos for Illumina (Dual Index Primers Sets) in an analogous two-step amplification process. In short, we replaced the standard NEB adapter with 2.5uL of a custom, pre-annealed TruSeq-compatibe adapter (sENG-021 and sENG-024, 15 uM) containing an 8N unique molecular identifier (UMI). We then performed an 0.3X/0.15X SPRI selection, eluting in 17uL of 0.1x TE buffer and transferring 15uL of this to a new tube. We set up 50uL enrichment PCR reactions with 15uL of adapter-ligated gDNA template, 5uL each of the forward and reverse primer (sENG-025 and sENG-027 for Tn-Seq donors, or sENG-030 and sENG-027, for RB-Tn-Seq donors), and 25uL of NEBNext Ultra II Q5 Master Mix. Using a T100 Thermal Cycler (Bio-Rad), the samples were incubated as follows (heated lid on): 98°C 30s, (98°C 10s, 65°C 1 min 15s) x 13 cycles, 65°C 5 mins, 4°C hold. We then performed a 0.9X SPRI selection, eluting in 17uL of 0.1x TE buffer and transferring 15uL of this to a new tube. We measured the concentration of DNA in a 1uL aliquot of each sample using a Qubit3 fluorimeter (ThermoFisher), and set up indexing PCR reactions with 10ng of template (made up to 15uL), 5uL each of a unique pair of i5 and i7 primers from a NEBNext Multiplex Oligos for Illumina (NEB), and 25uL NEBNext Ultra II Q5 Master Mix. The samples were incubated as follows (heated lid on): 98°C 30s, (98°C 10s, 65°C 1 min 30s) x 11 cycles, 65°C

5 mins, 4°C hold. We then performed two 0.9X SPRI size selection steps on each sample, eluting in 30uL 0.1x TE buffer. We used 2% and 4% TAE-agarose gels to check the size distributions of the amplified products relative to a low molecular weight DNA ladder (NEB), and verified the absence of primer-dimer bands (∼128bp). After measuring the concentrations of each sample, we then pooled them by equal mass and submitted them for sequencing by Quintara Biosciences (Cambridge, MA). Samples were run in paired-end mode on a MiSeq (2x150 cycles, 1M reads) or HiSeqX (2x150 cycles, 300M reads) with a 15% PhiX spike-in.

Rb-TnSeq analysis

We first filtered for read pairs in which the forward read contained the expected transposon terminus sequence, with 2 mismatches permitted, using the seqkit (215) v2.2.0 "grep" function with parameters "-s -i -m 2 -P -p CAGACCGGGGACTTATCAGCCAACCTGTTA". Since this process only returns forward reads, we obtained corresponding reverse reads using fastq-pair (ref: https://github.com/linsalrob/fastq-pair) v0.4. Next, we extracted RB-TnSeq barcodes by locating the sequence between expected barcode flanking regions using the "seqkit amplicon" function with parameters "-m 4 -P -s -F GATGTCCACGAGGTCTCT -R GCCGGCCGTCGACCTGCAGCGTACG -r 19:-26 âbed". In order to consolidate nearly-identical barcode sequences that differ because of sequencing errors, we clustered barcodes using the Bartender (216) "bartender_single_comâ function with parameter "-d 2". Only reads associated with barcodes that matched the expected length (20bp) were used in subsequent analyses.

After extracting transposon-derived sequences and their associated barcodes, reads were trimmed, quality filtered, and tagged for UMIs using fastp with parameters "-U –umi_loc=read2 –umi_len=8 –trim_front1=30 –trim_front2=0". We then mapped filtered reads using Bowtie2 and deduplicated alignments as described above. Every deduplicated alignment was then associated with its most proximal gene using the BEDTools "closest" function.

Barcodes were used to demultiplex samples into lineages (i.e., samples from different time points that contain the same RB-TnSeq barcode). We used barcode-gene associations to track genes that were enriched in transposon insertions throughout and between lineages.

References

[1] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in Escherichia coli," *Nature*, vol. 403, pp. 339–342, 1 2000.

[2] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, pp. 335–338, 1 2000.

[3] M. A. English, R. V. Gayet, and J. J. Collins, "Designing Biological Circuits: Synthetic Biology Within the Operon Model and Beyond," *Annual Review of Biochemistry*, vol. 90, pp. 013118–111914, 6 2021.

[4] A. S. Khalil and J. J. Collins, "Synthetic biology: applications come of age.," *Nature reviews. Genetics*, vol. 11, no. 5, pp. 367–79, 2010.

[5] J. Nielsen and J. D. Keasling, "Engineering Cellular Metabolism," *Cell*, vol. 164, no. 6, pp. 1185–1197, 2016.

[6] C. J. Bashor and J. J. Collins, "Understanding Biological Regulation Through Synthetic Biology," *Annual Review of Biophysics*, vol. 47, no. 1, pp. 399–423, 2018.

[7] C. J. Bashor, A. A. Horwitz, S. G. Peisajovich, and W. A. Lim, "Rewiring Cells: Synthetic Biology as a Tool to Interrogate the Organizational Principles of Living Systems," *Annual Review of Biophysics*, vol. 39, no. 1, pp. 515–537, 2010.

[8] G. Balázsi, A. Van Oudenaarden, and J. J. Collins, "Cellular decision making and biological noise: From microbes to mammals," *Cell*, vol. 144, no. 6, pp. 910–925, 2011.

[9] M. A. English, L. R. Soenksen, R. V. Gayet, H. de Puig, N. M. Angenent-Mari, A. S. Mao, P. Q. Nguyen, and J. J. Collins, "Programmable CRISPR-responsive smart materials," *Science*, vol. 365, no. 6455, pp. 780–785, 2019.

[10] R. V. Gayet, H. de Puig, M. A. English, L. R. Soenksen, P. Q. Nguyen, A. S. Mao, N. M. Angenent-Mari, and J. J. Collins, "Creating CRISPR-responsive smart materials for diagnostics and programmable cargo release," *Nature Protocols*, vol. 15, pp. 3030–3063, 9 2020.

[11] S. Slomovic, K. Pardee, and J. J. Collins, "Synthetic biology devices for in vitro and in vivo diagnostics," *Proceedings of the National Academy of Sciences*, vol. 112, no. 47, pp. 14429–14435, 2015.

[12] K. Pardee, A. A. Green, T. Ferrante, D. E. Cameron, A. DaleyKeyser, P. Yin, and J. J. Collins, "Paper-Based Synthetic Gene Networks," *Cell*, vol. 159, pp. 940–954, 11 2014.

[13] M. M. Kaminski, O. O. Abudayyeh, J. S. Gootenberg, F. Zhang, and J. J. Collins,

"CRISPR-based diagnostics," *Nature Biomedical Engineering*, vol. 5, no. 7, pp. 643–656, 2021.

[14] G. J. Knott and J. A. Doudna, "CRISPR-Cas guides the future of genetic engineering," *Science*, vol. 361, no. 6405, pp. 866–869, 2018.

[15] A. K. Cain, L. Barquist, A. L. Goodman, I. T. Paulsen, J. Parkhill, and T. van Opijnen, "A decade of advances in transposon-insertion sequencing," *Nature Reviews Genetics*, vol. 21, no. 9, pp. 526–540, 2020.

[16] A. J. Simon, S. DâOelsnitz, and A. D. Ellington, "Synthetic evolution," *Nature Biotechnology*, vol. 37, pp. 730–743, 7 2019.

[17] E. A. DeBenedictis, E. J. Chory, D. W. Gretton, B. Wang, S. Golas, and K. M. Esvelt, "Systematic molecular evolution enables robust biomolecule discovery," *Nature Methods*, vol. 19, no. 1, pp. 55–64, 2022.

