Scalable Platforms for Computation and Memory in Living Cells

by

Fahim Farzadfard

Submitted to the Microbiology Graduate Program in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2018

Author .....

Microbiology Graduate Program
January 11, 2017

Certified by .....

Timothy K. Lu
Associate Professor of Biological Engineering and
of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by .....

Kristala Jones Prather
Arthur D. Little Professor of Chemical Engineering,
Co-Director of the Microbiology Graduate Program

Scalable Platforms for Computation and Memory in Living Cells

by

Fahim Farzadfard

Submitted to the Microbiology Graduate Program
on January 11, 2018, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Living cells are biological computers – constantly sensing, processing and responding to biological cues they receive over time and space. Devised by evolution, these biological machines are capable of performing many computing and memory operations, some of which are analogous to and some are distinct from man-made computers. The ability to rationally design and dynamically control genetic programs in living cells in a robust and scalable fashion offers unprecedented capacities to investigate and engineer biological systems and holds a great promise for many biotechnological and biomedical applications. In this thesis, I describe foundational platforms for computation and memory in living cells and demonstrate strategies for investigating biology and engineering robust, scalable, and sophisticated cellular programs. These include platforms for genomically-encoded analog memory (SCRIBE – Chapter 2), efficient and generalizable DNA writers for spatiotemporal recording and genome engineering (HiSCRIBE – Chapter 3), single-nucleotide resolution digital and analog computing and memory (DOMINO – Chapter 4), concurrent, autonomous and high-capacity recording of signaling dynamics and events histories for cell lineage mapping with tunable resolution (ENGRAM – Chapter 5), continuous *in vivo* evolution and synthetic Lamarckian evolution (DRIVE – Chapter 6), tunable and multifunctional transcriptional factors for gene regulation in eukaryotes (crisprTF – Chapter 7), and an unbiased, high-throughput and combinatorial strategy for perturbing transcriptional networks for genetic screening (PRISM – Chapter 8). I envision the platforms and approaches described herein will enable broad applications for investigating basic biology and engineering cellular programs.

Thesis Supervisor: Timothy K. Lu
Title: Associate Professor of Biological Engineering and of Electrical Engineering and Computer Science

Acknowledgments

My time in graduate school has been a stimulating and rewarding journey, both personally and intellectually, and I am immensely grateful to all the people who contributed in different ways to writing this thesis.

First and foremost, I want to thank my mentor, Prof. Timothy Lu; for creating a multidisciplinary and intellectually-stimulating environment in his lab in which I could thrive and pursue my scientific curiosities, for giving me the freedom to explore my ideas and learn, for guiding me how to identify and critically approach fundamental scientific questions and technological needs, and for his guidance and constant support throughout my PhD.

I am grateful to my thesis committee members – Profs. Jim Collins, Michael Laub, David Bartel and Kevin Esvelt – for their mentorship and guidance and for making time in their busy schedules to meet me. Likewise, I am also grateful to Prof. George Church for serving on my thesis defense committee.

I am grateful to the MIT Microbiology Graduate Program (in particular, Profs. Alan Grossman, Michael Laub, and Kris Prather) which gave me the opportunity to join the MIT community and shaped my graduate education.

I am grateful to my friends and collaborators in the Lu lab for their camaraderie: Jacob Rubens, Robert Citorik, Mark Mimee, Ying-Chou Chen, Isaak Mueller, Ramez Daniel, Samuel Perli, Kevin Yehl, Yasutomi Higashikuni, Giyoung Jung, Zijay Tang, Eleonore Tham, William Chen, Jicong Cao and others. I also would like to thank Ky Lowenhaupt, for her tireless efforts to ensure everything runs smoothly in the lab.

I am grateful to my parents, Mohammad and Mansoureh, and my sisters, Fahimeh, Farzaneh, and Farahnaz for their endless love, support, and inspiration throughout different chapters of my life. I am also grateful to my in-laws, for their continuous support.

Finally, I am grateful to my wife, friend, and collaborator, Nava, who supported me constantly and unconditionally throughout this journey. Closing this chapter of my life, I feel immensely fortunate and elated knowing I have her by my side in my future journeys.

Table of Contents

Chapter 1 : Introduction

To describe life at the molecular level, Schrodinger postulated the idea of an "aperiodic crystal" that contained hereditary (i.e., genetic) information in its configuration (*1*). Avery, MacLeod, and McCarty showed that DNA is the molecule carrying genetic information (*2*). Watson, Crick, Wilkins, and Franklin deciphered the structure of DNA and demonstrated that genetic information is in fact encoded within the configuration of nucleobases of DNA molecules (*3, 4*). Subsequently, Sanger and Gilbert pioneered methods to sequence DNA molecules (*5, 6*), initiating efforts that eventually led to reading information encoded in the entire human genome.

Around the same time that the molecular nature of DNA was being deciphered, principles of computing machinery and intelligence were being formulated by Turing, Von-Neumann, and others (*7, 8*). These pioneering works initiated numerous theoretical and technological advances that led to the "Computer Age" and development of powerful man-made machines (computers), capable of processing and storing a massive amount of data based on a series of defined logics and rules (i.e., programs).

While comparing living cells with these man-made computers might seem far-fetched at first, and there are certainly major differences between the ways that these machines process and store information, one could see striking similarities between the two. Analogous to man-made computers, living cells constantly sense environmental cues, compute these external signals and internal cellular states (i.e., inputs) based on their genetic programs, and subsequently generate various phenotypic responses (i.e., outputs). Thus, living cells can be considered as biological computers – those that use biomolecules for computation and seek to optimize their fitness in a given environment.

Given this analogy, Turing's formulation of artificial intelligence and Darwin's formulation of evolution seem strikingly similar; they both describe a machine that perceives its environment and tries to maximize its chance of success at some goal(s) (learn/evolve). Both formulations imply the involvement of memory, time, and fitness and highlight striking analogies between concepts of life, evolution, intelligence, and learning. A man-made machine could use silicon-based memory as the information storage medium and perform fast computation using electronic circuits to maximize its fitness toward a defined goal. On the other hand, living cells use various biological

media for information storage and exploit different biomolecule-mediated strategies to compute and respond to inputs that they receive. For example, these machines use genomic DNA for long-term information storage. On the other hand, they use alternative media such as protein configuration, protein modification states, epigenetic marks and neural synapses for information storage in shorter timescales. Furthermore, short-term information processing and responses are often mediated via a variety of transient molecular events and programs ranging from neural pulses to protein modifications and transcriptional and post-translational programs while long-term responses often involve permanent genetic changes.

Living cells can be considered as evolvable functional memories. In these biological machines, genomic DNA encodes current (genetic) state and genetic programs required for short-term responses, and at the same time, records the history of adaptive changes over evolutionary timescales with the goal of optimizing cellular programs and cells overall fitness. Thus, genomic DNA provides the main medium for memory storage over long timescales and substrate for evolution. The ability to dynamically read and write information encoded in the genome and rationally design and control cellular programs in a robust and scalable fashion over space and time offers powerful strategies for investigating basic cellular biology and building sophisticated cellular programs for many biotechnological and biomedical applications.

In my thesis work, I have developed multiple scalable platforms for computation and memory in living cells. These platforms enable to dynamically read and write information stored in genomic DNA and thus control cellular programs in living cells. The next five chapters describe DNA writing technologies that enable to precisely manipulate information stored in genomic DNA along with several applications that can be achieved with these technologies. The following two chapters feature a platform for transcriptional regulation in living cells and its application for transcriptional perturbation and genetic screening.

Specifically, in the second chapter, I introduce the concept of analog memory in living cells and describe SCRIBE (Synthetic Cellular Recorders Integrating Biological Events), a modular and scalable platform for the continuous and long-term recording of molecular events directly into the genomic DNA of living cells. SCRIBE is an autonomous DNA writing platform that enables to record analog (i.e., continuous) information into distributed genomic DNA of cell populations. By converting genomic DNA into a “tape recorder”, SCRIBE opens up the entire genomic space for artificial

memory storage and overcomes the limited scalability and recording capacity of existing cellular memories.

The third chapter describes an improved and high-efficiency SCRIBE (HiSCRIBE) DNA writing platform that enables efficient manipulation of bacterial genomes. Using these efficient and generalizable DNA writers, I demonstrate that transient spatiotemporal molecular events, such as cellular connectome, can be recorded into genomic DNA for later retrieval by sequencing. I further show that these DNA writers can be used to efficiently edit genomic DNA within bacterial communities.

In the fourth chapter, I introduce a highly robust and scalable DNA writing platform for implementing computing and memory operations in living cells. This platform, dubbed DOMINO for DNA-based Ordered Memory and Iteration Network Operators, uses a single-nucleotide-resolution read-write head to efficiently manipulate genomic DNA and execute cascades of DNA writing events in living cells. Using genomic DNA as a medium for computation and memory, DOMINO enables to encode various forms of scalable order-independent, sequential and temporal digital logic and memory, analog computing and memory, and associative learning circuits in living cells. I further show that DOMINO operators can be used to autonomously and continuously record signaling dynamics and molecular events in living cells.

The fifth chapter features the ENGRAM molecular recording platform (ENGineered Random Accumulative Memory). Similar to DOMINO, ENGRAM enables concurrent recording of both signaling dynamics and cellular event histories (lineages) in a continuous and autonomous fashion. However, ENGRAM recorders offer a more compact design and much higher recording capacity than DOMINO, which makes them especially useful for recording cellular lineage maps with tunable resolution.

In the sixth chapter, I demonstrate the efficient and precise DNA writing enabled by HiSCRIBE and DOMINO DNA writers can be used to introduce targeted genetic diversity within cell populations. I show that these diversity generation strategies can be coupled to continuous selection, and be used to continuously tune cellular phenotypes with minimal human intervention. In a series of experiments, using a strategy referred to as DRIVE (for Directed and Recurring *In Vivo* Evolution), I further demonstrate how these writers can be used to introduce *de novo* targeted diversity into desired genomic loci, thus endowing cells with the ability to undergo synthetic Lamarckian evolution and evolve faster than possible by natural Darwinian evolution.

The seventh chapter features crisprTF, a tunable and multifunctional class of synthetic transcriptional factors based on CRISPR-Cas9 for regulation of gene expression in eukaryotic cells. These tunable and orthogonal transcription factors offer powerful tools for both building highly scalable synthetic gene programs and perturbing cells natural genetic programs for basic study of biology.

The eighth chapter describes the PRISM (Perturbing Regulatory Interactions by Synthetic Modulators), a crisprTF-based screening platform for unbiased, high-throughput and combinatorial perturbations of transcriptional networks to study complex, multilayered phenotypes. PRISM was used to screen for and identify genes involved in Parkinson disease models.

The last chapter outlines future research directions that can be achieved and explored by the platforms and approaches described in this thesis.

Chapter 2: SCRIBE

Genomically-encoded Analog Memory in Living Cells

This chapter is adapted from

Farzadfard, F. and Lu, T.K., 2014. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. Science, 346(6211), p.1256272. (http://dx.doi.org/10.1126/science.1256272).

to fit the format of this thesis (with permission from AAAS).

2.1 Abstract

Cellular memory is crucial to many natural biological processes and for sophisticated synthetic-biology applications. Existing cellular memories rely on epigenetic switches or recombinases, which are limited in scalability and recording capacity. Here, we use the DNA of living cell populations as genomic 'tape recorders' for the analog and distributed recording of long-term event histories. We describe a platform for generating single-stranded DNA (ssDNA) in vivo in response to arbitrary transcriptional signals. When co-expressed with a recombinase, these intracellularly expressed ssDNAs target specific genomic DNA addresses, resulting in precise mutations that accumulate in cell populations as a function of the magnitude and duration of the inputs. This platform could enable long-term cellular recorders for environmental and biomedical applications, biological state machines, and enhanced genome engineering strategies.

2.2 Introduction

Due to its high storage capacity, durability, ease of duplication, and high-fidelity maintenance of information, DNA has garnered much interest as an artificial storage medium (*9, 10*). However, existing technologies for *in vivo* autonomous recording of information in cellular memory are limited in their storage capacity and scalability (*11*). Epigenetic memory devices such as bi-stable toggle switches (*12-15*) and positive-feedback loops (*16*) require orthogonal transcription factors and can lose their digital state due to environmental fluctuations or cell death. Recombinase-based memory devices enable the writing and storage of digital information in the DNA of living cells (*17-20*), where binary bits of information are stored in the orientation of large stretches of DNA. However, these devices do not efficiently exploit the full capacity of DNA for

information storage – recording a single bit of information with these devices often requires at least a few hundred base-pairs of DNA, overexpression of a recombinase protein to invert the target DNA, and engineering recombinase recognition sites into target loci in advance. The scalability of this type of memory is further limited by the number of orthogonal recombinases that can be used in a single cell. Finally, epigenetic and recombinase-based memory devices described to-date store digital information and their recording capacity is exhausted within a few hours of induction. Thus, these devices have not been adapted to record analog information, such as magnitude and time course of inputs over extended periods of time (i.e., multiple days or more).

Here, we introduce SCRIBE (Synthetic Cellular Recorders Integrating Biological Events), a compact, modular strategy for producing single-stranded DNA (ssDNA) inside of living cells in response to a range of regulatory signals, such as small chemical inducers and light. These ssDNAs uniquely address specific target loci based on sequence homology and introduce precise mutations into genomic DNA. The memory device can be easily reprogrammed to target different genomic locations by changing the ssDNA template. SCRIBE memory does not just record the absence or presence of arbitrary inputs (digital signals represented as binary '0s' or '1s'). Instead, by encoding information into the collective genomic DNA of cell populations, SCRIBE can track the magnitude and long-term temporal behavior of inputs, which are analog signals since they can vary over a wide range of continuous values. This analog memory architecture leverages the large number of cells in bacterial cultures for distributed information storage and archives event histories in the fraction of cells in a population that carry specific mutations.

2.3 Results

Single-Stranded DNA Expression in Living Cells

Previously, it was shown that synthetic oligonucleotides delivered by electroporation into cells that overexpress Beta recombinase (from bacteriophage λ) in *Escherichia coli* (*E. coli*) are specifically and efficiently recombined into homologous genomic sites (*21-24*). Thus, oligonucleotide-mediated recombination offers a powerful way to introduce targeted mutations in a bacterial genome (*25, 26*). However, this technique requires the exogenous delivery of ssDNAs and cannot be used to couple arbitrary signals into genetic memory. To overcome these limitations, we developed a genome-editing platform based on expressing ssDNAs inside of living cells by taking advantage of a widespread class of bacterial reverse transcriptases called retrons (*27, 28*).

The wild-type retron cassette encodes three components in a single transcript – a reverse transcriptase protein (RT), and two RNA moieties, msr and msd, which act as the primer and the template for the reverse transcriptase, respectively (Fig. 2.1A, left). The *msr-msd* sequence in the retron cassette is flanked by two inverted repeats. Once transcribed, the msr-msd RNA folds into a secondary structure guided by the base-pairing of the inverted repeats and the msr-msd sequence. The RT recognizes this secondary structure and uses a conserved guanosine residue in the msr as a priming site to reverse transcribe the msd sequence and produce a hybrid RNA-ssDNA molecule called msDNA (*28, 29*). To couple the expression of ssDNA to an external input, the wild-type Ec86 retron cassette from *E. coli* BL21 (*29*) was placed under the control of an Isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible promoter (P_{lacO}) in *E. coli* DH5αPRO cells (*30*), which expresses high levels of the LacI and TetR repressors (Fig. 1A). The wild-type retron ssDNA (ssDNA(wt)) was readily detected in IPTG-induced cells while no ssDNA was detected in non-induced cells (Fig. 2.1B). The identity of the detected ssDNA band was further confirmed by DNA sequencing (Fig. S2.1). To verify that ssDNA expression depended on RT activity, point mutations (D197A and D198A) were introduced to the active site of the RT to make a catalytically dead RT (dRT) (*31*). This modification completely abolished ssDNA production (Fig. 2.1B).

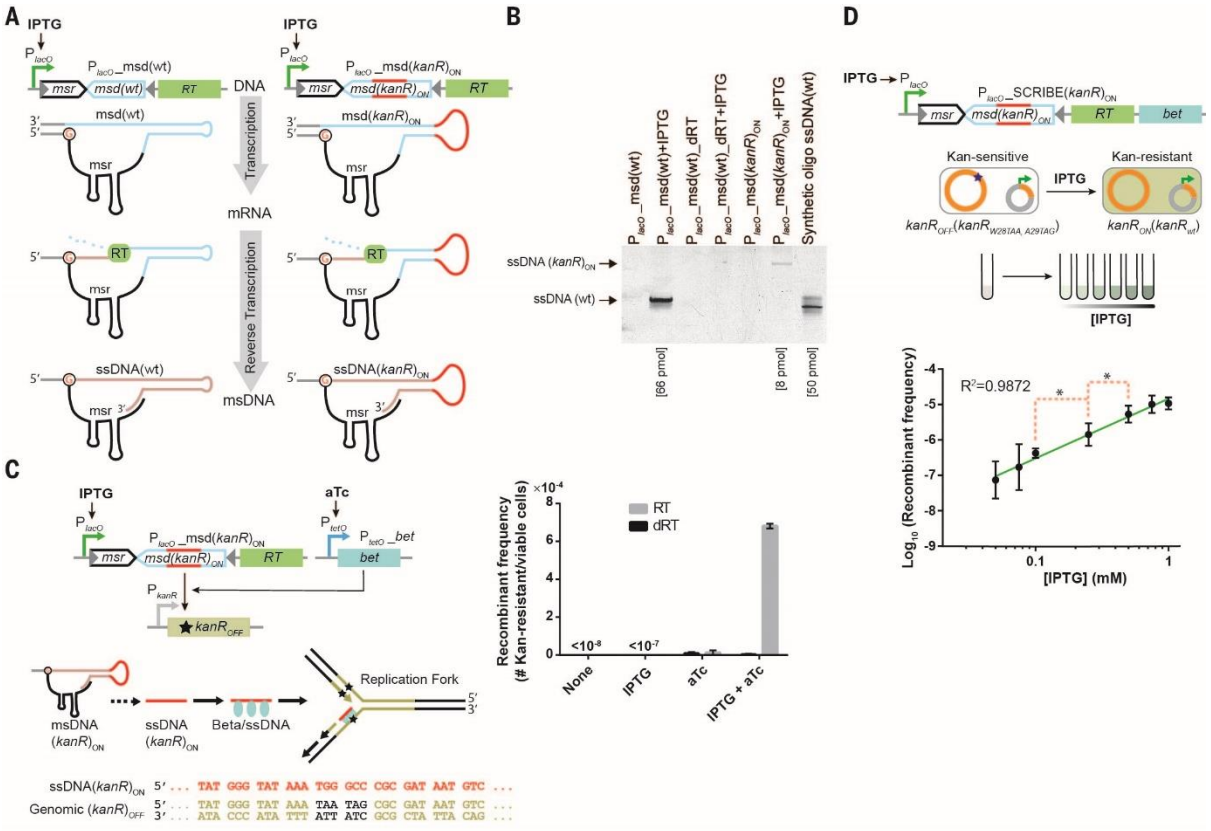


Fig. 2.1 | SCRIBE system for recording inputs in the distributed genomic DNA of bacterial populations. (A) Synthetic ssDNA (red line) generation inside of living cells by retrons. **(B)** Visualization of retron-mediated ssDNAs produced in living bacteria. The amount of ssDNA in each sample (shown in brackets) was calculated by densitometry. **(C)** A *kanR* reversion assay was used to measure the efficiency of DNA writing within living cells, where the msd(*kanR*)$_{ON}$ cassette and the *bet* gene were inducible by IPTG and aTc, respectively. **(D)** Demonstration of analog memory achieved via SCRIBE to record the magnitude of an input into genomic DNA. The green line is a linear regression fit. The red dashed brackets marked with asterisks connect the closest data points that are statistically significant with respect to each other (p-value < 0.05 based on one-tailed Welch's t-test). Error bars indicate the standard error of the mean for three independent biological replicates.

To engineer the msd template to express synthetic ssDNAs of interest, we initially tried to replace the whole *msd* sequence with a desired template. However, no ssDNA was detected, suggesting that some features of msd are required for ssDNA expression, as was previously noted for another retron (*32*). A variant in which the flanking regions of the msd stem remained intact (Fig. 2.1A, right) produced detectable amounts of ssDNA when induced by IPTG (Fig. 2.1B, P$_{lacO}$msd(*kanR*)$_{ON}$ + IPTG). The correct identity of the detected ssDNA band was further confirmed by DNA sequencing (Fig. S2.1). Thus, the lower part of the msd stem is essential for reverse transcription while the upper part of the stem and the loop are dispensable and can be replaced with desired templates to produce ssDNAs of interest *in vivo*.

Regulated Genome Editing with *In vivo* ssDNAs

To demonstrate that intracellularly expressed ssDNAs can be recombined into target genomic loci by concomitant expression of Beta, we developed a selectable marker reversion assay (Fig. 2.1C). The *kanR* gene, which encodes neomycin phosphotransferase II and confers resistance to kanamycin (Kan), was integrated into the *galK* locus. Two stop codons were then introduced into the genomic *kanR* to make a Kan-sensitive *kanR*$_{OFF}$ reporter strain (DH5αPRO *galK::kanR*$_{W28TAA,\ A29TAG}$). These premature stop codons could be reverted back to the wild-type sequence via recombination with engineered ssDNA(*kanR*)$_{ON}$, thus conferring kanamycin resistance (Fig. 2.1C). Specifically, ssDNA(*kanR*)$_{ON}$ contains 74 base-pairs (bps) of homology to the regions of the *kanR*$_{OFF}$ locus flanking the premature stop codons, and replaces the stop codons with the wild-type *kanR* gene sequence (Fig. 2.1C).

We cloned the Beta gene (*bet*) into a plasmid under the control of the anhydrotetracycline (aTc)-inducible P$_{tetO}$ promoter and introduced it along with the IPTG-inducible msd(*kanR*)$_{ON}$ construct into the *kanR*$_{OFF}$ strain (Fig. 2.1C). Induction of cultures harboring these two plasmids with either IPTG (1 mM) or aTc (100 ng/ml) resulted in a slight increase in the number of the Kan-resistant cells (Fig. 2.1C).

However, co-expression of both ssDNA($kanR$)$_{ON}$ and Beta with IPTG and aTc resulted in a >10^4-fold increase in the recombinant frequency relative to the non-induced cells. This corresponded to a >10^3-fold increase relative to cells induced with IPTG only and a 60-fold increase relative to cells induced with aTc only. This increase in the recombinant frequency was dependent on the RT activity, as it was largely abolished with dRT. The genotypes of randomly selected Kan-resistant colonies were further confirmed by DNA sequencing to contain precise reversions of the two codons to the wild-type sequence (Fig. S2.1). No Kan-resistant colonies were detected when a non-specific ssDNA (ssDNA(wt)) was co-expressed with Beta in the $kanR_{OFF}$ reporter cells, confirming that Kan-resistant cells were not produced due to spontaneous mutations. In additional experiments, high-throughput sequencing (Illumina HiSeq) on the bacterial populations was used to analyze the genomically encoded memory (Supplementary Materials, Fig. S2.2). Comparable recombinant frequencies were obtained from both the plating assay and sequencing, confirming that genomically encoded memory can be read without the need for functional assays and reporters.

Recording Input Magnitudes into Genomic Memory

We reasoned that the rate of recombination between engineered ssDNAs and genomic DNA could be effectively modulated by changing expression levels of the engineered retron cassette and Beta. This feature would enable the recording of analog information, such as the magnitude of an input signal, in the proportion of cells in a population with a specific mutation in genomic DNA. To demonstrate this, both the ssDNA($kanR$)$_{ON}$ expression cassette and *bet* were placed into a single synthetic operon (hereafter referred to as the SCRIBE($kanR$)$_{ON}$ cassette) under the control of P_{lacO} (Fig. 2.1D). The $kanR_{OFF}$ reporter cells harboring this synthetic operon were induced with different concentrations of IPTG. The fraction of Kan-resistant recombinants increased linearly with the input inducer concentration on a log-log plot over a range of ~10^{-7} to ~10^{-5} (Fig. 2.1D). Statistical tests showed that at least four different concentrations of the inducer (including 0 mM IPTG) could be resolved in this experiment. Thus, the efficiency of genome writing in a population can be quantitatively tuned with external inputs.

Writing and Rewriting Genomic Memory

We next created a complementary set of SCRIBE cassettes to write and erase (rewrite) information in the genomic *galK* locus using two different chemical inducers. Cells expressing *galK* can metabolize and grow on galactose as the sole carbon source. However, these *galK*-positive ($galK_{ON}$) cells cannot metabolize 2-deoxy-galactose

(2DOG) and cannot grow on plates containing glycerol (carbon source) + 2DOG. On the other hand, *galK*-negative ($galK_{OFF}$) cells cannot grow on galactose as the sole carbon source but can grow on glycerol + 2DOG plates (*33*). We transformed DH5αPRO $galK_{ON}$ cells with plasmids expressing IPTG-inducible SCRIBE($galK$)$_{OFF}$ and aTc-inducible SCRIBE($galK$)$_{ON}$ cassettes (Fig. 2.2A). Induction of SCRIBE($galK$)$_{OFF}$ by IPTG resulted in the writing of two stop codons into $galK_{ON}$, leading to $galK_{OFF}$ cells that could grow on glycerol + 2DOG plates (Fig. 2.2B). Induction of SCRIBE($galK$)$_{ON}$ in these $galK_{OFF}$ cells with aTc reversed the IPTG-induced modification, leading to $galK_{ON}$ cells that could grow on galactose plates (Fig. 2.2C). These results show that writing on genomic DNA with SCRIBE is reversible and that distinct information can be written and rewritten into the same locus.

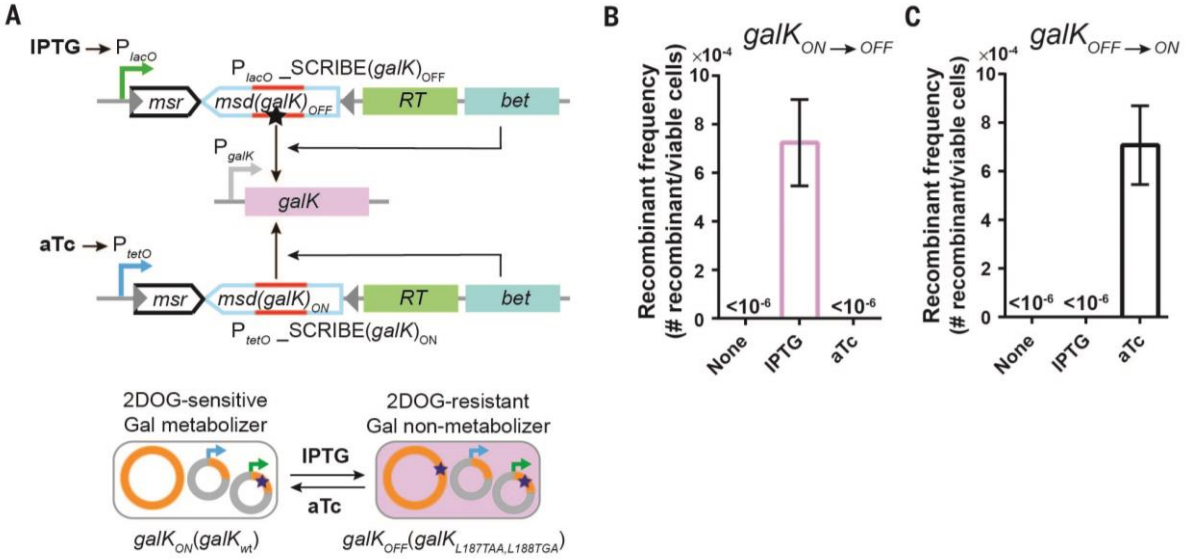


Fig. 2.2 | SCRIBE can write multiple different DNA mutations into a common target loci (*galK*). **(A)** Schematic of the procedure (see text for details). **(B)** $galK_{ON}$ cells harboring the circuits shown in A) were induced with either IPTG (1 mM) or aTc (100 ng/ml) for 24 hours and the $galK_{OFF}$ frequencies in the population were determined by plating the cells on appropriate selective conditions. **(C)** $galK_{OFF}$ cells (obtained from the experiment described in B)) were induced with IPTG (1 mM) or aTc (100 ng/ml) for 24 hours and the $galK_{ON}$ frequencies in the population were determined by plating the cells on appropriate selective conditions. Error bars indicate the standard error of the mean for three independent biological replicates.

Writing Multiple Mutations into Independent Loci

Scaling the capacity of previous memory devices is challenging since each additional bit of information requires new orthogonal proteins (e.g., recombinases or transcription factors). In contrast, orthogonal SCRIBE memory devices are potentially easier to scale because they can be built by simply changing the ssDNA template (*msd*). To

demonstrate this, we used SCRIBE to record multiple independent inputs into different genomic loci of bacterial population. We integrated the $kanR_{OFF}$ reporter gene into the *bioA* locus of DH5αPRO to create a $kanR_{OFF}$ $galK_{ON}$ strain. These cells were then transformed with plasmids expressing IPTG-inducible SCRIBE($kanR$)$_{ON}$ and aTc-inducible SCRIBE($galK$)$_{OFF}$ cassettes (Fig. 2.3A). Induction of these cells with IPTG or aTc resulted in the production of cells with phenotypes corresponding to $kanR_{ON}$ $galK_{ON}$ or $kanR_{OFF}$ $galK_{OFF}$ genotypes, respectively (Fig. 2.3B and C). Comparable numbers of $kanR_{ON}$ $galK_{ON}$ and $kanR_{OFF}$ $galK_{OFF}$ cells (~$2*10^{-4}$ and ~$3*10^{-4}$ recombinant/viable cells, respectively) were produced when the cultures were induced with both aTc and IPTG (Fig. 2.3C). Furthermore, very few individual colonies (~$3*10^{-7}$ recombinant/viable cells) containing both writing events ($kanR_{ON}$ $galK_{OFF}$) were obtained in the cultures that were induced with both aTc and IPTG (Figure 2.3C). These data suggest that while multiplexed writing at single-cell level is rare with SCRIBE's current level of recombination efficiency, multiple independent inputs can be successfully recorded into the distributed genomic DNA of bacterial subpopulations.

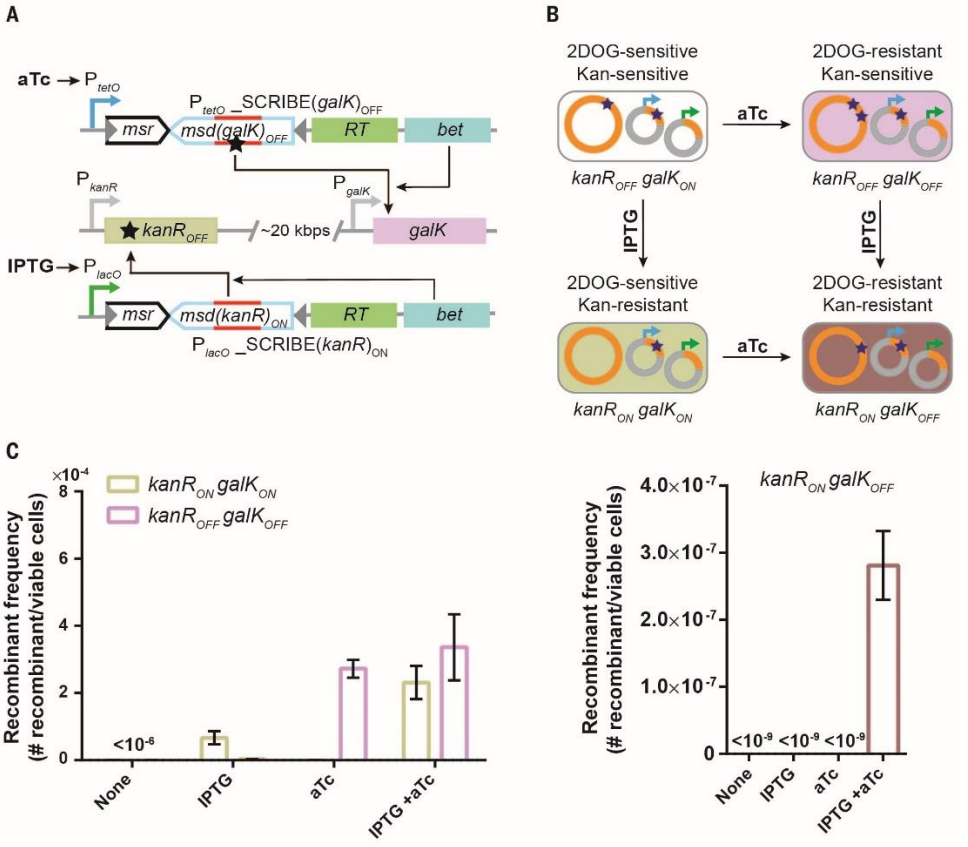


Fig. 2.3 | Writing multiple mutations into independent target loci within population. (A) Constructs used to target genomic $kanR_{OFF}$ and $galK_{ON}$ loci with IPTG-inducible and aTc-inducible SCRIBE cassettes, respectively. **(B)** Induction of $kanR_{OFF}$ $galK_{ON}$ cells with IPTG or aTc generates cells with the $kanR_{ON}$ $galK_{ON}$ or $kanR_{OFF}$ $galK_{OFF}$ genotypes, respectively.

Induction of *kanR*$_{OFF}$ *galK*$_{ON}$ cells with IPTG and aTc generates cells with the *kanR*$_{ON}$ *galK*$_{OFF}$ genotype. (C) *kanR*$_{OFF}$ *galK*$_{ON}$ reporter cells containing the circuits in A) were induced with different combinations of IPTG (1 mM) and aTc (100 ng/ml) for 24 h at 30°C and the fraction of cells with the various genotypes were determined by plating the cells on appropriate selective media. Error bars indicate the standard error of the mean for three independent biological replicates.

Optogenetic Genome Editing for Light-to-DNA Memory

In SCRIBE, the expression of each individual ssDNA can be triggered by any endogenous or exogenous signal that can be coupled into transcriptional regulation, thus recording these inputs into long-lasting DNA storage. In addition to small-molecule chemicals, we showed that light can be used to trigger specific genome editing for genomically encoded memory. We placed the SCRIBE(*kanR*)$_{ON}$ cassette under the control of a previously described light-inducible promoter (P$_{Dawn}$, (*34*)) within *kanR*$_{OFF}$ cells (Fig. 2.4A). These cultures were then grown for 4 days in the presence of light or in the dark (Fig. 2.4A). As Beta-mediated recombination is reportedly replication-dependent (*35-37*), dilutions of these cultures were made into fresh media at the end of each day to maintain active replication in the cultures. At the end of each day, samples were taken to determine the number of Kan-resistant and viable cells (Fig. 2.4A). Cultures grown in the dark yielded undetectable levels of Kan-resistant cells (Fig. 2.4A). In contrast, the number of Kan-resistant colonies increased steadily over time in the cultures that were grown in the presence of light, indicating the successful recording of light input into long-lasting DNA memory. The analog memory faithfully stored the total time of light exposure, rather than just the digital presence or absence of light.

Recording the Time Exposure of Inputs

The linear increase in the number of Kan-resistant colonies over time due to exposure to light indicates that the duration of inputs can be recorded into population-wide DNA memory using SCRIBE. To further explore population-wide genomically encoded memory whose state is a function of input exposure time, we used the *kanR*$_{OFF}$ strain harboring the constructs shown in Fig. 2.1C, where expression of ssDNA(*kanR*)$_{ON}$ and Beta are controlled by IPTG and aTc, respectively. These cells were subjected to four different patterns of the inputs for 12 successive days (patterns I-IV, Fig. 2.4B). Kan-resistant cells did not accumulate in the negative control (pattern I), which was never exposed to the inducers. The fraction of Kan-resistant cells in the three other patterns (II, III, and IV) increased linearly over their respective induction periods and remained relatively constant when the inputs were removed. These data indicate that the genomically encoded memory was stable in the absence of the inputs over the course

of the experiment. The recombinant frequencies in patterns III and IV, which were induced for the same total amount of time but with different temporal patterns, reached comparable levels at the end of the experiment. These data demonstrate that the genomic memory integrates over the total induction time and is independent of the input pattern, and therefore can be used to stably record event histories over many days.

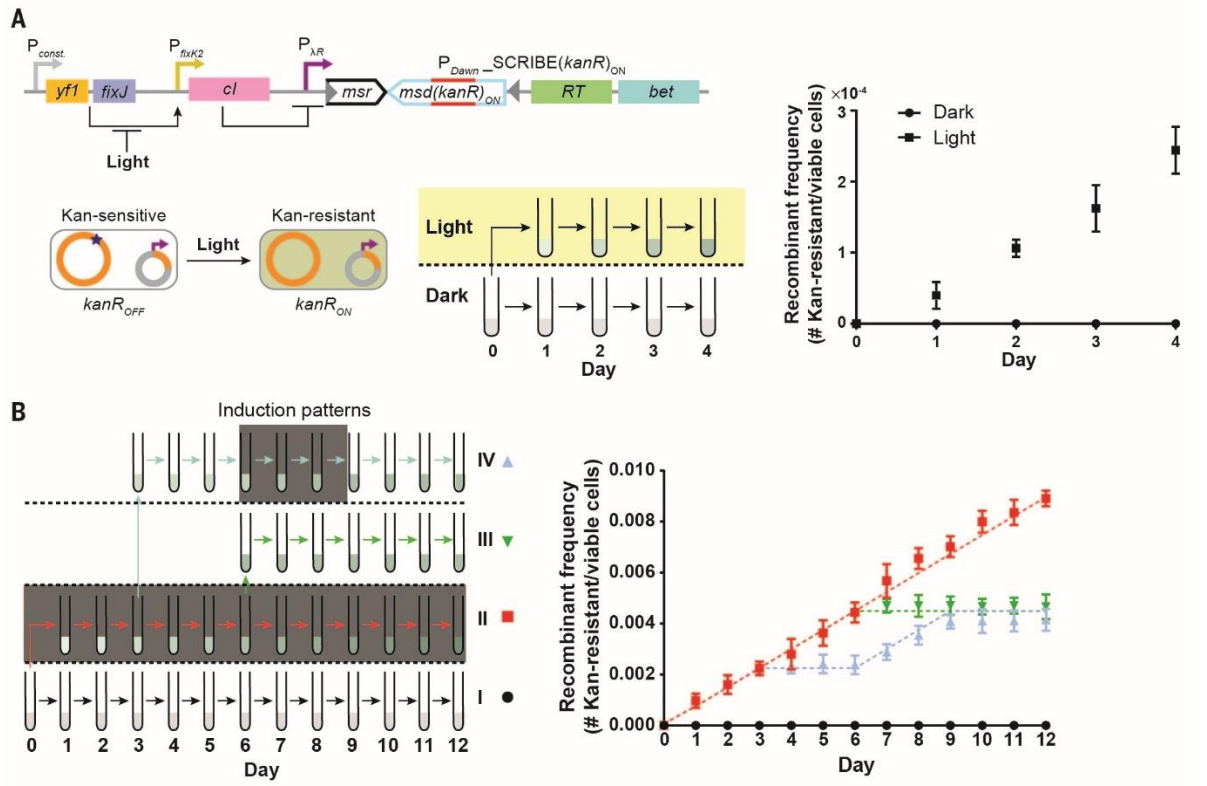


Fig. 4 | Optogenetic genome editing and analog memory for long-term recording of input signal exposure times in the genomic DNA of living cell populations. (A) We coupled expression of SCRIBE($kanR$)$_{ON}$ to an optogenetic system (P_{Dawn}). The *yf1/fixJ* synthetic operon was expressed from a constitutive promoter – its products cooperatively activate the P_{fixK2} promoter, which drives lambda repressor (*cI*) expression, which subsequently represses the SCRIBE($kanR$)$_{ON}$ cassette. Light inhibits the interaction between *yf1* and *fixJ*, leading to the generation of ssDNA($kanR$)$_{ON}$ and Beta expression, and thus the conversion of $kanR_{OFF}$ to $kanR_{ON}$. Cells harboring this circuit were grown overnight at 37°C in the dark, diluted 1:1000, and then incubated for 24 h at 30°C in the dark (no shading) or in the presence of light (yellow shading). Subsequently, cells were diluted by 1:1000 and grown for another 24 h at 30°C in the dark or in the presence of light. The dilution/regrowth cycle was performed for four consecutive days. The *kanR* allele frequencies in the populations were determined by sampling the cultures after each 24-hour period. **(B)** SCRIBE analog memory records the total time exposure to a given input, regardless of the underlying induction pattern. Cells harboring the circuit shown in Fig. 2.1C were grown in four different patterns (I-IV) over a twelve-day period, where induction by IPTG (1 mM) and aTc (100 ng/mL) is represented by dark gray shading. At the end of each 24 h incubation period, cells were diluted by 1:1000 into fresh media. The number of Kan-resistant cells in the cultures was determined at the end of each

day. Dashed lines represent the recombinant allele frequencies predicted by the model (see Supplementary Materials). Error bars indicate the standard error of the mean for three independent biological replicates.

The linear increase in the fraction of recombinants in the induced cell populations over time was consistent with a deterministic model (dashed lines in Fig. 2.4B, see Supplementary Materials). Specifically, when triggered by inputs, SCRIBE can significantly increase the rate of recombination events at a specific target site above the wild-type rate (which is reportedly $<10^{-10}$ events/generation in *recA*$^-$ background (*38*)). When recombination rates are ~10^{-4} events/generation, which is consistent with the recombination rate estimated for SCRIBE from data in Fig. 2.4B, a simple deterministic model as well as a detailed stochastic simulation both predict a linear increase in the total number of recombinant alleles in a population over time, as long as the frequency of recombinants in the population is less than a few percent and cells in the population are equally fit over the time-scale of interest (Supplementary Materials, Fig. S2.3 and S2.4). These models enable one to determine the ideal range of recombination efficiencies for a given application, which depends on parameters such as the frequency of dilution, the sensitivity of the method used for reading the memory, the desired input duration to be recorded, and so forth. For example, recombination rates that are too low would be challenging to quantify and could result in loss of memory if the cultures were diluted. Moreover, higher recombination rates lead to more rapid saturation of memory capacity in which the system is unable to provide a straightforward linear relationship between the time exposure of an input and the state of the memory (Fig. S2.3). Thus, intermediate levels of recombination efficiency are desirable for population-level analog memory units that can record the time-span of exposure to inputs (see Supplementary Materials).

Decoupling Memory Operations

SCRIBE memory can be used to create more complex synthetic memory circuits. To demonstrate this, we first built a synthetic gene circuit that can record different input magnitudes into DNA memory. The memory state can then be read out later (after the initial input is removed) upon addition of a secondary signal. Specifically, we built an IPTG-inducible $lacZ_{OFF}$ ($lacZ_{A35TAA,\ S36TAG}$) reporter construct in DH5αPRO cells (Fig. 2.5A). Expression of this reporter is normally repressed except when IPTG ("Read" signal, Fig. 2.5A) is added as an inducer, thus enabling a convenient and switchable population-level readout of the memory based on total LacZ activity (Fig. 2.5B). The $lacZ_{OFF}$ reporter cells were transformed with a plasmid encoding an aTc-inducible SCRIBE($lacZ$)$_{ON}$ cassette (Fig. 2.5A). Overnight cultures were diluted and

induced with various amounts of aTc to write the genomic memory (Fig. 2.5B). These cells were grown up to saturation and then diluted into fresh media in the presence or absence of IPTG to read the genomic memory (Fig. 2.5B). In the absence of IPTG, the total LacZ activity remained low, regardless of the aTc concentration. In the presence of IPTG, cultures that had been exposed to higher aTc concentrations had greater total LacZ activity. These results show that population-level reading of genomically encoded memory can be decoupled from writing and controlled externally. Furthermore, this circuit enables the magnitude of the inducer (aTc) to be stably recorded in the distributed genomic memory of a cellular population. Independent control over the "Read" memory operation as shown in this experiment could help to minimize fitness costs associated with the expression of reporter genes until needed.

We have shown that both ssDNA expression and Beta are required for writing into genomic memory (Fig. 2.1C), multiple ssDNAs can be used to independently address different memory units (Fig. 2.3), and genomic memory is stably recorded into DNA and can be used to modify functional genes whose expression can be controlled by external inducers (Fig. 2.1-2.4). Thus, SCRIBE memory units can be conceptually decomposed into separate "Input", "Write", and "Read" operations to facilitate greater control and the integration of logic with memory. The separation of these signals could enable master control over the writing of multiple independent inputs into genomic memory. To achieve this, we placed the msd($lacZ$)$_{ON}$ cassette under the control of an AHL-inducible promoter (P_{luxR}) (*39*) and co-transformed this plasmid with an aTc-inducible Beta-expressing plasmid into the $lacZ_{OFF}$ reporter strain (Fig. 2.5D). Using this design, information on the "Input" (ssDNA expression via addition of AHL) can be written into DNA memory only in the presence of the "Write" signal (Beta expression via addition of aTc). The information recorded in the memory register (i.e., the state of *lacZ* across the population) can be retrieved by adding the "Read" signal (IPTG).

To demonstrate this, overnight $lacZ_{OFF}$ cultures harboring the circuit shown in Fig. 2.5D were diluted and then grown to saturation in the presence of all four possible combinations of AHL and aTc (Fig. 2.5E). The saturated cultures were then diluted into fresh media in the absence or presence of IPTG. As shown in Fig. 2.5F, only cultures that had been exposed to both the "Input" and "Write" signals simultaneously showed significant LacZ activity, and only when they were induced with the "Read" signal. These results indicate that short stretches of DNA of living organisms can be used as addressable read/write memory registers to record transcriptional inputs.

Furthermore, SCRIBE memory can be combined with logic, such as the AND function between the "Input" and "Write" signals shown here. The logic in Fig. 2.5D enables this circuit to act as a "sample-and-hold" system in which information about an input can be recorded in the presence of another signal and read out at will. Additional "Inputs" in the form of orthogonal ssDNAs under the control of other inducible inputs (e.g., Fig. 2.3), could be written into genomic memory only when the "Write" signal (Beta expression) is present. Thus, SCRIBE memory units can be readily reprogrammed, integrated with logic circuits, and decomposed into independent "Input", "Write", and "Read" operations. We anticipate that more complex logic circuits could be combined with SCRIBE-based memory to create analog-memory-and-computation systems capable of storing the results of multi-input calculations (*40, 41*).

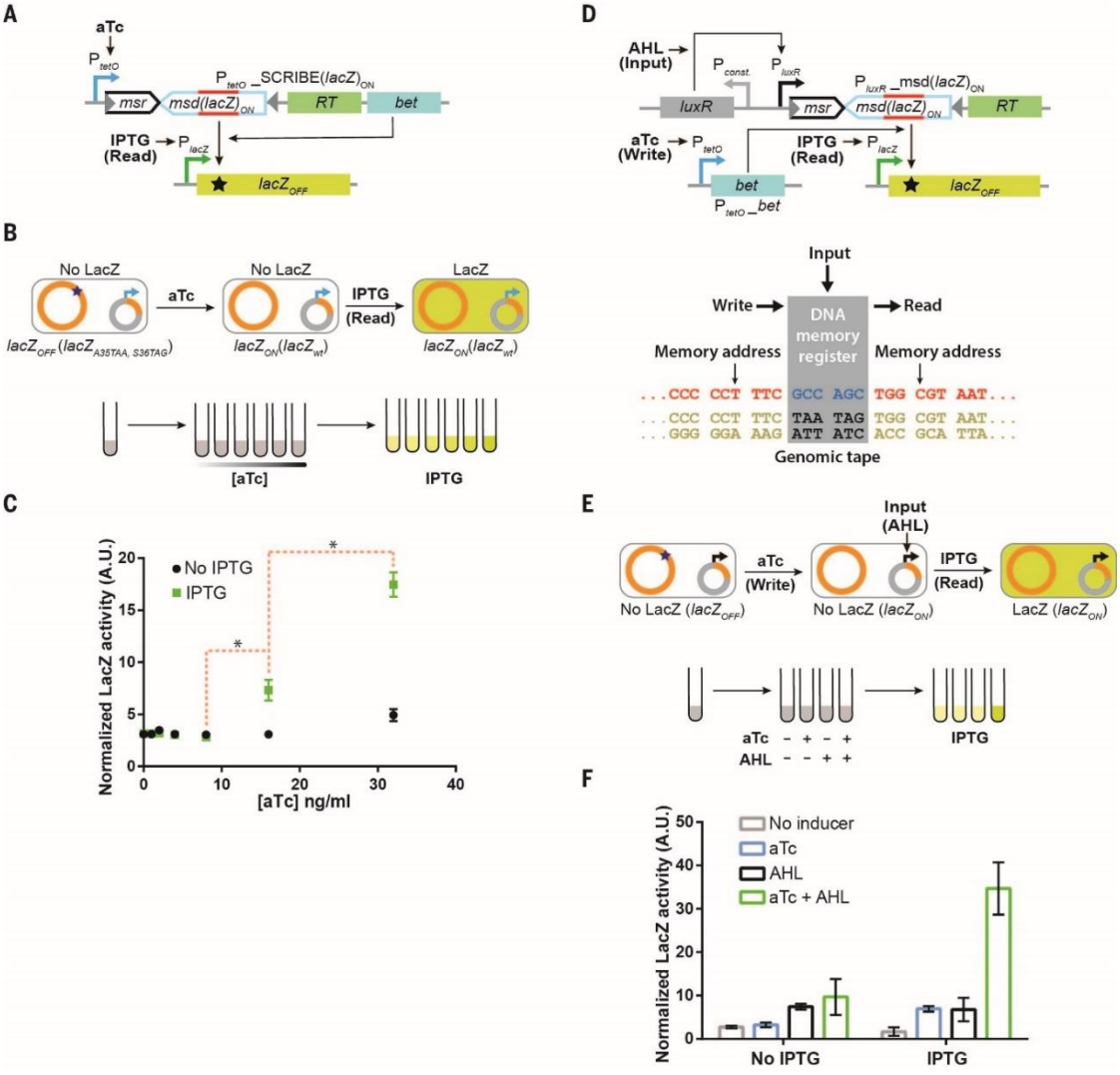


Fig. 2.5 | SCRIBE memory operations can be decoupled into independent Input, Write, and Read operations, thus facilitating greater control over addressable memory registers in genomic

tape recorders and the creation of sample-and-hold circuits. **(A)** We built a circuit where information about the first inducer (aTc) is recorded in the population, which can then be read later upon addition of a second inducer (IPTG) that triggers a "Read" operation. We created an IPTG-inducible $lacZ_{OFF}$ locus in the DH5α PRO background, which contains the full-length *lacZ* gene with two premature stop codons inside the open-reading frame. Expression of ssDNA$(lacZ)_{ON}$ from the aTc-inducible SCRIBE$(lacZ)_{ON}$ cassette results in the reversion of the stop codons inside $lacZ_{OFF}$ to yield the $lacZ_{ON}$ genotype. **(B)** Cells harboring the circuit shown in A) were grown in the presence of different levels of aTc for 24 h at 30°C to enable recording into genomic DNA. Subsequently, cell populations were diluted into fresh media without or with IPTG (1 mM) and incubated at 37°C for 8 hours. **(C)** Total LacZ activity in these cultures was measured using a fluorogenic *lacZ* substrate (FDG) assay. The red dashed brackets marked with asterisks connect the closest data points of IPTG-induced samples that are statistically significant (p-value < 0.05 based on one-tailed Welch's t-test). **(D)** We extended the circuit in A) to create a sample-and-hold circuit where "Input", "Write", and "Read" operations are independently controlled. This feature enables the creation of addressable Read/Write memory registers in the genomic DNA tape. Induction of cells with the "Input" signal (AHL) produces ssDNA$(lacZ)_{ON}$, which targets the genomic $lacZ_{OFF}$ locus for reversion to the wild-type sequence. In the presence of the "Write" signal (aTc), which expresses Beta, ssDNA$(lacZ)_{ON}$ is recombined into the $lacZ_{OFF}$ locus and produces the $lacZ_{ON}$ genotype. Thus, the "Write" signal enables the "Input" signal to be sampled and held in memory. The total LacZ activity in the cell populations is retrieved by adding the "Read" signal (IPTG). **(E)** Cells harboring the circuit shown in D) were induced with different combinations of aTc (100 ng/ml) and AHL (50 ng/ml) for 24 h, after which the cultures were diluted in fresh media with or without IPTG (1 mM). These cultures were then incubated at 37flC for 8 hours and assayed for total LacZ activity with the FDG assay. **(F)** Cell populations that received both the "Input" and "Write" signals, followed by the "Read" signal exhibited enhanced levels of total LacZ activity. Error bars indicate the standard error of the mean for three independent biological replicates.

2.4 Discussion

We described a scalable platform that uses genomic DNA for analog, rewritable, and flexible memory distributed across living cell populations. One current limitation is the number of orthogonal inducible promoters that can be used as inputs, but this could be addressed by coupling ssDNA expression to endogenous promoters to record native cellular events and the development of additional inducible transcriptional regulatory devices (*42*). Although we primarily targeted mutations into functional genes to facilitate convenient functional and reporter assays in this paper, natural or synthetic non-coding DNA segments could also be used to record memory within genomic DNA. The recorded memory could then be read by high-throughput sequencing (Fig. S2.2). A potential benefit of using synthetic DNA segments as memory registers is the ability to introduce mutations for memory storage that are neutral in terms of fitness costs.

SCRIBE enables conditional increases in the recombination rate at specific loci beyond background levels. The maximum observed recombination rate of the current SCRIBE

platform (~10^{-4} recombination events/generation) is suitable for long-term recording of analog memory distributed across the collective genomes of cellular populations (Fig. S2.3). However, it is not high enough to allow recording of digital information and efficient genome editing at the single-cell level. In principle, population-level analog memory could be achieved by other types of DNA memory switches, such as site-specific recombinases, if they were tuned to achieve intermediate recombination efficiencies. Further investigation is required to determine the exact mechanisms involved in processing retron-based ssDNAs for recombination into genomic DNA and the effects of different growth conditions on SCRIBE memory. Since Beta-mediated recombination is replication-dependent (*35-37*) and ssDNA is believed to be recombined into the genome during the passage of the replication fork (*35*), we speculate that only actively dividing cells are likely to participate in the described population-level memory. Future optimization of SCRIBE (e.g., by modulating the mismatch repair system (*22*) and cellular exonucleases (*43*)) could lead to more efficient single-cell digital memories. This could enable other useful applications, including recording extracellular and intracellular events at the single cell level for biological studies, dynamic engineering of cellular phenotypes, experimental evolution and population dynamics studies, single-cell computation and memory, the construction of complex cellular state machines and biological Turing machines, and enhanced genome engineering techniques.

Additionally, since retrons have been found in a diverse range of microorganisms (*28*), *in vivo* ssDNA expression could be extended to hard-to-transform organisms where SCRIBE plasmids could be introduced by conjugation or transduction. Since retrons have also been shown to be functional in eukaryotes (*32, 44, 45*), they could be potentially used with other genome-editing tools for memory. Moreover, by using error-prone RNA polymerases (*46*) and reverse transcriptases (*47, 48*), we anticipate that mutagenized ssDNA libraries could be generated inside cells for *in vivo* continuous evolution (*49*) and cellular barcoding applications. Finally, *in vivo* ssDNA generation could be potentially used to create DNA nanosystems (*50-56*) and ssDNA-protein hybrid nanomachines in living cells (*57*), or could be optimized and scaled-up to create an economical source of ssDNAs for DNA nanotechnology (*58*). In summary, we envision that *in vivo* ssDNA production and SCRIBE platforms will open up a broad range of new capabilities for engineering biology.

2.5 Supplementary Information

Modeling and Simulation

Deterministic Model

We sought to model the accumulation of recombinants in growing cell populations. The model assumes that clonal interference is negligible, and that the recombinant and wild-type alleles are equally fit. In other words, the model assumes that all the cells in the population have the same growth profile. It also assumes that the rate of recombination in the reverse direction (i.e., from the genome to the plasmid) is negligible (the rate of recombination in *recA*$^-$ background is $<10^{-10}$ (*54*)). The model also assumes that after each Beta-mediated recombination event, only one of the two daughter cells becomes recombinant (*27-29, 55*).

For a given time (t), the recombinant frequency (f_t) is defined as the ratio between the number of recombinants (m_t) to the total number of viable cells in the population (N_t).

$$f_t = \frac{m_t}{N_t}$$

The recombination rate (r) represents the frequency of recombination events that happen in one generation (dt). After one generation, the number of viable cells doubles ($N_{t+dt} = 2N_t$). The number of recombinants in the culture is the sum of the number of cells that are the progeny of pre-existing recombinants and new recombinants that are produced during that generation ($m_{t+dt} = 2m_t + (N_t - m_t)r$). Thus:

$$f_{t+dt} = \frac{2m_t + (N_t - m_t)r}{2N_t} = f_t + \frac{(1 - f_t)r}{2} \qquad where\ dt = one\ generation$$

$$\Rightarrow f_{t+dt} - f_t = \frac{(1 - f_t)r}{2} \quad \Rightarrow df = \frac{(1 - f_t)r}{2} dt$$

$$\Rightarrow \frac{df}{1 - f_t} = \frac{r}{2} dt$$

$$\Rightarrow f_t = 1 - (1 - f_0)e^{-\frac{r}{2}t} \qquad (1)$$

Similarly, for two constitutive generations (t and $t + 1$) we can write:

$$f_{t+1} - f_t = (1 - (1 - f_0)e^{-\frac{r}{2}(t+1)}) - (1 - (1 - f_0)e^{-\frac{r}{2}(t)}) = (1 - f_0)(e^{-\frac{r}{2}t} - e^{-\frac{r}{2}(t+1)})$$

$$f_{t+1} - f_t = (1 - f_0)e^{-\frac{r}{2}t}\left(1 - e^{-\frac{r}{2}}\right) = (1 - f_t)\left(1 - e^{-\frac{r}{2}}\right)$$

$$\Rightarrow f_{t+1} = f_t + (1 - f_t)(1 - e^{-\frac{r}{2}}) = 1 - (1 - f_t)e^{-\frac{r}{2}}$$

Equation (1) describes the frequency of recombinants in a growing bacterial population. In this equation, if $\left(\frac{r}{2}t\right)$ is very small, we have:

$$e^{-\frac{r}{2}t} \cong 1 - \frac{r}{2}t$$

$$f_t \cong 1 - (1 - f_0)\left(1 - \frac{r}{2}t\right) = \frac{r}{2}t + f_0 - \frac{r}{2}tf_0$$

And if f_0 is also very small, the last term is negligible, thus yielding:

$$f_t \cong \frac{r}{2}t + f_0 \qquad (2)$$

Equation (2) shows that when the initial frequency of recombinants (f_0) and the recombination rate (r) are very small, the recombinant frequency in the population increases linearly over time (as long as $\frac{r}{2}tf_0$ is relatively small) with a slope that is equal to half of the recombination rate. However, when those two quantities are relatively high or as the number of generations increases, the recombinant frequency will start to saturate and deviate from a straight line due to a significant drop in the number of cells that can be recombined (i.e. wild-type cells). Nonetheless, Equation (1) should still describe the accumulation of recombinants in the population.

Overall, our model predicts a linear increase (with a slope $= \frac{r}{2}$) in the recombinant frequency as long as the cells in the population are equally fit and as long as $\frac{r}{2}tf_0$ is relatively small. However, in reality, mutations can occur within populations over time, which can potentially affect the fitness of individual cells. In the absence of recombination in asexual populations, two beneficial mutations that arise independently cannot be combined into a single, superior genotype (*56, 57*). Hence, these carriers could compete with each other, a phenomenon known as clonal interference that is important in shaping the evolutionary trajectory of large asexual populations with high mutation rates over prolonged growth. Under these circumstances, the model assumption that all the cells in the population are equally fit does not hold and deviation from the model is expected. However, since the natural rate of beneficial mutations is low (~10^{-9} per bp per generation for *E. coli (57)*), the probability of mutations with significant fitness effects and clonal interference is

relatively low, at least over the timescales of our experiments. Similarly, a linear increase in mutant frequencies during exponential growth of a bacterial culture was previously predicted (*58, 59*).

Stochastic Simulations

To further validate the model, we performed stochastic simulations of a growing bacterial population with three different recombination rates (r=10^{-9}, 0.00015, or 0.005 events/generation) for 250 generations (Fig. S2.3). Growth was simulated for 25 serial iterations, with 10 generations in each iteration. The simulation started with a clonal population of bacteria (10^6 cells). During each generation, each cell could stochastically produce a recombinant allele with a likelihood equal to the recombination rate. The wild-type and recombinant cells were assumed to be equally fit. We also assumed that all the cells in the population followed the same growth profile (no clonal interference). After 10 generations, a sample of ~10^6 cells was taken from the population to start a new culture in order to simulate the serial batch culture procedure.

As shown in Fig. S2.3A, the model predicts a linear increase in the frequency of recombinants with a very low mutation rate ($r = 10^{-9}$). However, the simulation results were not consistent with the deterministic model; instead, the simulation showed stochastic fluctuations in the recombinant frequency since samples taken after 10 generations may not contain representative numbers of recombinants due to the low recombination rate. This condition is representative of the recombinant frequencies observed in the absence of SCRIBE. Major recombination pathways in *E. coli* are *recA*-dependent and knocking out RecA activity can severely affect the recombination rate (*30, 54*). In a recombination-deficient background (*recA*$^-$), such as DH5α, recombination is a very rare, stochastic event (<10^{-10} events/generation (*30, 54*)). These data are consistent with Fig. 2.4B, where no significant increase in recombinant frequencies was observed in the absence of SCRIBE activation (induction pattern I).

In contrast, at a higher targeted recombination rate ($r = 0.00015$ events/generation), a linear increase in the frequency of recombinants is predicted by both the model and simulation (Fig. S2.3B). This rate is representative of cells containing a specific locus targeted by SCRIBE memory. SCRIBE enables control over the recombination rate at a specific locus by external inputs, thus increasing the recombination rate by multiple orders of magnitude over the background rate. For example, using data shown in Fig. 2.4B for cells induced with both aTc and IPTG (induction pattern II), $r = 0.00015$ events/generation was calculated based on the linear regression of the recombination frequency versus generation (Fig. S2.4). This recombination rate ensures that samples

taken from an induced culture contain a representative number of recombinant cells. Thus, successive sampling and regrowth of cells result in the gradual accumulation of recombinants in the population over time in the presence of the inputs (Fig. S2.3B and Fig. 2.4B).

Finally, as the recombination rate increases ($r = 0.005$ events/generation, Fig. S2.3C), the model and simulation predict a linear increase in the recombination frequency at initial times. However, they both start to deviate from the linear approximation as the frequency of recombinants increases (above ~5%) since the cultures are increasingly depleted of the wild-type alleles. These models demonstrate that the upper and lower limits of recombination rates that can be used for analog memory depend on the time-span that is desired for recording. In our current configuration, recombination rates lower than 10^{-7} resulted in stochastic fluctuations in the recombination frequency while recombination rates higher than 10^{-2} quickly led to saturated memory that deviated from the linear regime within less than 10 generations (Fig. S2.5). Intermediate recombination rates (e.g., 10^{-5}-10^{-4} events/generation) enabled the recording of input exposures with a simple linear relationship for hundreds of generation without saturation.

Fig. S2.5 shows stochastic simulations of populations with 9 different recombination rates, with 10 independent runs for each recombination rate. At very low recombination rates, the number of recombinants in simulated populations increases in a noisy fashion. However, this increase becomes less noisy as the recombination rate increases. Thus, recombinant frequencies measured after exposure to low recombination rates (e.g., low inducer levels) are likely to have higher relative standard deviations than recombinant frequencies measured after exposure to higher recombination rates (e.g., higher inducer levels). This trend was observed in the data shown in Fig. 2.1D and is also shown in the simulation results of Fig. S2.6A. In other words, measurements at low recombination rates are inherently noisier and have lower signal-to-noise ratios than measurements at higher recombination rates (Fig. S2.5 and S2.6).

Finally, when starting from a clonal population (or when the initial number of recombinants in the population is negligible), for a limited number of generations, the model and simulation both predict a linear increase in the recombinant frequency as a function of the recombination rate (Equation 2, Fig. S2.6A). However, the recombination rate is not generally a linear function with respect to the concentration of an input inducer, but rather depends on the input-output transfer function of the

inducible system. The transfer functions of many commonly used inducible promoters have outputs that undergo a sharp transition within a narrow range of inputs (60). Using promoters that are can be titrated over a wider range of concentrations (*33*) in combination with more sensitive assays (e.g., high-throughput sequencing) could help to achieve a wider input dynamic range with SCRIBE.

High-Throughput Sequencing of Genomically Encoded Memory

In order to investigate whether SCRIBE's genomically encoded memory could be read out using high-throughput sequencing, we analyzed the genomic content of bacterial populations at the *kanR* locus using Illumina Hi-Seq. Overnight cultures of three independent colonies harboring the gene circuit shown in Fig. 2.1C were diluted into fresh media and then incubated with inducers (1 mM IPTG and 100 ng/ml aTc) or without inducers for 24 hours at 30°C. As an additional control, cells expressing ssDNA($kanR$)$_{\mathrm{OFF}}$ (which has the exact ssDNA template sequence as genomic $kanR_{\mathrm{OFF}}$) were included in this experiment and grown similarly. After 24 hours of induction, total genomic DNA was prepared from the samples using Zymo ZR Fungal/Bacterial DNA MiniPrep Kit. Using these genomic DNA preps as the template, the *kanR* locus was PCR-amplified by primers FF_oligo183 and FF_oligo185. After gel purification, another round of PCR was performed (using primers FF_oligo1291 and FF_oligo1292) to add Illumina adaptors as well as a 10 bp randomized nucleotide to increase the diversity of the library. Barcodes and Illumina anchors were then added using an additional round of PCR. Samples were then gel-purified, multiplexed, and run on a lane of Illumina HiSeq.

The obtained reads were processed and demultiplexed by the MIT BMC-BCC Pipeline. These reads then were trimmed to remove the added 10 bp randomized sequence. To filter out any reads that could have been produced by non-specific binding of primers during PCR, we discarded reads that lacked the expected "CGCGNNNNNATTT" motif, where "NNNNN" corresponds to the 5 base-pair *kanR* memory register. Furthermore, any reads that contained ambiguous bases within this 5 base-pair memory register were discarded. The frequencies of the obtained variants (either GGCCC ($kanR_{\mathrm{ON}}$) or CTATT ($kanR_{\mathrm{OFF}}$), which constitute the two states of the *kanR* memory register (Fig. 2.1C)), were then calculated for each sample.

As shown in Fig. S2.2A, the frequency of reads mapping to $kanR_{\mathrm{ON}}$ in the induced samples expressing ssDNA($kanR$)$_{\mathrm{ON}}$ was comparable to the frequency of Kan-resistant colonies obtained from the plating assay in the *KanR* reversion assay (Fig. 2.1C). Very

few reads mapping to ssDNA($kanR$)$_{ON}$ were observed in the non-induced samples. Interestingly, a few reads mapping to ssDNA($kanR$)$_{ON}$ were observed in induced samples expressing ssDNA($kanR$)$_{OFF}$. To better understand the source of these reads we analyzed the variants observed in the 5 bp *kanR* memory register. These variants and their corresponding frequencies are shown for one representative sample for P$_{lacO}$_msd($kanR$)$_{OFF}$ + P$_{tetO}$_*bet* + IPTG + aTc Rep#1 in Fig. S2.2B. In all the samples, less than 25 variants out of the total 4^5=1024 possible variants were observed. Reads mapping exactly to $kanR_{OFF}$ constituted the majority of reads, as expected. Reads with single or two bp mutations relative to $kanR_{OFF}$ were observed in all the samples, with frequencies ranging from 10^{-7}-10^{-3}. These reads were likely produced by the relatively high mutation rate of high-throughput sequencing (*52*) or during library preparation steps. We did not observe any reads with more than 2 bps of mismatch to both $kanR_{ON}$ and $kanR_{OFF}$. In the negative control sample of Fig. S2B (in which ssDNA($KanR$)$_{OFF}$ was expressed and no $kanR_{ON}$ sequence was present), the absence of reads with 3 or 4 mismatches to $kanR_{OFF}$ suggests that the observed $kanR_{ON}$ reads were likely an artifact of multiplexed sequencing, such as barcode misassignment or recombination during the sequencing protocol.

Overall, these results indicate that high-throughput sequencing can be used to readout genomically encoded memory. The occurrence of false-positive reads (due to sequencing errors) can be effectively avoided by having multiple mismatches (3 bps or more) between the different memory states. Furthermore, improved library preparation (*53*) methods could be used to reduce the error rate of sequencing, thus enhancing readout accuracy.

Materials and Methods

Strains and Plasmids

Conventional cloning methods were used to construct the plasmids. Lists of strains and plasmids used in this study and the construction procedures are provided in Tables S2.1 and S2.2, respectively. The sequences for the synthetic parts and primers are provided in Tables S2.3 and S2.4.

Cells and Antibiotics

Chemically competent *E. coli* DH5α was used for cloning. Unless otherwise noted, antibiotics were used at the following concentrations: carbenicillin (50 μg/ml), kanamycin (20 μg/ml), chloramphenicol (30 μg/ml) and spectinomycin (100 μg/ml).

In the experiment shown in Fig. 2.2, kanamycin (15 μg/ml), and chloramphenicol (15 μg/ml) were used.

Detection of Single-Stranded DNA

Overnight cultures harboring IPTG-inducible plasmids encoding msd(wt), msd(wt) with deactivated RT (msd(wt)_dRT), or msd(*kanR*)$_{ON}$ were grown overnight with or without IPTG (1 mM). Total RNA samples were prepared from non-induced or induced cultures using TRIzol reagent (Invitrogen) according to the manufacturer's protocol. 10 μg total RNA from each sample was treated with RNase A (37°C, 2 hours) to remove RNA species and the msr moiety. The samples were then resolved on 10% TBE-Urea denaturing gel and visualized with SYBR-Gold. 50 pmol of a PAGE-purified synthetic oligo (FF_oligo347) with the same sequence as ssDNA(wt) was used as a molecular size marker. The band intensities were measured by Fiji software (*59*). The intensities were normalized to the intensity of the marker oligo and normalized intensities were used to calculate the amount of ssDNA in each sample.

Induction of Cells and Plating Assays

For each experiment, three transformants were separately inoculated in Luria Broth (LB) media + appropriate antibiotics and grown overnight (37°C, 700 RPM) to obtain seed cultures. Unless otherwise noted, inductions were performed by diluting the seed cultures (1:1000) in 2 ml of pre-warmed LB + appropriate antibiotics ± inducers followed by 24 hours incubation (30°C, 700 RPM). Aliquots of the samples were then serially diluted and appropriate dilutions were plated on selective media to determine the number of recombinants and viable cells in each culture. For each sample, the recombinant frequency was reported as the mean of the ratio of recombinants to viable cells for three independent replicates.

In all the experiments, the number of viable cells was determined by plating aliquots of cultures on LB + spectinomycin plates. LB + kanamycin plates were used to determine the number of recombinants in the *kanR* reversion assay. For the *galK* reversion assay (Fig. 2.2), the numbers of *galK*$_{ON}$ recombinants were determined by plating the cells on MOPS EZ rich defined media (Teknova) + galactose (0.2%). The numbers of *galK*$_{OFF}$ recombinants were determined by plating the cells on MOPS EZ rich defined media + glycerol (0.2%) + 2-DOG (2%). For the experiment shown in Fig. 2.3, the numbers of *kanR*$_{ON}$ *galK*$_{ON}$ and *kanR*$_{OFF}$ *galK*$_{OFF}$ cells were determined by using LB + kanamycin plates and MOPS EZ rich defined media + glycerol (0.2%) +

2-DOG (2%) + D-biotin (0.01%), respectively. The numbers of *kanR*$_{ON}$ *galK*$_{OFF}$ cells were determined by plating the cells on MOPS EZ rich defined media + glycerol (0.2%) + 2-DOG (2%) + kanamycin + D-biotin (0.01%).

For the light-inducible SCRIBE experiment (Fig. 2.4A), induction was performed with white light (using the built-in fluorescent lamp in a VWR 1585 shaker incubator). The "dark" condition was achieved by wrapping aluminum foil around the tubes. Growth of these cultures and sampling from these cultures were performed as described earlier.

LacZ Assay

Overnight seed cultures were diluted (1:1000) in pre-warmed LB + appropriate antibiotics and inducers (with different concentrations of aTc or without aTc in Fig. 2.5A-C, and with all the four possible combinations of aTc (100 ng/ml) and AHL (50 ng/ml) in Fig. 2.5D-F) and incubated for 24 hours (30°C, 700 RPM). These cultures then were diluted (1:50) in pre-warmed LB + appropriate antibiotics with or without IPTG (1 mM) and incubated for 8 hours (37°C, 700 RPM). To measure LacZ activity, 60 µl of each culture was mixed with 60 µl of B-PER II reagent (Pierce Biotechnology) and Fluorescein Di-ß-D-Galactopyranoside (FDG, 0.05 mg/ml final concentration). The fluorescence signal (absorption/emission: 485/515) was monitored in a plate reader with continuous shaking for 2 hours. The LacZ activity was calculated by normalizing the rate of FDG hydrolysis (obtained from fluorescence signal) to the initial OD. For each sample, LacZ activity was reported as the mean of three independent biological replicates.

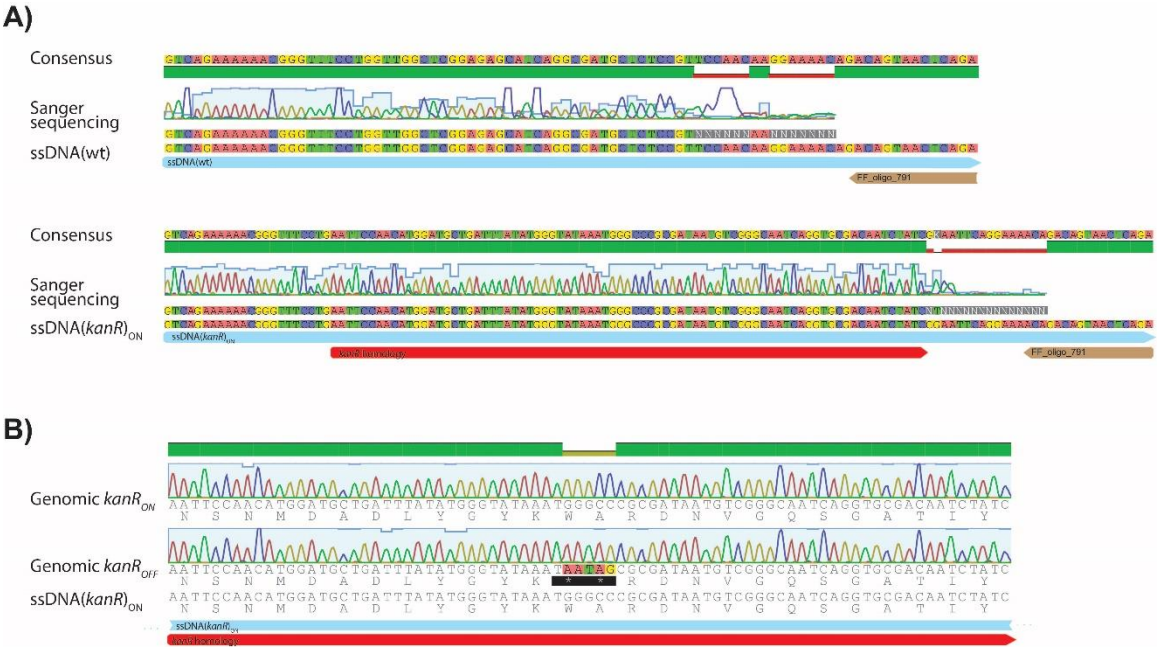


Fig. S2.1 | Sanger sequencing results for *in vivo* expressed ssDNAs and for recombined target loci. (A) ssDNA bands shown in Fig. 2.1B were purified and sequenced. The ssDNA($kanR$)$_{ON}$ construct contained the expected engineered DNA sequence (red arrow). **(B)** Sanger sequencing for the genomic *kanR* locus. The *kanR* locus in $kanR_{\mathrm{OFF}}$ cells, as well as the kanR locus in Kan-resistant (kanRON) cells obtained from induction of ssDNA(kanR)ON in kanROFF reporter cells (Fig. 2.1C), were PCR amplified and sequenced.