[18] B. G. Wong, C. P. Mancuso, S. Kiriakov, C. J. Bashor, and A. S. Khalil, "Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER," *Nature Biotechnology*, vol. 36, no. 7, pp. 614–623, 2018.

[19] E. J. Chory, D. W. Gretton, E. A. DeBenedictis, and K. M. Esvelt, "Enabling high-âthroughput biology with flexible openâsource automation," *Molecular Systems Biology*, vol. 17, no. 3, pp. 1–10, 2021.

[20] M. Lynch, "The evolution of genetic networks by non-adaptive processes," *Nature Reviews Genetics*, vol. 8, no. 10, pp. 803–813, 2007.

[21] I. Nocedal and A. D. Johnson, "How transcription networks evolve and produce biological novelty," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 80, pp. 265–274, 2016.

[22] C. Feschotte, "Transposable elements and the evolution of regulatory networks," *Nature Reviews Genetics*, vol. 9, no. 5, pp. 397–405, 2008.

[23] E. B. Chuong, N. C. Elde, and C. Feschotte, "Regulatory activities of transposable elements: From conflicts to benefits," *Nature Reviews Genetics*, vol. 18, no. 2, pp. 71–86, 2017.

[24] E. V. Koonin, "Viruses and mobile elements as drivers of evolutionary transitions," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, no. 1701, 2016.

[25] E. V. Koonin, K. S. Makarova, Y. I. Wolf, and M. Krupovic, "Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire," *Nature Reviews Genetics*, vol. 21, no. 2, pp. 119–131, 2020.

[26] H. Altae-Tran, S. Kannan, F. E. Demircioglu, R. Oshiro, S. P. Nety, L. J. McKay, M. Dlakić, W. P. Inskeep, K. S. Makarova, R. K. Macrae, E. V. Koonin, and F. Zhang, "The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases," *Science*, vol. 374, no. 6563, pp. 57–65, 2021.

[27] S. Kannan, H. Altae-Tran, X. Jin, V. J. Madigan, R. Oshiro, K. S. Makarova, E. V. Koonin, and F. Zhang, "Compact RNA editors with small Cas13 proteins," *Nature Biotechnology*, 2021.

[28] M. Segel, B. Lash, J. Song, A. Ladha, C. C. Liu, X. Jin, S. L. Mekhedov, R. K. Macrae, E. V. Koonin, and F. Zhang, "Mammalian retrovirus-like protein PEG10 packages its own mRNA and can be pseudotyped for mRNA delivery," *Science*, vol. 373, no. 6557, pp. 882–889, 2021.

[29] J. Strecker, A. Ladha, Z. Gardner, J. L. Schmid-Burgk, K. S. Makarova, E. V. Koonin, and F. Zhang, "RNA-guided DNA insertion with CRISPR-associated transposases," *Science*, vol. 364, pp. 48–53, 7 2019.

[30] S. E. Klompe, P. L. H. Vo, T. S. Halpin-Healy, and S. H. Sternberg, "Transposon-encoded CRISPRâCas systems direct RNA-guided DNA integration," *Nature*, vol. 571, pp. 219–225, 7 2019.

[31] P. L. H. Vo, C. Ronda, S. E. Klompe, E. E. Chen, C. Acree, H. H. Wang, and S. H. Sternberg, "CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering," *Nature Biotechnology*, vol. 39, pp. 480–489, 4 2021.

[32] I. Tagkopoulos, Y.-C. Liu, and S. Tavazoie, "Predictive Behavior Within Microbial Genetic Networks," *Science*, vol. 320, pp. 1313–1317, 6 2008.

[33] M. Mimee, P. Nadeau, A. Hayward, S. Carim, S. Flanagan, L. Jerger, J. Collins, S. McDonnell, R. Swartwout, R. J. Citorik, V. Bulović, R. Langer, G. Traverso, A. P. Chandrakasan, and T. K. Lu, "An ingestible bacterial-electronic system to monitor gastrointestinal health," *Science*, vol. 360, pp. 915–918, 5 2018.

[34] N. Mao, A. Cubillos-Ruiz, D. E. Cameron, and J. J. Collins, "Probiotic strains detect and suppress cholera in mice," *Science Translational Medicine*, vol. 10, p. eaao2586, 6 2018.

[35] H. de Puig, I. Bosch, J. J. Collins, and L. Gehrke, "Point-of-Care Devices to Detect Zika and Other Emerging Viruses," *Annual Review of Biomedical Engineering*, vol. 22, pp. 371–386, 6 2020.

[36] S. Baker, N. Thomson, F.-X. Weill, and K. E. Holt, "Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens," *Science*, vol. 360, no. May, pp. 733–738, 2017.

[37] J. Ji, J. Shi, A. Budhu, Z. Yu, M. Forgues, S. Roessler, S. Ambs, Y. Chen, P. S. Meltzer, C. M. Croce, L. X. Qin, K. Man, C. M. Lo, J. Lee, I. O. Ng, J. Fan, Z. Y. Tang, H. C. Sun, and X. W. Wang, "MicroRNA expression, survival, and response to interferon in liver cancer," *New England Journal of Medicine*, vol. 361, no. 15, pp. 1437–1447, 2009.

[38] K. Pardee, A. A. Green, M. K. Takahashi, D. Braff, G. Lambert, J. W. Lee, T. Ferrante, D. Ma, N. Donghia, M. Fan, N. M. Daringer, I. Bosch, D. M. Dudley, D. H. O'Connor, L. Gehrke, and J. J. Collins, "Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components," *Cell*, vol. 165, no. 5, pp. 1255–1266, 2016.

[39] M. K. Takahashi, X. Tan, A. J. Dy, D. Braff, R. T. Akana, Y. Furuta, N. Donghia, A. Ananthakrishnan, and J. J. Collins, "A low-cost paper-based synthetic biology platform for analyzing gut microbiota and host biomarkers," *Nature Communications*, vol. 9, p. 3347, 12 2018.

[40] J. S. Gootenberg, O. O. Abudayyeh, J. W. Lee, P. Essletzbichler, A. J. Dy, J. Joung, V. Verdine, N. Donghia, N. M. Daringer, C. A. Freije, C. Myhrvold, R. P. Bhattacharyya, J. Livny, A. Regev, E. V. Koonin, D. T. Hung, P. C. Sabeti, J. J. Collins, and F. Zhang, "Nucleic acid detection with CRISPR-Cas13a/C2c2," *Science*, vol. 356, pp. 438–442, 4 2017.

[41] X. Tan, J. H. Letendre, J. J. Collins, and W. W. Wong, "Synthetic biology in the clinic: engineering vaccines, diagnostics, and therapeutics," *Cell*, vol. 184, no. 4, pp. 881–898, 2021.

[42] K. J. Land, D. I. Boeras, X. S. Chen, A. R. Ramsay, and R. W. Peeling, "REASSURED diagnostics to inform disease control strategies, strengthen health systems and improve patient outcomes," *Nature Microbiology*, vol. 4, no. 1, pp. 46–54, 2019.

[43] W. Jiang and L. A. Marraffini, "CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems," *Annual Review of Microbiology*, vol. 69, no. 1, pp. 209–228, 2015.

[44] J. S. Gootenberg, O. O. Abudayyeh, M. J. Kellner, J. Joung, J. J. Collins, and F. Zhang, "Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6," *Science*, vol. 360, pp. 439–444, 4 2018.

[45] S. Y. Li, Q. X. Cheng, J. K. Liu, X. Q. Nie, G. P. Zhao, and J. Wang, "CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA," *Cell Research*, vol. 28, no. 4, pp. 491–493, 2018.