A)

	Frequency of reads that perfectly match to							
	***kanR*$_{OFF}$ (CTATT)**				***kanR*$_{ON}$ (GGCCC)**			
	Rep #1	Rep #2	Rep #3	**Mean**	Rep #1	Rep #2	Rep #3	**Mean**
P_{lacO}_msd(*kanR*)$_{ON}$ + P_{tetO}_*bet* + IPTG + aTc	$9.98*10^{-1}$	$9.98*10^{-1}$	$9.98*10^{-1}$	**$9.98*10^{-1}$**	$4.35*10^{-4}$	$4.10*10^{-4}$	$3.87*10^{-4}$	**$4.11*10^{-4}$**
P_{lacO}_msd(*kanR*)$_{ON}$ + P_{tetO}_*bet*	$9.98*10^{-1}$	$9.98*10^{-1}$	$9.98*10^{-1}$	**$9.98*10^{-1}$**	0	$8.88*10^{-7}$	0	**$2.96*10^{-7}$**
P_{lacO}_msd(*kanR*)$_{OFF}$ + P_{tetO}_*bet* + IPTG + aTc	$9.98*10^{-1}$	$9.98*10^{-1}$	$9.98*10^{-1}$	**$9.98*10^{-1}$**	$6.26*10^{-7}$	0	$3.33*10^{-7}$	**$3.20*10^{-7}$**

B)

Row	Variants observed in the 5 bp *kanR* memory register	# of reads mapped to the variant	Frequency	# of mismatches relative to *kanR*$_{OFF}$ (CTATT)	# of mismatches relative to *kanR*$_{ON}$ (GGCCC)
1	CTATT	11155669	$9.98*10^{-1}$	0	5
2	CTACT	3782	$3.38*10^{-4}$	1	4
3	CTATC	1615	$1.45*10^{-4}$	1	4
4	GTATT	175	$1.57*10^{-5}$	1	4
5	CTCTT	113	$1.01*10^{-5}$	1	4
6	CGATT	75	$6.71*10^{-6}$	1	4
7	ATATT	6797	$6.08*10^{-4}$	1	5
8	CCATT	2804	$2.51*10^{-4}$	1	5
9	CTAAT	1289	$1.15*10^{-4}$	1	5
10	CTATA	1097	$9.82*10^{-5}$	1	5
11	CTTTT	508	$4.55*10^{-5}$	1	5
12	CAATT	473	$4.23*10^{-5}$	1	5
13	CTGTT	338	$3.02*10^{-5}$	1	5
14	TTATT	336	$3.01*10^{-5}$	1	5
15	CTAGT	120	$1.07*10^{-5}$	1	5
16	CTATG	105	$9.40*10^{-6}$	1	5
17	CTACC	11	$9.84*10^{-7}$	2	3
18	CAACT	6	$5.37*10^{-7}$	2	4
19	ATATC	2	$1.79*10^{-7}$	2	4
20	CTAAA	4	$3.58*10^{-7}$	2	5
21	GGCCC	7	$6.26*10^{-7}$	5	0
22	AGCCC	107	$9.57*10^{-6}$	5	1

Fig. S2.2 | Using high-throughput Illumina HiSeq sequencing to read out the genomically encoded memory at the *kanR* memory register. (A) The frequency of reads that perfectly match to *kanR*$_{ON}$ or *kanR*$_{OFF}$ after writing with SCRIBE. Note that the sequences attributed to *kanR*$_{ON}$ and *kanR*$_{OFF}$ here are reverse complemented with respect to the sequences in Fig. 2.1C. **(B)** Sequencing variants and their corresponding frequencies observed in the 5 bp *kanR* memory register in one representative sample from cells induced to express ssDNA(*kanR*)$_{OFF}$ within a genomic *kanR*$_{OFF}$ background (P_{lacO}_msd(*kanR*)$_{OFF}$ + P_{tetO}_*bet* + IPTG + aTc Rep#1).

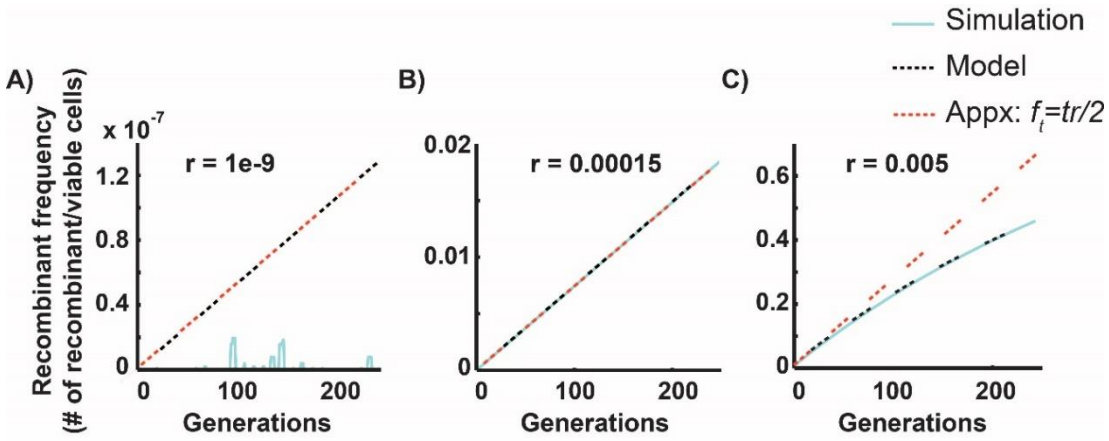


Fig. S2.3 | Deterministic model, stochastic simulations, and mathematical approximations describing the long-term recording of information into genomically encoded memory with the SCRIBE system at three different recombination rates. (A) $r = 10^{-9}$ (B) $r = 0.00015$, and (C) $r = 0.005$. The parameters in the approximation, f_t, r, and t, correspond to the recombinant frequency, recombination rate, and number of generations, respectively.

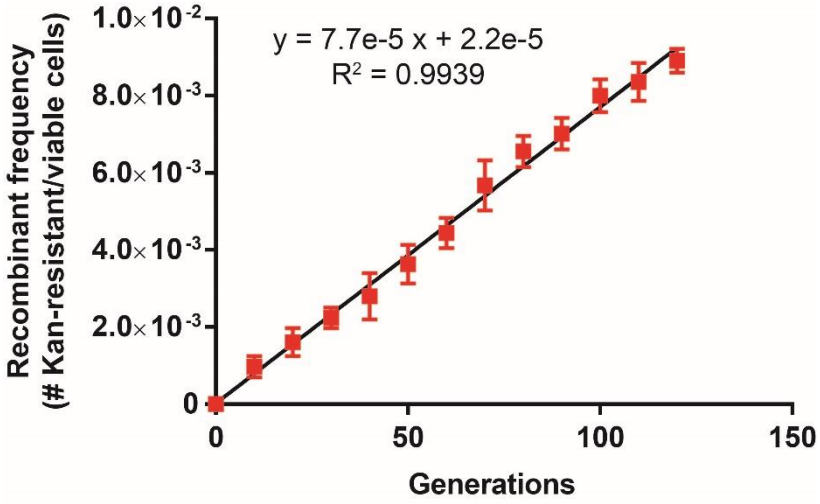


Fig. S2.4 | Estimating the recombination rate of the SCRIBE system. The recombination rate for the SCRIBE circuit (shown in Fig. 2.1C) when the system is induced with both IPTG (1 mM) and aTc (100 ng/ml) was estimated by calculating the slope of the regression line for the data shown in Fig. 2.4B (induction pattern II) and multiplying that slope by a factor of two as described in the deterministic model ($r = 2\frac{df}{dt} = 2(7.7 * 10^{-5}) = 1.54 * 10^{-4}$). In the experiment shown in Fig. 2.4B, the cultures were diluted 1:1000 at the beginning of each day and grown to saturation by the end of the day. Thus, the unit of the x-axis in Fig. 2.4B corresponds to $\log_2(1000) \approx 10$ generations per day.

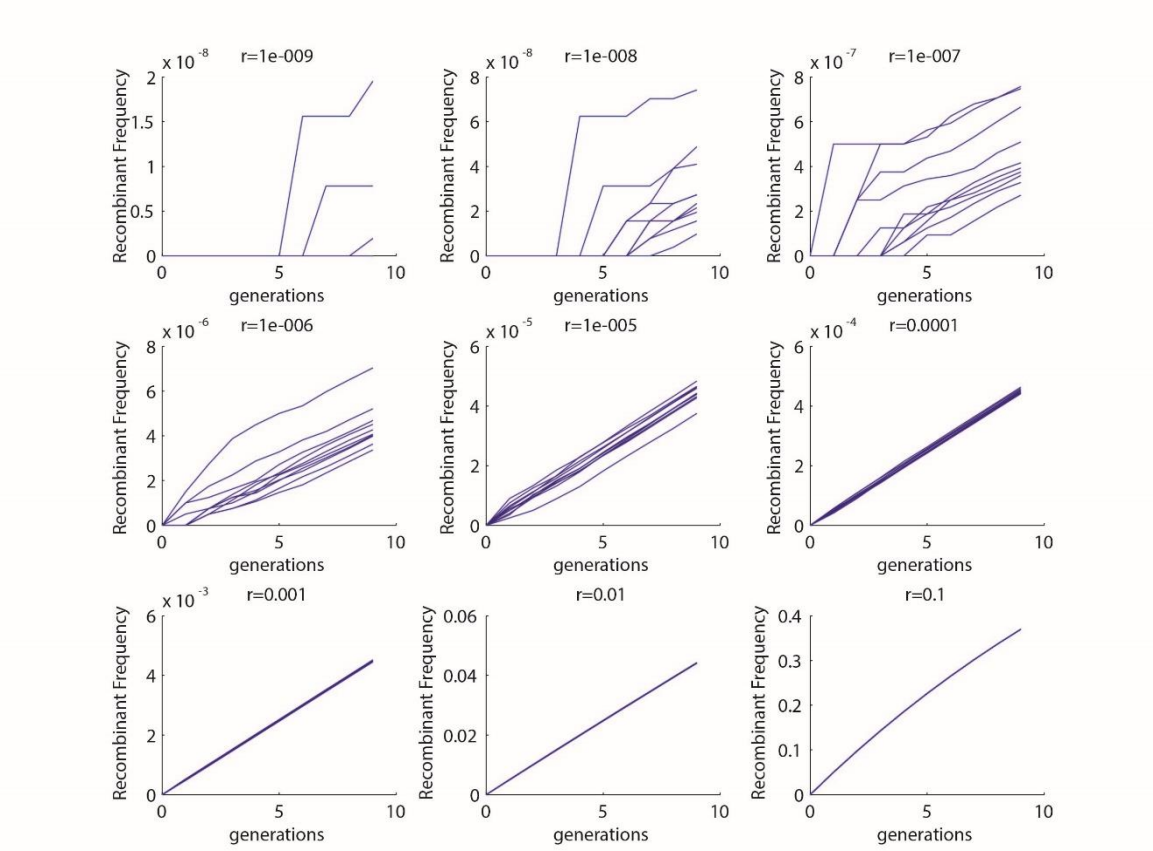


Fig. S2.5 | Stochastic simulations of analog memory with different recombination rates. Recombinant frequencies within populations with 9 different recombination rates are shown, with 10 independent runs for each recombination rate. As the recombination rate increases, the increase in the frequency of recombinants in the population becomes less variable.

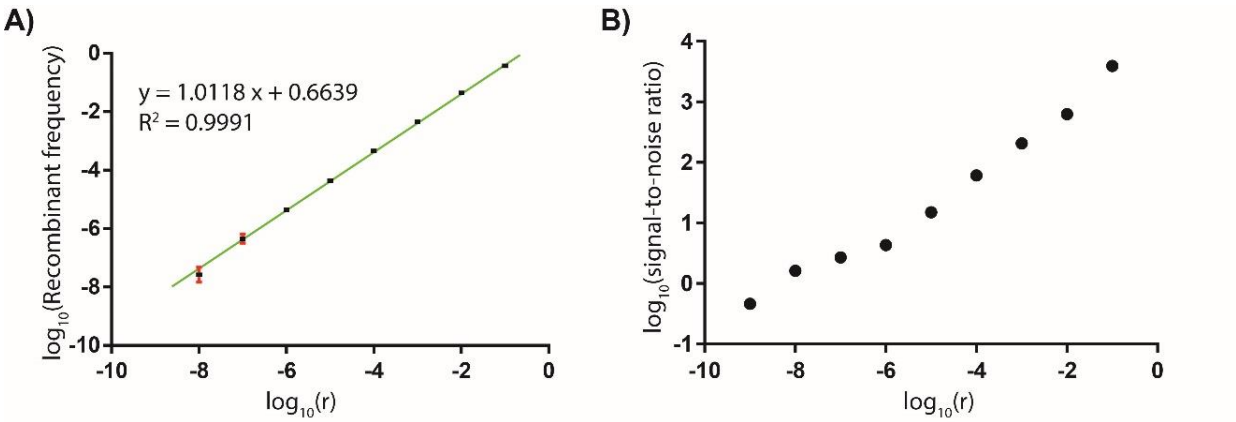


Fig. S2.6 | The signal-to-noise ratio of the population-level analog memory increases as the recombination rate increases. (A) The mean of recombinant frequencies *(f)* obtained from the simulation results shown in Fig. S2.5 (after 9 generations) were plotted on a log-log scale. The error bars (red bars) indicate the standard deviation of the recombinant frequency for 10 independent replicate simulations ($n = 10$). As the recombination rate increases, the relative standard deviation of the measurements is reduced. **(B)** The signal-to-noise ratio ($\frac{mean(f)}{std(f)}$) of the recombinant frequency increases as the recombination rate increases. Due to the inherent noisiness of the system at low recombination rates, the first three data points are highly variable and new iterations may result in different values.

Table S2.1 | List of the reporter strains used in this study.

Name	Strain Code	Construction method	Genotype	Used In
kanR$_{OFF}$ reporter strain	FFF144	The *kanR* cassette was PCR amplified from the pBT3-SUC (Dualsystems Biotech) plasmid using FF_oligo183 and FF_oligo184 primers followed by a second round of PCR with FF_oligo185 and FF_oligo186 to add additional sequences with homology to the sequences flanking the *galK* locus. The fragment then was integrated into the *galK* locus of a DH5α strain (with an integrated PRO cassette (*22*)) by recombineering. Two premature stop codons then were introduced into this *kanR* cassette using oligo-mediated recombineering with FF_oligo187 to make the *kanR*$_{OFF}$ strain.	DH5αPRO *galK::kanR*$_{W28TAA,\ A29TAG}$	Fig. 2.1 Fig. 2.4
kanR$_{OFF}$ *galK*$_{ON}$ reporter strain	FFF774	The *kanR*$_{OFF}$ cassette was PCR amplified from FFF144 and integrated into the *bioA* locus of DH5α. The cells were then transformed with the PRO plasmid (pZS4Int-LacI/TetR (*22*)).	DH5α *bioA::kanR*$_{W28TAA,\ A29TAG}$ + PRO plasmid	Fig. 2.3
galK reporter strain	FFF762	DH5α cells transformed with the PRO plasmid (*22*).	DH5α + PRO plasmid	Fig. 2.2
lacZ$_{OFF}$ reporter strain	FFF798	The *lacZ* α-fragment was introduced into the DH5α *lacZ* locus by recombineering using a PCR fragment amplified from *E. coli* MG1655 (using FF_oligo1069 and FF_oligo1070). Two premature stop codons were then introduced into the *lacZ* ORF using oligo-mediated recombineering with FF_oligo220 to make the *lacZ*$_{OFF}$ strain. This strain was then transformed with the PRO plasmid (*22*).	DH5α *lacZ*$_{A35TAA,\ S36TAG}$ + PRO plasmid	Fig. 2.5

Table S2.2 | List of the plasmids used in this study.

Name	Plasmid Code	Construction method	Used in
P_{lacO}_msd(wt)	pFF753	The wild-type retron Ec86 cassette was PCR-amplified from *E. coli* BL21 and cloned downstream of the P_{lacO} promoter (PacI and BamHI sites) in the pZE32 (*22*) plasmid.	Fig. 2.1B
P_{lacO}_msd(wt) _dRT	pFF758	This plasmid was produced by QuikChange site-directed mutagenesis (using FF_oligo912 and FF_oligo913) to mutate the YADD active site of the RT to YAAA (D197A and D198A mutations) in the P_{lacO}_msd(wt) plasmid.	Fig. 2.1B
P_{lacO}_msd(*kanR*)$_{ON}$	pFF530	This plasmid was produced by introducing a 79-bp fragment with homology to the *kanR* ORF (template for ssDNA(*kanR*)$_{ON}$) and flanked by EcoRI sites into the P_{lacO}_msd(wt) plasmid using QuikChange site-directed mutagenesis.	Fig. 2.1B Fig. 2.1C Fig. 2.4B
P_{lacO}_msd(*kanR*)$_{ON}$_dRT	pFF749	This plasmid was produced by QuikChange site-directed mutagenesis (using FF_oligo912 and FF_oligo913) primers to mutate the YADD active site of the RT to YAAA (D197A and D198A mutations) in the P_{lacO}_msd(*kanR*)$_{ON}$ plasmid.	Fig. 2.1C
P_{tetO}_*bet*	pFF145	This plasmid was constructed by cloning the *bet* ORF from pKD46 (*61*) plasmid downstream of the P_{tetO} promoter (KpnI and BamHI sites) in the pZA11 (*22*) plasmid.	Fig. 2.1C Fig. 2.4B Figs. 2.5D-F
P_{lacO}_SCRIBE(*kanR*)$_{ON}$	pFF745	This plasmid was constructed by cloning *bet* and its natural ribosome binding site (RBS) downstream of the RT in the P_{lacO}_msd(*kanR*)$_{ON}$ plasmid (BamH1, MluI sites). 18 bp upstream of the *bet* start codon in the pKD46 (*61*) plasmid was used as the *bet* RBS.	Fig. 2.1D Fig. 2.3
P_{Dawn}_SCRIBE(*kanR*)$_{ON}$ (light-inducible)	pFF706	This plasmid was constructed by replacing the P_{lacO} promoter in SCRIBE(*kanR*)$_{ON}$ with a PCR fragment containing the light-regulated cassettes (*yf1/fixJ* operon and *cI* and their corresponding promoters as shown in Fig. 2.4A) from pDawn plasmid (Addgene # 43796 (*26*)).	Fig. 2.4A
P_{lacO}_SCRIBE(*galK*)$_{OFF}$	pFF714	This plasmid was constructed by replacing the 79-bp *kanR* homology in P_{lacO}_SCRIBE(*kanR*)$_{ON}$ with a 78-bp fragment containing two stop codons flanked by 72 bp homology to *galK* using QuikChange site-directed mutagenesis.	Fig. 2.2
P_{tetO}_SCRIBE(*galK*)$_{OFF}$	pFF761	This plasmid was constructed by cloning the SCRIBE(*galK*)$_{OFF}$ cassette into the pZA11	Fig. 2.3

		plasmid (*22*) downstream of the P_{tetO} promoter.	
P_{tetO}_SCRIBE(*galK*)$_{ON}$	pFF746	This plasmid constructed by cloning SCRIBE(*galK*)$_{OFF}$ in the pZA21 backbone (*22*) (downstream of P_{tetO}) followed by a QuikChange *in vitro* mutagenesis step to revert the two stop codons in the msd(*galK*)$_{OFF}$ back to the wild-type sequence.	Fig. 2.2
P_{tetO}_SCRIBE(*lacZ*)$_{ON}$	pFF838	This plasmid was made by cloning a 78-bp fragment from the *lacZ* ORF into EcoRI sites of the SCRIBE cassette in P_{tetO}_SCRIBE(*galK*)$_{ON}$, replacing the *galK* homology with *lacZ* homology. The obtained SCRIBE cassette then was cloned into the pZA31 (*22*) backbone.	Figs. 2.5A-C
P_{luxR}_msd(*lacZ*)$_{ON}$	pFF828	This plasmid was made by replacing the P_{lacO} in the P_{lacO}_msd(*kanR*)$_{ON}$ plasmid with an AHL-inducible promoter (*luxR* cassette and P_{luxR} promoter (*31*)) followed by the replacement of the ssDNA(*kanR*)$_{ON}$ template with a 78-bp fragment from the *lacZ* ORF.	Figs. 2.5D-F

Table S2.3 | List of the synthetic parts and their corresponding sequences used in this study.

Part name	Type	Sequence	Ref
P_{lacO}	Promoter	AATTGTGAGCGGATAACAATTGACATTGTGAGCGGATA ACAAGATACTGAGCACATCAGCAGGACGCACTGACC	(*22*)
P_{tetO}	Promoter	TCCCTATCAGTGATAGAGATTGACATCCCTATCAGTGA TAGAGATACTGAGCACATCAGCAGGACGCACTGACC	(*22*)
P_{luxR}	Promoter	ACCTGTAGGATCGTACAGGTTTACGCAAGAAAATGGTT TGTTATAGTCGAATA	(*31*)
$P_{\lambda R}$	Promoter	TAACACCGTGCGTGTTGACTATTTTACCTCTGGCGGTG ATAATGGTTGC	(*26*)
P_{fixK2}	Promoter	ACGCCCGTGATCCTGATCACCGGCTATCCGGACGAAAA CATCTCGACCCGGGCCGCCGAGGCCGGCGTAAAAGACG TGGTTTTGAAGCCGCTTCTCGACGAAAACCTGCTCAAG CGTATCCGCCGCGCCATCCAGGACCGGCCTCGGGCATG ACCTACGGGGTTCTACGTAAGGCACCCCCCTTAAGATA TCGCTCGAAATTTTCGAACCTCCCGATACCGCGTACCA ATGCGTCATCACAACGGAG	(*26*)
msr	Primer for the RT	ATGCGCACCCTTAGCGAGAGGTTTATCATTAAGGTCAA CCTCTGGATGTTGTTTCGGCATCCTGCATTGAATCTGA GTTACT	This work
msd(*wt*)	Template for the RT	GTCAGAAAAAACGGGTTTCCTGGTTGGCTCGGAGAGCA TCAGGCGATGCTCTCCGTTCCAACAAGGAAAACAGACA GTAACTCAGA	This work
$msd(kanR)_{ON}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCCAACATGGATG CTGATTTATATGGGTATAAATGGGCCCGCGATAATGTC GGGCAATCAGGTGCGACAATCTATCGGAATTCAGGAAA ACAGACAGTAACTCAGA	This work
$msd(galK)_{OFF}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCCAGCTAATTTC CGCGCTCGGCAAGAAAGATCATGCCTAATGAATCGATT GCCGCTCACTGGGGACCAAAGCAGTTTCCGAATTCAGG AAAACAGACAGTAACTCAGA	This work
$msd(galK)_{ON}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCCAGCTAATTTC CGCGCTCGGCAAGAAAGATCATGCCCTCTTGATCGATT GCCGCTCACTGGGGACCAAAGCAGTTTCCGAATTCAGG AAAACAGACAGTAACTCAGA	This work
$msd(lacZ)_{ON}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCACCCAACTTAA TCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTA ATAGCGAAGAGGCCCGCACCGATCGCCCTGAATTCAGG AAAACAGACAGTAACTCAGA	This work
RT Ec86	Reverse Transcriptase	ATGAAATCCGCTGAATATTTGAACACTTTTAGATTGAG AAATCTCGGCCTACCTGTCATGAACAATTTGCATGACA TGTCTAAGGCGACTCGCATATCTGTTGAAACACTTCGG TTGTTAATCTATACAGCTGATTTTCGCTATAGGATCTA CACTGTAGAAAAGAAAGGCCCAGAGAAGAGAATGAGAA CCATTTACCAACCTTCTCGAGAACTTAAAGCCTTACAA GGATGGGTTCTACGTAACATTTTAGATAAACTGTCGTC ATCTCCTTTTTCTATTGGATTTGAAAAGCACCAATCTA TTTTGAATAATGCTACCCCGCATATTGGGGCAAACTTT ATACTGAATATTGATTTGGAGGATTTTTTCCCAAGTTT AACTGCTAACAAAGTTTTTGGAGTGTTCCATTCTCTTG GTTATAATCGACTAATATCTTCAGTTTTGACAAAAATA TGTTGTTATAAAAATCTGCTACCACAAGGTGCTCCATC ATCACCTAAATTAGCTAATCTAATATGTTCTAAACTTG	This work

		ATTATCGTATTCAGGGTTATGCAGGTAGTCGGGGCTTG ATATATACGAGATATGCCGATGACCTCACCTTATCTGC ACAGTCTATGAAAAAGGTTGTTAAAGCACGTGATTTTT TATTTTCTATAATCCCAAGTGAAGGATTGGTTATTAAC TCAAAAAAAACTTGTATTAGTGGGCCTCGTAGTCAGAG GAAAGTTACAGGTTTAGTTATTTCACAAGAGAAAGTTG GGATAGGTAGAGAAAAATATAAAGAAATTAGAGCAAAG ATACATCATATATTTTGCGGTAAGTCTTCTGAGATAGA ACACGTTAGGGGATGGTTGTCATTTATTTTAAGTGTGG ATTCAAAAAGCCATAGGAGATTAATAACTTATATTAGC AAATTAGAAAAAAAAATATGGAAAGAACCCTTTAAATAA AGCGAAGACCTAA	
Beta	ssDNA-specific recombinase protein	ATGAGTACTGCACTCGCAACGCTGGCTGGGAAGCTGGC TGAACGTGTCGGCATGGATTCTGTCGACCCACAGGAAC TGATCACCACTCTTCGCCAGACGGCATTTAAAGGTGAT GCCAGCGATGCGCAGTTCATCGCATTACTGATCGTTGC CAACCAGTACGGCCTTAATCCGTGGACGAAAGAAATTT ACGCCTTTCCTGATAAGCAGAATGGCATCGTTCCGGTG GTGGGCGTTGATGGCTGGTCCCGCATCATCAATGAAAA CCAGCAGTTTGATGGCATGGACTTTGAGCAGGACAATG AATCCTGTACATGCCGGATTTACCGCAAGGACCGTAAT CATCCGATCTGCGTTACCGAATGGATGGATGAATGCCG CCGCGAACCATTCAAAACTCGCGAAGGCAGAGAAATCA CGGGGCCGTGGCAGTCGCATCCCAAACGGATGTTACGT CATAAAGCCATGATTCAGTGTGCCCGTCTGGCCTTCGG ATTTGCTGGTATCTATGACAAGGATGAAGCCGAGCGCA TTGTCGAAAATACTGCATACACTGCAGAACGTCAGCCG GAACGCGACATCACTCCGGTTAACGATGAAACCATGCA GGAGATTAACACTCTGCTGATCGCCCTGGATAAAACAT GGGATGACGACTTATTGCCGCTCTGTTCCCAGATATTT CGCCGCGACATTCGTGCATCGTCAGAACTGACACAGGC CGAAGCAGTAAAAGCTCTTGGATTCCTGAAACAGAAAG CCGCAGAGCAGAAGGTGGCAGCATGA	(*61*)
cI	λ repressor	ATGAGCACAAAAAAGAAACCATTAACACAAGAGCAGCT TGAGGACGCACGTCGCCTTAAAGCAATTTATGAAAAAA AGAAAAATGAACTTGGCTTATCCCAGGAATCTGTCGCA GACAAGATGGGGATGGGGCAGTCAGGCGTTGGTGCTT TATTTAATGGCATCAATGCATTAAATGCTTATAACGCC GCATTGCTTGCAAAAATTCTCAAAGTTAGCGTTGAAGA ATTTAGCCCTTCAATCGCCAGAGAAATCTACGAGATGT ATGAAGCGGTTAGTATGCAGCCGTCACTTAGAAGTGAG TATGAGTACCCTGTTTTTTCTCATGTTCAGGCAGGGAT GTTCTCACCTGAGCTTAGAACCTTTACCAAAGGTGATG CGGAGAGATGGGTAAGCACAACCAAAAAAGCCAGTGAT TCTGCATTCTGGCTTGAGGTTGAAGGTAATTCCATGAC CGCACCAACAGGCTCCAAGCCGAGCTTTCCTGACGGAA TGTTAATTCTCGTTGACCCTGAGCAGGCTGTTGAGCCA GGTGATTTCTGCATAGCCAGACTTGGGGGTGATGAGTT TACCTTCAAGAAACTGATCAGGGATAGCGGTCAGGTGT TTTTACAACCACTAAACCCACAGTACCCAATGATCCCA TGCAATGAGAGTTGTTCCGTTGTGGGGAAAGTTATCGC TAGTCAGTGGCCTGAAGAGACGTTTGGCGCTGCAAACG ACGAAAACTACGCTTTAGTAGCTTAA	(*26*)
$kanR_{OFF}$	Reporter gene (premature stop codons are highlighted)	ATGAGCCATATTCAACGGGAAACGTCTTGCTCGAGGCC GCGATTAAATTCCAACATGGATGCTGATTTATATGGGT ATAAA**TAATAG**CGCGATAATGTCGGGCAATCAGGTGCG ACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGA GTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATG ATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACG GAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCG TACTCCTGATGATGCATGGTTACTCACCACTGCGATCC CCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCT GATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTT CCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTC CTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCG	This work

		CAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGA TTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAG TCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCG GATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAA CCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTG ATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGAT CTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCC TTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTG ATAATCCTGATATGAATAAATTGCAGTTTCATTTGATG CTCGATGAGTTTTTCTAA	
galK$_{OFF}$	Reporter gene (premature stop codons are highlighted)	ATGAGTCTGAAAGAAAAAACACAATCTCTGTTTGCCAA CGCATTTGGCTACCCTGCCACTCACACCATTCAGGCGC CTGGCCGCGTGAATTTGATTGGTGAACACACCGACTAC AACGACGGTTTCGTTCTGCCCTGCGCGATTGATTATCA AACCGTGATCAGTTGTGCACCACGCGATGACCGTAAAG TTCGCGTGATGGCAGCCGATTATGAAAATCAGCTCGAC GAGTTTTCCCTCGATGCGCCCATTGTCGCACATGAAAA CTATCAATGGGCTAACTACGTTCGTGGCGTGGTGAAAC ATCTGCAACTGCGTAACAACAGCTTCGGCGGCGTGGAC ATGGTGATCAGCGGCAATGTGCCGCAGGGTGCCGGGTT AAGTTCTTCCGCTTCACTGGAAGTCGCGGTCGGAACCG TATTGCAGCAGCTTTATCATCTGCCGCTGGACGGCGCA CAAATCGCGCTTAACGGTCAGGAAGCAGAAAACCAGTT TGTAGGCTGTAACTGCGGGATCATGGATCAGCTAATTT CCGCGCTCGGCAAGAAAGATCATGCC**TAATGA**ATCGAT TGCCGCTCACTGGGGACCAAAGCAGTTTCCATGCCCAA AGGTGTGGCTGTCGTCATCATCAACAGTAACTTCAAAC GTACCCTGGTTGGCAGCGAATACAACACCCGTCGTGAA CAGTGCGAAACCGGTGCGCGTTTCTTCCAGCAGCCAGC CCTGCGTGATGTCACCATTGAAGAGTTCAACGCTGTTG CGCATGAACTGGACCCGATCGTGGCAAAACGCGTGCGT CATATACTGACTGAAAACGCCCGCACCGTTGAAGCTGC CAGCGCGCTGGAGCAAGGCGACCTGAAACGTATGGGCG AGTTGATGGCGGAGTCTCATGCCTCTATGCGCGATGAT TTCGAAATCACCGTGCCGCAAATTGACACTCTGGTAGA AATCGTCAAAGCTGTGATTGGCGACAAAGGTGGCGTAC GCATGACCGGCGGCGGATTTGGCGGCTGTATCGTCGCG CTGATCCCGGAAGAGCTGGTGCCTGCCGTACAGCAAGC TGTCGCTGAACAATATGAAGCAAAAACAGGTATTAAAG AGACTTTTTACGTTTGTAAACCATCACAAGGAGCAGGA CAGTGCTGA	
lacZ$_{OFF}$	Reporter gene (premature stop codons are highlighted)	ATGACCATGATTACGGATTCACTGGCCGTCGTTTTACA ACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTA ATCGCCTTGCAGCACATCCCCCTTTC**TAATAG**TGGCGT AATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACA GTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGGT TTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAG TGCGATCTTCCTGAGGCCGATACTGTCGTCGTCCCCTC AAACTGGCAGATGCACGGTTACGATGCGCCCATCTACA CCAACGTGACCTATCCCATTACGGTCAATCCGCCGTTT GTTCCCACGGAGAATCCGACGGGTTGTTACTCGCTCAC ATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGA CGCGAATTATTTTTGATGGCGTTAACTCGGCGTTTCAT CTGTGGTGCAACGGGCGCTGGGTCGGTTACGGCCAGGA CAGTCGTTTGCCGTCTGAATTTGACCTGAGCGCATTTT TACGCGCCGGAGAAAACCGCCTCGCGGTGATGGTGCTG CGCTGGAGTGACGGCAGTTATCTGGAAGATCAGGATAT GTGGCGGATGAGCGGCATTTTCCGTGACGTCTCGTTGC TGCATAAACCGACTACACAAATCAGCGATTTCCATGTT GCCACTCGCTTTAATGATGATTTCAGCCGCGCTGTACT GGAGGCTGAAGTTCAGATGTGCGGCGAGTTGCGTGACT ACCTACGGGTAACAGTTTCTTTATGGCAGGGTGAAACG CAGGTCGCCAGCGGCACCGCGCCTTTCGGCGGTGAAAT TATCGATGAGCGTGGTGGTTATGCCGATCGCGTCACAC TACGTCTGAACGTCGAAAACCCGAAACTGTGGAGCGCC	This work

		GAAATCCCGAATCTCTATCGTGCGGTGGTTGAACTGCA CACCGCCGACGGCACGCTGATTGAAGCAGAAGCCTGCG ATGTCGGTTTCCGCGAGGTGCGGATTGAAAATGGTCTG CTGCTGCTGAACGGCAAGCCGTTGCTGATTCGAGGCGT TAACCGTCACGAGCATCATCCTCTGCATGGTCAGGTCA TGGATGAGCAGACGATGGTGCAGGATATCCTGCTGATG AAGCAGAACAACTTTAACGCCGTGCGCTGTTCGCATTA TCCGAACCATCCGCTGTGGTACACGCTGTGCGACCGCT ACGGCCTGTATGTGGTGGATGAAGCCAATATTGAAACC CACGGCATGGTGCCAATGAATCGTCTGACCGATGATCC GCGCTGGCTACCGGCGATGAGCGAACGCGTAACGCGAA TGGTGCAGCGCGATCGTAATCACCCGAGTGTGATCATC TGGTCGCTGGGGAATGAATCAGGCCACGGCGCTAATCA CGACGCGCTGTATCGCTGGATCAAATCTGTCGATCCTT CCCGCCCGGTGCAGTATGAAGGCGGCGGAGCCGACACC ACGGCCACCGATATTATTTGCCCGATGTACGCGCGCGT GGATGAAGACCAGCCCTTCCCGGCTGTGCCGAAATGGT CCATCAAAAAATGGCTTTCGCTACCTGGAGAGACGCGC CCGCTGATCCTTTGCGAATACGCCCACGCGATGGGTAA CAGTCTTGGCGGTTTCGCTAAATACTGGCAGGCGTTTC GTCAGTATCCCCGTTTACAGGGCGGCTTCGTCTGGGAC TGGGTGGATCAGTCGCTGATTAAATATGATGAAAACGG CAACCCGTGGTCGGCTTACGGCGGTGATTTTGGCGATA CGCCGAACGATCGCCAGTTCTGTATGAACGGTCTGGTC TTTGCCGACCGCACGCCGCATCCAGCGCTGACGGAAGC AAAACACCAGCAGCAGTTTTTCCAGTTCCGTTTATCCG GGCAAACCATCGAAGTGACCAGCGAATACCTGTTCCGT CATAGCGATAACGAGCTCCTGCACTGGATGGTGGCGCT GGATGGTAAGCCGCTGGCAAGCGGTGAAGTGCCTCTGG ATGTCGCTCCACAAGGTAAACAGTTGATTGAACTGCCT GAACTACCGCAGCCGGAGAGCGCCGGGCAACTCTGGCT CACAGTACGCGTAGTGCAACCGAACGCGACCGCATGGT CAGAAGCCGGGCACATCAGCGCCTGGCAGCAGTGGCGT CTGGCGGAAAACCTCAGTGTGACGCTCCCCGCCGCGTC CCACGCCATCCCGCATCTGACCACCAGCGAAATGGATT TTTGCATCGAGCTGGGTAATAAGCGTTGGCAATTTAAC CGCCAGTCAGGCTTTCTTTCACAGATGTGGATTGGCGA TAAAAAACAACTGCTGACGCCGCTGCGCGATCAGTTCA CCCGTGCACCGCTGGATAACGACATTGGCGTAAGTGAA GCGACCCGCATTGACCCTAACGCCTGGGTCGAACGCTG GAAGGCGGCGGGCCATTACCAGGCCGAAGCAGCGTTGT TGCAGTGCACGGCAGATACACTTGCTGATGCGGTGCTG ATTACGACCGCTCACGCGTGGCAGCATCAGGGGAAAAC CTTATTTATCAGCCGGAAAACCTACCGGATTGATGGTA GTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCG AGCGATACACCGCATCCGGCGCGGATTGGCCTGAACTG CCAGCTGGCGCAGGTAGCAGAGCGGGTAAACTGGCTCG GATTAGGGCCGCAAGAAAACTATCCCGACCGCCTTACT GCCGCCTGTTTTGACCGCTGGGATCTGCCATTGTCAGA CATGTATACCCCGTACGTCTTCCCGAGCGAAAACGGTC TGCGCTGCGGGACGCGCGAATTGAATTATGGCCCACAC CAGTGGCGCGGCGACTTCCAGTTCAACATCAGCCGCTA CAGTCAACAGCAACTGATGGAAACCAGCCATCGCCATC TGCTGCACGCGGAAGAAGGCACATGGCTGAATATCGAC GGTTTCCATATGGGGATTGGTGGCGACGACTCCTGGAG CCCGTCAGTATCGGCGGAATTCCAGCTGAGCGCCGGTC GCTACCATTACCAGTTGGTCTGGTGTCAAAAATAA	
SCRIBE(*kanR*)$_{ON}$	The synthetic operon for writing into the *kanR* locus.	ATGCGCACCCTTAGCGAGAGGTTTATCATTAAGGTCAA CCTCTGGATGTTGTTTCGGCATCCTGCATTGAATCTGA GTTACTGTCTGTTTTCCTGAATTCCGATAGATTGTCGC ACCTGATTGCCCGACATTATCGCGGGCCCATTTATACC CATATAAATCAGCATCCATGTTGGAATTCAGGAAACCC GTTTTTTCTGACGTAAGGGTGCGCAACTTTCATGAAAT CCGCTGAATATTTGAACACTTTTAGATTGAGAAATCTC GGCCTACCTGTCATGAACAATTTGCATGACATGTCTAA GGCGACTCGCATATCTGTTGAAACACTTCGGTTGTTAA	This work

	The msd(*kanR*)$_{ON}$ region is highlighted. The region flanked by EcoRI sites (red) can be replaced with a template for ssDNA of interest.	TCTATACAGCTGATTTTCGCTATAGGATCTACACTGTA GAAAAGAAAGGCCCAGAGAAGAGAATGAGAACCATTTA CCAACCTTCTCGAGAACTTAAAGCCTTACAAGGATGGG TTCTACGTAACATTTTAGATAAACTGTCGTCATCTCCT TTTTCTATTGGATTTGAAAAGCACCAATCTATTTTGAA TAATGCTACCCCGCATATTGGGGCAAACTTTATACTGA ATATTGATTTGGAGGATTTTTTCCCAAGTTTAACTGCT AACAAAGTTTTTGGAGTGTTCCATTCTCTTGGTTATAA TCGACTAATATCTTCAGTTTTGACAAAAATATGTTGTT ATAAAAATCTGCTACCACAAGGTGCTCCATCATCACCT AAATTAGCTAATCTAATATGTTCTAAACTTGATTATCG TATTCAGGGTTATGCAGGTAGTCGGGGCTTGATATATA CGAGATATGCCGATGACCTCACCTTATCTGCACAGTCT ATGAAAAAGGTTGTTAAAGCACGTGATTTTTTATTTTC TATAATCCCAAGTGAAGGATTGGTTATTAACTCAAAAA AAACTTGTATTAGTGGGCCTCGTAGTCAGAGGAAAGTT ACAGGTTTAGTTATTTCACAAGAGAAAGTTGGGATAGG TAGAGAAAAATATAAAGAAATTAGAGCAAAGATACATC ATATATTTTGCGGTAAGTCTTCTGAGATAGAACACGTT AGGGGATGGTTGTCATTTATTTTAAGTGTGGATTCAAA AAGCCATAGGAGATTAATAACTTATATTAGCAAATTAG AAAAAAAATATGGAAAGAACCCTTTAAATAAAGCGAAG ACCTAAGGATCCGGTTGATATTGATTCAGAGGTATAAA ACGAATGAGTACTGCACTCGCAACGCTGGCTGGGAAGC TGGCTGAACGTGTCGGCATGGATTCTGTCGACCCACAG GAACTGATCACCACTCTTCGCCAGACGGCATTTAAAGG TGATGCCAGCGATGCGCAGTTCATCGCATTACTGATCG TTGCCAACCAGTACGGCCTTAATCCGTGGACGAAAGAA ATTTACGCCTTTCCTGATAAGCAGAATGGCATCGTTCC GGTGGTGGGCGTTGATGGCTGGTCCCGCATCATCAATG AAAACCAGCAGTTTGATGGCATGGACTTTGAGCAGGAC AATGAATCCTGTACATGCCGGATTTACCGCAAGGACCG TAATCATCCGATCTGCGTTACCGAATGGATGGATGAAT GCCGCCGCGAACCATTCAAAACTCGCGAAGGCAGAGAA ATCACGGGGCCGTGGCAGTCGCATCCCAAACGGATGTT ACGTCATAAAGCCATGATTCAGTGTGCCCGTCTGGCCT TCGGATTTGCTGGTATCTATGACAAGGATGAAGCCGAG CGCATTGTCGAAAATACTGCATACACTGCAGAACGTCA GCCGGAACGCGACATCACTCCGGTTAACGATGAAACCA TGCAGGAGATTAACACTCTGCTGATCGCCCTGGATAAA ACATGGGATGACGACTTATTGCCGCTCTGTTCCCAGAT ATTTCGCCGCGACATTCGTGCATCGTCAGAACTGACAC AGGCCGAAGCAGTAAAAGCTCTTGGATTCCTGAAACAG AAAGCCGCAGAGCAGAAGGTGGCAGCATGA	

Table S2.4 | List of the synthetic oligos used in this study.

Name	Sequence
FF_oligo183	GCGATATCCATTTTCGCGAATCCGGAGTGTAAGAAGAGCTCCTGACTCCCCGTCGTGTAG
FF_oligo184	GACCGCAGAACAGGCAGCAGAGCGTTTGCGCGCAGTCAGCGATATCCATTTTCGCGAATC
FF_oligo185	CGGCTGACCATCGGGTGCCAGTGCGGGAGTTTCGTGACGTCGTTAAGCCAGCCCCGACAC
FF_oligo186	ACTACCATCCCTGCGTTGTTACGCAAAGTTAACAGTCGGTACGGCTGACCATCGGGTGCC
FF_oligo187 (* shows phosphorothioate bond)	C*G*CGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATAATAGCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCG*A*T
FF_oligo220	CAACTTAATCGCCTTGCAGCACATCCCCCTTTCTAATAGTGGCGTAATAGCGAAGAGGCCCGCACCGATCGC
FF_oligo912	GATATATACGAGATATGCCGCTGCTCTCACCTTATCTGCAC
FF_oligo913	GTGCAGATAAGGTGAGAGCAGCGGCATATCTCGTATATATC
FF_oligo1069	AATACGCAAACCGCCTCTCC
FF_oligo1070	CGGCGGATTGACCGTAATGG
FF_oligo1291	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNGCCCGACATTATCGCG
FF_oligo1292	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTCAACATGGATGCTGATTTATATGGGT
FF_oligo347 (PAGE purified, used as ssDNA size marker in Fig. 2.1B)	GTCAGAAAAAACGGGTTTCCTGGTTGGCTCGGAGAGCATCAGGCGATGCTCTCCGTTCCAACAAGGAAAACAGACAGTAACTCAGA

Chapter 3: HiSCRIBE

Efficient and Generalizable DNA Writers for Spatiotemporal Recording and Genome Editing in Bacterial Communities

3.1 Abstract

The ability to efficiently and dynamically change information stored in genomes would enable powerful strategies for studying, modifying, and controlling cellular phenotypes. Current platforms for engineering bacterial genomes are limited to a few model strains and laboratory conditions, often suffer from suboptimal editing efficiencies, and are not suitable for *in situ* applications. Furthermore, these techniques require *cis*-encoded elements on the target for efficient editing and cannot be readily linked to biological signals for dynamic and autonomous genome engineering. To overcome these limitations, we created genetically encoded DNA writers that enable efficient, precise, and dynamic editing of bacterial genomes without the requirement for target-specific elements or double-strand DNA breaks. We demonstrate that this DNA writing platform enables a broad range of applications that were not previously possible, including conditional, efficient, scarless and *cis*-element-independent genome editing and spatiotemporal molecular recording, the editing of targeted microbial genomes within complex communities, and the high-throughput mapping of spatial information and cellular interactions into DNA memory. We envision that this DNA writing technology will accelerate our understanding and facilitate engineering of biological systems over space and time.

3.2 Introduction

Platforms that enable efficient targeted modifications of genomic DNA are essential for studying and engineering living cells. An ideal DNA writer (a molecular device for targeted editing of DNA in living cells) should enable one to introduce any desired mutation to any desired genomic target with high efficiency and without the requirement for specific *cis*-encoded elements or the generation of double-strand DNA breaks. However, current genome editing platforms (*22, 23, 25, 60-62*) are not ideal for many applications (Table S3.1). For example, recombineering-based approaches enable targeted modification of bacterial genomes but they are restricted to a few laboratory model strains and specific conditions in which efficient transformation is possible, are

often limited by suboptimal editing rates, and are not applicable to complex environments, such as bacterial communities (*22, 23, 62*). In addition, recombineering events cannot be linked to cellular regulatory networks and thus cannot be used for continuous and dynamic manipulation of cellular phenotypes and autonomous recording of cellular events histories. Recently, recombineering efficiencies have been improved by using CRISPR-Cas9 counter-selection (*63, 64*); however, this approach requires the presence of *cis*-encoded elements (i.e., the PAM domain) on the target and cytotoxic double-stranded breaks (*65, 66*), which limits its applicability.

To circumvent some of these limitations, we previously developed SCRIBE (Synthetic Cellular Recorders Integrating Biological Events), a platform for converting transcriptional signals into DNA memory via conditional and targeted editing of bacterial genomes (*67*). In this DNA writing system, single-stranded DNAs are expressed intracellularly from an engineered retron cassette via reverse transcription and recombined into homologous sites on the genome by Beta-mediated recombination. The moderate recombination rate (~10^{-4} recombination events per generation) achieved by the original SCRIBE system enables recording of the duration and magnitude of exposure (which are analog properties) of an input in the form of mutations that accumulate in the genomic DNA of bacterial populations. However, this level of recombination is not adequate for many applications that require a much more efficient and dynamic DNA writing system.

In the present study, we significantly improved SCRIBE efficiency to create HiSCRIBE (High-efficiency SCRIBE), a genetically encoded DNA writing system that enables autonomous, dynamic, and transcriptionally controlled modification of bacterial genomes with high efficiency and precision. HiSCRIBE writers achieve up to ~100% editing efficiency in a scarless fashion without the need for *cis*-encoded sequences on the target, generation of double-strand DNA breaks or selection. We demonstrate that this high-efficiency and dynamic DNA-writing platform opens up an unprecedented plethora of new applications for study and engineering biology (Fig. 3.1). Specifically, we demonstrate that HiSCRIBE DNA writers can be introduced into cells via different delivery mechanisms, including transduction and conjugation, enabling efficient and specific genome writing in bacteria within communities that is not possible with traditional recombineering approaches. Furthermore, we show that high-efficiency genome editing can be used to record transient spatial information into genomic DNA, allowing one to reduce multidimensional interactomes into a one-dimensional DNA sequence space, thus facilitating the study of complex cellular interactions.

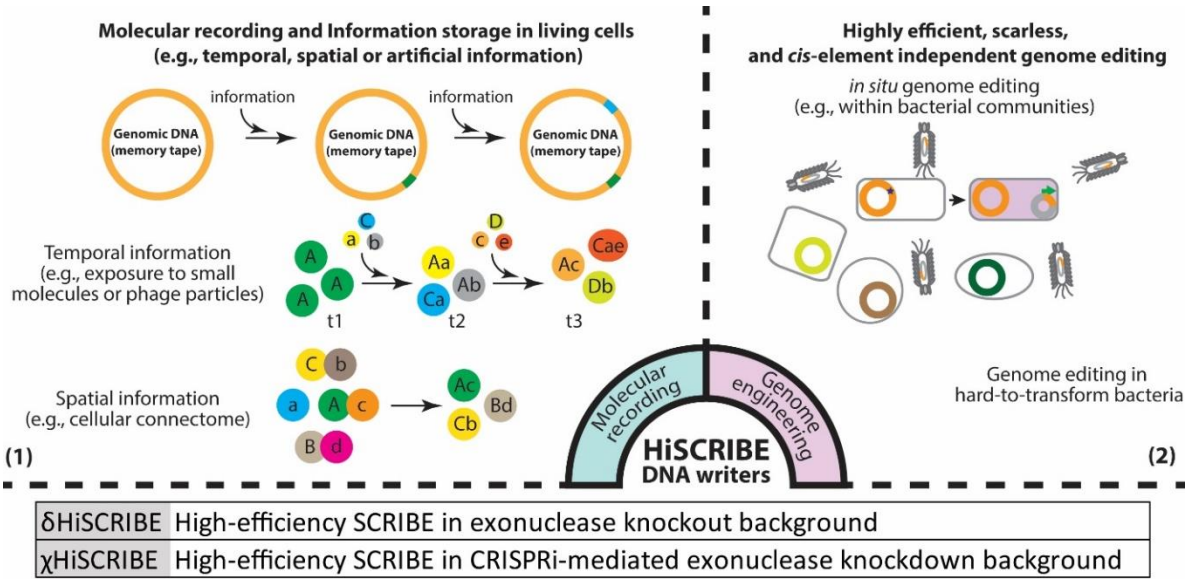


Figure 3.1 | Distinctive applications enabled by high-efficiency SCRIBE (HiSCRIBE) DNA writers. (1) Recording spatiotemporal information into genomic DNA by writing unique barcodes into genomic DNA in response to stimuli. (2) Highly efficient and scarless genome editing without the requirement for double-strand DNA breaks and *cis*-encoded elements for genome editing of bacteria within communities.

3.3 Results

Optimizing SCRIBE for Molecular Recording and High-efficiency Genome Writing

The moderate genome editing efficiency of the original SCRIBE system limits its utility to population-level molecular recording (*67*) and makes it unsuitable for many dynamic genome-editing applications which could benefit higher editing efficiencies. To create a generalizable and highly efficient DNA writer, we sought to identify cellular factors that limit SCRIBE's DNA writing efficiency. To this end, we knocked out genes involved in mismatch repair system (MMR) (*22*) and exonucleases (*24*) that are thought to affect recombination efficiency and intracellular stability of ssDNAs, respectively. We measured SCRIBE genome editing efficiency in these different cellular knockout backgrounds in DH5αPRO cells, which overexpress *lacI* and *tetR*, using a *kanR* reversion assay (hereafter referred to as $kanR_{OFF}$ cells) (*67*). In this assay, two premature stop codons within a genomic *kanR* cassette ($kanR_{OFF}$) are reverted back to the wild-type (WT) sequence by recombineering of intracellularly expressed ssDNAs (ssDNA($kanR$)$_{ON}$). The retron cassette, which expresses ssDNA($kanR$)$_{ON}$, as well as the Beta protein, which promotes ssDNA-mediated recombination, were placed in a synthetic operon (dubbed SCRIBE($kanR$)$_{ON}$) under the control of an isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible promoter and expressed from a plasmid (Fig. 3.2A) (*67*). The SCRIBE writing efficiency in cells harboring the SCRIBE($kanR$)$_{ON}$

plasmid was assessed by measuring the recombinant frequency (ratio of Kanamycin resistant (Kan^R) cells to total viable cells) in the population in the presence or absence of IPTG induction. As shown in Fig. 3.2A, deactivating the MMR system (Δ*mutS*) resulted in only a modest increase in the recombination efficiency. This slight increase may reflect the fact that mismatches longer than three base pairs are poorly recognized by the MMR system (*24*). Knocking out *xseA*, an ssDNA-specific exonuclease that converts large ssDNA substrates into smaller oligonucleotides (*68*), slightly reduced recombination efficiency. On the other hand, knocking out *recJ* or *xonA*, which respectively encode 5'- and 3'-specific ssDNA exonucleases, significantly increased the recombinant frequency, suggesting that SCRIBE performance is limited by the availability of intracellular recombinogenic ssDNAs. Knocking out both *recJ* and *xonA* simultaneously increased the recombinant frequency even further, resulting in a $>10^4$-fold increase over the wild-type background. The editing efficiency of this improved DNA writing system (hereafter referred to as δHiSCRIBE to reflect the *recJ* and *xonA* knockout background) is comparable with the highest efficiencies reported for oligo-mediated recombineering (~10%) (*24, 61*). Furthermore, consistent with oligo-mediated recombineering, we found that the optimum length of homology arm between HiSCRIBE-generated ssDNA template and its target to be ~35 bps (Fig. S3.1).

Knocking out cellular exonucleases also increased the background recombination rate to some extent (Fig. 3.2A), which we speculate is likely due to recombination of ssDNA intermediates generated by the degradation of the template plasmid that persists in cells in the absence of cellular exonucleases (see Supplementary Materials and Fig. S3.1). To reduce the basal activity of the DNA writer, and to demonstrate that high-efficiency DNA recording is not limited to a specific genetic background, we conditionally knocked-down *recJ* and *xonA* exonucleases using CRISPR interference (CRISPRi) in the WT background, rather than using a knockout background (*69*). We cloned dCas9 and guide RNAs (gRNAs) targeting these two exonucleases under the control of anhydrotetracycline (aTc)-inducible promoters (Fig. 3.2B). We then co-transformed this plasmid along with the IPTG-inducible SCRIBE($kanR$)$_{ON}$ plasmid into the DH5αPRO $kanR_{OFF}$ reporter strain. Induction of either the SCRIBE or CRISPRi system resulted in a modest increase in the recombination efficiency, while co-induction of both systems resulted in an increase in recombination efficiency of $>10^4$-fold over uninduced cells (Fig. 3.2B). The recombinant frequency was significantly reduced when cells were transformed with a CRISPRi system lacking the gRNAs. No recombinants were detected in cells that were transformed with a non-targeting

SCRIBE(NS) plasmid. These results further support that cellular exonucleases limit SCRIBE genome editing efficiency and demonstrate that efficient DNA recording can be performed in genomically unmodified cells by combining SCRIBE and CRISPRi (hereafter referred to as the χHiSCRIBE system), with significantly less background compared to the exonuclease knockout (δHiSCRIBE) system. These features enable more efficient DNA recorders and other computing-and-memory circuits that use DNA as the computing substrate, without the need to genetically engineer target cells beforehand.

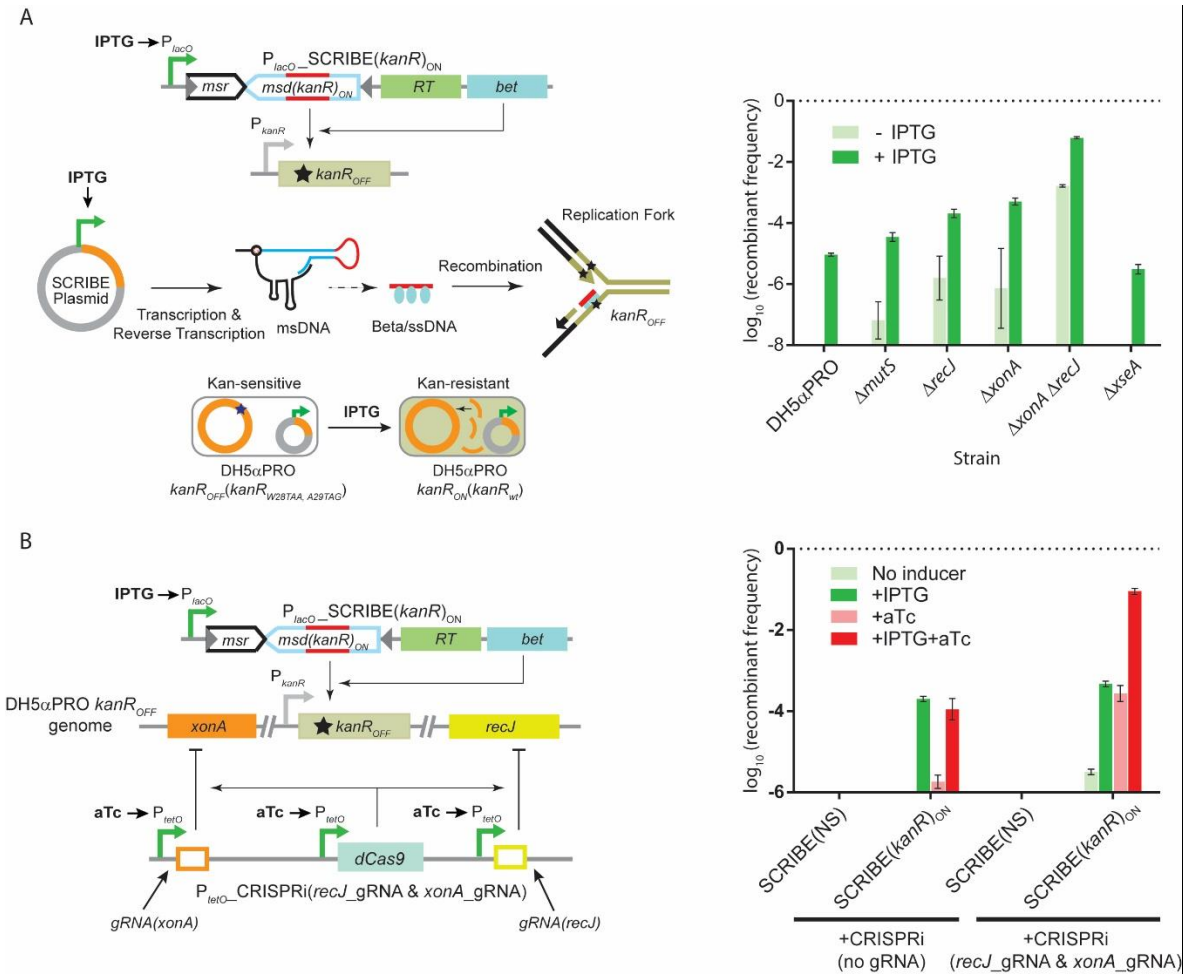


Figure 3.2 | Optimizing SCRIBE DNA Writing Efficiency. (A) SCRIBE DNA writing efficiency in different knockout backgrounds in *E. coli* DH5αPRO determined by a *kanR* reversion assay (see Methods). DNA writing efficiency in the Δ*xonA* Δ*recJ* was increased >10^4-fold relative to the wild-type background. This background thus was chosen to be used in the δHiSCRIBE system. **(B)** Combining IPTG-inducible SCRIBE and aTc-inducible CRISPRi system (to knockdown cellular exonucleases (*xonA* and *recJ*)) in the WT DH5αPRO strain enables efficient DNA memory recording and dynamic genome engineering. We refer to this system, which allows DNA writing in the WT background by knocking down the *xonA* and *recJ* genes,

as χHiSCRIBE to distinguish it from the δHiSCRIBE system (DNA writing in the Δ*xonA* Δ*recJ* background). Error bars indicate standard errors for three biological replicates.

In addition to molecular recording applications, such as linking a transcriptional signal to a precise mutation in the genome, HiSCRIBE DNA writers can be used for high-efficiency genome editing where maximal efficiencies are desired. Oligo-mediated recombineering can introduce desired modifications into a bacterial genome, but in this technique, synthetic oligos are introduced to target cells transiently (via electroporation) and have a very short intracellular half-life. Due to these shortcomings and the simultaneous presence of multiple replication forks in bacteria, the theoretical editing efficiency of oligo-mediated recombineering is limited to 25%, while the practical editing efficiency is often limited to a few percent (*24, 61*). Thus, multiple rounds of recombineering are needed to improve efficiency and additional screening steps are required to obtain desired mutants. In addition, to achieve such efficiencies, it is often necessary to modify the host by knocking out the MMR system, which may not be possible across a broad range of target strains. Knocking out MMR, in turn, elevates the global mutation rate and leads to undesirable genome-wide off-target mutations (*70*). On the other hand, HiSCRIBE provides a persistent intracellular source of recombinogenic oligos over many generations and can be introduced to cells even with low-efficiency delivery methods, thus bypassing the abovementioned limitations.

To demonstrate high-efficiency genome editing by HiSCRIBE writers, we optimized the δHiSCRIBE system by using a stronger Ribosome Binding Site (RBS) to overexpress Beta (Fig. S3.2). Using this system, we sought to change two consecutive leucine codons in the *galK* ORF in the MG1655 Δ*recJ* Δ*xonA* strain (hereafter referred to as MG1655 *exo*⁻ strain) to synonymous codons. Cells were transformed with the δHiSCRIBE(*galK*)$_{SYN}$ plasmid, which encodes an ssDNA with mismatches in three nucleotides to *galK* in order to write synonymous leucine codons into *galK* while effectively avoid the MMR system (*24*). We plated these cells on agar and then monitored the conversion of the genomic *galK*$_{WT}$ allele to the *galK*$_{SYN}$ allele in transformants at 24 hours after transformation (~30 generations) by colony PCR followed by Sanger sequencing as well as Illumina sequencing. As shown in Fig. 3.3A (middle panel), Sanger chromatograms obtained from colonies at this stage showed mixed peaks at the targeted nucleotides, indicating the presence of both *galK*$_{WT}$ and *galK*$_{SYN}$ alleles within single colonies. Sequencing these amplicons using Illumina MiSeq indicated that ~60% of the *galK*$_{WT}$ allele was converted to *galK*$_{SYN}$ after one day (Fig.

3.3A, bottom panel). Since Beta-mediated recombineering is a replication-dependent process (*35, 67*), the frequency of recombinants in HiSCRIBE-expressing populations should increase with more generations. We thus re-streaked the transformants on new plates to allow additional time for writing. After an additional day of growth, Sanger sequencing of *galK* PCR amplicons from these new colonies showed that the conversion of the $galK_{WT}$ allele to $galK_{SYN}$ was so efficient that the $galK_{WT}$ allele was below the limit of detection (Fig. 3.3A). Illumina sequencing of the amplicons further confirmed that ~100% of $galK_{WT}$ allele within individual colonies was converted to $galK_{SYN}$. When cells were transformed with a non-specific δHiSCRIBE (δHiSCRIBE(NS)) plasmid, no modified alleles were detected by sequencing.

We further assessed the DNA writing frequency in the entire population using a screenable plating assay, and observed that more than 99% of transformants (colony forming units (CFUs)) in the population underwent successful DNA editing after receiving the δHiSCRIBE plasmid (see Supplementary Materials and Fig. S3.3). Similar to the previous experiment, more than 99% of WT alleles within each CFU were converted into mutated alleles within 2 days (~60 generations). These results demonstrate that δHiSCRIBE is a highly efficient, broadly applicable, and scarless genome writing platform that can achieve ~100% editing efficiency at both the single-cell and population level without requiring any *cis*-encoded sequence on the target, double-strand DNA breaks, or selection.

The enrichment of a mutant allele within a population directly correlates with its fitness. In the absence of a selective advantage, it may take many generations for a neutral allele to accumulate within a population (*67*). As demonstrated in Fig. 3.3A, δHiSCRIBE by itself can achieve ~100% editing efficiency over the course of two days (~60 generations) during which the desired mutation accumulates in a replication-dependent fashion. We sought to increase the rate of this gene conversion process by putting a selective pressure against the WT allele at the nucleotide level using a CRISPR-Cas9 nuclease. We first constructed a $galK_{OFF}$ reporter strain by introducing two premature stop codons into the MG1655PRO Δ*recJ* Δ*xonA* strain (hereafter referred to as MG1655PRO *exo*⁻ $galK_{OFF}$ reporter strain). We encoded an aTc-inducible gRNA against the $galK_{OFF}$ allele into the δHiSCRIBE($galK$)$_{ON}$ plasmid, which expresses ssDNA with the same sequence as the WT *galK*. We transformed the δHiSCRIBE($galK$)$_{ON}$ plasmid into MG1655PRO *exo*⁻ $galK_{OFF}$ reporter cells containing either aTc-inducible Cas9 or dCas9 (as a negative control) plasmids. Single colonies of

transformants were grown, diluted, and regrown for multiple cycles in the presence or absence of aTc. The *galK* allele frequencies within the population were monitored by PCR amplification of the *galK* locus from the cultures followed by Illumina sequencing at different time points. As shown in Fig. 3B, $galK_{ON}$ alleles were enriched in all the cultures over time, further confirming that genome editing via HiSCRIBE is a replication-dependent process. However, upon induction with aTc, the $galK_{ON}$ alleles quickly enriched in cells expressing Cas9 compared to cells expressing dCas9, and comprised ~99% of *galK* alleles 12 hours after induction. These results demonstrate CRISPR-Cas9 nuclease activity, which is deleterious by itself if targeted against a bacterium's own genome (*65, 71*), can be combined with HiSCRIBE genome writing to induce selective sweeps and accelerate the enrichment of desired alleles in a population.

Efficient and Specific Genome Editing of Bacteria within Communities

Traditional recombineering techniques rely on high-efficiency delivery methods, such as electroporation or natural competence, for the introduction of synthetic oligos to the cells, and thus are limited to organisms where these with high-efficiency delivery methods are possible. Furthermore, these techniques are limited to certain laboratory conditions (e.g., highly electrocompetent cells grown to mid-log phage in test tubes) and cannot be applied to edit bacterial genome within complex communities or *in situ*. Unlike these traditional techniques, HiSCRIBE can be encoded on plasmids and delivered to cells via low transformation efficiency methods, such as chemical transformation, and alternative delivery methods, such as transduction and conjugation (Fig. S3.4), thus greatly expanding applicability of recombineering techniques to non-traditional hosts (Fig. S3.4) and complex bacterial communities.

To demonstrate the latter, we encoded HiSCRIBE on an M13 phagemid and used it to target and edit specific cells within a synthetic bacterial community. We used *E. coli* MG1655 $galK_{OFF}$ F^+ Str^R strain, which encodes the receptor for M13 bacteriophage, as our target strain. We introduced this target strain into an undefined bacterial community derived from mouse stool at a 1:100 ratio to make a synthetic bacterial community. To reduce the number of plasmids that needed to be delivered into target cells, we placed both HiSCRIBE($galK$)$_{ON}$ and the CRISPRi cassette targeting *recJ* and *xonA* exonucleases into a single synthetic operon, referred to as the χHiSCRIBE($galK$)$_{ON}$ operon, Fig. 3.3C). To allow for *in vivo* processing and release of these gRNAs from the synthetic operon transcripts, gRNAs were flanked by a Hammerhead Ribozyme (*HHR*) and a hepatitis delta virus Ribozyme (*HDVR*) (*72*).

We cloned this synthetic operon into a plasmid harboring the M13 bacteriophage packaging signal. χHiSCRIBE-encoding M13 phagemid particles were produced in an M13 packaging strain, purified, and added to the synthetic community or the reporter strain alone. The target cells were then scored on MacConkey + Streptomycin (Str) + galactose (gal) plates for the ability to metabolize galactose (*galK* reversion assay, see Methods). As shown in Fig. 3.3C, more than 99% of the reporter cells that received χHiSCRIBE($galK$)$_{ON}$ phagemids formed pink colonies on the indicator plates, demonstrating successful editing of targeted cells within a complex community. No pink colonies were observed in the negative control, in which the bacterial community was transduced with χHiSCRIBE(NS) phagemid particles.

As an additional control, we introduced $galK_{ON}$ oligo into reporter cells harboring the pKD46 recombineering plasmid in a clonal population or within the synthetic community using an established recombineering protocol (*24*). Consistent with previous reports, we observed ~10% recombineering efficiencies in the clonal population of the reporter strain. However, no recombinant pink colonies were obtained when reporter cells were contained within the synthetic community, further confirming that highly efficient delivery of oligos, as needed for traditional recombineering, are not achievable in bacterial communities. We further showed that conjugation, a common strategy for horizontal gene transfer in natural bacterial communities, can be used to deliver the χHiSCRIBE plasmid for genome editing within bacterial communities (Fig. S3.4B). However, the efficiency of plasmid delivery by conjugation was lower than transduction (Fig. S3.4C). These results demonstrate that diverse strategies can be used to deliver HiSCRIBE constructs into complex bacterial communities with the potential for *in situ* genome editing applications.

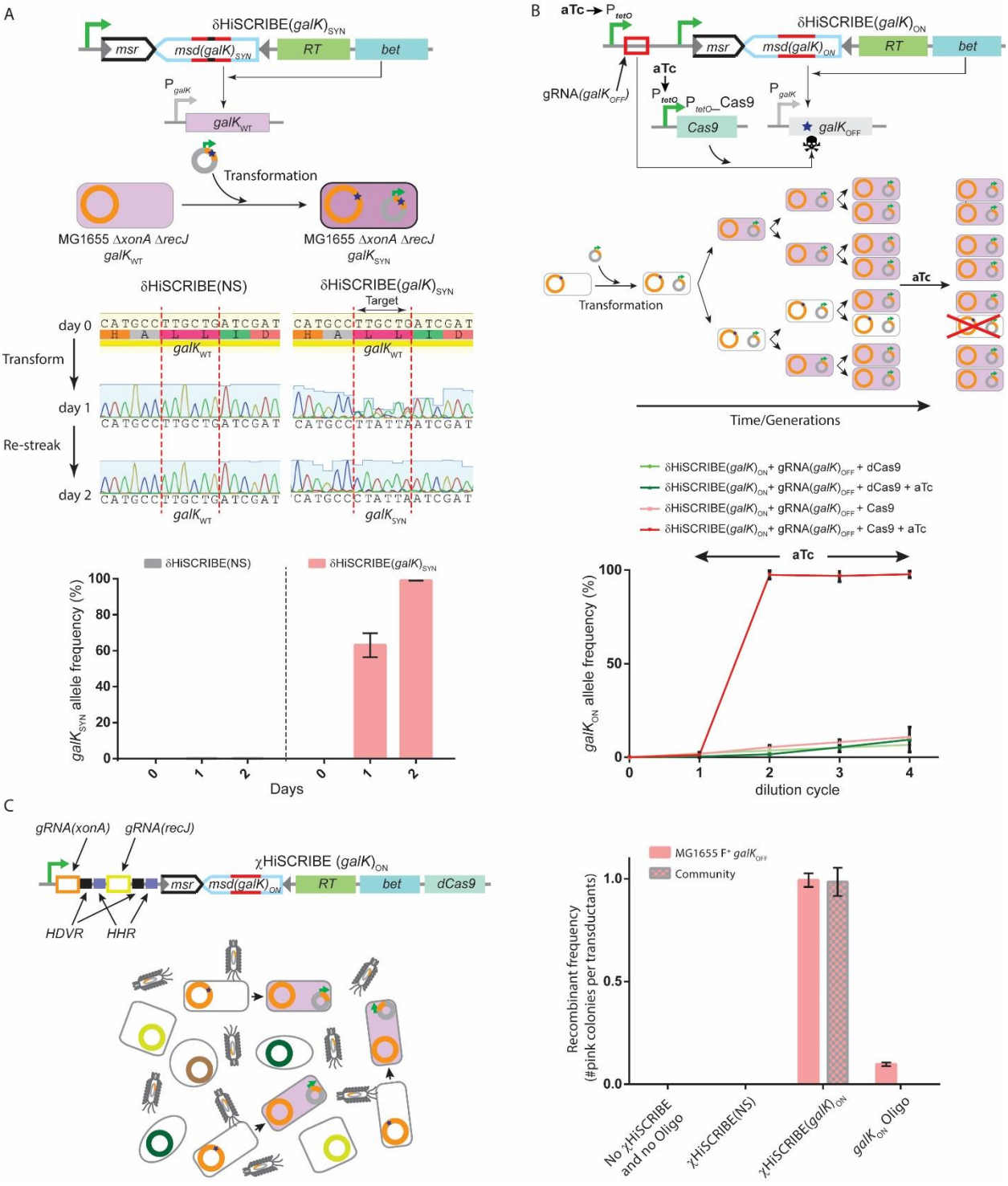


Figure 3.3 | Highly efficient, specific, scarless, and *cis*-element-independent editing of bacterial genomes by HiSCRIBE in clonal populations and within synthetic communities. (A) HiSCRIBE enables highly efficient genome editing in clonal populations. δHiSCRIBE was used to convert two consecutive leucine codons in the *galK* locus of MG1655 *exo* cells to synonymous codons. Cells were transformed with the δHiSCRIBE($galK$)$_{SYN}$ plasmid and the conversion of the $galK_{WT}$ to $galK_{SYN}$ was monitored 24 hours after transformation by PCR amplification of the *galK* locus of the transformants followed by Sanger sequencing (middle

panel) as well as Illumina sequencing (bottom panel). Re-streaking the transformants on new plates and growing cells for an additional 24 hours led to the ~100% conversion of the $galK_{WT}$ allele to $galK_{SYN}$ in all the tested transformants (also see Supplementary Materials and Fig. S3.3). No allele conversion was observed in cells that had been transformed with the non-specific δHiSCRIBE(NS) plasmid. **(B)** Combining δHiSCRIBE DNA writing with aTc-inducible CRISPR-Cas9 nuclease-mediated counterselection of unedited wild-type alleles increases the rate of enrichment of modified alleles within a population (see Methods). Error bars indicate standard deviation for three biological replicates. **(C)** Genome editing within a bacterial community via phagemid-mediated delivery of the χHiSCRIBE system. Target cells (*E. coli* MG1655 $galK_{OFF}$ F^+ Str^R) either as a clonal bacterial population or mixed with a stool-derived bacterial community were incubated with χHiSCRIBE($galK_{ON}$) or χHiSCRIBE(NS) phagemid particles and DNA writing efficiency in the *galK* locus was assessed by the *galK* reversion assay (see Methods). Recombinant frequency was calculated as the ratio of pink (galactose fermenting) colonies to target cell transductants. As an additional control, we used oligo-mediated recombineering with a synthetic $galK_{ON}$ oligo to edit reporter cells harboring a recombineering pKD46 plasmid either as a clonal population or in the context of a bacterial community. Recombinant frequency was calculated as the ratio of pink (galactose fermenting) colonies to total viable reporter cells. Transduction efficiencies of the χHiSCRIBE phagemids are presented in Fig. S3.4. Error bars indicate standard errors for three biological replicates.

Recording Spatial Information into DNA Memory

A useful feature of the HiSCRIBE system is that, unlike oligo-mediated recombineering, high-efficiency DNA writing can be linked to and triggered by biological processes. This feature could be especially useful to study events and interactions that occur in biological systems, such as cell-cell interactions within bacterial communities and biofilms, that are transient and thus hard to study in high throughput or with high resolution, especially in their native contexts. Enabled by efficient HiSCRIBE DNA writers, we devised a barcode joining strategy to uniquely mark and permanently record such transient interactions into DNA. This strategy reduces a transient multidimensional interactome space into a one-dimensional DNA sequence space. The recorded memory can be retrieved via high-throughput sequencing to study and map the interactome with high resolution and throughput, even after samples and interactions are disrupted.

We sought to demonstrate this concept by mapping conjugation events between bacterial populations. To this end, we first designated two neighboring 6 bp sequences on the *galK* locus as memory registers. We then constructed a series of δHiSCRIBE$(Reg1)_{r\text{-}barcode}$ and δHiSCRIBE$(Reg2)_{d\text{-}barcode}$ plasmids, each encoding a different barcoded ssDNA template. These plasmids each write a unique 7 bp DNA sequence (1 bp writing control + 6 bp barcode) on the first and the second registers, respectively (Fig. 3.4A). The writing control nucleotide was designed as a mismatch to the unedited register and used to selectively amplify edited registers (see Methods).