[46] J. S. Chen, E. Ma, L. B. Harrington, M. Da Costa, X. Tian, J. M. Palefsky, and J. A. Doudna, "CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity," *Science*, vol. 360, pp. 436–439, 4 2018.

[47] M. M. Kaminski, M. A. Alcantar, I. T. Lape, R. Greensmith, A. C. Huske, J. A. Valeri, F. M. Marty, V. Klämbt, J. Azzi, E. Akalin, L. V. Riella, and J. J. Collins, "A CRISPR-based assay for the detection of opportunistic infections post-transplantation and for the monitoring of transplant rejection.," *Nature Biomedical Engineering*, vol. 4, pp. 601–609, 6 2020.

[48] H. De Puig, R. A. Lee, D. Najjar, X. Tan, L. R. Soekensen, N. M. Angenent-Mari, N. M. Donghia, N. E. Weckman, A. Ory, C. F. Ng, P. Q. Nguyen, A. S. Mao, T. C. Ferrante, G. Lansberry, H. Sallum, J. Niemi, and J. J. Collins, "Minimally instrumented SHERLOCK (miSHERLOCK) for CRISPR-based point-of-care diagnosis of SARS-CoV-2 and emerging variants," *Science Advances*, vol. 7, no. 32, pp. 23–26, 2021.

[49] R. Hajian, S. Balderston, T. Tran, T. DeBoer, J. Etienne, M. Sandhu, N. A. Wauford, J.-Y. Chung, J. Nokes, M. Athaiya, J. Paredes, R. Peytavi, B. Goldsmith, N. Murthy, I. M. Conboy, and K. Aran, "Detection of unamplified target genes via CRISPRâCas9 immobilized on a graphene field-effect transistor," *Nature Biomedical Engineering*, 3 2019.

[50] Y. Lu, A. A. Aimetti, R. Langer, and Z. Gu, "Bioresponsive materials," *Nature Reviews Materials*, vol. 2, p. 16075, 1 2017.

[51] J. Li and D. J. Mooney, "Designing hydrogels for controlled drug delivery," *Nature Reviews Materials*, vol. 1, no. 12, 2016.

[52] H. W. Ooi, S. Hafeez, C. A. van Blitterswijk, L. Moroni, and M. B. Baker, "Hydrogels that listen to cells: a review of cell-responsive strategies in biomaterial design for tissue regeneration," *Mater. Horiz.*, vol. 4, no. 6, pp. 1020–1040, 2017.

[53] E. S. Place, N. D. Evans, and M. M. Stevens, "Complexity in biomaterials for tissue engineering," *Nature Materials*, vol. 8, pp. 457–470, 6 2009.

[54] L. Yan, Z. Zhu, Y. Zou, Y. Huang, D. Liu, S. Jia, D. Xu, M. Wu, Y. Zhou, S. Zhou, and C. J. Yang, "Target-responsive "sweet" hydrogel with glucometer readout for portable and quantitative detection of non-glucose targets," *Journal of the American Chemical Society*, vol. 135, no. 10, pp. 3748–3751, 2013.

[55] X. Wei, T. Tian, S. Jia, Z. Zhu, Y. Ma, J. Sun, Z. Lin, and C. J. Yang, "Target-responsive DNA hydrogel mediated stop-flow microfluidic paper-based analytic device for rapid, portable and visual detection of multiple targets," *Analytical Chemistry*, vol. 87, no. 8, pp. 4275–4282, 2015.

[56] G. Kocak, C. Tuncer, and V. Bütün, "pH-Responsive polymers," *Polymer Chemistry*, vol. 8, no. 1, pp. 144–176, 2017.

[57] S. Mura, J. Nicolas, and P. Couvreur, "Stimuli-responsive nanocarriers for drug delivery," *Nature Materials*, vol. 12, pp. 991–1003, 11 2013.

[58] F. D. Jochum and P. Theato, "Temperature- and light-responsive smart polymer materials," *Chem. Soc. Rev.*, vol. 42, no. 17, pp. 7468–7483, 2013.

[59] W. Na, D. Nam, H. Lee, and S. Shin, "Rapid molecular diagnosis of infectious viruses in microfluidics using DNA hydrogel formation," *Biosensors and Bioelectronics*, vol. 108, no. November 2017, pp. 9–13, 2018.

[60] S. Venkatesh, J. Wower, and M. E. Byrne, "Nucleic acid therapeutic carriers with on-demand triggered release," *Bioconjugate Chemistry*, vol. 20, no. 9, pp. 1773–1782, 2009.

[61] J. Li, L. Mo, C. H. Lu, T. Fu, H. H. Yang, and W. Tan, "Functional nucleic acid-based hydrogels for bioanalytical and biomedical applications," *Chemical Society Reviews*, vol. 45, no. 5, pp. 1410–1431, 2016.

[62] R. Carlson, "The changing economics of DNA synthesis," *Nature Biotechnology*, vol. 27, no. 12, pp. 1091–1094, 2009.

[63] M. Gao, K. Gawel, and B. T. Stokke, "Toehold of dsDNA exchange affects the hydrogel swelling kinetics of a polymer-dsDNA hybrid hydrogel," *Soft Matter*, vol. 7, no. 5, pp. 1741–1746, 2011.

[64] G. Sicilia, C. Grainger-Boultby, N. Francini, J. P. Magnusson, A. O. Saeed, F. Fernández-Trillo, S. G. Spain, and C. Alexander, "Programmable polymer-DNA hydrogels with dual input and multiscale responses," *Biomaterials Science*, vol. 2, no. 2, pp. 203–211, 2014.

[65] J. Fern and R. Schulman, "Modular DNA strand-displacement controllers for directing material expansion," *Nature Communications*, vol. 9, no. 1, pp. 1–8, 2018.

[66] M. Arthur, C. Molinas, C. Mabilat, and P. Courvalin, "Detection of erythromycin resistance by the polymerase chain reaction using primers in conserved regions of erm rRNA methylase genes," *Antimicrobial Agents and Chemotherapy*, vol. 34, no. 10, pp. 2024–2026, 1990.

[67] B. Strommenger, C. Kettlitz, G. Werner, and W. Witte, "Multiplex PCR assay for simultaneous detection of nine clinically relevant antibiotic resistance genes in Staphylococcus aureus," *Journal of Clinical Microbiology*, vol. 41, no. 9, pp. 4089–4094, 2003.

[68] C. E. Okolie, K. G. Wooldridge, D. P. Turner, A. Cockayne, and R. James, "Development of a heptaplex PCR assay for identification of Staphylococcus aureus and CoNS with simultaneous detection of virulence and antibiotic resistance genes," *BMC Microbiology*, vol. 15, no. 1, pp. 1–7, 2015.

[69] N. K. Qureshi, S. Yin, and S. Boyle-Vavra, "The role of the staphylococcal VraTSR regula-

tory system on vancomycin resistance and vana operon expression in vancomycin-resistant staphylococcus aureus," *PLoS ONE*, vol. 9, no. 1, pp. 1–7, 2014.

[70] K. W. Wang, T. Betancourt, and C. K. Hall, "Computational Study of DNA-Cross-Linked Hydrogel Formation for Drug Delivery Applications," *Macromolecules*, vol. 51, no. 23, pp. 9758–9768, 2018.

[71] J. S. Kahn, A. Trifonov, A. Cecconello, W. Guo, C. Fan, and I. Willner, "Integration of Switchable DNA-Based Hydrogels with Surfaces by the Hybridization Chain Reaction," *Nano Letters*, vol. 15, no. 11, pp. 7773–7778, 2015.

[72] D. C. Lin, B. Yurke, and N. A. Langrana, "Mechanical Properties of a Reversible, DNA-Crosslinked Polyacrylamide Hydrogel," *Transactions of the ASME*, vol. 126, pp. 104–110, 3 2004.