We introduced the δHiSCRIBE(Reg1)$_{\text{r-barcode}}$ plasmids into the MG1655 *exo*⁻ strain to make a set of conjugation recipient populations. Upon transformation, these plasmids write a unique barcode in the first genomic register in these cells (Register 1), and uniquely mark these recipient populations. δHiSCRIBE(Reg2)$_{\text{d-barcode}}$ plasmids, harboring an RP4 origin of transfer, were transformed into MFDpirPRO cells to make a set of conjugation donor populations. Upon successful conjugation and transfer from donor to recipient, these plasmids write a unique barcode in the Register 2 in recipient cells. Thus, sequencing the consecutive Register 1 and Register 2 in recipient genomes yield a record of this interaction (Fig. 3.4A). Using this barcode joining strategy, we first demonstrated that the interaction between a barcoded donor population and a barcoded recipient population could be successfully recorded and faithfully retrieved by allele-specific PCR of conjugation mixtures followed by Sanger sequencing (Fig. S3.5). To this end, we spotted a donor population with a single donor barcode on filter paper, overlapped it with another filter paper with a recipient population containing a single recipient barcode, and then confirmed that our retrieval process was correct (Fig. S3.5A). We then constructed more complex spatial layouts by overlapping multiple different barcoded donor populations and barcoded recipient populations. We demonstrated that allele-specific PCR combined with high-throughput sequencing could faithfully retrieve conjugative interactions between the distinct barcoded donor and recipient populations laid down in different patterns (Fig. 3.4B and Fig. S3.5B).

After validating that the barcode joining strategy can be used to map interactions between barcoded bacterial populations, we next sought to map cell-cell interactions at single-cell resolution as an example of a "cellular connectome". In this experiment, we used donor and recipient populations harboring pooled randomized barcodes that uniquely barcode individual cells in each population. Specifically, we constructed a pooled recipient population, harboring δHiSCRIBE(Reg1)$_{\text{r-rand}}$ plasmid library that encoded an ssDNA library with 6 randomized nucleotides targeting Register 1 in the *galK* locus. We also created a pooled donor library by transforming MFDpirPRO cells with a conjugative δHiSCRIBE(Reg2)$_{\text{d-rand}}$ plasmid library that encoded an ssDNA library with 6 randomized nucleotides targeting Register 1 in the *galK* locus. To test this method of recording mating interactions at the single-cell level, donor and recipient populations were mixed and spotted on filter paper on a solid agar surface to allow for conjugation of the δHiSCRIBE(Reg2)$_{\text{d-rand}}$ library from donors to recipients (Fig. 3.4C). Samples were then disrupted and grown in liquid cultures to allow for propagation and amplification of rare conjugated alleles. The two neighboring DNA memory registers

were amplified as a single amplicon by PCR, enzymatically digested to remove non-edited registers that contained parental restriction sites, and deep-sequenced (Fig. S3.6A, see Methods). Connectivity matrices between members of donor and recipient populations were then deduced based on the DNA barcodes obtained in the two specified memory registers (Fig. 3.4C).

To better understand the frequency of conjugation events between different population members, we analyzed the frequency of interacting donor and recipient barcodes in three parallel experiments. The degree of donor barcodes, which was defined as the number of different connections that each unique donor barcode makes with recipient barcodes, was well correlated between the three parallel experiments (Fig. S3.7). This suggests that the number of conjugation events that each donor barcode makes is independent of the identity of their interacting partners (i.e., recipient barcodes) and likely depends on the rate of transfer of donor barcodes, which itself is likely to be a function of the frequency of these barcodes in the population and the efficiency of transfer of each individual barcode. On the other hand, we observed a weak correlation between the degree of recipient barcodes, which was defined as the number of different connections that each unique recipient barcode makes with donor barcodes, in the three parallel experiments. This indicates that the number of donor barcodes that interact with each unique recipient barcode is different in each sample and suggests that other factors, such as the identity and frequency of donors in each iteration of conjugation, could affect the rate of successful conjugation and barcode transfer / writing in recipients.

To estimate the rate of false positives due to sequencing errors or spontaneous mutations, we calculated connectivity matrices for two other 6-bp regions within the *galK* locus that were not targeted by HiSCRIBE. Unique variants in the HiSCRIBE-targeted registers (both Register 1 and Register 2) were three orders of magnitude higher than in randomly chosen non-targeted regions (see Supplementary Materials and Fig. S3.6C). Further inspection of these mutated non-targeted regions revealed that they were mostly comprised of single base pair differences with the wild-type sequences, suggesting that these arose from sequencing errors, which are reportedly ~10^{-3}-10^{-2} mutations per nucleotide (*73*). False positives could be further reduced by using error-reducing library preparations, computational correction methods, and/or more accurate sequencing platforms (*73-75*).

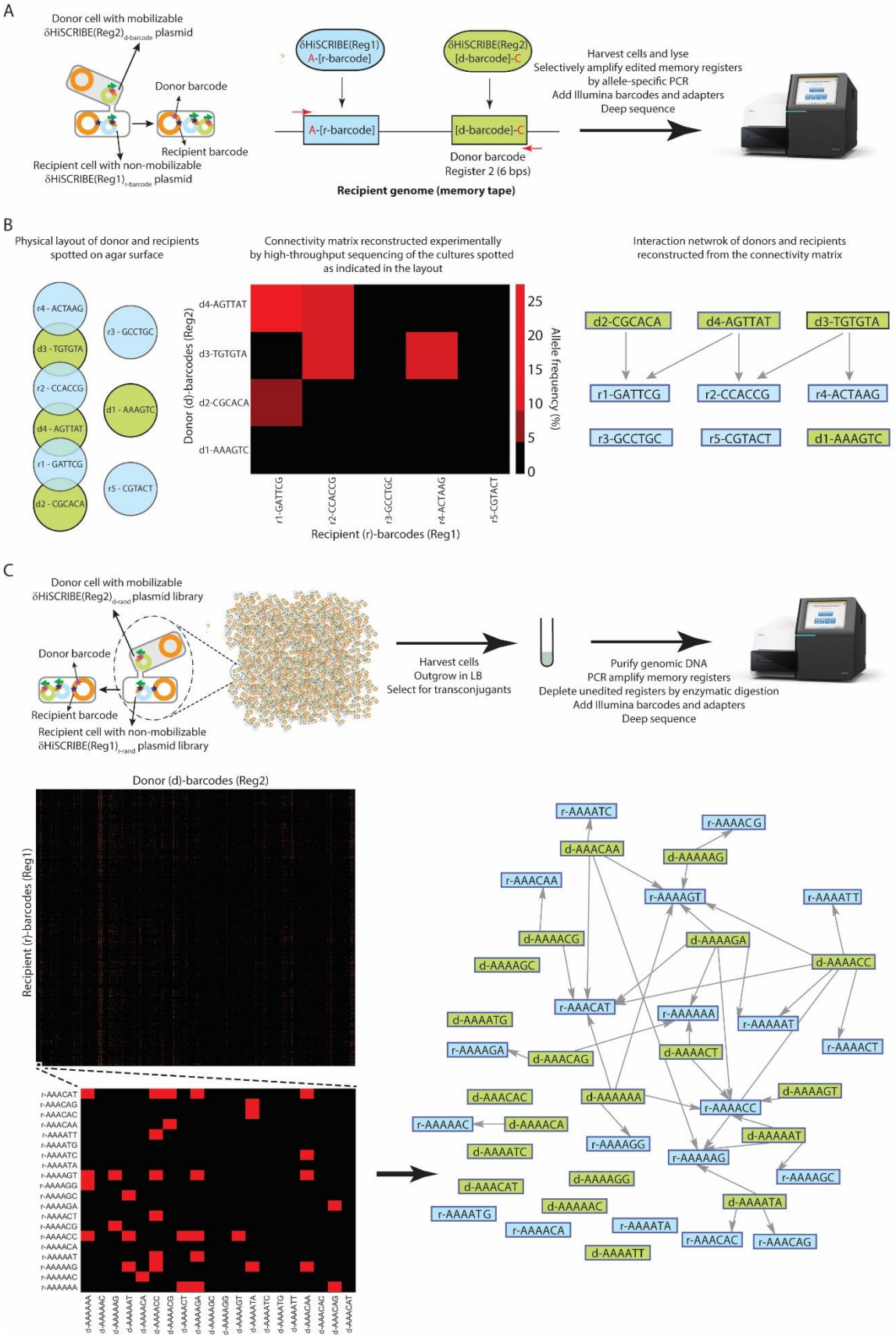

A
Donor cell with mobilizable δHiSCRIBE(Reg2)d-barcode plasmid
Donor barcode
Recipient barcode
Recipient cell with non-mobilizable δHiSCRIBE(Reg1)r-barcode plasmid
δHiSCRIBE(Reg1) A-[r-barcode]
δHiSCRIBE(Reg2) [d-barcode]-C
A-[r-barcode]
[d-barcode]-C
Donor barcode Register 2 (6 bps)
Recipient genome (memory tape)
Harvest cells and lyse
Selectively amplify edited memory registers by allele-specific PCR
Add Illumina barcodes and adapters
Deep sequence
B
Physical layout of donor and recipients spotted on agar surface
Connectivity matrix reconstructed experimentally by high-throughput sequencing of the cultures spotted as indicated in the layout
Interaction netwrok of donors and recipients reconstructed from the connectivity matrix
r4 - ACTAAG
d3 - TGTGTA
r2 - CCACCG
d4 - AGTTAT
r1 - GATTCG
d2 - CGCACA
r3 - GCCTGC
d1 - AAAGTC
r5 - CGTACT
d4-AGTTAT
d3-TGTGTA
d2-CGCACA
d1-AAAGTC
Donor (d)-barcodes (Reg2)
r1-GATTCG
r2-CCACCG
r3-GCCTGC
r4-ACTAAG
r5-CGTACT
Recipient (r)-barcodes (Reg1)
25
20
15
10
5
0
Allele frequency (%)
d2-CGCACA
d4-AGTTAT
d3-TGTGTA
r1-GATTCG
r2-CCACCG
r4-ACTAAG
r3-GCCTGC
r5-CGTACT
d1-AAAGTC
C
Donor cell with mobilizable δHiSCRIBE(Reg2)d-rand plasmid library
Donor barcode
Recipient barcode
Recipient cell with non-mobilizable δHiSCRIBE(Reg1)r-rand plasmid library
Harvest cells
Outgrow in LB
Select for transconjugants
Purify genomic DNA
PCR amplify memory registers
Deplete unedited registers by enzymatic digestion
Add Illumina barcodes and adapters
Deep sequence
Donor (d)-barcodes (Reg2)
Recipient (r)-barcodes (Reg1)
r-AAACAT
r-AAACAG
r-AAACAC
r-AAACAA
r-AAAATT
r-AAAATG
r-AAAATC
r-AAAATA
r-AAAAGT
r-AAAAGG
r-AAAAGC
r-AAAAGA
r-AAAACT
r-AAAACG
r-AAAACC
r-AAAACA
r-AAAAAT
r-AAAAAG
r-AAAAAC
r-AAAAAA
d-AAAAAA
d-AAAAAC
d-AAAAAG
d-AAAAAT
d-AAAACA
d-AAAACC
d-AAAACG
d-AAAACT
d-AAAAGA
d-AAAAGC
d-AAAAGG
d-AAAAGT
d-AAAATA
d-AAAATC
d-AAAATG
d-AAAATT
d-AAACAA
d-AAACAC
d-AAACAG
d-AAACAT
r-AAAATC
r-AAAACG
d-AAACAA
d-AAAAAG
r-AAACAA
r-AAAAGT
r-AAAATT
d-AAAACG
d-AAAAGA
d-AAAAGC
d-AAAACC
r-AAACAT
r-AAAAAA
d-AAAATG
r-AAAAAT
r-AAAAGA
d-AAACAG
d-AAAACT
r-AAAACT
d-AAACAC
d-AAAAAA
d-AAAAGT
r-AAAACC
r-AAAAAC
d-AAAACA
d-AAAAAT
r-AAAAGG
d-AAAATC
r-AAAAAG
d-AAACAT
d-AAAAGG
r-AAAAGC
d-AAAAAC
d-AAAATA
r-AAAATG
r-AAAACA
r-AAAATA
r-AAACAC
r-AAACAG
d-AAAATT

Figure 3.4 | Mapping spatial patterns and the connectome of conjugative mating pairs in bacterial populations. (A) Schematic representation of the barcode joining strategy used to record pairwise interactions (conjugation events) between conjugative pairs of bacteria using HiSCRIBE-based DNA writing. Upon successful conjugation, the interactions between a recipient cell and donor cell are recorded into neighboring DNA memory registers in the recipient cell genome. The edited registers are then amplified using allele-specific PCR (to deplete non-edited registers) and the identity of the interacting partners are retrieved by sequencing. The nucleotide shown in red in each register represents a single nucleotide that was included in each barcode to distinguish between unedited and edited registers. These "writing control" nucleotides were then used to selectively amplify edited registers by allele-specific PCR using primers that match these nucleotides but not to unedited registers. **(B)** Detecting the spatial organization of clonal bacterial populations. Clonal populations of donors and recipients harboring δHiSCRIBE-encoded "d-*barcode*" (green circles) and "r-*barcode*" (blue circles), respectively, were spotted on nitrocellulose filters that were then placed on the agar surface in the patterns shown in the left panel. Conjugation mixtures were harvested and the memory registers were amplified by allele-specific PCR and sequenced by Illumina sequencing (see Methods). Recorded barcodes in the two consecutive memory registers were parsed and the donor-recipient population connectivity matrix was calculated based on the percentage of reads corresponding to each possible pair-wise interaction of donors and recipient barcodes. The heatmap representation of the retrieved connectivity matrix (middle panel), as well as the corresponding interaction network (right panel), are shown. Red boxes in the heatmap depict connected barcodes, indicating that a conjugation event from the corresponding donor population resulted in δHiSCRIBE transfer and subsequent recording of the donor barcode into the specific recipient genome. In the interaction network, donor and recipient barcodes are indicated by green ("d-*barcode*") and blue ("r-*barcode*") rectangles, respectively. Data obtained from additional spatial patterns for bacterial populations are provided in Fig. S3.5. **(C)** The strategy used to map conjugation events between individual pairs of donor and recipient cells as a proxy for a conjugative connectome using randomized δHiSCRIBE libraries. The connectivity matrix was obtained using the method described in **B**. Due to the large size of the connectivity matrix (˜16 million elements), a submatrix for the first 20 (alphabetically sorted) barcodes of donors and recipients in one of the samples is shown in the inset. The y- and x-axis show recipient genomic barcodes (recorded in Register 1) and donor barcodes (recorded in Register 2), respectively. The corresponding interaction subnetwork for the presented connectivity submatrix is shown on the right.

With the described barcode joining strategy, up to 4^{12} ~ 16 million unique interactions can be theoretically recorded using two 6-nucleotide memory registers. The recording capacity can be increased by using larger barcodes. In our experiment, we could detect ~1% of the theoretical recording capacity (Fig. S3.6C). Increasing the sequencing depth or coverage of individual barcodes in the library could help to increase barcode recovery.

With these proof-of-concept experiments, we demonstrate that an efficient DNA writer that is coupled with biological processes can be used to memorize transient information, such as spatial patterns and cell-cell mating events between bacterial strains, into genomic DNA in their native context, which can then be later retrieved by sequencing. While only pairwise interactions were recorded in this proof-of-concept experiment, in

principle, multiple interactions can be recorded into adjacent DNA registers to map multidimensional interactomes with high-throughput sequencing, particularly as sequencing fidelity and read lengths continue to improve. Analogously to traditional imaging techniques that can take images or videos of spatial interactions, this "DNA imaging" strategy enables information capture into genomic DNA for later retrieval. We envision that DNA imaging by HiSCRIBE and other analogous efficient DNA writing systems could be used for high-throughput and high-resolution mapping of cellular organization and connectomes, as well as other types of transient interactions, in environments and conditions where traditional imaging techniques cannot work.

3.4 Discussion

Recently, different DNA writing technologies for recording molecular events into the DNA of living cells have been described (*76, 77*). Memory recording using site-specific recombinases (*76*), CRISPR spacer acquisition (*77*) and CRISPR-Cas9 (*78*) requires *cis*-encoded elements (e.g., recombinase sites, CRISPR repeats, PAM domains, etc.) and is confined to predefined sequences. In contrast, HiSCRIBE writers do not require any *cis*-encoded element on the target and thus are the first class of DNA writers that open up the entire genome for high-efficiency editing and memory applications. Furthermore, HiSCRIBE enables active and dynamic modification of bacterial genomes without the need to introduce double-strand DNA breaks, which may help to reduce associated cytotoxicity and unwanted chromosomal rearrangements. This feature is especially important for genome editing in the context of bacterial communities and evolutionary engineering applications, where fitness costs could be deleterious for the targeted population. Additionally, unlike genome editing strategies that rely on counterselection by CRISPR-Cas9 nucleases (*63*), HiSCRIBE does not require the presence of a PAM sequence on the target, thereby allowing one to perform multiple rounds of allele replacement on the same target, a property which is especially important for evolutionary engineering applications. Thus, HiSCRIBE DNA writers could enable temporally- and spatially-regulated continuous targeted diversity generation, which would be useful for evolutionary engineering of cellular phenotypes if coupled to continuous selection. Furthermore, HiSCRIBE template plasmids can serve as unique barcodes to identify and track mutations and their enrichment in genome-wide trait optimization scenarios, a challenge for traditional recombineering-based approaches (*79*). Finally, by providing a sustainable source of mutagenic oligos *in vivo*, the HiSCRIBE system can help to bypass current limitations in performing recombineering in non-traditional hosts where transformation is difficult (*80*) and expand the applicability of recombineering-based techniques to *in situ* conditions.

Thus, HiSCRIBE DNA writers offer unprecedented and desirable features (summarized in Table S3.1) for dynamic recording and genome engineering in bacteria and open up a plethora of new avenues for investigating and engineering biology.

To highlight the power of dynamic recording and autonomous genome engineering enabled by an efficient *in vivo* DNA writing system such as HiSCRIBE, we demonstrated that spatial information such as cellular patterns and cell-cell interactions can be mapped with high resolution and throughput using efficient DNA writers with next-generation sequencing. This approach could be used to study bacterial spatial organization within biofilms, which has been challenging to study with traditional techniques (*81*). In future work, HiSCRIBE could be encoded in phages, conjugative plasmids, or other mobile genetic elements and designed to write similar barcodes near identifiable genomic signatures (e.g., 16S rRNA gene) to assess the *in situ* host range of these mobile elements. In addition, efficient and conditional DNA writers could be used to record other types of transient spatiotemporal events, such as protein-protein interactions, into DNA for high-throughput studies. Furthermore, extending this DNA imaging approach to multicellular organisms and mammalian cells using analogous high-efficiency DNA writing technologies could be used to record cellular interactions such as neural connectomes (*82-84*). Although we only used our system to record spatial and temporal biological events in this paper, in principle, arbitrary information can be encoded and written into the genomic DNA of living cells (*85*). For example, digital information (e.g., documents, images, videos, etc.) could be encoded into HiSCRIBE ssDNA templates and written across various genomic regions in living cell populations. The recorded memory could then be retrieved by sequencing the genomic memory registers.

In summary, our work sheds light into various factors that modulate the efficiency of *in vivo* ssDNA-mediated recombineering, circumvents limitations imposed by other genome editing strategies, offers a framework for dynamic engineering of bacterial genomes with high efficiency and precision, and demonstrates and foreshadows multiple useful applications that are enabled by efficient and dynamic *in vivo* DNA writing. Inspired by the approaches described here, we anticipate that HiSCRIBE DNA writing technology, along with analogous DNA writers, will have broad biomedical utility, including single-cell memory and computing, *in situ* engineering of bacterial genomes within communities, spatiotemporal molecular recording and connectome mapping, continuous *in vivo* evolution of single-gene (e.g., protein function) or multi-gene (e.g., metabolic network) traits, evolutionary dynamics and gene resurrection studies, and other yet to be defined applications.

3.5 Supplementary Information

A Model for HiSCRIBE DNA Writing

Knocking out cellular exonucleases increases the background recombinant frequency in the uninduced δHiSCRIBE system (Fig. 3.2A). This could be due to the leakiness of the promoter expressing the δHiSCRIBE operon (P_{lacO}) and/or an elevated recombination rate between the double-stranded DNA plasmid template and its target site in the Δ*recJ* Δ*xonA* background, as reported previously (*38*). To investigate these two possibilities, we measured the recombinant frequency in the presence and absence of reverse transcriptase (RT) activity in different genetic backgrounds. An elevated recombinant frequency was observed even in the presence of inactive RT (Fig. S3.1A). However, in all of the tested conditions, cells expressing an active RT showed about two orders of magnitude greater recombinant frequencies compared to those expressing an inactive RT (Fig. 3.2A and Fig. S3.1A). Furthermore, the recombinant frequency in the absence of Beta in the Δ*recJ* Δ*xonA* background was also elevated (Fig. S3.2) compared to the WT background, which was calculated to be ~10^{-9} events per generation (*67*).

These results are consistent with a previously proposed model in which double-stranded plasmid templates can be recombined into a genomic target site via ssDNA intermediates through a RecA-independent process normally repressed by cellular exonucleases (*38*). We speculate that even in the absence of an active retron system, recombinogenic oligonucleotides are produced *in vivo*, likely due to plasmid degradation by cellular nucleases. This intracellular ssDNA pool could then be processed and further degraded by cellular exonucleases, thus limiting the efficiency of recombination in the WT background. However, when cellular exonucleases (*recJ* and *xonA*) are knocked out (in the δHISCRIBE system), the intermediate degradation products of retron-encoded ssDNAs, as well as the template double-stranded DNA, could accumulate and contribute to the intracellular ssDNA pool, thereby increasing recombination efficiency (Fig. S3.1B). This model is further supported by previous observations in which the efficiency of oligo-mediated recombination directly correlated with the concentration of transformed oligos (*24, 86*). The addition of non-specific carrier ssDNAs can also compensate for low concentrations of specific ssDNA, potentially by transiently saturating cellular nucleases (*24*). In this working model (Fig. S3.1B), Beta recombinase protects the intracellular oligonucleotide pool from cellular exonucleases and facilitates recombination between the ssDNAs and their corresponding genomic target loci.

In our experiments, ssDNAs were specifically designed to have at least three mismatches to the target in order to efficiently suppress MMR system (*24*) and achieve high-efficiency writing. Inefficient recognition of mismatched lesions, which is likely to occur in the absence of ssDNA expression in an exonuclease knockout background, could also contribute to the increased background observed in the δHiSCRIBE system.

Knocking out *xseA*, which encodes one of the two subunits of ExoVII, slightly reduced the recombination efficiency (Fig. 3.2A). ExoVII is an ssDNA-specific exonuclease that converts large ssDNA substrates into smaller oligonucleotides (*68*) and has been shown to be responsible for the removal of phosphorothioated nucleotides from the flanking ends of recombineering oligos (*43*), as well as the removal of the *msr* moiety from the msDNA of RNA-less retrons (*87*). Based on these observations, we speculate that ExoVII, among other cellular factors, may be involved in generating recombinogenic ssDNA intermediates. It is also possible that RecBCD-mediated processing of double-stranded breaks could provide another source for the intracellular recombinogenic ssDNA pool (*88*).

Lastly, the optimal length of the flanking ssDNA homology arms that result in maximal HiSCRIBE editing efficiency was found to be around 35 bps (Fig. S3.1C). Increasing the size of homology arm to 80 bp reduced the recombination efficiency, which we speculate could be due to secondary structures that prevent efficient recombination and/or inefficient ssDNA production by the retron system. These results are consistent with previous reports for recombineering with synthetic oligos (*24*), and further confirm the involvement of a RecA-independent, Beta-mediated process in high-efficiency DNA writing by HiSCRIBE.

Measuring δHiSCRIBE DNA Writing Efficiency with a Screenable Phenotype and High-throughput Sequencing

To systematically assess δHiSCRIBE writing efficiency in an entire population, we used a screening assay with colorimetric readout. We introduced two stop codons into the *galK* ORF of the MG1655 Δ*recJ* Δ*xonA* (*exo*⁻ $galK_{OFF}$) reporter strain. These reporter cells were transformed with δHiSCRIBE$(galK)_{ON}$ (δHiSCRIBE plasmid encoding ssDNA identical to the WT *galK*). These cells were recovered for one hour in LB (37°C, 300 RPM) and plated on MacConkey + galactose (gal) + antibiotic plates in order to select for transformants. The conversion of the $galK_{OFF}$ allele to $galK_{ON}$ (i.e., the WT allele) was monitored by scoring the color of transformant colonies. As shown in Fig. S3.3, all the $galK_{OFF}$ (white) cells transformed with the

δHiSCRIBE(*galK*)$_{ON}$ plasmid formed galactose-fermenting *galK*$_{ON}$ (pink) colonies on the indicator plates. No pink colonies were detected when cells were transformed with a non-specific δHiSCRIBE (δHiSCRIBE(NS)) plasmid. These results demonstrate that in the entire population of cells that received the δHiSCRIBE(*galK*)$_{ON}$ plasmid, *galK*$_{OFF}$ alleles were converted to *galK*$_{ON}$ over the course of colony growth, resulting in a phenotypic change in colony color.

Since Beta-mediated recombineering is a replication-dependent process (*35, 86*), the conversion of *galK*$_{OFF}$ to *galK*$_{ON}$ occurs over the course of growth of the colonies, and a single pink colony observed on a transformation plate may contain a heterogeneous population of both edited and non-edited alleles. We measured the frequency of these alleles within single colonies by PCR amplification of the *galK* locus followed by Sanger sequencing as well as high-throughput sequencing. To avoid any difference in fitness between the two alleles in the presence of galactose, after we transformed the δHiSCRIBE(*galK*)$_{ON}$ plasmid into *exo galK*$_{OFF}$ reporter cells, we selected transformants on LB plates, instead of MacConkey + gal plates. Sanger sequencing of PCR amplicons of the *galK* locus obtained from these transformants showed a mixture of peaks in the target site, suggesting that each colony on these plates may have contained a mixture of edited and non-edited alleles (Fig. S3.3). To give the replication-dependent δHiSCRIBE writing system additional time to work, we re-streaked the colonies on fresh plates. Sanger sequencing of *galK* locus amplicons obtained from these colonies indicated the full conversion of *galK*$_{OFF}$ allele to *galK*$_{ON}$, to the extent that the *galK*$_{OFF}$ allele was below the limit of detection (Fig. S3.3). These results were further quantified and validated by high-throughput sequencing of *galK* amplicons (Fig. S3.3). These results indicate that δHiSCRIBE system can be used to edit a desired genomic locus up to homogeneity (~100% efficiency) in an entire population, and without the requirement for any double-strand DNA breaks and *cis*-encoded elements on the target.

Delivering HiSCRIBE via Different Strategies for Editing Bacteria within Bacterial Communities and Editing Non-traditional Hosts

To facilitate the delivery of HiSCRIBE for DNA writing in non-modified hosts, we placed the HiSCRIBE and CRISPRi systems into a single synthetic operon (referred to as χHiSCRIBE operon as shown in Fig. 3.3C and S3.4A), cloned it into a high-copy-number plasmid, and assessed its performance in the WT MG1655 *galK*$_{OFF}$ reporter strain, which harbors two stop codons within the *galK* locus. Cells were chemically transformed with either χHiSCRIBE(*galK*)$_{ON}$ or χHiSCRIBE(NS), which expressed a

galK$_{ON}$ ssDNA or a non-specific ssDNA, respectively. The cells were recovered in LB for an hour, then plated on MacConkey + gal + antibiotic plates to select for χHiSCRIBE plasmid delivery and screen for *galK*$_{OFF}$ to *galK*$_{ON}$ editing. More than 99% of cells transformed with the χHiSCRIBE(*galK*)$_{\text{ON}}$ plasmid formed pink colonies on these plates, indicating successful writing in the *galK* locus in all cells that received this plasmid (Fig. S3.4A). No pink colonies were detected in the samples transformed with the χHiSCRIBE(NS) plasmid. The frequency of editing within individual colonies was assessed by PCR amplification of *galK* locus followed by high-throughput sequencing at 24 hours after transformation, as well as after a re-streaking step as described before (Fig. S3.4A).

Similar to transduction, conjugation is a common strategy for horizontal gene transfer in natural bacterial communities. In addition to using transduction for delivering χHiSCRIBE plasmids (Fig. 3.3C), we tested whether conjugation can be used to deliver and edit cells within a complex bacterial community. We encoded the origin of transfer from RP4 (oriT) into the χHiSCRIBE(*galK*)$_{\text{ON}}$ plasmid and then introduced this plasmid into MFDpirPRO cells (that harbor RP4 conjugation machinery) to produce a donor strain. We showed that these cells could conjugate the χHiSCRIBE(*galK*)$_{\text{ON}}$ plasmid into recipient cells (MG1655 Str$^{\text{R}}$ *galK*$_{\text{OFF}}$). More than 99% of transconjugants formed pink colonies on MacConkey + gal + antibiotic plates (Fig. S3.4B), while no pink colonies were obtained in recipients that had been conjugated with the non-specific χHiSCRIBE(NS) plasmid. We then conjugated the χHiSCRIBE(*galK*)$_{\text{ON}}$ plasmid into a stool-derived bacterial community containing MG1655 Str$^{\text{R}}$ *galK*$_{\text{OFF}}$, analogously to the transduction experiments (Fig. 3.3C). More than 99% of transconjugants that received the χHiSCRIBE(*galK*)$_{\text{ON}}$ plasmid formed pink colonies on the screening plates and no pink colonies were detected in cells conjugated with the non-specific χHiSCRIBE(NS) plasmid (Fig. S3.4B). However, the efficiency of delivery via conjugation was significantly lower than phagemid transduction (Fig. S3.4C). We anticipate that more specific transduction delivery mechanisms are better suited for editing specific species within a community, while more generalized (albeit less efficient) conjugation delivery mechanism is better suited for situations where editing a larger subpopulation of bacteria in the community are desired.

Next, to demonstrate the applicability of our system for DNA writing in non-traditional hosts, we used this system for genome editing in *Pseudomonas putida* (*P. putida*). To this end, we designed SCRIBE(*upp*)$_{\text{OFF}}$ plasmids targeting either the

lagging strand or the leading strand of the uracil phosphoribosyltransferase (*upp*) ORF to introduce two premature stop codons into this ORF, thus making cells insensitive to 5-fluorouracil (5-FU). SCRIBE cassettes were cloned into a broad-host-range plasmid (harboring the pBBR1 origin of replication) and transformed into the *P. putida* KT2440 strain. Recombinant frequency was assayed by measuring the ratio of cells resistant to 5-FU to viable cells. While targeting the leading strand did not result in a significant increase in the editing efficiency, targeting the lagging strand improved the editing efficiency by about two orders of magnitude, demonstrating that SCRIBE is functional in *P. putida* (Fig. S3.4D). The editing efficiency may be further improved by using strategies described in this work, including knocking out homologs of *recJ* and *xonA* in *P. putida* (or knocking down these genes using CRISPRi), counterselection by CRISPR-Cas9 nucleases, and using homologs of Beta that are more active in *Pseudomonas* (*80*).

Materials and Methods

Strains and Plasmids

Conventional cloning methods, Gibson assembly (*89*) and Golden Gate assembly (*90*) were used to construct the plasmids. Lists of strains and plasmids used in this study are provided in Tables S3.2 and S3.3, respectively. The sequences for the synthetic parts and primers are provided in Tables S3.4 and S3.5, respectively. Constructs will be available on Addgene.

Cells and Antibiotics

Chemically competent *E. coli* DH5α F' $lacI^q$ (NEB) was used for cloning. Unless otherwise noted, antibiotics and small molecule inducers were used at the following concentrations: Carbenicillin (Carb, 50 μg/mL), Kanamycin (Kan, 20 μg/mL), Chloramphenicol (Cam, 30 μg/mL), Streptomycin (Str, 50 μg/mL), 5-fluorouracil (5-FU, 20 μg/mL), anhydrotetracycline (aTc, 200 ng/mL) and Isopropyl β-D-1-thiogalactopyranoside (IPTG, 1mM).

Induction of Cells and Plating Assays

The *kanR* reversion assay was performed as described previously (*67*). Briefly, for each experiment, single colony transformants were separately inoculated into LB broth + appropriate antibiotics and grown overnight (37°C, 300 RPM) to obtain seed cultures. Unless otherwise noted, inductions were performed by diluting the seed cultures (1:1000) in LB + antibiotics ffi inducers followed by 24 h (corresponding to $\log_2(1000)$ ~10 generations of growth) incubation (37°C, 700 RPM) in 96-well plates. Cultures

were then serially diluted and spotted on selective media to determine the number of recombinant and viable cells in each culture. The number of viable cells was determined by plating serial dilutions of the cultures on LB plates with antibiotics corresponding to the marker present on the δHiSCRIBE plasmid (Carb or Cam). LB + Kan plates were used to determine the number of recombinants. For each sample, the recombinant frequency was reported as the mean of the ratio of recombinants to viable cells for three independent replicates.

In the *galK* conversion assays, HiSCRIBE plasmids were delivered to reporter cells (with either chemical transformation, transduction or conjugation) and cells were recovered in LB for one hour without selection and plated on LB + appropriate antibiotics for HiSCRIBE plasmid selection. Allele frequency was measured by MiSeq sequencing of colonies obtained on these plates after 24 h (corresponding to $\log_2(10^9)$ ~30 generations of growth (*91*)). Additionally, for $galK_{OFF}$ to $galK_{ON}$ reversion experiments, cells were plated on MacConkey agar base (without carbon source) + galactose (1%) + appropriate antibiotics (for HiSCRIBE plasmid selection). The ratio of pink colonies ($galK_{ON}$) to transformants (pink + white colonies) was used as a measure of recombinant frequency. For each sample, the recombinant frequency was reported as the mean of the ratio of recombinants to viable transformants for three independent replicates.

In the CRISPR-Cas9 counter-selection experiment (Fig. 3.3B), a gRNA against the $galK_{OFF}$ locus (gRNA($galK_{OFF}$)) was placed under the control of an aTc-inducible promoter and cloned into the δHiSCRIBE($galK$)$_{ON}$ plasmid. This plasmid was transformed into a $galK_{OFF}$ reporter strain harboring an aTc-inducible Cas9 (or dCas9 as a negative control) plasmid. Single transformant colonies were diluted to ~10^6 CFU/mL in LB + Carb + Cam in the presence or absence of aTc and grown for 12 hours. These cultures were diluted and regrown for two additional cycles at the presence or absence of the inducer. The allele frequencies were determined by PCR amplification of the *galK* locus followed by high-throughput sequencing.

Phagemid Packaging

χHiSCRIBE plasmids were packaged into M13 phagemid particles as described previously (*92*). Briefly, χHiSCRIBE plasmids with the M13 origin of replication were transformed into an M13 packaging strain (DH5αPRO F^+ harboring the M13cp helper plasmid (*92*)). Single colony transformants were grown overnight in 2 mL LB + antibiotics. The cultures were then diluted (1:100) in 50 mL fresh media and grown up

to saturation with selection. Phagemid particles were purified from the culture supernatants by PEG/NaCl precipitation (*93*), passed through a 0.2-μm filter and stored in SM buffer (50 mM Tris-HCl [pH 7.5]), 100 mM NaCl, 10 mM $MgSO_4$) at 4°C for later use.

Delivery by Transduction and Conjugation

For transduction experiments, overnight cultures of the reporter strains harboring an F-plasmid were diluted (1:1000) in fresh media and transduced by adding purified phagemid particles encoding χHiSCRIBE at a Multiplicity of Infection (MOI) of 50, unless otherwise noted. After one hour incubation (37°C, 700 RPM), serial dilutions of the cultures were spotted on MacConkey + gal + antibiotics plates and recombinant frequency was calculated as described above (*galK* reversion assay).

For conjugation delivery, the MFDpirPRO strain was first produced by transforming the PRO plasmid (pZS4Int-*lacI*/*tetR*, Expressys) into the diaminopimelic acid (DAP)-auxotrophic MFDpir strain (*94*) that encodes RP4 conjugation machinery. Media for the donor strains was supplemented with 0.3 mM DAP throughout the experiments. δHiSCRIBE or χHiSCRIBE plasmids harboring RP4 origin of transfer were transformed into the MFDpirPRO strain to produce donor strains. Overnight cultures of donor and recipient strains were diluted (1:100) in fresh media and grown to an OD_{600} ~1. Cells were pelleted and resuspended in LB, and mating pairs were mixed at a donor to recipient ratio of 100:1 and spotted onto nitrocellulose filters placed on LB agar supplemented with 0.3 mM DAP. The plates were incubated at 37°C for 6 h to allow conjugation. Conjugation mixtures were collected by vigorously vortexing the filters in 1 mL PBS, then serially diluted and spotted on MacConkey + gal + antibiotics plates as described in the *galK* reversion assay. The ratio of pink colonies per transconjugants was used as a measure of recombinant frequency.

For experiments showing genome editing in bacterial communities (Fig. 3.3C and S3.4B), an overnight culture of an undefined bacterial community was obtained by inoculating mouse stool in LB. This bacterial community was mixed (100:1) with a spontaneous Str^R resistant mutant of the MG1655 $galK_{OFF}$ reporter strain to build a synthetic bacterial community that served as the recipient cell population in these experiments. For transduction experiments, the F plasmid (from DH5α F^+ (NEB)) was introduced to the reporter strain via conjugation. The transduction and conjugation

protocols were performed as described above, using the synthetic community as the recipient population.