[73] B. Wei, I. Cheng, K. Q. Luo, and Y. Mi, "Capture and release of protein by a reversible DNA-induced sol-gel transition system," *Angewandte Chemie - International Edition*, vol. 47, no. 2, pp. 331–333, 2008.

[74] K. Barker, S. K. Rastogi, J. Dominguez, T. Cantu, W. Brittain, J. Irvin, and T. Betancourt, "SI - Biodegradable DNA-Enabled Poly (ethylene glycol) Hydrogels by Copper-Free Click Chemistry," *Journal of Biomaterials Science, Polymer Edition*, pp. 3–5, 2016.

[75] Z. Zhu, C. Wu, H. Liu, Y. Zou, X. Zhang, H. Kang, C. J. Yang, and W. Tan, "SI - An aptamer cross-linked hydrogel as a colorimetric platform for visual detection," *Angewandte Chemie - International Edition*, vol. 49, no. 6, pp. 1052–1056, 2010.

[76] K. Pardee, S. Slomovic, P. Q. Nguyen, J. W. Lee, N. Donghia, D. Burrill, T. Ferrante, F. R. McSorley, Y. Furuta, A. Vernet, M. Lewandowski, C. N. Boddy, N. S. Joshi, and J. J. Collins, "Portable, On-Demand Biomolecular Manufacturing," *Cell*, vol. 167, pp. 248–259, 9 2016.

[77] X. He, J. Cai, B. Liu, Y. Zhong, and Y. Qin, "Cellular magnetic resonance imaging contrast generated by the ferritin heavy chain genetic reporter under the control of a Tet-On switch," *Stem Cell Research & Therapy*, vol. 6, no. 1, p. 207, 2015.

[78] A. K. Badu-Tawiah, S. Lathwal, K. Kaastrup, M. Al-Sayah, D. C. Christodouleas, B. S. Smith, G. M. Whitesides, and H. D. Sikes, "Polymerization-based signal amplification for paper-based immunoassays," *Lab on a Chip*, vol. 15, no. 3, pp. 655–659, 2015.

[79] L. Magro, B. Jacquelin, C. Escadafal, P. Garneret, A. Kwasiborski, J. C. Manuguerra, F. Monti, A. Sakuntabhai, J. Vanhomwegen, P. Lafaye, and P. Tabeling, "Paper-based RNA detection and multiplexed analysis for Ebola virus diagnostics," *Scientific Reports*, vol. 7, no. 1, pp. 1–9, 2017.

[80] E. Fu, B. Lutz, P. Kauffman, and P. Yager, "Controlled reagent transport in disposable 2D paper networks," *Lab on a Chip*, vol. 10, no. 7, pp. 918–920, 2010.

[81] J. Monod and F. Jacob, "General Conclusions: Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 26, pp. 389–401, 1 1961.

[82] M. Costanzo, E. Kuzmin, J. van Leeuwen, B. Mair, J. Moffat, C. Boone, and B. Andrews, "Global Genetic Networks and the Genotype-to-Phenotype Relationship," *Cell*, vol. 177, no. 1, pp. 85–100, 2019.

[83] M. M. Babu and S. A. Teichmann, "Evolution of transcription factors and the gene regulatory network in Escherichia coli," *Nucleic Acids Research*, vol. 31, no. 4, pp. 1234–1244, 2003.

[84] M. M. Babu, N. M. Luscombe, L. Aravind, M. Gerstein, and S. A. Teichmann, "Structure and evolution of transcriptional regulatory networks," *Current Opinion in Structural Biology*, vol. 14, no. 3, pp. 283–291, 2004.

[85] J. C. Perez and E. A. Groisman, "Evolution of Transcriptional Regulatory Circuits in Bacteria," *Cell*, vol. 138, no. 2, pp. 233–244, 2009.

[86] D. Thompson, A. Regev, and S. Roy, *Comparative Analysis of Gene Regulatory Networks: From Network Reconstruction to Evolution*, vol. 31. 2015.

[87] J. D. Bloom and F. H. Arnold, "In the light of directed evolution: Pathways of adaptive protein evolution," *In the Light of Evolution*, vol. 3, pp. 149–163, 2009.

[88] M. K. Gouda, M. Manhart, and G. Balázsi, "Evolutionary regain of lost gene circuit function," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 50, pp. 25162–25171, 2019.

[89] N. Philippe, E. Crozat, R. E. Lenski, and D. Schneider, "Evolution of global regulatory networks during a long-term experiment with Escherichia coli," *BioEssays*, vol. 29, no. 9, pp. 846–860, 2007.

[90] J. E. Barrick, D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim, "Genome evolution and adaptation in a long-term experiment with Escherichia coli," *Nature*, vol. 461, no. 7268, pp. 1243–1247, 2009.

[91] N. C. Elde, S. J. Child, M. T. Eickbush, J. O. Kitzman, K. S. Rogers, J. Shendure, A. P. Geballe, and H. S. Malik, "Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses," *Cell*, vol. 150, no. 4, pp. 831–841, 2012.

[92] J. Näsvall, L. Sun, J. R. Roth, and D. I. Andersson, "Real-time evolution of new genes by innovation, amplification, and divergence," *Science*, vol. 338, no. 6105, pp. 384–387, 2012.

[93] T. B. Taylor, G. Mulley, A. H. Dills, A. S. Alsohim, L. J. McGuffin, D. J. Studholme, M. W. Silby, M. A. Brockhurst, L. J. Johnson, and R. W. Jackson, "Evolutionary resurrection of flagellar motility via rewiring of the nitrogen regulation system," *Science*, vol. 347, pp. 1014–1017, 2 2015.

[94] M. Isalan, C. Lemerle, K. Michalodimitrakis, C. Horn, P. Beltrao, E. Raineri, M. Garriga-Canut, and L. Serrano, "Evolvability and hierarchy in rewired bacterial gene networks," *Nature*, vol. 452, pp. 840–845, 4 2008.

[95] T. Kitada, B. DiAndreth, B. Teague, and R. Weiss, "Programming gene and engineered-cell therapies with synthetic biology," *Science*, vol. 359, no. 6376, 2018.

[96] D. H. Erwin and E. H. Davidson, "The evolution of hierarchical gene regulatory networks," 2 2009.

[97] K. Pougach, A. Voet, F. A. Kondrashov, K. Voordeckers, J. F. Christiaens, B. Baying, V. Benes, R. Sakai, J. Aerts, B. Zhu, P. Van Dijck, and K. J. Verstrepen, "Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network," *Nature Communications*, vol. 5, 2014.

[98] D. I. Andersson and D. Hughes, "Gene Amplification and Adaptive Evolution in Bacteria," *Annual Review of Genetics*, vol. 43, no. 1, pp. 167–195, 2009.

[99] M. Fondi, G. Emiliani, and R. Fani, "Origin and evolution of operons and metabolic pathways," *Research in Microbiology*, vol. 160, no. 7, pp. 502–512, 2009.

[100] S. Nagarajan, D. M. Sherman, I. Shaw, and L. A. Sherman, "Functions of the duplicated hik31 operons in central metabolism and responses to light, dark, and carbon sources in synechocystis sp. strain PCC 6803," *Journal of Bacteriology*, vol. 194, no. 2, pp. 448–459, 2012.

[101] I. Wapinski, A. Pfeffer, N. Friedman, and A. Regev, "Natural history and evolutionary principles of gene duplication in fungi," *Nature*, vol. 449, no. 7158, pp. 54–61, 2007.

[102] S. A. Teichmann and M. M. Babu, "Gene regulatory network growth by duplication," *Nature Genetics*, vol. 36, no. 5, pp. 492–496, 2004.

[103] A. B. Reams and J. R. Roth, "Mechanisms of gene duplication and amplification," *Cold Spring Harbor Perspectives in Biology*, vol. 7, no. 2, pp. 1–26, 2015.

[104] K. Voordeckers, K. Pougach, and K. J. Verstrepen, "How do regulatory networks evolve and

expand throughout evolution?," *Current Opinion in Biotechnology*, vol. 34, pp. 180–188, 2015.

[105] C. R. Baker, V. Hanson-Smith, and A. D. Johnson, "Following Gene Duplication, Paralog Interference Constrains Transcriptional Circuit Evolution," *Science*, vol. 342, pp. 104–108, 10 2013.

[106] J. González, G. López, S. Argueta, X. Escalera-Fanjul, M. E. Hafidi, C. Campero-Basaldua, J. Strauss, L. Riego-Ruiz, and A. González, "Diversification of transcriptional regulation determines subfunctionalization of paralogous branched chain aminotransferases in the yeast saccharomyces cerevisiae," *Genetics*, vol. 207, no. 3, pp. 975–991, 2017.

[107] M. Colón, F. Hernández, K. López, H. Quezada, J. González, G. López, C. Aranda, and A. González, "Saccharomyces cerevisiae Bat1 and Bat2 aminotransferases have functionally diverged from the ancestral-like Kluyveromyces lactis orthologous enzyme," *PLoS ONE*, vol. 6, no. 1, pp. 1–13, 2011.

[108] P. C. Kirchberger, M. Schmidt, and H. Ochman, "The Ingenuity of Bacterial Genomes," *Annual Review of Microbiology*, vol. 74, no. 1, 2020.

[109] A. R. Carvunis, T. Rolland, I. Wapinski, M. A. Calderwood, M. A. Yildirim, N. Simonis, B. Charloteaux, C. A. Hidalgo, J. Barbette, B. Santhanam, G. A. Brar, J. S. Weissman, A. Regev, N. Thierry-Mieg, M. E. Cusick, and M. Vidal, "Proto-genes and de novo gene birth," *Nature*, vol. 487, no. 7407, pp. 370–374, 2012.

[110] G. Bourque, K. H. Burns, M. Gehring, V. Gorbunova, A. Seluanov, M. Hammell, M. Imbeault, Z. Izsvák, H. L. Levin, T. S. Macfarlan, D. L. Mager, and C. Feschotte, "Ten things you should know about transposable elements," *Genome Biology*, vol. 19, no. 1, pp. 1–12, 2018.

[111] A. L. Gomes, N. I. Johns, A. Yang, F. Velez-Cortes, C. S. Smillie, M. B. Smith, E. J. Alm, and H. H. Wang, "Genome and sequence determinants governing the expression of horizontally acquired DNA in bacteria," *ISME Journal*, vol. 14, pp. 2347–2357, 9 2020.

[112] Y. Oren, M. B. Smith, N. I. Johns, M. K. Zeevi, D. Biran, E. Z. Ron, J. Corander, H. H. Wang, E. J. Alm, and T. Pupko, "Transfer of noncoding DNA drives regulatory rewiring in Bacteria," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 45, pp. 16112–16117, 2014.

[113] T. Y. Pang and M. J. Lercher, "Each of 3,323 metabolic innovations in the evolution of E. Coli arose through the horizontal transfer of a single DNA segment," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, pp. 187–192, 1 2019.

[114] R. K. Aziz, M. Breitbart, and R. A. Edwards, "Transposases are the most abundant, most ubiquitous genes in nature," *Nucleic Acids Research*, vol. 38, no. 13, pp. 4207–4217, 2010.

[115] R. L. Cosby, J. Judd, R. Zhang, A. Zhong, N. Garry, E. J. Pritham, and C. Feschotte, "Recurrent evolution of vertebrate transcription factors by transposase capture," *Science*, vol. 371, no. 6531, 2021.

[116] J. Judd, H. Sanderson, and C. Feschotte, "Evolution of mouse circadian enhancers from transposable elements," *bioRxiv*, vol. 0205, pp. 0–2, 2020.

[117] A. Szitenberg, S. Cha, C. H. Opperman, D. M. Bird, M. L. Blaxter, and D. H. Lunt, "Genetic drift, not life history or RNAi, determine long-Term evolution of transposable elements," *Genome Biology and Evolution*, vol. 8, no. 9, pp. 2964–2978, 2016.

[118] N. Kleckner, "Regulation of Transposition in Bacteria," *Annual Review of Cell Biology*, vol. 6, pp. 297–327, 11 1990.

[119] J. Vandecraen, M. Chandler, A. Aertsen, and R. Van Houdt, "The impact of insertion sequences on bacterial genome plasticity and adaptability," *Critical Reviews in Microbiology*, vol. 43, pp. 709–730, 11 2017.

[120] K. Umenhoffer, T. Fehér, G. Balikó, F. Ayaydin, J. Pósfai, F. R. Blattner, and G. Pósfai, "Reduced evolvability of Escherichia coli MDS42, an IS-less cellular chassis for molecular and synthetic biology applications," *Microbial Cell Factories*, vol. 9, p. 38, 12 2010.

[121] G. PoÌsfai, G. Plunkett, T. FeheÌr, D. Frisch, G. M. Keil, K. Umenhoffer, V. Kolisnychenko, B. Stahl, S. S. Sharma, M. de Arruda, V. Burland, S. W. Harcum, and F. R. Blattner, "Emergent Properties of Reduced-Genome Escherichia coli," *Science*, vol. 312, pp. 1044–1046, 5 2006.

[122] V. Vernyik, I. Karcagi, E. Tímár, I. Nagy, Ã. Györkei, B. Papp, Z. Györfy, and G. Pósfai, "Exploring the fitness benefits of genome reduction in Escherichia coli by a selection-driven approach," *Scientific Reports*, vol. 10, 12 2020.

[123] B. McClintock, "The origin and behavior of mutable loci in maize," *Proceedings of the National Academy of Sciences*, vol. 36, pp. 344–355, 6 1950.

[124] M. N. Price, K. M. Wetmore, R. J. Waters, M. Callaghan, J. Ray, H. Liu, J. V. Kuehl, R. A. Melnyk, J. S. Lamson, Y. Suh, H. K. Carlson, Z. Esquivel, H. Sadeeshkumar, R. Chakraborty, G. M. Zane, B. E. Rubin, J. D. Wall, A. Visel, J. Bristow, M. J. Blow, A. P. Arkin, and A. M. Deutschbauer, "Mutant phenotypes for thousands of bacterial genes of unknown function," *Nature*, vol. 557, no. 7706, pp. 503–509, 2018.

[125] M. Santiago, W. Lee, A. A. Fayad, K. A. Coe, M. Rajagopal, T. Do, F. Hennessen, V. Srisuknimit, R. Müller, T. C. Meredith, and S. Walker, "Genome-wide mutant profiling

predicts the mechanism of a Lipid II binding antibiotic article," *Nature Chemical Biology*, vol. 14, no. 6, pp. 601–608, 2018.

[126] H. Wang, D. Claveau, J. P. Vaillancourt, T. Roemer, and T. C. Meredith, "High-frequency transposition for determining antibacterial mode of action," *Nature Chemical Biology*, vol. 7, no. 10, pp. 720–729, 2011.