Bacterial Connectome Mapping

To demonstrate that spatial information can be recorded into DNA memory, we mapped the pairwise connectome network of mating pairs in conjugating bacterial populations (Fig. 3.4C). The δHiSCRIBE(Reg1)$_{\text{r-rand}}$ library (overexpresses an ssDNA library with 6 randomized nucleotides targeting Register 1 in the *galK* locus, pZA11 backbone) was transformed into MG1655 Δ*recJ* Δ*xonA* *galK*$_{\text{OFF}}$ to make a barcoded recipient population. A mobilizable δHiSCRIBE(Reg2)$_{\text{d-rand}}$ library (overexpressing an ssDNA library with 6 randomized nucleotides targeting Register 2 in the *galK* locus, pZE32 backbone) was transformed into MFDpirPRO cells to serve as the donor population. The donor and recipient populations were mixed at a 10:1 ratio (three parallel experiments) and conjugated as described above (37°C for 6 h). Conjugation mixtures were collected by vigorously vortexing nitrocellulose filters in 3 mL LB (without DAP) and recovered for 1 hour, after which antibiotics (Carb + Cam) were added to select for transconjugants harboring δHiSCRIBE(Reg1)$_{\text{r-rand}}$ and δHiSCRIBE(Reg2)$_{\text{d-rand}}$ plasmids. Samples were grown at 37°C overnight in the absence of DAP to selectively remove donor cells and allow HiSCRIBE writing and propagation of the edited alleles. Genomic DNA was prepared from the overnight cultures and the contents of memory registers were analyzed by high-throughput sequencing as described below.

For the bacterial organization mapping experiment (Fig. 3.4B), barcoded clonal donor and recipient populations harboring δHiSCRIBE(Reg1)$_{\text{r-barcode}}$ and δHiSCRIBE(Reg2)$_{\text{d-barcode}}$ were spotted as indicated patterns and conjugated as described above (37°C for 6 h). After conjugation, allele-specific PCR was used (see below) to amplify the edited registers directly from conjugation mixtures (without any outgrow).

High-throughput Sequencing

Allele frequencies of the HiSCRIBE target sites were measured by sequencing amplicons obtained from corresponding genomic sites using Illumina MiSeq. Target loci were amplified using 1 μL of liquid culture (or colony resuspension) as template. Barcodes and Illumina adapters were then added in an additional round of PCR. Samples were gel-purified, multiplexed, and sequenced by Illumina MiSeq. The

obtained reads were demultiplexed based on attached barcodes and mapped to the reference sequence.

For *galK* conversion experiments, any reads that lacked the expected "ATGCCXXXXXXATCGAT" motif, where "XXXXXX" corresponds to the 6-bp variable site in the *galK* alleles (TTGCTG for $galK_{WT}$, CTATTA for $galK_{SYN}$, CTCTTG for $galK_{ON}$, and TAATGA for $galK_{OFF}$), or that contained ambiguous nucleotides within this region were discarded. For $galK_{WT}$ to $galK_{SYN}$ experiment, editing efficiency was reported as the ratio of $galK_{SYN}$ reads to the total number of $galK_{SYN}$ + $galK_{WT}$ reads. For *galK* reversion experiments, editing efficiency was calculated as the ratio of $galK_{ON}$ reads to the total number of $galK_{ON}$ + $galK_{OFF}$ reads The enrichment of recombinant alleles in the WT *E. coli* MG1655 background (Fig. S3.4A) was investigated similarly. Single colonies of transformants were picked 24 h (or 48 h) after transformation, resuspended in water, and used as templates for PCR. The samples were processed as described above.

For the bacterial spatial organization recording and connectome mapping experiments (shown in Fig. 3.4B and Fig. 3.4C, respectively), barcoded donor and recipient populations were conjugated as described above. For the former experiment, conjugation mixtures were resuspended in LB and the memory registers in the *galK* locus were amplified by allele-specific PCR to deplete unedited registers. As shown in Fig. S3.5A, we designed primers that specifically bind to the writing control nucleotide of edited alleles but form a mismatch (at the 3'-end position) with the unedited registers. We then used these primers and HiDi DNA polymerase (a selective variant of DNA polymerase that can only amplify templates that are perfectly matched at the 3'-end with a given primer, myPLOS Biotec, DE) to specifically amplify edited registers from 1 μL of conjugation mixtures while depleting the unedited registers. Illumina barcodes and adapters were then added to the samples by a second round of PCR. Samples were gel-purified, multiplexed, and sequenced by Illumina MiSeq. Samples were then computationally demultiplexed, and any reads that contained non-edited registers, which lacked any of the two expected motifs flanking the two memory registers (ATGCCTMMMMMMMTCGATT and AGTGCGNNNNNNGTGCGC, where "MMMMMM" and "NNNNNN" correspond to positions of the memory Registers 1 and 2, respectively), or that contained ambiguous nucleotides within this region were discarded. The frequencies of variants that were observed simultaneously in a single read in the two registers were then calculated and presented as weighted connectivity matrices (Figs. 3.4B and S3.5B).

For the latter experiment, an alternative depletion strategy was used. Specifically, genomic DNA was purified from overnight cultures of the conjugation mixtures using the ZR Fungal/Bacterial DNA MiniPrep kit (Zymo Research). A DNA fragment including Registers 1 and 2 in the *galK* locus was PCR amplified from purified genomic DNA and gel purified. The samples were depleted of non-edited (i.e., WT) sequences by enzymatic digestion with ClaI and AgeI, since these sites are present in non-edited Register 1 and 2, but are removed after HiSCRIBE recording. Samples were subsequently run on TBE gels (6%) and uncut fragments (edited in both Registers) (Fig. S3.6A) were extracted for purification. Mixed sequence populations were detected in the two memory registers by Sanger sequencing, indicating successful writing in both registers (Fig. S3.6B). Illumina barcodes and adapters were added to the purified sample by a second round of PCR followed by enzymatic digestion as described above to remove residual non-edited registers. Samples were gel-purified, multiplexed, and sequenced by Illumina MiSeq (300 bps, single-end). Any reads that contained non-edited registers, that lacked any of the two expected motifs flanking the two memory registers (ATGCCTMMMMMMMTCGATT and AGTGCGNNNNNNNGTGCGC, where "MMMMMM" and "NNNNNN" correspond to positions of the memory Registers 1 and 2, respectively) were discarded, or that contained ambiguous nucleotides within this region were discarded. The connectivity matrices were deduced by linking variants that were observed simultaneously in a single read in the two registers and presented as heatmaps. To capture as many interactions as possible, we used an inclusive approach and did not filter out infrequent reads, which could potentially result in false positives due to the relatively high error rate of MiSeq. As an additional control, and in order to estimate the false positive discovery rates, we calculated a connectivity matrix for two randomly chosen (non-targeted) 6-bp regions within the *galK* amplicon. Only a limited number of connections were detected (Fig. S3.6C). Further inspection of the barcodes revealed mostly single base pair differences relative to the non-edited register, suggesting that these variants were likely produced due to library amplification and sequencing errors. False positives could be minimized by using sequencing platforms with higher accuracy or other techniques that facilitate error reduction (*74, 75*).

δHiSCRIBE Library Construction

Randomized δHiSCRIBE libraries (for experiments shown in Figs. 3.4C) were constructed by a modified Quik-Change (Agilent) protocol. Briefly, δHiSCRIBE plasmids (with or without the RP4 origin of transfer) were PCR amplified using primers containing the randomized regions within the desired target site in the

overhangs. The primers also contained compatible sites for the type IIS enzyme Esp3I. PCR products were used in a Golden Gate assembly (*90*) using this cut site to circularize the vector amplicon. Circularized vector libraries were amplified by transformation into Electro-ten Blue electrocompetent cells (Agilent). Amplified libraries were then transformed into donor and recipient strains and used in the mating pair connectome mapping experiment as described above.

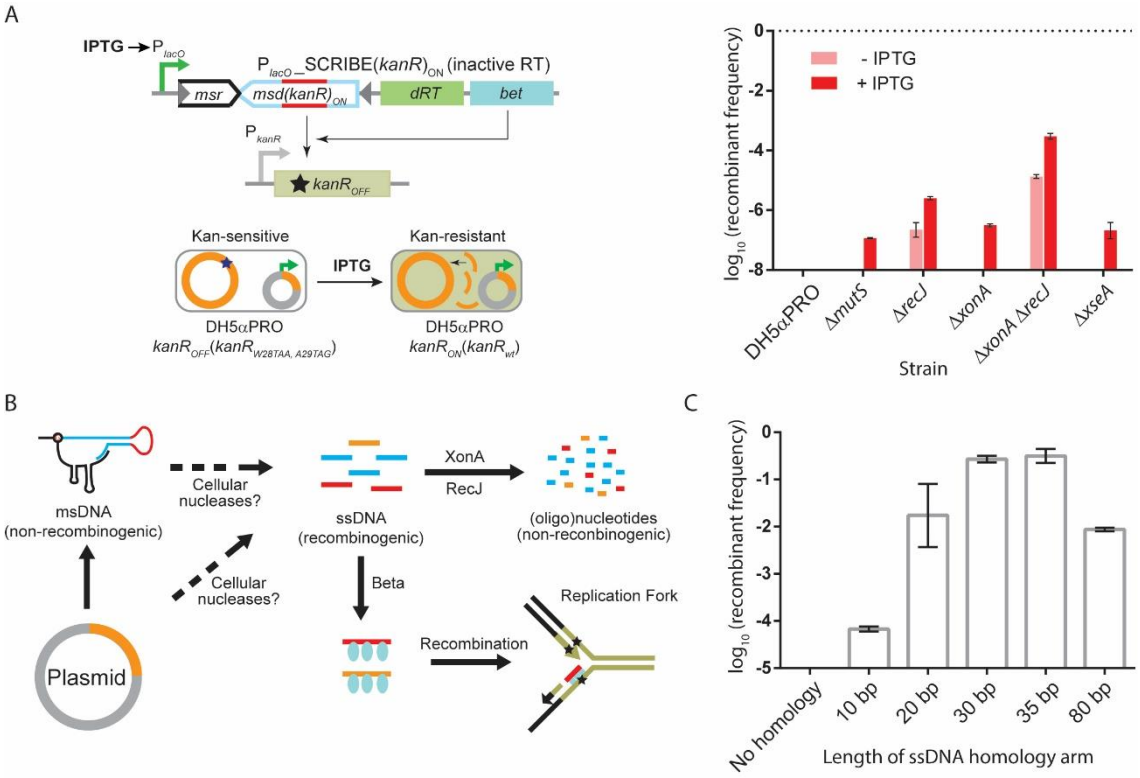


Figure S3.1 | A model for HiSCRIBE-mediated recombineering. (A) Genome editing efficiencies of SCRIBE harboring a catalytically inactive reverse transcriptase (*dRT*, in which the conserved YADD motif in the active site of the RT is replaced with YAAA (*67*)) was determined by the *kanR* reversion assay in different knockout backgrounds. Error bars indicate standard error of the mean for three biological replicates. **(B)** Proposed model for retron-mediated recombineering. Intracellular recombinogenic oligonucleotides are likely generated due to degradation of template plasmid as well as msDNA (retron product). ssDNA-specific cellular exonucleases (XonA and RecJ) can process these oligonucleotides into smaller, non-recombinogenic (oligo)nucleotides. Alternatively, Beta can bind to, protect, and recombine these oligonucleotides into their genomic target loci. **(C)** Effect of ssDNA homology length on HiSCRIBE DNA writing efficiency. Different δHiSCRIBE(*kanR*)$_{ON}$ plasmids expressing ssDNAs with different lengths of homology to the *kanR*$_{OFF}$ target were tested by the *kanR* reversion assay in DH5αPRO Δ*recJ* Δ*xonA* *kanR*$_{OFF}$ reporter strain. Maximal editing efficiency was observed with ssDNAs encoding 35 bp homology arms. Error bars indicate standard errors for three biological replicates.

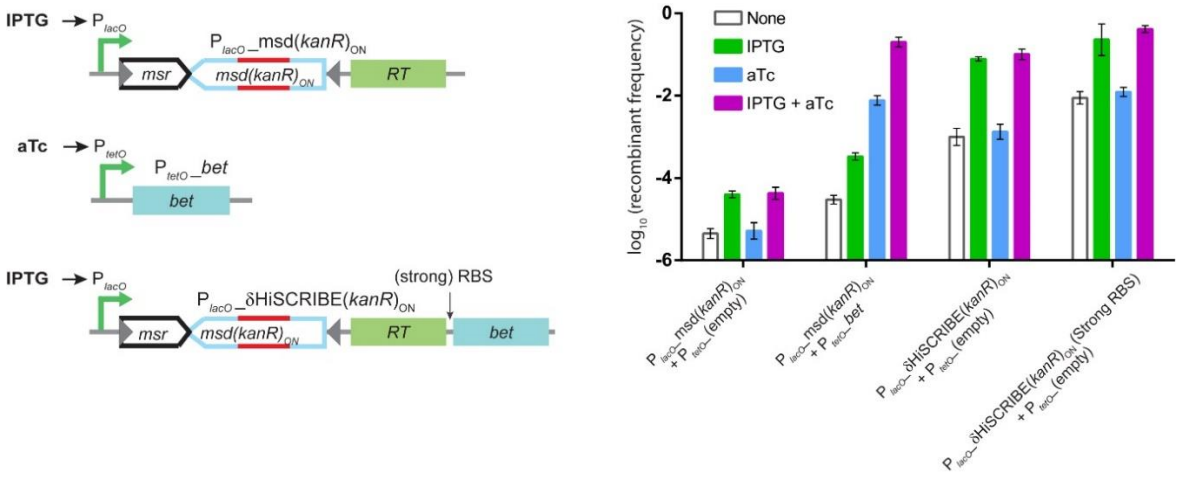


Figure S3.2 | Optimizing δHiSCRIBE efficiency by tuning the expression level of Beta. DH5αPRO Δ*recJ* Δ*xonA* $kanR_{OFF}$ reporter cells were transformed with the constructs shown above and the recombinant frequency was measured using the *kanR* reversion assay. Using δHiSCRIBE with a stronger RBS upstream of Beta (P_{lacO}·δHiSCRIBE(*kanR*)$_{ON}$ (Strong RBS) + P_{tetO}·(empty) + IPTG) resulted in the highest recombinant frequency (~36% after 24 hours induction in LB). Error bars indicate standard errors for three biological replicates.

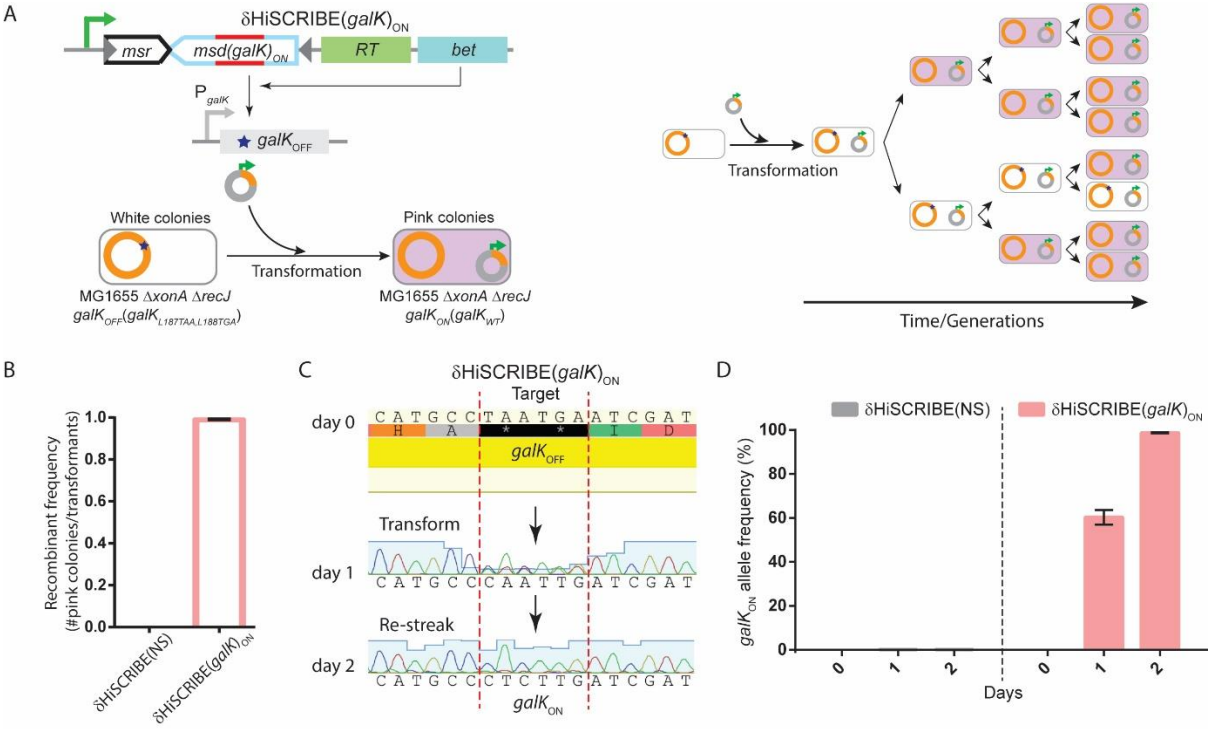


Figure S3.3 | Assessing population-wide δHiSCRIBE writing efficiency using plating assay and sequencing. **(A)** The genetic circuit used to assess writing efficiency (left panel) as well as schematic representation of the enrichment of mutant alleles within a single transformant colony (right panel). **(B)** MG1655 *exo galK*$_{\mathrm{OFF}}$ reporter cells were transformed with the δHiSCRIBE(*galK*)$_{\mathrm{ON}}$ plasmid and population-wide recombinant frequency was measured by the *galK* reversion assay. The frequencies of *galK*$_{\mathrm{ON}}$ and *galK*$_{\mathrm{OFF}}$ alleles in individual transformants colonies obtained on LB plates were assessed one and two days after transformation using **(C)** Sanger sequencing as well as **(D)** high-throughput Illumina sequencing.

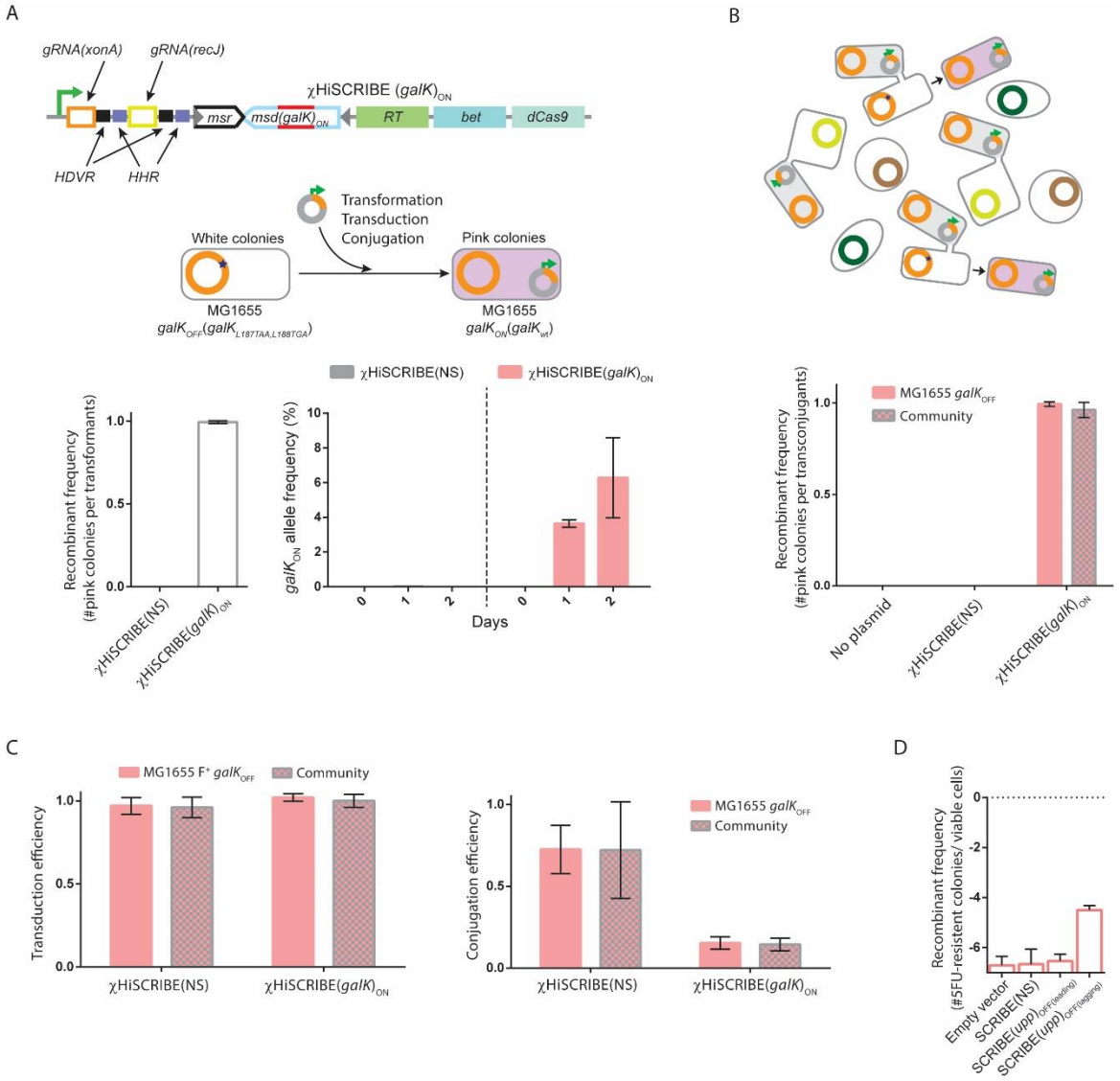


Figure S3.4 | Efficient editing of bacterial genomes in clonal populations as well as within bacterial communities. (A) χHiSCRIBE(*galK*)$_{ON}$ was cloned into a ColE1 plasmid encoding both the M13 origin and RP4 origin of transfer and delivered into the MG1655 *galK*$_{OFF}$ reporter strain via chemical transformation, transduction, and conjugation. Recombinant frequencies in cells that received χHiSCRIBE(*galK*)$_{ON}$ or χHiSCRIBE(NS) by chemical transformation were assessed using the *galK* reversion assay. Allele frequencies of individual transformant colonies obtained on LB with appropriate selection were measured by Illumina sequencing 24 hours after transformation, as well as after 24 hours additional growth. **(B)** Using a conjugative χHiSCRIBE plasmid (harboring RP4 origin of transfer) to edit the MG1655 galK$_{OFF}$ Str^R reporter strain in clonal population as well as within a synthetic bacterial community. **(C)** The efficiency of delivery of χHiSCRIBE plasmid by transduction and conjugation (for the experiments shown in Fig. 3.3C and S3.4B, respectively). To assess transduction efficiency of χHiSCRIBE phagemids, transduction mixtures were serially diluted and plated on LB + Str and LB + Str + Carb plates, to measure the number of viable target cells and transductants, respectively. The ratio between the transductants and viable target cells was reported as transduction efficiency. To measure the conjugation efficiency of delivering the χHiSCRIBE plasmids, conjugation mixtures were serially diluted and plated on LB + Str and LB + Str +

Carb plates, to measure the number of viable target cells and transconjugants, respectively. The ratio between the transconjugants and recipient cells was reported as conjugation efficiency. (D) Genome editing in *Pseudomonas putida* using SCRIBE. In this assay, two stop codons were introduced into the *upp* ORF of *P. putida* by SCRIBE, rendering these cells insensitive to the toxic compound 5-FU. The editing efficiency was reported as the ratio of 5-FU resistant to viable cells. Consistent with oligo-mediated recombineering results, higher recombination efficiency was achieved when targeting the lagging strand. Error bars indicate standard errors for three biological replicates.

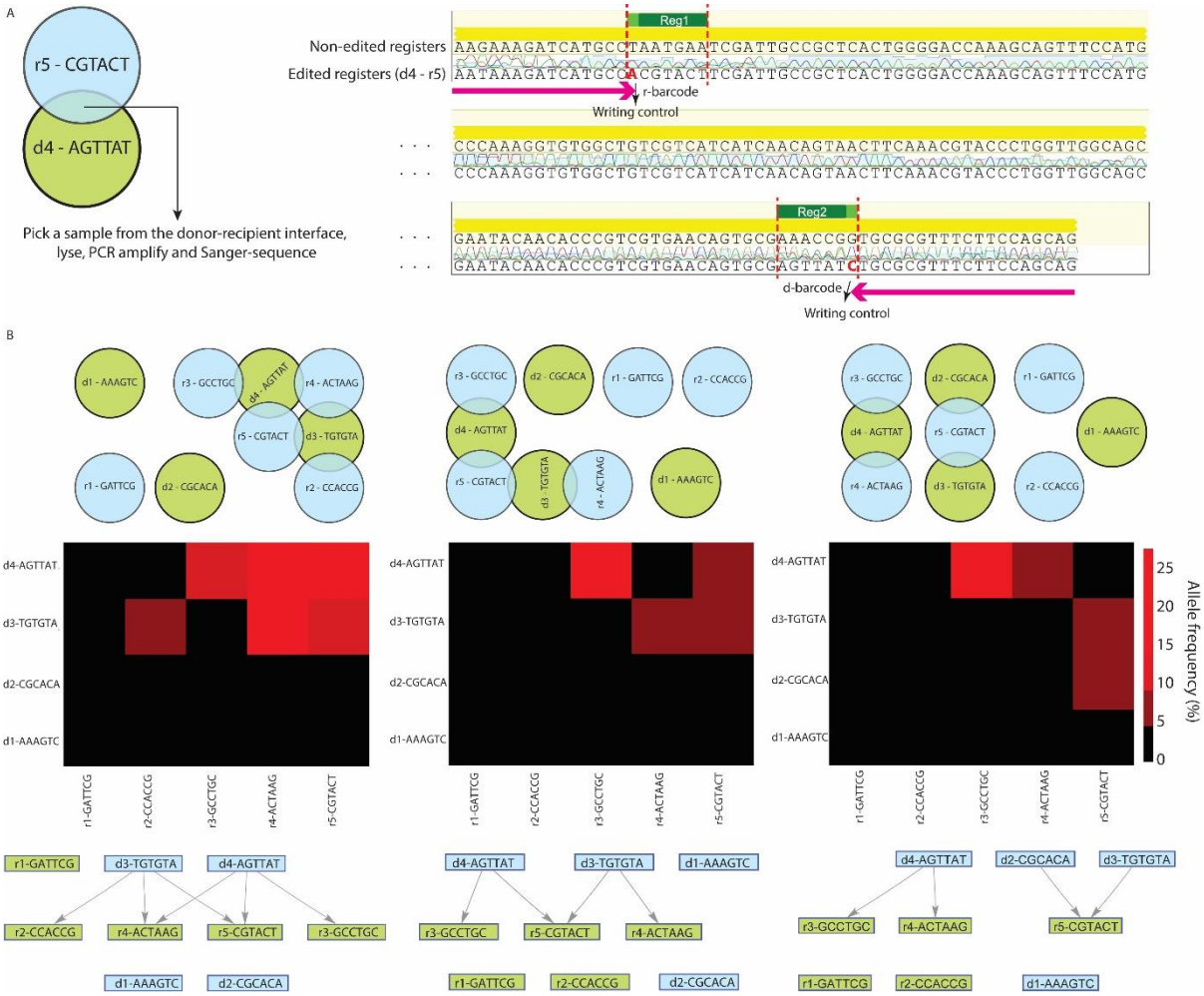
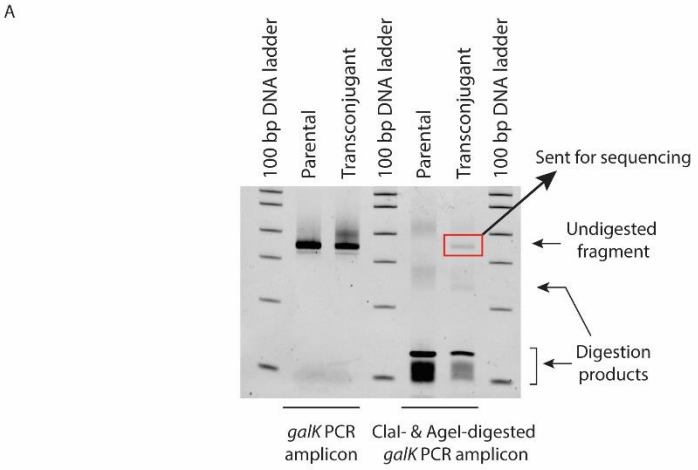
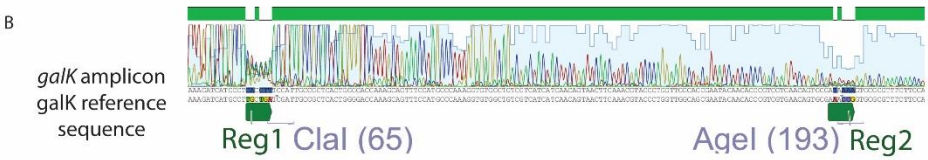


Figure S3.5 | Recording spatial patterns of bacterial cell populations into DNA. (A) Conjugation donor and recipient populations harboring δHiSCRIBE-encoded "*d-barcode*" and "*r-barcode*" were spotted on nitrocellulose filters placed on agar surface as indicated by the green and blue circles, respectively. These plasmids were designed to introduce unique 6 bp barcodes, as well as additional mismatches (which serve as "writing control nucleotides" to discriminate between edited and unedited memory registers when selectively PCR amplifying the edited registers) into two adjacent memory register on the *galK* locus, once inside the recipient cells. Samples taken from the intersection of the donor and recipient populations were lysed and used as templates in allele-specific PCR. Allele-specific PCR using primers that bind to the "writing control nucleotides" (but not to the non-edited registers) was used to selectively amplify the edited registers and deplete non-edited registers. The identities of the two barcodes corresponding to the interacting donor and recipient populations were then retrieved by Sanger sequencing. **(B)** Additional examples of cellular patterns that were recorded by the barcode joining approach described in Fig. 3.4A and 3.4B, and their corresponding weighted connectivity matrices and interaction networks that were faithfully retrieved using high-throughput sequencing.





C

	Targeted registers (Register 1 and Register 2)			Non-targeted regions		
	Rep#1	Rep#2	Rep#3	Rep#1	Rep#2	Rep#3
# of unique interactions (i.e., joined barcodes)	132853	171819	133932	159	185	173
# of total reads	3826347	5480449	4892306	3826347	5480449	4892306
# of unique interactions (per million reads)	34721	31351	27376	42	34	35

Figure S3.6 | Strategy used to deplete unedited memory registers from dual-register amplicons and the frequency of cell-cell interactions recovered by high-throughput sequencing in the connectome mapping experiment. **(A)** Using restriction digestion as an alternative strategy to remove unedited registers from the PCR amplified amplicons instead of allele-specific PCR. Genomic DNA samples were purified from the parental recipient cells (MG1655 Δ*recJ* Δ*xonA* $galK_{OFF}$), as well as cultures obtained after conjugation (transconjugants) in the experiment described in Fig. 4C. The *galK* locus was PCR amplified from the purified genomic DNA samples and run on a 6% TBE gel before and after digestion with ClaI and AgeI enzymes (which cut unedited Register 1 and Register 2, respectively) and stained by SYBR gold. The *galK* amplicon obtained from the parental sample was completely digested after enzymatic digestion. In contrast, the *galK* amplicon obtained from the transconjugant sample was not completely digested by ClaI and AgeI. The undigested band, corresponding to edited registers, comprised ~3.9% of the signal in this lane (measured by densitometry). **(B)** This band was subsequently excised, purified and Sanger-sequenced. Drops in the quality of sequencing in Register 1 and 2 indicate the presence of mixed DNA populations containing variations in these two regions in these samples. Subsequently, Illumina adaptors and barcodes were added to this undigested amplicon using an additional round of PCR and the obtained amplicon was sequenced by Illumina MiSeq (see Methods). **(C)** Number of unique variants (interactions) per million reads obtained from sequencing the two target registers in the genomes of recipient cells after conjugation with donor cells, as well as two randomly selected non-targeted regions within the *galK* amplicon (used as a negative control and to assess the rate of false-positives), for the experiment shown in Fig. 3.4C.

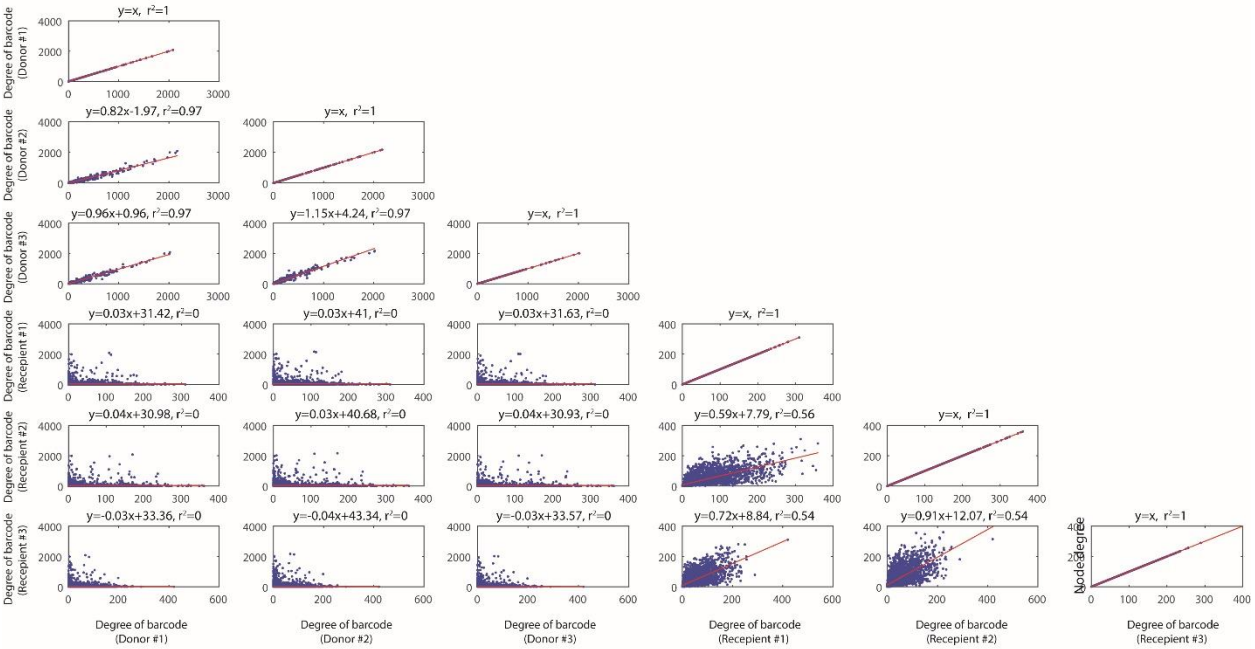


Figure S3.7 | Correlation between degree of nodes for donor and recipients for three parallel conjugation mixtures. Correlations between degrees of donor barcodes and degrees of recipient barcodes for three parallel conjugation experiments. The degree of donor barcodes is defined as the number of unique interactions that each donor barcode makes with recipient barcodes, which is equal to the sum of elements of the column corresponding to that barcode in the presented connectivity matrix. The degree of recipient barcodes is defined as the number of unique interaction that each recipient barcode makes with donor barcodes, which is equal to the sum of elements of the row corresponding to that barcode in the presented connectivity matrix. The strong correlation between the degree of donor barcodes in the parallel conjugation experiments suggests that transfer of barcodes from donors is not dependent on the identity of their conjugation partners (i.e., recipients). On the hand, the relatively weak correlation between the degree of recipient barcodes suggests that other factors, such as the identity of recipient cells partners (i.e., donor cells), could affect the frequency of successful conjugation.

Table S3.1 | Side-by-side comparison of different features of currently available DNA writing systems.

	HiSCRIBE (this work)	Oligo-mediated recombineering (e.g. MAGE) (*22, 62*)	Cas9-nuclease assisted genome editing (*63*)	Base editing (*95*)	Cas1-Cas2 spacer acquisition (*77*)	Site-specific recombinases (*76, 96*)
Requires presence of *cis*-elements on the target	No	No	Yes (PAM)	Yes (PAM and dC residues within editing window)	Yes (CRISPR array leader sequence and repeats)	Yes
Requires introduction of dsDNA breaks	No	No	Yes	No	No	No
~100% DNA writing efficiency	Yes	No	Yes	No	No	Yes
Editing can be linked to biological events	Yes	No (requires exogenous delivery of ssDNA donors)	Yes (but requires delivery of DNA donors)	Yes	Yes (but requires exogenous delivery of ssDNA donors)	Yes
Memory recording	Single-cell	NA	NA	NA	Population-level	Single-cell but binary states
Types of small modifications (insertions, deletions, or base-substitution mutations)	Any	Any	Any	dC to dT or dG to dA	Small fixed-size insertions	Flipping or excising DNA located between recombinase sites

Table S3.2 | List of the reporter strains used in this study.

Name	Strain Code	Genotype	Used in
kanR$_{OFF}$ reporter strain	FFF144	DH5αPRO *galK::kanR*$_{W28TAA, A29TAG}$	Fig. 3.2 Fig. S3.1
kanR$_{OFF}$ Δ*mutS*	FFF524	DH5αPRO Δ*mutS galK::kanR*$_{W28TAA, A29TAG}$	Fig. 3.2A Fig. S3.1
kanR$_{OFF}$ Δ*recJ*	FFF525	DH5αPRO Δ*recJ galK::kanR*$_{W28TAA, A29TAG}$	Fig. 3.2A Fig. S3.1
kanR$_{OFF}$ Δ*xonA*	FFF527	DH5αPRO Δ*xonA galK::kanR*$_{W28TAA, A29TAG}$	Fig. 3.2A Fig. S3.1
kanR$_{OFF}$ Δ*xseA*	FFF590	DH5αPRO Δ*xseA galK::kanR*$_{W28TAA, A29TAG}$	Fig. 3.2A Fig. S3.1
kanR$_{OFF}$ Δ*recJ* Δ*xonA*	FFF589	DH5αPRO Δ*recJ* Δ*xonA galK::kanR*$_{W28TAA, A29TAG}$	Fig. 3.2A Fig. S3.1 Fig. S3.2
MG1655 *exσ* reporter strain	FFF964	MG1655 Δ*recJ* Δ*xonA*	Fig. 3.3A
MG1655 *galK*$_{OFF}$ reporter strain	FFF1086	MG1655 *galK*$_{L187TAA, L188TGA}$ (For transduction experiments, the F-plasmid (from DH5α F$^+$(NEB)) was introduced to this strain via conjugation)	Fig. 3.3C Fig. S3.4
MG1655 *exσ* *galK*$_{OFF}$ reporter strain	FFF1087	MG1655 Δ*recJ* Δ*xonA galK*$_{L187TAA, L188TGA}$ (For transduction experiments, the F-plasmid (from DH5α F$^+$ (NEB)) was introduced to this strain via conjugation). PRO plasmid (pZS4Int-*lacI*/*tetR*, Expressys) was transformed to this strain to make a PRO version.	Fig. 3.3B Fig. 3.4 Fig. S3.3 Figs. S3.5-S3.7
MG1655 *galK*$_{OFF}$ Str^R reporter strain	FFF1296	MG1655 Str^R *galK*$_{L187TAA, L188TGA}$ (For transduction experiments, the F-plasmid (from DH5α F$^+$(NEB)) was introduced to this strain via conjugation)	Fig. 3.3C Fig. S3.4
MFDpir	FFF1040	MG1655 RP4-2-Tc::[Δ*Mu1*::aac(3)IV-Δ*aphA*-Δ*nic35*-Δ*Mu2*::*zeo*] Δ*dapA*::(*erm-pir*) Δ*recA*. PRO plasmid (pZS4Int-lacI/tetR, Expressys) was transformed to this strain to make a PRO version.	(*94*)
Pseudomonas putida KT2440	FFF480		Fig. S3.4D

Table S3.3 | List of the plasmids used in this study.