[127] R. Rad, L. Rad, W. Wang, A. Strong, H. Ponstingl, I. F. Bronner, M. Mayho, K. Steiger, J. Weber, M. Hieber, C. Veltkamp, S. Eser, U. Geumann, R. Öllinger, M. Zukowska, M. Barenboim, R. Maresch, J. Cadiñanos, M. Friedrich, I. Varela, F. Constantino-Casas, A. Sarver, J. Ten Hoeve, H. Prosser, B. Seidler, J. Bauer, M. Heikenwälder, E. Metzakopian, A. Krug, U. Ehmer, G. Schneider, T. Knösel, P. Römmele, D. Aust, R. Grötzmann, C. Pilarsky, Z. Ning, L. Wessels, R. M Schmid, M. A Quail, G. Vassiliou, I. Esposito, P. Liu, D. Saur, and A. Bradley, "A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer," *Nature Genetics*, vol. 47, no. 1, pp. 47–56, 2015.

[128] N. Judson and J. J. Mekalanos, "TnAraOut, A transposon-based approach to identify and characterize essential bacterial genes," *Nature Biotechnology*, vol. 18, pp. 740–745, 7 2000.

[129] M. Santiago, L. M. Matano, S. H. Moussa, M. S. Gilmore, S. Walker, and T. C. Meredith, "A new platform for ultra-high density Staphylococcus aureus transposon libraries," *BMC Genomics*, vol. 16, no. 1, pp. 1–18, 2015.

[130] H. L. Levin and J. V. Moran, "Dynamic interactions between transposable elements and their hosts," *Nature Reviews Genetics*, vol. 12, no. 9, pp. 615–627, 2011.

[131] A. V. Wright, J. K. Nuñez, and J. A. Doudna, "Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering," *Cell*, vol. 164, no. 1-2, pp. 29–44, 2016.

[132] A. Vigouroux and D. Bikard, "CRISPR Tools To Control Gene Expression in Bacteria," *Microbiology and Molecular Biology Reviews*, vol. 84, pp. 1–18, 4 2020.

[133] A. V. Anzalone, L. W. Koblan, and D. R. Liu, "Genome editing with CRISPRâCas nucleases, base editors, transposases and prime editors," *Nature Biotechnology*, 6 2020.

[134] F. Schmidt, M. Y. Cherepkova, and R. J. Platt, "Transcriptional recording by CRISPR spacer acquisition from RNA," *Nature*, vol. 562, no. 7727, pp. 380–385, 2018.

[135] A. A. Nielsen and C. A. Voigt, "Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks," *Molecular Systems Biology*, vol. 10, p. 763, 11 2014.

[136] B. Jusiak, S. Cleto, P. Perez-Piñera, and T. K. Lu, "Engineering Synthetic Gene Circuits

in Living Cells with CRISPR Technology," *Trends in Biotechnology*, vol. 34, no. 7, pp. 535–547, 2016.

[137] A. Chavez, M. Tuttle, B. W. Pruitt, B. Ewen-Campen, R. Chari, D. Ter-Ovanesyan, S. J. Haque, R. J. Cecchi, E. J. Kowal, J. Buchthal, B. E. Housden, N. Perrimon, J. J. Collins, and G. Church, "Comparison of Cas9 activators in multiple species," *Nature Methods*, vol. 13, no. 7, pp. 563–567, 2016.

[138] J. Fontana, C. Dong, C. Kiattisewee, V. P. Chavali, B. I. Tickman, J. M. Carothers, and J. G. Zalatan, "Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements," *Nature Communications*, vol. 11, p. 1618, 12 2020.

[139] W.-c. Ho, D. Li, Q. Zhu, and J. Zhang, "Phenotypic plasticity as a long-term memory easing readaptations to ancestral environments," *Science Advances*, vol. 6, p. eaba3388, 5 2020.

[140] G. Wang, Z. Zhao, J. Ke, Y. Engel, Y. M. Shi, D. Robinson, K. Bingol, Z. Zhang, B. Bowen, K. Louie, B. Wang, R. Evans, Y. Miyamoto, K. Cheng, S. Kosina, M. De Raad, L. Silva, A. Luhrs, A. Lubbe, D. W. Hoyt, C. Francavilla, H. Otani, S. Deutsch, N. M. Washton, E. M. Rubin, N. J. Mouncey, A. Visel, T. Northen, J. F. Cheng, H. B. Bode, and Y. Yoshikuni, "CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria," *Nature Microbiology*, vol. 4, no. 12, pp. 2498–2510, 2019.

[141] J. R. Rybarski, K. Hu, A. M. Hill, C. O. Wilke, and I. J. Finkelstein, "Metagenomic discovery of CRISPR-associated transposons," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 49, 2021.

[142] L. Trujillo Rodríguez, A. J. Ellington, and C. R. Reisch, "Broad-host-range mutagenesis with CRISPR-1 associated transposase," *bioRxiv*, 2022.

[143] B. J. Hussey and D. R. McMillen, "Programmable T7-based synthetic transcription factors," *Nucleic Acids Research*, vol. 46, no. 18, pp. 9842–9854, 2018.

[144] C. J. Tou, B. Orr, and B. P. Kleinstiver, "Cut-and-Paste DNA Insertion with Engineered Type V-K CRISPR-associated Transposases," *bioRxiv*, p. 2022.01.07.475005, 2022.

[145] C. M. Schmidt, D. L. Shis, T. D. Nguyen-Huu, and M. R. Bennett, "Stable maintenance of multiple plasmids in E. coli using a single selective marker.," *ACS synthetic biology*, vol. 1, no. 10, pp. 445–450, 2012.

[146] R. Lutz and H. Bujard, "Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements," *Nucleic Acids Research*, vol. 25, pp. 1203–1210, 3 1997.

[147] A. E. Friedland, T. K. Lu, X. Wang, D. Shi, G. Church, and J. J. Collins, "Synthetic gene networks that count.," *Science (New York, N.Y.)*, vol. 324, no. 5931, pp. 1199–202, 2009.

[148] N. Roquet, A. P. Soleimany, A. C. Ferris, S. Aaronson, and T. K. Lu, "Synthetic recombinase-based State machines in living cells," *Science*, vol. 353, no. 6297, 2016.

[149] A. Wagner, "Cooperation is fleeting in the world of transposable elements," *PLoS Computational Biology*, vol. 2, no. 12, pp. 1522–1529, 2006.

[150] J. Santos-Moreno, E. Tasiudi, J. Stelling, and Y. Schaerli, "Multistable and dynamic CRISPRi-based synthetic circuits," *bioRxiv*, pp. 1–17, 2020.

[151] E. D. Vaishnav, C. G. de Boer, J. Molinet, M. Yassour, L. Fan, X. Adiconis, D. A. Thompson, J. Z. Levin, F. A. Cubillos, and A. Regev, "The evolution, evolvability and engineering of gene regulatory DNA," *Nature*, vol. 603, no. 7901, pp. 455–463, 2022.

[152] A. J. Lopatkin and J. J. Collins, "Predictive biology: modelling, understanding and harnessing microbial complexity," *Nature Reviews Microbiology*, vol. 18, pp. 507–520, 9 2020.

[153] Y. Yokobayashi, R. Weiss, and F. H. Arnold, "Directed evolution of a genetic circuit.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 26, pp. 16587–91, 2002.

[154] K. M. Esvelt, J. C. Carlson, and D. R. Liu, "A system for the continuous directed evolution of biomolecules," *Nature*, vol. 472, no. 7344, pp. 499–503, 2011.

[155] A. Ravikumar, G. A. Arzumanyan, M. K. Obadi, A. A. Javanpour, and C. C. Liu, "Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds," *Cell*, vol. 175, no. 7, pp. 1946–1957, 2018.

[156] S. O. Halperin, C. J. Tou, E. B. Wong, C. Modavi, D. V. Schaffer, and J. E. Dueber, "CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window," *Nature*, vol. 560, pp. 248–252, 8 2018.

[157] J. W. Ellefson, M. P. Ledbetter, and A. D. Ellington, "Directed evolution of a synthetic phylogeny of programmable Trp repressors article," *Nature Chemical Biology*, vol. 14, no. 4, pp. 361–367, 2018.

[158] M. S. Morrison, C. J. Podracky, and D. R. Liu, "The developing toolkit of continuous directed evolution," *Nature chemical biology*, vol. 16, no. 6, pp. 610–619, 2020.

[159] T. Wang, A. H. Badran, T. P. Huang, and D. R. Liu, "Continuous directed evolution of proteins with improved soluble expression," *Nature Chemical Biology*, vol. 14, no. October, 2018.

[160] J. L. Payne and A. Wagner, "The causes of evolvability and their evolution," *Nature Reviews Genetics*, vol. 20, no. 1, pp. 24–38, 2019.

[161] J. Zheng, J. L. Payne, and A. Wagner, "Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks," *Science*, vol. 365, no. 6451, pp. 347–353, 2019.

[162] S. Zhou, Y. Wu, Z.-X. Xie, B. Jia, and Y.-J. Yuan, "Directed genome evolution driven by structural rearrangement techniques," *Chemical Society Reviews*, vol. 50, no. 22, pp. 12788–12807, 2021.

[163] W. J. Blake, M. KÆrn, C. R. Cantor, and J. J. Collins, "Noise in eukaryotic gene expression," *Nature*, vol. 422, pp. 633–637, 4 2003.

[164] J. M. Pedazra and A. V. Oudenaarden, "Noise Propagation in Gene Networks," *Science*, vol. 307, no. March, pp. 1965–1969, 2005.

[165] Z. Luo, L. Wang, Y. Wang, W. Zhang, Y. Guo, Y. Shen, L. Jiang, Q. Wu, C. Zhang, Y. Cai, and J. Dai, "Identifying and characterizing SCRaMbLEd synthetic yeast using ReSCuES," *Nature Communications*, vol. 9, no. 1, pp. 1–10, 2018.

[166] O. P. Windram, R. T. Rodrigues, S. Lee, M. Haines, and T. S. Bayer, "Engineering microbial phenotypes through rewiring of genetic networks," *Nucleic Acids Research*, vol. 45, no. 8, pp. 4984–4993, 2017.

[167] X. Wang and T. K. Wood, "Toxin-Antitoxin Systems Influence Biofilm and Persister Cell Formation and the General Stress Response," *Applied and Environmental Microbiology*, vol. 77, pp. 5577–5583, 8 2011.

[168] D. J. Lampe, M. E. Churchill, and H. M. Robertson, "A purified mariner transposase is sufficient to mediate transposition in vitro," *EMBO Journal*, vol. 15, no. 19, pp. 5470–5479, 1996.

[169] H. Liu, M. N. Price, R. J. Waters, J. Ray, H. K. Carlson, J. S. Lamson, R. Chakraborty, A. P. Arkin, and A. M. Deutschbauer, "Magic Pools: Parallel Assessment of Transposon Delivery Vectors in Bacteria," *mSystems*, vol. 3, no. 1, pp. 1–17, 2018.

[170] J. Santos-Moreno and Y. Schaerli, "A Framework for the Modular and Combinatorial Assembly of Synthetic Gene Circuits," *ACS Synthetic Biology*, vol. 8, no. 7, pp. 1691–1697, 2019.

[171] K. M. Wetmore, M. N. Price, R. J. Waters, J. S. Lamson, J. He, C. A. Hoover, M. J. Blow, J. Bristow, G. Butland, A. P. Arkin, and A. Deutschbauer, "Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons," *mBio*, vol. 6, no. 3, pp. 1–15, 2015.

[172] B. Szappanos, J. Fritzemeier, B. Csörgő, V. Lázár, X. Lu, G. Fekete, B. Bálint, R. Herczeg, I. Nagy, R. A. Notebaart, M. J. Lercher, C. Pál, and B. Papp, "Adaptive evolution of complex innovations through stepwise metabolic niche expansion," *Nature Communications*, vol. 7, p. 11607, 9 2016.

[173] M. Startek, A. Le Rouzic, P. Capy, D. Grzebelus, and A. Gambin, "Genomic parasites or symbionts? Modeling the effects of environmental pressure on transposition activity in asexual populations," *Theoretical Population Biology*, vol. 90, pp. 145–151, 12 2013.

[174] E. J. OâBrien, J. Utrilla, and B. O. Palsson, "Quantification and Classification of E. coli Proteome Utilization and Unused Protein Costs across Environments," *PLoS Computational Biology*, vol. 12, no. 6, pp. 1–22, 2016.

[175] T. E. Sandberg, C. J. Lloyd, B. O. Palsson, and A. M. Feist, "Laboratory evolution to alternating substrate environments yields distinct phenotypic and genetic adaptive strategies," *Applied and Environmental Microbiology*, vol. 83, 7 2017.

[176] T. Fehér, B. Cseh, K. Umenhoffer, I. Karcagi, and G. Pósfai, "Characterization of cycA mutants of Escherichia coli: An assay for measuring in vivo mutation rates," *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 595, no. 1-2, pp. 184–190, 2006.

[177] C. E. Armbruster, V. Forsyth-DeOrnellas, A. O. Johnson, S. N. Smith, L. Zhao, W. Wu, and H. L. T. Mobley, "Genome-wide transposon mutagenesis of Proteus mirabilis: Essential genes, fitness factors for catheter-associated urinary tract infection, and the impact of polymicrobial infection on fitness requirements," *PLOS Pathogens*, vol. 13, p. e1006434, 6 2017.

[178] R. J. Wargel, C. A. Hadur, and F. C. Neuhaus, "Mechanism of D-cycloserine action: transport mutants for D-alanine, D-cycloserine, and glycine.," *Journal of Bacteriology*, vol. 105, no. 3, pp. 1028–1035, 1971.

[179] B. G. Hall, "Activation of the bgl operon by adaptive mutation," *Molecular Biology and Evolution*, vol. 15, no. 1, pp. 1–5, 1998.

[180] K. Schentz, "Silencing of Escherichia coli bgl promoter by flanking sequence elements," *EMBO Journal*, vol. 14, no. 11, pp. 2545–2550, 1995.

[181] L. Barquist, M. Mayho, C. Cummins, A. K. Cain, C. J. Boinett, A. J. Page, G. C. Langridge, M. A. Quail, J. A. Keane, and J. Parkhill, "The TraDIS toolkit: Sequencing and analysis for dense transposon mutant libraries," *Bioinformatics*, vol. 32, no. 7, pp. 1109–1111, 2016.

[182] A. T. Hagemann and N. L. Craig, "Tn7 transposition creates a hotspot for homologous recombination at the transposon donor site," *Genetics*, vol. 133, no. 1, pp. 9–16, 1993.

[183] J. L. Garland and A. L. Mills, "Classification and characterization of heterotrophic microbial communities on the basis of patterns of community-level sole-carbon-source utilization," *Applied and Environmental Microbiology*, vol. 57, no. 8, pp. 2351–2359, 1991.

[184] L. M. Durso, D. Smith, and R. W. Hutkins, "Measurements of fitness and competition in commensal Escherichia coli and E. coli O157:H7 strains," *Applied and Environmental Microbiology*, vol. 70, no. 11, pp. 6466–6472, 2004.

[185] S. Schaefler and A. Malamy, "Taxonomic investigations on expressed and cryptic phospho-beta-glucosidases in Enterobacteriaceae.," *Journal of bacteriology*, vol. 99, no. 2, pp. 422–433, 1969.

[186] M. AbuOun, P. F. Suthers, G. I. Jones, B. R. Carter, M. P. Saunders, C. D. Maranas, M. J. Woodward, and M. F. Anjum, "Genome scale reconstruction of a salmonella metabolic model: Comparison of similarity and differences with a commensal escherichia coli strain," *Journal of Biological Chemistry*, vol. 284, no. 43, pp. 29480–29488, 2009.