Name	Plasmid Code	Maker	Used in	Ref
PRO plasmid (pZS4Int-*lacI*/*tetR*)	pFF187	Spe/Str	Fig. 3.3B Fig. 3.4 Fig. S3.5-7	Expressys (*30*)
pKD46	pFF59	Carb	Fig. 3.3C	(*60*)
P_{lacO}_msd(*kanR*)$_{ON}$	pFF530	Cam	Fig. S3.2	(*67*)
P_{tetO}_*bet*	pFF145	Carb	Fig. S3.2	(*67*)
P_{lacO}_SCRIBE(*kanR*)$_{ON}$	pFF745	Cam	Fig. 3.2A	(*67*)
P_{lacO}_SCRIBE(*kanR*)$_{ON}$_dRT	pFF755	Cam	Fig. S3.1	(*67*)
P_{lacO}_δHiSCRIBE(*kanR*)$_{ON}$ (Strong RBS)	pFF804	Cam	Fig. S3.2	This work
P_{lacO}_δHiSCRIBE(*kanR*)$_{ON}$ (Strong RBS)	pFF944	Carb	Fig. S3.1C	This work
P_{tetO}_CRISPRi(no gRNA) [or P_{tetO}_dCas9]	pFF1156	Cam	Fig. 3.2B Fig. 3.3B	(*69*) Addgene #44249
P_{tetO}_CRISPRi(*recJ*_gRNA & *xonA*_gRNA)	pFF1165	Cam	Fig. 3.2B	This work
δHiSCRIBE(*galK*)$_{SYN}$ (Strong RBS)	pFF1493	Carb	Fig. 3.3A	This work
δHiSCRIBE(*galK*)$_{ON}$ (Strong RBS)	pFF1081	Carb	Fig. S3.3	This work
δHiSCRIBE(*galK*)$_{ON}$-P_{tetO}_gRNA(*galK*$_{OFF}$)	pFF1220	Carb	Fig. 3.3B	This work
P_{tetO}_Cas9	pFF1172	Cam	Fig. 3.3B	This work
χHiSCRIBE(*galK*)$_{ON}$	pFF1298	Carb	Fig. 3.3C Fig. S3.4	This work
SCRIBE(*upp*)$_{OFF(leading)}$	pFF1113	Kan	Fig. S3.4D	This work
SCRIBE(*upp*)$_{OFF(lagging)}$	pFF1114	Kan	Fig. S3.4D	This work

Table S3.4 | List of the synthetic parts and their corresponding sequences used in this study.

Part name	Type	Sequence	Ref
P_{lacO} ($P_{LlacO\text{-}1}$)	Promoter	AATTGTGAGCGGATAACAATTGACATTGTGAGCG GATAACAAGATACTGAGCACATCAGCAGGACGCA CTGACC	(*30*)
P_{tetO} ($P_{LtetO\text{-}1}$)	Promoter	TCCCTATCAGTGATAGAGATTGACATCCCTATCA GTGATAGAGATACTGAGCACATCAGCAGGACGCA CTGACC	(*30*)
msr	Primer for the RT	ATGCGCACCCTTAGCGAGAGGTTTATCATTAAGG TCAACCTCTGGATGTTGTTTCGGCATCCTGCATT GAATCTGAGTTACT	(*67*)
msd(*kanR*)$_{ON}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCCAACATG GATGCTGATTTATATGGGTATAAATGGGCCCGCG ATAATGTCGGGCAATCAGGTGCGACAATCTATCG GAATTCAGGAAAACAGACAGTAACTCAGA	(*67*)
msd(*galK*)$_{ON}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCCAGCTAA TTTCCGCGCTCGGCAAGAAAGATCATGCCCTCTT GATCGATTGCCGCTCACTGGGGACCAAAGCAGTT TCCGAATTCAGGAAAACAGACAGTAACTCAGA	(*67*)
Ec86 RT	Reverse Transcriptase	As described in (*67*)	(*67*)
bet	ssDNA-specific recombinase protein	As described in (*67*)	(*67*)
kanR$_{OFF}$	Reporter gene	As described in (*67*)	(*67*)
galK$_{OFF}$	Reporter gene The two premature stop codons in this ORF are underlined. The location of Reg1 and Reg2 in this ORF are highlighted. ClaI and AgeI sites are shown in bold.	ATGAGTCTGAAAGAAAAAACACAATCTCTGTTTG CCAACGCATTTGGCTACCCTGCCACTCACACCAT TCAGGCGCCTGGCCGCGTGAATTTGATTGGTGAA CACACCGACTACAACGACGGTTTCGTTCTGCCCT GCGCGATTGATTATCAAACCGTGATCAGTTGTGC ACCACGCGATGACCGTAAAGTTCGCGTGATGGCA GCCGATTATGAAAATCAGCTCGACGAGTTTTCCC TCGATGCGCCCATTGTCGCACATGAAAACTATCA ATGGGCTAACTACGTTCGTGGCGTGGTGAAACAT CTGCAACTGCGTAACAACAGCTTCGGCGGCGTGG ACATGGTGATCAGCGGCAATGTGCCGCAGGGTGC CGGGTTAAGTTCTTCCGCTTCACTGGAAGTCGCG GTCGGAACCGTATTGCAGCAGCTTTATCATCTGC CGCTGGACGGCGCACAAATCGCGCTTAACGGTCA GGAAGCAGAAAACCAGTTTGTAGGCTGTAACTGC GGGATCATGGATCAGCTAATTTCCGCGCTCGGCA AGAAAGATCATGCCTAATGAATCGATTGCCGCTC ACTGGGGACCAAAGCAGTTTCCATGCCCAAAGGT GTGGCTGTCGTCATCATCAACAGTAACTTCAAAC GTACCCTGGTTGGCAGCGAATACAACACCCGTCG TGAACAGTGCGAAACCGGTGCGCGTTTCTTCCAG CAGCCAGCCCTGCGTGATGTCACCATTGAAGAGT TCAACGCTGTTGCGCATGAACTGGACCCGATCGT GGCAAAACGCGTGCGTCATATACTGACTGAAAAC GCCCGCACCGTTGAAGCTGCCAGCGCGCTGGAGC AAGGCGACCTGAAACGTATGGGCGAGTTGATGGC GGAGTCTCATGCCTCTATGCGCGATGATTTCGAA ATCACCGTGCCGCAAATTGACACTCTGGTAGAAA TCGTCAAAGCTGTGATTGGCGACAAAGGTGGCGT ACGCATGACCGGCGGCGGATTTGGCGGCTGTATC GTCGCGCTGATCCCGGAAGAGCTGGTGCCTGCCG	(*67*)

		TACAGCAAGCTGTCGCTGAACAATATGAAGCAAA AACAGGTATTAAAGAGACTTTTTACGTTTGTAAA CCATCACAAGGAGCAGGACAGTGCTGA	
*bet*_RBS	Natural RBS of *bet*	GGTTGATATTGATTCAGAGGTATAAAACGA	(*67*)
RBS_A	Strong RBS	AGGAGGTTTGGA	(*97*)
msd(*kanR*)$_{ON}$ (10 bp homology arm)	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCGGGTATA AATGGGCCCGCGATAATGGAATTCAGGAAAACAG ACAGTAACTCAGA	This work
msd(*kanR*)$_{ON}$ (20 bp homology arm)	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCTGATTTA TATGGGTATAAATGGGCCCGCGATAATGTCGGGC AATCGAATTCAGGAAAACAGACAGTAACTCAGA	This work
msd(*kanR*)$_{ON}$ (30 bp homology arm)	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCACATGGA TGCTGATTTATATGGGTATAAATGGGCCCGCGAT AATGTCGGGCAATCAGGTGCGACAGAATTCAGGA AAACAGACAGTAACTCAGA	This work
msd(*kanR*)$_{ON}$ (35 bp homology arm)	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCCAACATG GATGCTGATTTATATGGGTATAAATGGGCCCGCG ATAATGTCGGGCAATCAGGTGCGACAATCTATCG GAATTCAGGAAAACAGACAGTAACTCAGA	(*67*)
msd(*kanR*)$_{ON}$ (80 bp homology arm)	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCGAGCCAT ATTCAACGGGAAACGTCTTGCTCGAGGCCGCGAT TAAATTCCAACATGGATGCTGATTTATATGGGTA TAAATGGGCCCGCGATAATGTCGGGCAATCAGGT GCGACAATCTATCGATTGTATGGGAAGCCCGATG CGCCAGAGTTGTTTCTGAAACAGAATTCAGGAAA ACAGACAGTAACTCAGA	This work
msd(*upp*)$_{OFF(leading)}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCGGTGATC TTCTTGCCGGCGATTTTTTCAACCGAGACTCACT AACACCAGCCGTCGATCTCGTAGGTTTCGAGGGG CAGGAATTCAGGAAAACAGACAGTAACTCAGA	This work
msd(*upp*)$_{OFF(lagging)}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCCTGCCCC TCGAAACCTACGAGATCGACGGCTGGTGTTAGTG AGTCTCGGTTGAAAAAATCGCCGGCAAGAAGATC ACCGAATTCAGGAAAACAGACAGTAACTCAGA	This work
msd(Reg1) (highlighted region indicates positions in the msd corresponding to the randomized Register 1	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCGCTAATT TCCGCGCTCGGCAAGAAAGATCATGCCTNNNNNN TCGATTGCCGCTCACTGGGGACCAAAGCAGTTTC CATGCGAATTCAGGAAAACAGACAGTAACTCAGA	This work

msd(Reg2) (highlighted region indicates positions in the msd corresponding to the randomized Register 2	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCGTTGGCAGCGAATACAACACCCGTCGTGAACAGTGCGNNNNNNGTGCGCGTTTCTTCCAGCAGCCAGCCCTGCGTGATGTGAATTCAGGAAAACAGACAGTAACTCAGA	This work
galK$_{OFF}$_gRNA	gRNA protospacer	TGAGCGGCAATCGATTCATT	This work
*recJ*_gRNA	gRNA protospacer	TCACGCGAATTATTTACCGC	This work
*recJ*_gRNA (14 bps)	gRNA protospacer (used in the χHiSCRIBE cassette)	GGAGGCAATTCAGC	This work
*xonA*_gRNA	gRNA protospacer	GCTTACCGTCATTCATCATT	This work
*xonA*_gRNA (14 bps)	gRNA protospacer (used in the χHiSCRIBE cassette)	GGCGATCTAACGCG	This work
galK(ON) synthetic oligo (FF_oligo_2304)	Used for recombineering. Asterisks show phosphorothioate bonds added to oligo to increase its intracellular stability.	C*A*GCTAATTTCCGCGCTCGGCAAGAAAGATCATGCCCTCTTGATCGATTGCCGCTCACTGGGGACCAAAGCAGTTT*C*C	

Table S3.5 | List of the sequencing primers used in this study.

Primer code	Name	Sequence
FF_oligo_1890	galK(+)	ACACGACGCTCTTCCGATCTNNNNNGTTTGTAGGCTG TAACTGCGGGATCATGG
FF_oligo_1891	galK(-)	CGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNT CACGCAGGGCTGGCTGCTG
FF_oligo_2444	galK_1n(+)	ACACGACGCTCTTCCGATCTNNNNNGCTCGGCAAGAA AGATCATGCCa
FF_oligo_2445	galK_1n(-)	CGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNC TGCTGGAAGAAACGCGCAg

Chapter 4: DOMINO

Single-Nucleotide-Resolution Computing and Memory in Living Cells

4.1 Abstract

Computing and memory in living cells are central to encoding next-generation therapies and studying *in situ* biology, but existing strategies have limited encoding capacity and are challenging to scale. To overcome this bottleneck, we describe a highly scalable, robust and compact platform for encoding logic and memory operations in living bacterial and human cells. This platform, named DOMINO for DNA-based Ordered Memory and Iteration Network Operator, converts DNA in living cells into an addressable, readable, and writable computation and storage medium via a single-nucleotide resolution read-write head that enables dynamic and highly efficient DNA manipulation. We demonstrate that the order and combination of DNA writing events can be programmed by biological cues to encode a wide range of order-independent, sequential, and temporal logic and memory operations. Furthermore, we show that these operators can be used to perform both digital and analog computation, and record signaling dynamics and cellular states in a long-term, autonomous, and minimally disruptive fashion. Finally, we show that the platform can be functionalized with gene regulatory modules and interfaced with cellular circuits to control and monitor cellular phenotypes. We envision that highly scalable, compact, and modular DOMINO operators will lay the foundation for building robust and sophisticated synthetic gene circuits for numerous biotechnological and biomedical applications.

4.2 Introduction

Robust and scalable computation and memory platforms in living cells are key to enabling a broad range of bioengineering and biomedical applications. Unlike their silicon-based counterparts that have access to large capacities of addressable memory registers, synthetic genetic circuits currently have very limited information storage capacities. Existing methods for encoding information into cellular memory, as well as strategies for integrating such memory with logic operations, are challenging to scale.

Genomic DNA is an ideal medium for biological memory since it is ubiquitously present, naturally replicated at high fidelity within cells, and compatible with natural

biological operations. In recent years, several strategies for encoding information into DNA and integrating these memories with cellular computers have been described (*76, 78, 98, 99*). However, these methods remain limited in their encoding capacity and scalability. For example, site-specific recombinases that flip or excise targeted DNA segments have been used to create digital memory, sequential logic, and biological state machines in living cells (*76, 98*). However, a different recombinase is required for every unique event that one wishes to record, thus limiting the number of potential states that can be encoded into DNA memory. Furthermore, distances between recombinase-recognition sites usually need to be several hundred base pairs to achieve efficient recombination, thus increasing circuit size (*100, 101*). Furthermore, recombinase sites must be pre-engineered into desired target sites, which is time- and labor-intensive, especially if they are to be used in the genomic context. To address these limitations, we previously developed the SCRIBE DNA writing system, which uses in vivo single-stranded DNA expression to generate precise mutations that accumulate into target genomic loci as a function of the magnitude and duration of exposure to an input (*99*). However, this approach has been limited to bacteria thus far due to the requirement for specific recombination mechanisms.

To overcome these bottlenecks, we describe a platform called DOMINO (for DNA-based Ordered Memory and Iteration Network Operator) that uses highly efficient and precise DNA writing with CRISPR base editors (*95, 102*) to manipulate DNA dynamically and efficiently with single-nucleotide resolution in living cells. DOMINO enables the use of DNA as a uniquely addressable, readable, and writable information storage and computation medium. We show that the order and combinations of these DNA writing events can be tuned by external inputs, allowing one to execute order-independent (e.g., IF EVER A AND IF EVER B), sequential (e.g., A AND THEN B), and temporal (e.g., A AND THEN B after time X) logic and memory operations. DOMINO operators enable highly compact and scalable logic and memory operators that, unlike previous strategies, can be used to realize both digital and analog computation in living cells. Various orthogonal DOMINO operators can be simply created by changing guide RNA (gRNA) sequences, thus making the system highly scalable. These operators can then be layered and interfaced with synthetic or natural regulatory circuits to build more sophisticated genetic programs. Finally, we demonstrate that DOMINO can be combined with established CRISPR-based gene regulation platforms, such as CRISPR interference (CRISPRi) (*69*) and CRISPR activator (CRISPRa) (*103, 104*), to achieve modular and versatile memory and gene regulation programs. Thus, by enabling cascades of DNA writing events in a

continuous and autonomous fashion, DOMINO addresses many limitations of current in vivo computing and memory technologies.

4.3 Results

Engineering an Efficient Read-Write Head for Genomic DNA

In order to efficiently manipulate genomic DNA in living cells, we sought to build a single-nucleotide resolution "read-write head" for this medium. To this end, we fused Cas9 nickase (nCas9, an addressable DNA "reader" module that is directed by gRNA to bind to specific DNA targets and nicks them) to cytidine deaminase (CDA, a DNA "writer" module that edits the DNA) and uracil DNA glycosylase inhibitor (*ugi*, a peptide which has been shown to improve the DNA writing efficiency by blocking cellular repair machinery) to create CDA-nCas9-ugi (*95*). Once localized to the target based on the 12 bp gRNA seed sequence ("READ" address), the writer module can deaminate dC positions in the vicinity of 5'-end of the target ("WRITE" address), thus resulting in DNA lesions that are preferentially repaired as dT (*95, 102*). Using cytidine deaminase as the DNA writer module enables dC to dT mutations (or dG to dA mutations if the reverse complement strand is targeted) to be introduced to the WRITE address, resulting in permanent records in DNA. In this memory scheme, an individual mutation or a group of mutations in a target site can be designated as a unique memory state for the corresponding memory register, and mutations introduced by DNA writing events can be considered as transitions between DNA memory states (Fig. 4.1A). DNA writing events can be controlled by internal or external inputs by placing both the gRNA expression and CDA-nCas9-ugi under regulation by inducible promoters.

We demonstrated that this approach enables highly efficient, robust and scalable DNA writing in *E. coli*. We first placed CDA-nCas9-ugi under the control of anhydrotetracycline (aTc)-inducible promoter. Using an Isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible gRNA as an input, we demonstrated efficient and inducible DNA writing (dC to dT mutations) at desired target sites in the presence of aTc and IPTG induction (Fig. 4.1A). In this design, which forms the basis of DOMINO operators, the signal controlling the expression of CDA-nCas9-ugi (aTc) that is required for the overall circuit to function can be considered as the "operational signal", while the signals controlling the expression of individual gRNAs can be considered as independently controllable "inputs".

Order-independent DOMINO Logic

DOMINO operators can be arrayed and interconnected in a highly scalable fashion to build robust and complex forms of computing and memory circuits that execute a series of order-independent and/or sequential unidirectional DNA writing events. The frequency and order of these DNA writing events can be controlled by internal and external cues, as well as by carefully selecting the position of mutable residues within the target. For example, by layering two DOMINO operators, we built a two-input order-independent AND logic gate, where the A AND B logic is executed independent of the order of addition of the inputs (Fig. 4.1B). In this design, two distinct gRNAs were placed under the control of IPTG- and Arabinose (Ara)-inducible promoters, respectively. In the presence of its corresponding inducer, each gRNA is expressed and directs the DNA read-write module (which itself is expressed in the presence of the operational signal, aTc) to its cognate target site, resulting in precise dC to dT mutations (or dG to dA mutations in cases where the gRNA targets the reverse-complement strand) within the WRITE address.

To assess the performance of the order-independent DOMINO AND gate, we induced cells harboring this circuit with different combinations of the inducers for multiple days and analyzed dynamics of allele frequencies at the target locus by high-throughput sequencing (HTS) over multiple time points. As shown in Fig. 4.1C, in the presence of the operational signal (aTc) and each of the two inputs (IPTG or Ara), mutations were accumulated in the target sites of the induced gRNA in a linear fashion within the population and comprised ˜100% of the population after 72 hours of induction. This corresponds to transitions from the unmodified state (state S0) to either of the two singly modified states (state S1 or S2). The time required for transitioning between the two states can be considered as the "propagation delay" of the corresponding DOMINO operator. On the other hand, when cells were induced with both inputs (IPTG AND Ara), the target sites for both gRNAs were edited, resulting in the accumulation of doubly edited sites (state S3) in the target locus. We defined states S0, S1, and S2 as the OFF states and S3 as the ON state, which means that this system implements AND logic. In this experiment, low levels of a singly mutated allele (state S2) accumulated in the absence of any induction, likely due to leakiness of the Ara-inducible promoter (pBAD) in these cells and/or high binding efficiency of its corresponding gRNA. The performance of the circuit should be improved by lowering the Ara-inducible promoter's basal activity, for example, by overexpressing pBAD repressor (*araC*) or using tighter promoters, or alternatively, by lowering the copy

numbers of DOMINO operators. Nevertheless, the doubly-edited allele (state S3) only accumulated in the presence of both IPTG and Ara.

Notably, these results show that in DOMINO operators, the accumulation of the singly mutated alleles in the presence of the operational signal and individual inducer inputs follows a linear trend over the course of few days. About 3 days were required for the unmodified allele to be fully converted into the modified allele(s), thus indicating the propagation delays of the corresponding operators. This feature enables one to use DOMINO to implement both analog and digital computing, since continuous changes that occur within the propagation delay window can be used to implement analog computation, while fully converted states can be considered as transitions between digital states and thus used for digital computation.

The states designated in the AND gate logic described in this example are arbitrary defined; for example, the doubly mutated allele (state 3) was defined as the ON state. The same circuit can be defined, for example, as a NAND gate if the unmodified state (state S0) is designated as ON ("1") output and states S1 through S3 are designated as OFF ("0") outputs. Alternatively, each of the four different mutational states can be defined as distinct outputs, in which case the circuit can be considered as a 2-input/4-output decoder.

In this experiment, two mutable residues within the editing window of each gRNA were used, and the memory states were defined so that mutations in both of these residues were required to be considered as a state transition. One could define mutations in only one of the two nucleotides available for editing as intermediate states (that can be discarded), or if desired, as usable transient memory states. Furthermore, the number of memory states as well as the response dynamics (e.g., propagation delay) for each DOMINO operator can be tuned by using different numbers of mutable residues (dC or dG) within the WRITE window, or adjusting the position of these residues within this window.

While HTS offers a powerful way to quantify the outcome of DOMINO circuits, its relatively high cost inspired us to develop a strategy for using Sanger sequencing chromatograms to quantify position-specific mutant frequencies within a mixture of DNA species. This algorithm, named Sequalizer (for Sequence equalizer), normalizes Sanger chromatogram signals and calculates the difference between the normalized signals from a test sample and an unmodified reference to identify position-specific mutations. It then uses this calculated difference to estimate position-specific mutant

frequencies at any given target position. We validated the accuracy of this method by constructing a standard curve based on known ratios of mutant and wild-type (WT) sequences, and comparing the Sequalizer results with next-generation sequencing (see Supplementary Materials and Fig. S4.1). The Sequalizer output, which is based on population-averaged Sanger sequencing results, provides an estimate of position-specific mutant frequencies in an entire population. Though Sequalizer does not always provide accurate absolute values of mutant frequencies, fold changes in estimated mutant frequencies are accurate (see Supplementary Materials and Fig. S4.1C). Additionally, unlike HTS, Sequalizer output does not provide insights into the identities and frequencies of individual alleles in the population. Nevertheless, given the high specificity of the DNA writers and predefined target sites for DNA writing, this approach can be used as a low-cost alternative to HTS to assess the performance of DOMINO and other precise genome-editing platforms.

In addition to HTS, we analyzed the samples obtained from the experiment shown in Fig. 1B by Sanger sequencing and Sequalizer. As shown in Fig. 4.1D and Fig. S4.1C, the Sequalizer results were consistent with and could accurately estimate changes in position-specific mutant frequencies obtained by HTS. Specifically, in samples induced with either of the two inputs, the frequencies of mutants in positions corresponding to the cognate target sites of the induced gRNA increased in the population. In addition, in samples that were induced with both gRNAs, the mutation frequencies in the target sites of both gRNAs were increased (state S3).

In addition to AND gate, other logic can be readily implemented by carefully positioning mutable residues on the targets, as well as designing the combinations and order of DNA writing events. Furthermore, additional input gRNAs can be incorporated to achieve operators with more than two inputs, thus demonstrating the scalability of this approach (Fig. S4.2).

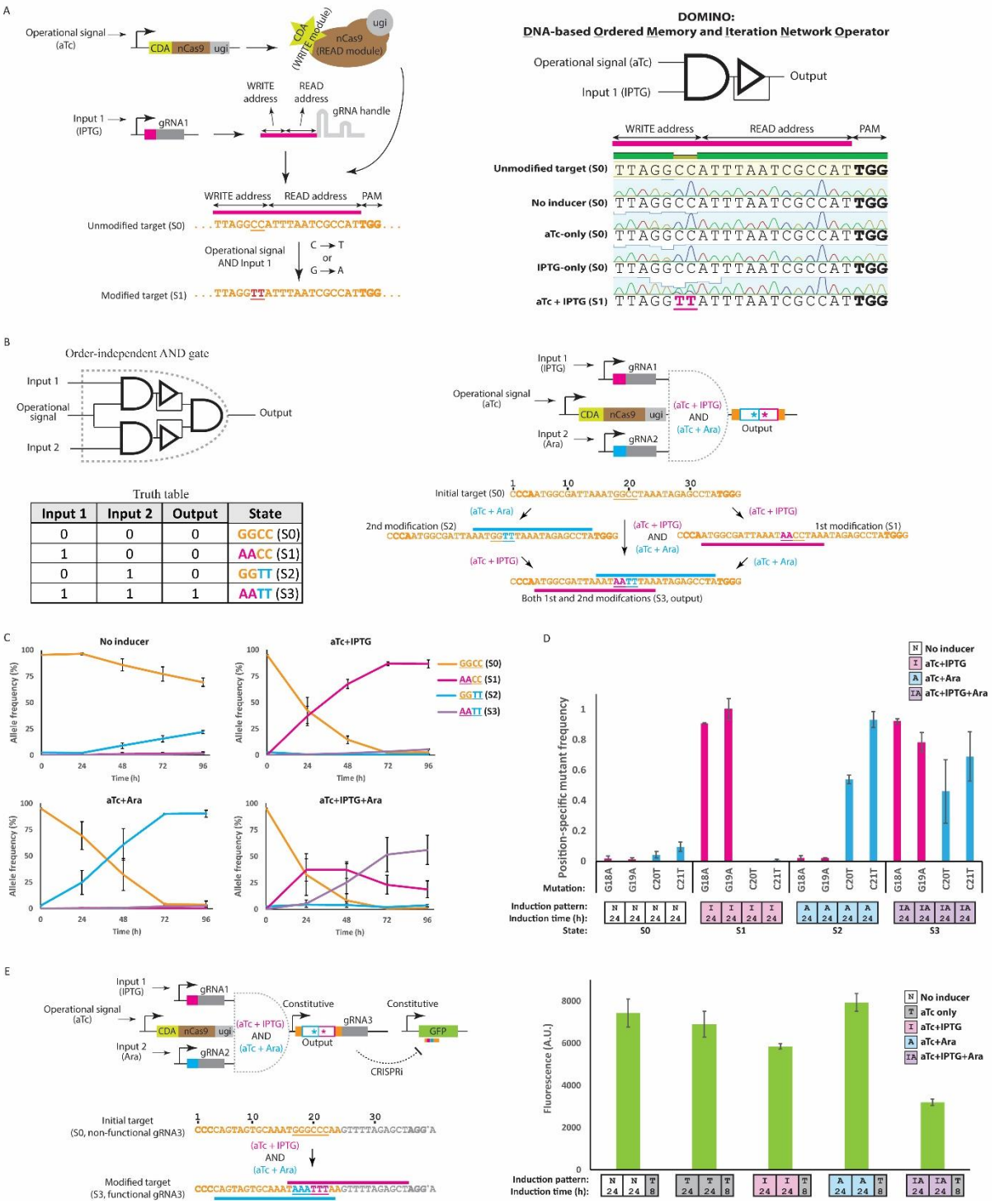


Figure 4.1 | Incorporating memory and logic in living cells by DOMINO. (A) Schematic representation of DOMINO operators. DOMINO operators are enabled by a DNA read-write head that performs efficient and precise manipulation of genomic DNA with single-nucleotide resolution. In this device, nCas9 (READ module), along with cytidine deaminase (CDA, WRITE module) and a uracil DNA glycosylase (*ugi*, WRITE enhancer) are addressed to a desired genomic loci using gRNA with a complementary seed region (READ address). Localization of the CDA write module to the target results in the deamination of cytidine (dC) residues in the vicinity of the 5' -end of the gRNA (WRITE address) and their conversion to dU residues, which are then preferentially repaired by the cellular machinery to dT (or dG to

dA mutation if the negative strand of DNA is targeted by gRNA). By placing the DNA read-write module and the gRNA under the control of inducible signals, DNA writing for DOMINO operators can be tuned and controlled by external cues. Here, we schematize the basic DOMINO operator as an AND gate since it requires the expression of both the DNA read-write head (i.e., CDA-nCas9-ugi controlled by the " operational signal") as well as the gRNA (regulated by " Input 1") with a downstream feedback delay operator (to illustrate the unidirectional and memory aspect of the operator). DOMINO operators can be layered to a wide variety of memory and logic functions. Bold nucleotides on the target show the location of NGG PAM sequence. Targeted nucleotides are underlined. **(B)** Order-independent AND gate enabled by DOMINO where the output is ON only when both inputs have been present with any possible order. Induction of the circuit with either of the two inducers (IPTG or Ara), results in the editing of the target and transition to an intermediate state (states S1 or S2, respectively). Induction of the circuit with both gRNAs results in the generation of the doubly edited DNA sequence (state S3), which is designated as ON state. **(C)** Dynamics of allele frequencies obtained by Illumina High-Throughput Sequencing (HTS) for the circuit shown in (B). *E. coli* cells were exposed to different inducer combinations for four days with serial dilution after every 24 hours. Error bars indicate the standard deviation of three biological replicates. **(D)** Position-specific mutant allele frequencies for the last time point (96 h) of the experiment shown in (C) estimated from Sanger sequencing analysis by Sequalizer (see Supplementary Materials). This data demonstrates the expected outcomes of AND gate behavior at the population level. The x-axis shows dC to dT or dG to dA mutations in the specified positions. For example, the G18A mutation means a dG to dA mutation in position 18 of the target sequence. Small boxes along the x-axis show the induction patterns and duration of induction used in each experiment. For example, the induction pattern of the last sample set ([IA][IA][IA][IA]) means that the samples were induced with aTc + IPTG + Ara for four days with dilutions every 24 hours. Error bars indicate the standard deviation of three biological replicates. **(E)** The output of DOMINO operators, which is in the form of DNA mutations, can be converted to a gRNA by flanking the target DNA sequence with a desired promoter and gRNA handle. This allows DOMINO operators to be linked to other DOMINO operators or host regulatory networks. To demonstrate this concept, we designed an order-independent DOMINO AND gate with a target sequence flanked by a constitutive promoter and a modified gRNA handle. The modified gRNA handle harbored a dA to dG mutation in a position that was not essential for gRNA function (*105*). This modification (shown by an asterisk) was required to generate an NGG PAM motif for binding of one of the input gRNAs. Upon induction by both inducers, the input gRNAs edit the Specificity-Determining Sequence (SDS) of the output gRNA. The doubly edited output gRNA can then bind to the GFP ORF and repress it via CRISPRi in *E. coli*. In this example, AND logic is realized on the target DNA register (i.e., the output gRNA) while NAND logic is achieved on the output GFP reporter. Error bars indicate standard deviation for three biological replicates.

The output of DOMINO operators takes the form of DNA mutations that accumulate at a target site. One can flank this target site with a desired promoter and a gRNA handle to convert the output of a given DOMINO operator into downstream gRNA expression. The output gRNA can then be interconnected with other DOMINO operators to build more complex circuits. In addition, it can be combined with CRISPR-based gene regulation platforms such as CRISPRi and CRISPRa to dynamically regulate cellular phenotypes. To demonstrate this, we engineered an AND operator by layering two DOMINO operators under the control of inducible promoters to edit a third gRNA as the output (Fig. 4.1E). The input gRNAs were controlled by

IPTG- and Ara-inducible promoters, respectively. In the presence of both inducers, the output gRNA was modified by both input gRNAs such that it could then bind to and repress a downstream reporter gene (GFP) (Fig. 4.1E, aTc + IPTG + Ara co-induction for two 8-hour periods followed by aTc-induction for 8 hours: [IA][IA][T] induction pattern). When targeting gRNA as an output, both the Specificity Determining Sequence (SDS) of the output gRNA as well as its constant region (handle) can be modified. Mutating the SDS is useful when the creation of a unique gRNA is the desired output. On the other hand, mutating the gRNA handle enables one to activate/deactivate an entire set of gRNAs. Furthermore, one can also target gene regulatory and functional elements, such as promoters, ribosome binding sites, start/stop codons, as well as active sites within proteins to tune the expression or activity of downstream components as shown in Fig. S4.3.

Sequential DOMINO Logic

In addition to realizing order-independent logic, one can carefully control the sequence and timing of DNA writing events executed by DOMINO operators to achieve sequential logic, where desired outputs are generated only when the correct order of inducers is added. To achieve this, for example, one can design the gRNA output of one operator to be used as the input for a downstream operator (Fig. S4.2C). This design can be used to functionally connect DOMINO operators that are not physically co-located, and offers control over the individual DOMINO operators. Alternatively, sequential logic can be achieved by overlapping mutable residues in the WRITE address of one operator with the READ address of a downstream operator (Fig. 4.2). This design uses DNA mutations rather than cascades of gRNAs as a way to interconnect *cis*-encoded DOMINO operators, thus offering a highly compact and scalable strategy for encoding sequential logic.

To demonstrate the latter strategy, we first constructed an asynchronous sequential AND gate, where sequential addition of the two inputs in the correct order (IPTG AND THEN Ara) leads to mutation of a cryptic start codon (ACG) into the canonical (and more efficient) start codon (ATG) in the GFP ORF, thus increasing the GFP signal (Figs. 4.2A and 4.2B). We observed slight increases in GFP signal in cells that had been induced with the first inducer (i.e., IPTG) or those that had been co-induced with both inducers (Fig. 4.2B). The former was likely caused by the leakiness of the second (Ara-inducible) promoter while the latter was likely due to the simultaneous presence of both inducers in the media, which could result in the execution of sequential DNA mutations in the correct order to some extent. Nevertheless, the GFP signal was significantly higher when cells were exposed to the correct order of the inducers. We

further confirmed these results by analyzing Sanger sequencing chromatograms by Sequalizer (Fig. 4.2C). Consistent with flow cytometry data, samples induced with the correct order of the inputs showed the highest level of the dC to dT mutation in the position corresponding to the cryptic start codon (Fig. 4.2C), indicating the execution of a cascade of DNA writing events that lead to the execution of sequential AND logic.

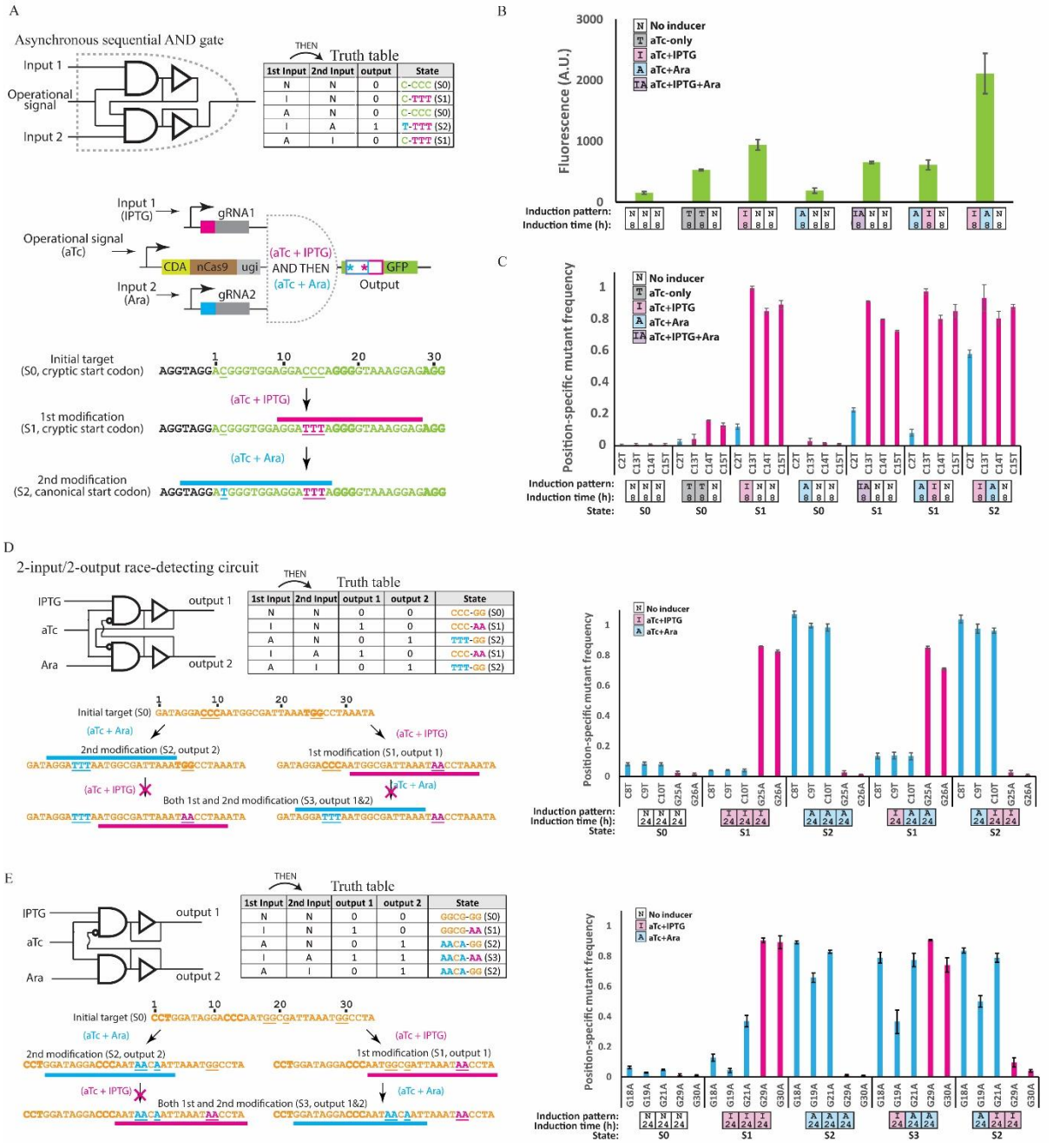


Figure 4.2 | Building sequential logic by DOMINO operators. (A) Sequential AND gate encoded with DOMINO operators. The output of a DOMINO operator was used as an input for another operator, which in turn mutates a non-canonical start codon (ACG) within the GFP ORF into a canonical (efficient) start codon (ATG), thus increasing GFP signal. The

second gRNA (induced by Ara) can bind to and enact the start-codon mutation only after the first gRNA (induced by IPTG) has edited its target. **(B)** GFP signal measured by flow cytometry for the circuit shown in (A). Only when IPTG AND THEN Ara were applied was the sequential logic satisfied, thus resulting in increased GFP signal. Error bars indicate the standard deviation of three biological replicates. **(C)** Position-specific mutation frequency obtained from Sequalizer analysis for the experiment shown in (A). Consistent with GFP data, the highest frequency of ACG to ATG conversion (blue bars) was achieved when the samples were induced with IPTG AND THEN Ara. Error bars indicate standard deviation for three biological replicates. **(D)** Two-input/two-output race-detecting circuit. Two gRNAs were designed so that editing by one gRNA destroys the PAM domain for the other gRNA, thus inhibiting its binding. Sequential expression of each gRNA resulted in an output corresponding to the output of the first gRNA, independent of whether the second gRNA was expressed or not. Error bars indicate standard deviation for three biological replicates. **(D)** Another example of sequential DOMINO logic, where sequential induction of cells with IPTG AND THEN Ara results in the sequential transition between two modified states (states S1 and S3, respectively). However, induction of cells with the reverse order (Ara AND THEN IPTG) only results in a one-step transition to state S2. Error bars indicate standard deviation for three biological replicates.