[187] R. M. Cicchillo, M. A. Baker, E. J. Schnitzer, E. B. Newman, C. Krebs, and S. J. Booker, "Escherichia coli L-serine deaminase requires a [4Fe-4S] cluster in catalysis," *Journal of Biological Chemistry*, vol. 279, no. 31, pp. 32418–32425, 2004.

[188] E. B. Newman, D. Dumont, and C. Walker, "In vitro and in vivo activation of L-serine deaminase in Escherichia coli K-12," *Journal of Bacteriology*, vol. 162, no. 3, pp. 1270–1275, 1985.

[189] T. L. Ruegg, J. H. Pereira, J. C. Chen, A. DeGiovanni, P. Novichkov, V. K. Mutalik, G. P. Tomaleri, S. W. Singer, N. J. Hillson, B. A. Simmons, P. D. Adams, and M. P. Thelen, "Jungle Express is a versatile repressor system for tight transcriptional control," *Nature Communications*, vol. 9, 12 2018.

[190] I. Prasad, B. Young, and S. Schaefler, "Genetic determination of the constitutive biosynthesis of phospho β glucosidase A in Escherichia coli K-12," *Journal of Bacteriology*, vol. 114, no. 3, pp. 909–915, 1973.

[191] A. S. Wong, G. C. Choi, and T. K. Lu, "Deciphering Combinatorial Genetics," *Annual Review of Genetics*, vol. 50, no. 1, pp. 515–538, 2016.

[192] D. Thibault, P. A. Jensen, S. Wood, C. Qabar, S. Clark, M. G. Shainheit, R. R. Isberg, and T. van Opijnen, "Droplet Tn-Seq combines microfluidics with Tn-Seq for identifying complex single-cell phenotypes," *Nature Communications*, vol. 10, no. 1, 2019.

[193] T. Woyke, D. F. Doud, and F. Schulz, "The trajectory of microbial single-cell sequencing," *Nature Methods*, vol. 14, no. 11, pp. 1045–1054, 2017.

[194] A. Kuchina, L. M. Brettner, L. Paleologu, C. M. Roco, A. B. Rosenberg, A. Carignano, R. Kibler, M. Hirano, R. W. DePaolo, and G. Seelig, "Microbial single-cell RNA sequencing by split-pool barcoding," *Science*, vol. 371, no. 6531, 2021.

[195] K. L. Frieda, J. M. Linton, S. Hormoz, J. Choi, K.-H. K. Chow, Z. S. Singer, M. W. Budde, M. B. Elowitz, and L. Cai, "Synthetic recording and in situ readout of lineage information in single cells," *Nature*, vol. 541, pp. 107–111, 1 2017.

[196] R. Kalhor, K. Kalhor, L. Mejia, K. Leeper, A. Graveline, P. Mali, and G. M. Church, "Developmental barcoding of whole mouse via homing CRISPR," *Science*, vol. 361, 8 2018.

[197] S. Bhattarai-Kline, S. K. Lear, C. B. Fishman, S. C. Lopez, E. R. Lockshin, M. G. Schubert, J. Nivala, G. M. Church, and S. L. Shipman, "Recording gene expression order in DNA by CRISPR addition of retron barcodes," *Nature*, vol. 608, pp. 217–225, 8 2022.

[198] A. Mitchell, G. H. Romano, B. Groisman, A. Yona, E. Dekel, M. Kupiec, O. Dahan, and Y. Pilpel, "Adaptive prediction of environmental changes by microorganisms," *Nature*, vol. 460, pp. 220–224, 7 2009.

[199] E. Toprak, A. Veres, J. B. Michel, R. Chait, D. L. Hartl, and R. Kishony, "Evolutionary paths to antibiotic resistance under dynamically sustained drug selection," *Nature Genetics*, vol. 44, no. 1, pp. 101–105, 2012.

[200] P. L. Freddolino and S. Tavazoie, "Beyond Homeostasis: A Predictive-Dynamic Framework for Understanding Cellular Behavior," *Annual Review of Cell and Developmental Biology*, vol. 28, pp. 363–384, 11 2012.

[201] P. L. Freddolino and S. Tavazoie, "Beyond Homeostasis: A Predictive-Dynamic Framework for Understanding Cellular Behavior," *Annual Review of Cell and Developmental Biology*, vol. 28, pp. 363–384, 11 2012.

[202] P. E. M. Purnick and R. Weiss, "The second wave of synthetic biology: From modules to systems," *Nature Reviews Molecular Cell Biology*, vol. 10, no. 6, pp. 410–422, 2009.

[203] N. Nandagopal and M. B. Elowitz, "Synthetic biology: Integrated gene circuits," *Science*, vol. 333, no. 6047, pp. 1244–1248, 2011.

[204] S. Mukherji and A. Van Oudenaarden, "Synthetic biology: Understanding biological design from synthetic circuits," *Nature Reviews Genetics*, vol. 10, no. 12, pp. 859–871, 2009.

[205] V. M. Isabella, B. N. Ha, M. J. Castillo, D. J. Lubkowicz, S. E. Rowe, Y. A. Millet, C. L. Anderson, N. Li, A. B. Fisher, K. A. West, P. J. Reeder, M. M. Momin, C. G.

Bergeron, S. E. Guilmain, P. F. Miller, C. B. Kurtz, and D. Falb, "Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria," *Nature Biotechnology*, vol. 36, no. 9, 2018.

[206] A. Mitchell and Y. Pilpel, "A mathematical model for adaptive prediction of environmental changes by microorganisms," *Proceedings of the National Academy of Sciences*, vol. 108, pp. 7271–7276, 4 2011.

[207] Y. Ram, E. Dellus-Gur, M. Bibi, K. Karkare, U. Obolski, M. W. Feldman, T. F. Cooper, J. Berman, and L. Hadany, "Predicting microbial growth in a mixed culture from growth curve data," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 29, pp. 14698–14707, 2019.

[208] N. Palani, "Transposon insertion sequencing (Tn-seq) library preparation protocol - includes UMI for PCR duplicate removal," *protocols.io*, 2019.

[209] S. Chen, Y. Zhou, Y. Chen, and J. Gu, "Fastp: An ultra-fast all-in-one FASTQ preprocessor," *Bioinformatics*, vol. 34, no. 17, pp. i884–i890, 2018.

[210] B. Langmead and S. L. Salzberg, "Fast gapped-read alignment with Bowtie 2," *Nature Methods*, vol. 9, no. 4, pp. 357–359, 2012.

[211] T. Smith, A. Heger, and I. Sudbery, "UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy," *Genome Research*, vol. 27, no. 3, pp. 491–499, 2017.

[212] Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nussbaum, R. M. Myers, M. Brown, W. Li, and X. S. Shirley, "Model-based analysis of ChIP-Seq (MACS)," *Genome Biology*, vol. 9, no. 9, 2008.

[213] A. R. Quinlan and I. M. Hall, "BEDTools: A flexible suite of utilities for comparing genomic features," *Bioinformatics*, vol. 26, no. 6, pp. 841–842, 2010.

[214] K. Blighe, S. Rana, and M. Lewis, "EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling.," 2018.

[215] W. Shen, S. Le, Y. Li, and F. Hu, "SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation," *PLoS ONE*, vol. 11, no. 10, pp. 1–10, 2016.

[216] L. Zhao, Z. Liu, S. F. Levy, and S. Wu, "Bartender: A fast and accurate clustering algorithm to count barcode reads," *Bioinformatics*, vol. 34, no. 5, pp. 739–747, 2018.