As another example, we built an asynchronous 2-input/2-output race-detecting circuit, where the output of the circuit is determined by the inducer added first and not the other inducer added second (Fig. 4.2D). In this design, the PAM domain for each gRNA is placed within the WRITE window of the other, in a way that editing mediated by one gRNA destroys the PAM domain for the other gRNA, thus preventing binding and subsequent editing by that gRNA. As shown in Fig. 4.2D, Sequalizer analysis of cells induced with different combinations of inducers showed that the output of the circuit depends on the identity of the first inducer. Specifically, cells that were first induced with IPTG were converted to state S1, independent of the addition of the second inducer (Ara) at a later stage, and those cells that were first induced with Ara were converted to state S2 independent of IPTG induction.

When cells were induced with IPTG AND THEN Ara (Fig. 4.2D, IPTG induction for one day AND THEN Ara induction for two days ([I][A][A] induction pattern)), we observed a slight increase in the mutant frequency in the positions corresponding to targets of the Ara-inducible gRNA. We suspected this to be due to leakiness of the Ara-inducible promoter during IPTG induction period (i.e., before ending the propagation delay of the first operator), which would lead to the expression of gRNA2 and aberrant transition of a small subpopulation of cells to state S2. Nevertheless, since editing by one gRNA should destroy the PAM domain for the second gRNA, the race-detecting logic should still hold within each single DNA molecule. High-throughput sequencing of these samples revealed that indeed this was the case since doubly edited allele (i.e., state S3, corresponding to editing events by both gRNAs) were extremely rare (Fig. S4.4A).

This experiment indicates that the ratio between edited alleles in a population can be tuned by controlling the induction time of each of the inputs, while ensuring that the desired logic is applied at the level of each individual DNA molecule. Alternatively, if the conversion of the whole population to a final state is desired, one can perform each induction step for periods longer than operator's propagation delay (i.e., multiple days) to allow the full conversion of cells to a given state before moving to the next induction step. This control over the degree of commitment of cells to different states could be useful for dividing biological tasks between different subpopulations in a community. For example, one subpopulation of cells could be edited to activate metabolic pathway 1 and the other subpopulation of cells could be edited activate metabolic pathway 2; the relative ratio of activation could be tuned using our DOMINO circuits to control the overall population performance.

Finally, we constructed a 2-input/2-output sequential logic circuit, where induction with IPTG AND THEN Ara results in step-wise transition between two modified states (a sequential AND gate) while induction in the opposite direction (i.e., Ara AND THEN IPTG) results in the transition to a different state. In this circuit, editing mediated by one gRNA destroys the binding site of the other gRNA, while editing mediated by the second gRNA does not interfere with the binding or editing of the first gRNA. As shown in Fig. 4.2E, this circuit is an intermediate circuit between the sequential AND gate (Fig. 4.2A) and the race-detecting circuit (Fig. 4.2D). Induction of this circuit with IPTG resulted in the transition of the target register from the initial unmodified state (state S0) to the first modified state (state S1). Subsequent induction of these cells with the second inducer (Ara) led to the transition of these cells to the doubly-mutated state (state S3). On the other hand, when cells were first induced with Ara, they were converted to an alternative singly modified state (state S2). However, subsequent induction of these cells with IPTG did not result in a transition, thus realizing the expected behavior. Using high-throughput sequencing, we confirmed that expected transitions between the states, and thus the circuit logic, held at the single-molecule level (Fig. S4.4B).

Temporal DOMINO Logic

The above examples demonstrate that the sequence and timing of DNA writing events mediated by DOMINO operators can be controlled by external cues. In addition to building sequential logic, where the execution of events in a specified order leads to a desired output, the propagation delay in DOMINO operators can be exploited to incorporate temporal logic into circuits, where a desired output is produced only after a certain period of time has passed. In a simple form, DOMINO delay operators can

be built by constructing a series of overlapping repeats to act as target sites for a desired gRNA (Fig. 4.3A). This repeat configuration allows one to overlap the READ address of each gRNA operator site with the WRITE address of the previous gRNA. Initially, the gRNA can bind to the first (i.e., 3'-end) repeat, but not to the upstream copies of the repeat that harbor dC residues (instead of dT) in the sequence corresponding to the gRNA READ address (i.e., the gRNA seed sequence). Upon binding to the first repeat, the gRNA can mutate the dC residues in the repeat immediately upstream of its binding site (i.e., the second repeat), thus converting that repeat to a new binding site for another copy of the same gRNA. This process is sequentially repeated to generate new binding sites for the gRNA. Much like an array of physical domino pieces that fall down one by one, each genome-editing event is initiated only after editing in the previous repeat has occurred, thus ensuring a sequential cascade of DNA writing events. The total delay can be tuned by changing the number of the repeats, modifying the overlapping distance between the repeats, or adjusting the distance of mutable residues from their corresponding PAM sequences.

In addition, the output of the delay elements can be combined with additional logic operators and internal or external cues to create more complex forms of temporal logic. To demonstrate this concept, we placed three DOMINO delay elements into an array and linked the output of the array to a second DOMINO operator that implements sequential AND logic (Fig. 4.3A). This design achieves temporal and sequential AND logic since the first (IPTG-inducible) gRNA has to execute three consecutive DNA writing events before the Ara-inducible gRNA corresponding to the last operator can bind to and edit its target. We induced cells harboring this circuit with different IPTG concentrations for 4 consecutive days followed by a final day of induction with Ara. Using Sanger sequencing on the population and Sequalizer analysis, we observed a time- and IPTG-dosage-dependent accumulation of mutations in the target sites within repeats, corresponding to the propagation of the signal through the repeat array (Fig. 4.3B). The rate of propagation of the mutation cascade through the delay elements correlated with both the concentration and duration of exposure to IPTG. By the end of the experiment, mutations in the position corresponding to the target site of the second gRNA (shown by the blue arrow in Fig. 4.3B) were detected only in conditions in which mutations had accumulated through the entire cascade, corresponding to the samples that had been induced with the highest IPTG concentrations.

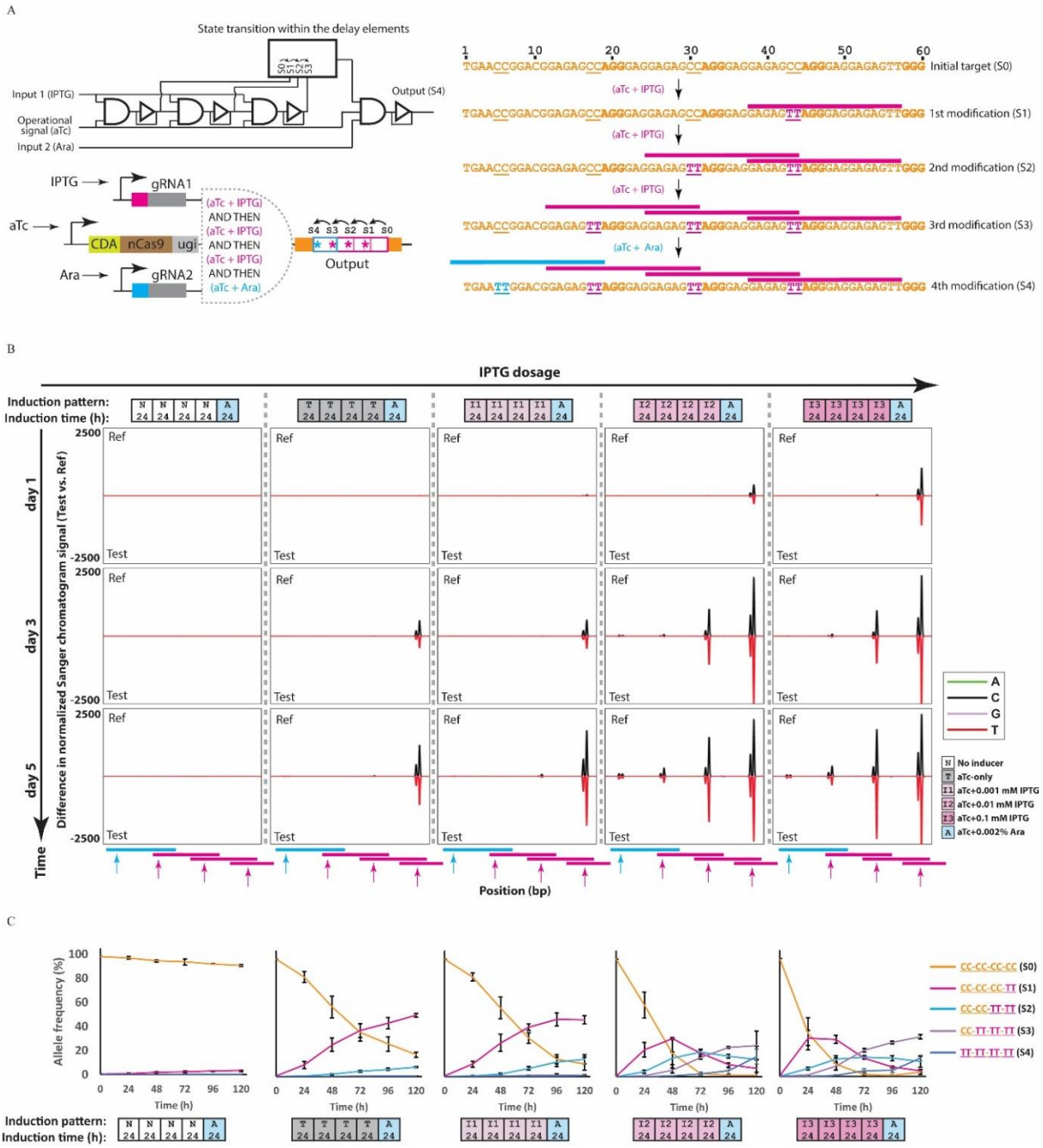


Figure 4.3 | Incorporating propagation delay and temporal logic into living cells. (A) Time-dependent logic and tunable propagation delay can be programmed by DOMINO operator cascades. DOMINO operators possess an inherent propagation delay (the time required for the transition from a non-modified state to modified state) that can be modulated in an analog fashion (stronger induction results in a shorter delay). Multiple DOMINO operators can be placed sequentially in an array to build longer delays and then coupled with other logic operators to build temporal logic. We constructed a series of overlapping repeats to serve as gRNA binding sites. Once expressed, the first gRNA (IPTG-inducible, pink) can bind to the downstream repeat, but not to the other instances of the repeats due to the presence of dC residues in these repeats that form mismatches with the gRNA READ address. Upon binding the downstream repeat, the DNA read-write head can mutate these dC residues to dT in the immediately adjacent upstream repeat, thus creating a new binding site for this gRNA. In

turn, this event recruits the read-write head once again and makes the third repeat available for binding. The second gRNA, which is under control of Ara, is only able to bind to and edit its target when the third copy of the repeat is edited by the first gRNA, thus encoding time-dependent sequential logic. **(B)** *E. coli* cells harboring the circuit shown in (A) were exposed to different concentrations of the first inducer (IPTG) for 4 days with serial dilution after each day, followed by a one-day exposure to the second inducer (Ara). The propagation of the signal as manifested by sequential mutations in the repeat array was monitored by analyzing Sanger chromatograms with Sequalizer. Transitions between states occurred in a time- and IPTG-dosage dependent fashion, and only cells exposed to higher concentrations of IPTG (0.1 mM and 0.01 mM) accumulated mutations to the level that enabled a response to the second inducer (Ara) by the last day of the experiment. **(C)** Transitions between the memory states for samples shown in (B) assessed by HTS. Error bars indicate standard deviation for three biological replicates.

We further confirmed these results by analyzing these samples with HTS. This analysis also showed time- and IPTG-dosage-dependent mutation accumulation within the repeats (Fig. 4.3C). Furthermore, the mutation corresponding to the target of the Ara-inducible gRNA only accumulated in the later time points and only in cultures induced with high concentrations of IPTG. Upon induction of the samples by Ara, the frequency of the allele corresponding to the final output of the circuit (i.e., state S4) only increased in samples that had been previously induced with high IPTG concentrations (i.e., 0.01 mM and 0.1 mM). These results further demonstrate that, in addition to enacting delays in gene circuits, an array of DOMINO delay elements can be used as a multi-state memory register that undergoes transitions between different discrete states (i.e., sequential mutations) in a time- and dosage-dependent fashion. In this design, the number of memory states can be tuned by changing the number of repeats. Moreover, the timing and probability of transitions between repeats can be adjusted by changing the position of mutable residues within the repeat overlaps, or tuned dynamically by external cues.

Finally, to demonstrate the power of the technique, we used DOMINO delay elements to build a gene expression program in which the conversion of cryptic ACG start codons into canonical ATG start codons in three different ORFs was temporally controlled by a single input (Fig. S4.5). We envision that more complex versions of temporal logic, such as counters, can be constructed by integrating delay elements into multiple-input DOMINO operators.

Associative Learning Circuits and Online DNA-State Reporters

A unique feature of DOMINO operators compared to other memory platforms is that the DOMINO DNA read-write head can be further functionalized with additional effector domains, such as transcriptional activators and repressors, to achieve combined DNA writing and transcriptional regulation. This offers the unprecedented capacity to

perform both genetic and epigenetic modulation and thus combine DNA memory states with functional outcomes. For example, this feature enables the construction of circuits that can learn and remember. Specifically, we devised a synthetic gene circuit that undergoes associative learning (*106-109*) such that its gene expression output is reinforced by a given stimulus (Figure 4.4A). While transcriptional positive feedback loop can also be used to implement synthetic self-reinforcing circuits, the state of such circuits can fluctuate due to their reliance on continuous transcription for state maintenance. In contrast, an associative learning circuit that uses genetically encoded memory to gradually reinforce a response remains intact and stable even after the initial stimuli is removed.

To demonstrate this concept, we first made an array of overlapping repeats (operators) composed of four WT repeats (4xOp) and a downstream mutant repeat (1xOp*) which harbored a dC to dT mutation. We then placed this repeat array upstream of a minimal promoter driving GFP to build the 4xOp_1xOp*_GFP reporter construct. Additionally, we built a second reporter (1xOp*_GFP) by placing a single Op* repeat upstream of the minimal promoter driving GFP. We also functionalized the DNA read-write head (nCas9-CDA-ugi) with a transcriptional activator domain (VP64) and cloned the nCas9-CDA-ugi-VP64 fusion construct along with either of the two reporter constructs into lentiviral vectors, which were subsequently introduced into the human HEK 293T cell line. We then delivered a second lentiviral vector encoding an Op*-specific gRNA (gRNA(Op*)) (or a non-specific gRNA (gRNA(NS)) as a negative control) to these cells. Upon binding, gRNA(Op*) could mutate the critical dC residue in the WT Op repeat immediately upstream of its binding site, thus converting the Op repeat to a new Op* sequence that could serve as a new binding site for the same gRNA; this strategy enables sequential rounds of mutations (i.e., Op to Op* conversion) and gRNA binding events (Fig. 4.4A). We sequentially passaged cells harboring these circuits every three days for fifteen days (Fig. 4.4B) and observed GFP expression and the genotype of the cells by microscopy (Figs. 4C-D and S4.6A) and HTS (Figs. 4.4E-F), respectively. As shown in Fig. 4.4C, the frequency of GFP-positive cells in cultures harboring the 4xOp_1xOp*_GFP reporter and gRNA(Op*) increased over time, indicating the gradual activation of the reporter in the population. On the other hand, the frequency of GFP-positive cells did not change significantly in cultures that were transfected with gRNA(NS), or those that contained the 1xOp*_GFP reporter.

In addition to observing an increased frequency of GFP-positive cells, we observed that the intensity of the GFP signal in GFP-positive cells increased in cultures that harbored the 4xOp_1xOp*_GFP reporter and gRNA(Op*) over time (Fig. 4.4D). This data suggests that the number of bound transactivators, and thus, the number of activated (i.e., Op*) repeats that can serve as operator sites for the chimeric read-write-transactivator protein increased in these cells. On the other hand, no significant increase was observed in negative controls that harbored gRNA(NS) or those that that contained the 1xOp*_GFP reporter.

These results were further confirmed by analysis of the allele frequencies throughout the experiment by HTS. As shown in Fig. 4.4E, the frequency of the WT allele (state S0) in cells containing the repeat array and gRNA(Op*) decreased linearly with time over the course of the experiment. On the other hand, the frequency of intermediate states (S1 through S4) gradually increased and reached a plateau towards the end of the experiment, suggesting that these intermediate states reached steady state (Fig. 4.4F). The allele frequency of the final state (S5) gradually increased over the course of the experiment. No significant change in allele frequency was observed in cells that were transduced with a non-specific gRNA (Fig. S4.6B). Together with the microscopy data, these results show that the analog properties of a signal, such as the duration of exposure to gRNA(Op*), can be faithfully and permanently recorded within the distribution of memory states of the DNA recorder within the population. On the other hand, at the single cell level, each repeat forms a multi-bit digital recorder that associates longer or higher intensity of exposures to an incoming signal with transitions to higher memory states in the form of more accumulated mutations.

In samples harboring the gRNA(Op*) and either the 1xOp*_GFP or 4xOp_1xOP*_GFP reporters, we also observed dC to dG and dC to dA mutations, albeit with lower frequencies than for dC to dT mutations (Fig. S4.6C). This is consistent with previous results reported in mammalian cell lines (*95, 102*), and reflects the promiscuous outcome of repair of deaminated dC (dU) lesions in these cells. Notably, in samples containing the 1xOp*_GFP reporter, the frequency of the WT allele (state S0) decreased and the frequency of the mutant alleles increased linearly over time (Fig. S4.6C). Thus, even without having a repeat array, the accumulation of mutations in a specific target site can be used as an analog readout of an incoming signal.

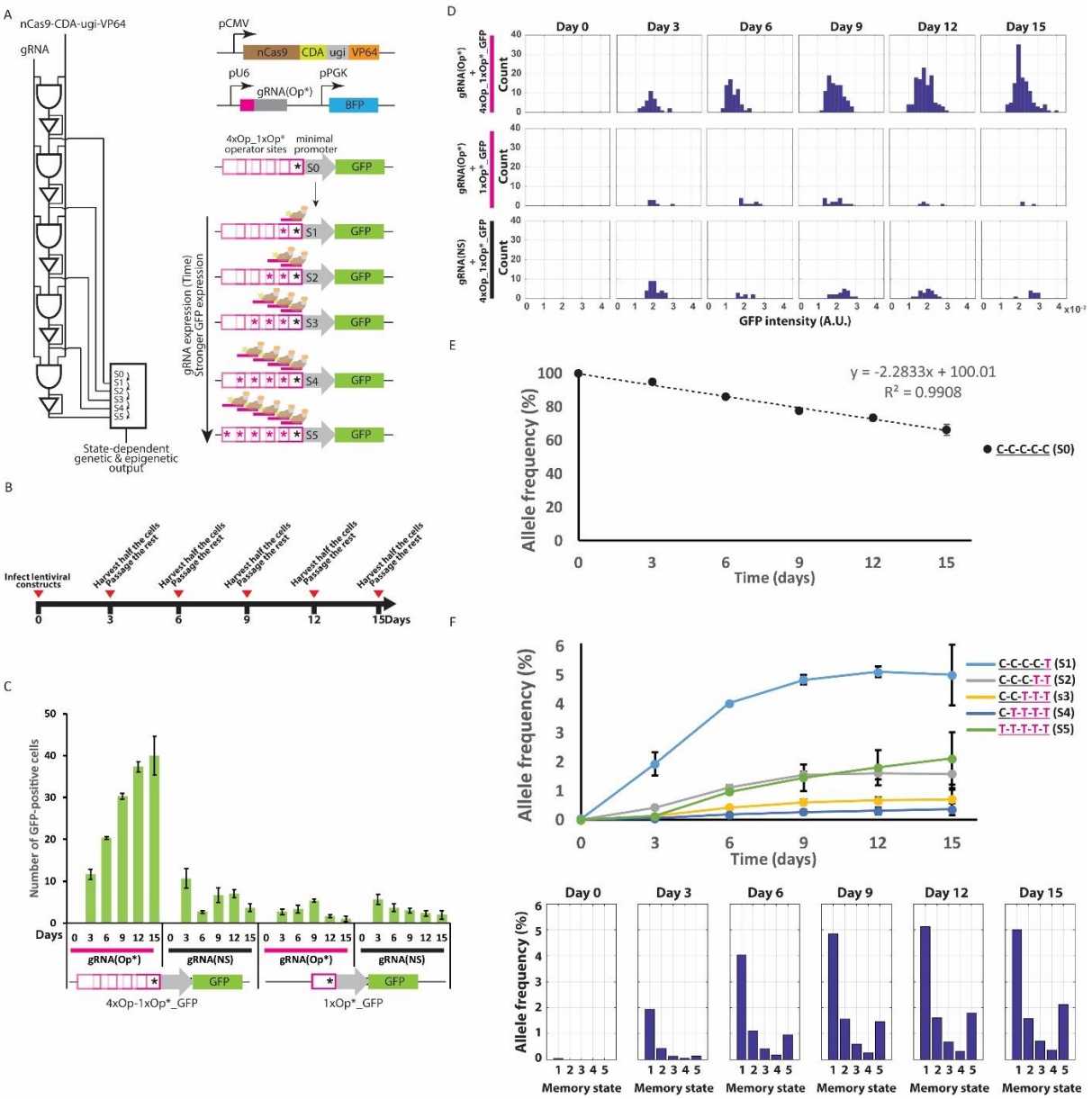


Figure 4.4 | Associative learning and online DNA-state reporting circuits in human cells. (A) Being CRISPR-Cas9-based, DOMINO operators can be functionalized with transcriptional and epigenetic modules to implement gene regulation integrated with computing and memory. As an example, we functionalized the read-write head with a transcriptional activator (VP64) and used it to sequentially edit and activate multiple operator sites that were arrayed in overlapping repeats (composed of four copies WT unmutated repeats (Op) followed by a downstream mutated repeat (Op*)) upstream of a minimal promoter (4xOp_1xOp*_GFP). In the presence of an Op*-specific gRNA (gRNA(Op*)), this system allows for the sequential conversion of Op sites to Op* and binding of the transactivator to the progressively mutated operator sites in the promoter, which in turn results in GFP signal increases. Therefore, cells harboring this circuit manifest sequential and permanent transitions between DNA states and increases in GFP in response to increased gRNA expression over time. Thus, the circuit can be considered as an example of associative learning. **(B)** HEK 293T cells were transfected with the circuit shown in (A) via a two-step lentiviral delivery protocol and were grown with serial passaging every three days as indicated. At the end of each passage, GFP signal was assessed by microscopy and DNA memory state was assessed by HTS. **(C)** The mean number of GFP-

positive cells in different samples harboring either the Op*-specific gRNA (gRNA(Op*)) or a non-specific gRNA (gRNA(NS)) and either 4xOp_1xOp*_GFP or 1xOp*_GFP as the reporter. The number of GFP-positive cells harboring 4xOp_1xOp*_GFP and gRNA(Op*) increased over time. In contrast, the number of GFP-positive cells in cultures harboring gRNA(NS) or 1xOp*_GFP and gRNA(Op*) did not change and remained at background levels. **(D)** Histogram of signal intensities for GFP-positive cells shown in (C). Over time, the intensity of GFP-positive cells increased in samples harboring 4xOp_1xOP*_GFP and gRNA(Op*) gradually increased, reflected as a shift to the right in the histograms, indicating multi-stage GFP activation in these cells. The signal intensities in cells harboring gRNA(NS) or those that had 1xOp*_GFP and gRNA(Op*) remained at the background level. **(E)** Dynamics of the frequency of the WT unmodified allele (state S0) in cultures harboring 4xOp_1xOp*_GFP and gRNA(Op*) assessed by HTS. The frequency of the unedited allele decreased linearly over time, indicating that the DNA writing circuit can be used as an analog recorder for the input gRNA. **(F)** Dynamics of mutant allele frequencies (memory states S1 through S5) for the same samples as (E), shown as time-series data and histograms. Consistent with the GFP data, the first four memory states (S1 through S4) started to accumulate sequentially (state S1, then state S2, then S3 and then S4) until they reached a plateau. Moreover, memory state S5, which corresponds to the highest GFP expression state, increased steadily over time, as is expected from the terminal product of the DNA memory circuit.

Besides serving as a proof of concept for associative learning, the synthetic genetic circuit described in this experiment can be used as an online functional reporter for DNA memory states. Unlike existing DNA-based molecular recording technologies that rely on DNA sequencing to be read, the precise and sequential DNA writing achieved by DOMINO enables one to correlate the DNA memory state (i.e., the number of edited repeats) with the intensity of a fluorescence reporter signal that can be monitored in living cells without disrupting the cells (Fig. 4.4). This feature makes DOMINO recorders especially useful for studying biological events in an online fashion in their native context.

In this experiment, we used VP64 as an activator domain. However, the activation level and dynamic range of the reporter output can be tuned by using stronger activator domains such as VPR (*110*). Alternatively, other effector domains (such as repressors (*103*), DNA methyltransferases (*111*), acetyltransferases (*112*), or other types of histone modification domains) could be used to implement more sophisticated forms of gene regulation programs.

4.4 Discussion

Our DOMINO platform addresses many limitations of current DNA writing platforms by using a DNA read-write head that converts the genomic DNA of living cells into a readable and writable medium that can be manipulated with single-nucleotide resolution. Orthogonal DOMINO operators can be built by simply changing the sequence of gRNAs, making the system highly scalable. Furthermore, due to the ability

to manipulate DNA with single-nucleotide resolution within a defined window, compact multi-input operators can be readily created by targeting multiple gRNAs to nearby registers. By leveraging DNA as the computing and storage medium, we anticipate that this approach will be more stable than transcriptional memory strategies. Unlike other systems that require multiple recombinases to encode memory, DOMINO uses small gRNAs and only one protein moiety. The CRISPR-Cas9-based nature of this system and the absence of any requirement for double-strand DNA breaks or special repair mechanisms (such as Non-Homologous End Joining (NHEJ)) enables this system to be functional in both prokaryotic and eukaryotic cells. As a result, DOMINO offers a highly modular, robust and scalable strategy for dynamic programming of memory as well as order-independent, sequential and temporal logic operations in living cells. Furthermore, we show that DOMINO can be used to record both analog and digital signals, depending on the temporal nature of the circuits constructed. DOMINO circuits can be readily interfaced with other gene regulatory mechanisms to modulate gene expression and provide online readouts of cellular memory. Thus, we anticipate that DOMINO will allow for new strategies and unprecedented capacities to control cellular phenotypes and study biological phenomena in their native contexts.

In this paper, we focused on executing unidirectional DNA writing events by using a cytidine deaminase as the DNA writing module. Very recently, an adenosine deaminase DNA writing module that allows for dA to dG and dT to dC mutations was described (*113*). Incorporating this new DNA writing module (or other orthogonal writer modules) into DOMINO should make reversible DNA writing possible, which has been challenging to achieve with previous DNA memory platforms. This will enable bidirectional cellular programs and thus pave the way for sophisticated biological state machines, cellular automata, and Turing machines that use the genomic DNA of living cells as a rewritable memory tape to perform advanced memory and computation operations.

In addition to digital computation, DOMINO operators can be used to perform analog memory and computation in living cells when propagation delays are taken into account. Furthermore, as shown in Figs. 4.3 and 4.4, analog properties (i.e. duration and magnitude) of an incoming signal can be recorded within the mutation states of the DOMINO operators. In these examples, recording capacity can be increased by extending the number of repeat elements or tuning the overlapping distance between the repeats. On the other hand, the input-output transfer function (i.e., the relationship

between gRNA expression level and degree of mutation) can be tuned by adjusting the position of mutable residues within the gRNA WRITE window.

Deterministic DOMINO operators and cascades rely on precise base editing events for proper function. Our results show that using the CDA-nCas9-ugi head, the outcome of these operators in *E. coli* are almost exclusively in the form of dC to dT mutations. However, in human cells, other nucleotides (dG, and to a lesser extent, dA) are also generated, albeit with a lower rate than dT (Fig. S4.6C). In human cells, this issue could generate undesirable memory states that could reduce the performance of deterministic DOMINO operators. This can be addressed by implementing strategies that favor dC to dT mutations over the other possible outcomes to improve the efficiency of correct outcomes (*114*) or using alternative DNA writing modules that generate more pure editing products (*113*).

Several CRISPR-Cas9 based strategies for recording information, such as signaling dynamics and cellular lineage histories, into DNA have been recently described (*115-117*). These approaches rely on stochastic DNA memory states (i.e., indel mutations) that are generated by Cas9-mediated double-strand DNA breaks and subsequent repair of these breaks by NHEJ. However, the recording capacity of these recorders are exhausted within a few generations or after recording a few molecular events due to loss of gRNA target sites and are therefore not ideal for long-term recording of signaling dynamics and event histories. Moreover, since indel mutations (memory states) are stochastically generated due to NHEJ, new mutations could destroy the previous mutations and thus overwrite the previous memory states, making tracing lineage histories challenging. In addition, none of these strategies can be used in organisms without an efficient NHEJ repair pathway, such as prokaryotes.

In contrast, mutational memory states generated by DOMINO are precise, unidirectional, position-specific, and minimally-disruptive. The features ensure that previous mutations are preserved after each editing step and can be accurately traced. The precise and predictable memory state transitions in DOMINO recorders enables one to couple memory states to functional biological outcomes, such as changes in gene expression (Fig. 4.4). Furthermore, DOMINO does not require double-strand DNA breaks or NHEJ, thus enabling it to function in both bacterial and mammalian cells in an autonomous and continuous fashion over many generations. We envision that the DNA record generated by the DOMINO recording system could be used to study signaling dynamics and event histories over many generations in their native contexts. The promiscuous repair of dC lesions in mammalian cells could actually be beneficial

for lineage tracking applications, as it can increase the number of potential memory states. Moreover, signal-responsive lineage maps with tunable resolution can be generated because the activity of DOMINO recorder can be modulated by internal or external signals of interest. Combining these recorders with single-cell sequencing, advanced barcoding schemes, and self-targeting guide RNAs (*78*) should pave the way toward more advanced recorders for long-time monitoring of signaling dynamics and cellular lineages.

We envision that our long-term, compact, scalable, modular, and minimally disruptive DNA writers will enable an unprecedented set of applications for both building genetic programs and the recording of spatiotemporal molecular events in their native contexts. These applications could be highly impactful across many different fields, including development, cancer, stem cell differentiation, brain mapping, and many other areas. For example, DOMINO can be used to design and program the progression of developmental stages within living animals, or to perform long-term lineage tracking experiments in mammals, which has been impossible to date due to the lack of scalable and long-term methodologies. DOMINO recorders could be adapted to map neural activity by driving the activity of DNA writers with regulators that respond to neural activity. One could study the order and temporal nature of signaling events in their native contexts and robustly control cellular differentiation cascades *ex vivo* and *in vivo*. Our DNA writers could be programmed to investigate tumor development and unveil the cellular and environmental cues involved in tumor heterogeneity. Arbitrary information could be programmed into the DNA of living cells for DNA storage applications. Finally, living sensors could be designed to sense pathogens, toxins, or other signals within the body or in the environment and then later report on this information in detail.

4.5 Supplementary Information

Estimating Position-Specific Mutant Frequencies by Sequalizer

We developed a MATLAB program, dubbed Sequalizer (for Sequence equalizer), to calculate the frequency of base-pair substitutions in specific positions in a mixture of DNA species from Sanger sequencing chromatograms. Analyzing Sanger chromatograms by Sequalizer offers a low-cost strategy to HTS for assessing and quantifying the frequency of precise mutations (i.e. nucleotide substitutions) that are generated by base-editing and other targeted genome engineering platforms.

Sequalizer uses a previously described algorithm (SeqDoC (*118*)) to normalize and compute the difference between Sanger chromatogram of a reference (unmodified)

sequence and a test sample (which is expected to contain a mixture of DNA species containing mutations in specific positions). It then overlays the computed difference for all the four nucleotides (A, C, G, and T) on a single plot for the reference (top) and test sample (inverted, bottom) as a function of nucleotide position (x-axis) (Fig. S4.1A). A peak in this plot indicates a difference in the normalized chromatogram signal between the reference and the test sample, and thus a mutation (i.e. base substitution) in that specific mutation. Sequlizer then estimates the frequency of mutants in each specific (targeted) position in the test sample using the difference between the heights of peaks corresponding to the reference and test samples in that position and reports that frequency as a number on top of the corresponding peaks. A test sample that has the same position-specific mutant frequency as the reference would result in no peaks in the Sequalizer plots (Fig. S4.1A, top panel). On the other hand, base-substitutions in the test sample compared to the reference sample can be detected as a peak in the Sequalizer plots (Fig. S4.1A, bottom panel). If a pure WT sample is used as the reference sample, the number printed on top of the peak estimates the frequency of molecules with mutation in that specific position in the test sample.

Since there is a high degree of variation between the height of peaks between different positions along a Sanger chromatogram, for each position Sequlizer normalizes the computed difference to the height of the peak for the reference chromatogram in that specific position. However, the height of the Sanger chromatogram containing 100% mutant alleles in a position could be different from the reference in that position, which could result in under- or over-estimation of mutant frequencies by Sequalizer. Since the Sanger chromatogram, and thus the height of peaks for samples with the 100% mutant alleles are not always known, Sequlizer uses an experimentally determined parameter to account for the difference in height of peaks of Sanger chromatogram in each position. This parameter was calculated by mixing pure WT and pure mutant samples with different ratios, sequencing the mixtures, and using the Sequalizer output of the corresponding chromatograms to calculate a standard curve. As shown in Fig. S4.1B, the Sequalizer algorithm is able to compute frequencies of mutants at different positions solely based on Sanger chromatogram data, which correlates well with the mutant ratios in the mixtures.

We further verified Sequalizer by measuring position-specific mutant frequencies and comparing the output with the HTS for samples obtained from the order-independent AND gate circuit for the experiment described in Fig 4.1B. As shown in Fig. S4.1C, we observed high correlation (R^2 values) between mutant frequencies measured by both methods in all the targeted positions, indicating that Sequalizer output can be

used as a low-cost alternative to HTS. Deviation of the regression slope from unity (e.g., for the C20 position) could be partially due to variations in the height of peaks of Sanger chromatograms between pure WT and pure mutant at different positions. As mentioned above, the Sequalizer algorithm tries to minimize the effect of such variations by normalizing the differences to the height of the WT peak in corresponding positions. However, since the heights of Sanger chromatograms for a pure mutant species also could affect the Sequalizer and this value is often unknown, it could cause the Sequalizer to underestimate or overestimate mutant frequencies compared to those measured by HTS. Nevertheless, the high correlation between Sequalizer outputs and HTS results indicate that changes in Sequalizer output can be used as a quantitative measure of changes in allele frequencies in a given position, even if they are not used for absolute measurements. The MATLAB script for Sequalizer is provided in Supplementary Materials.

Materials and Methods

Strains and Plasmids

Standard molecular biology and cloning techniques, including ligation, Gibson assembly (*89*) and Golden Gate assembly (*90*) were used to construct the plasmids. Chemically competent *E. coli* DH5α F' $lacI^q$ (NEB) and E. cloni 10G (Lucigen) were used for cloning. MG1655 PRO strain (MG1655 strain that harbors PRO cassette (pZS4Int-*lacI*/*tetR*, Expressys) and expresses *lacI* and *tetR* at high levels) (*30*) was used for all the bacterial experiments. HEK 293T cells (ATCC CRL-11268) were purchased from and authenticated by ATCC and were used for mammalian cell experiments. Lists of plasmids, synthetic parts and sequencing primers used in this study are provided in Tables S4.1, S4.2, and S4.3, respectively. Plasmids and their corresponding maps will be available on Addgene.

Antibiotics and Inducers

For bacterial selection, antibiotics were used at the following concentrations: Carbenicillin (Carb, 50 μg/mL), and Chloramphenicol (Cam, 25-30 μg/mL).

For the experiments shown in Figs. 4.1E, 4.2D, 4.2E, S4.2C, and S4.4 different combinations of 200 ng/ml anhydrotetracycline (aTc), 0.1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and 0.2% Arabinose (Ara) were used to induce the corresponding circuits. For the experiments shown in Figs. S4.3 and S4.5, 250 ng/ml aTc and 0.005% Ara were used. For the experiment shown in Fig. 4.2A, 150 ng/ml aTc, 0.1 mM IPTG, and 0.2% Ara were used. For all the other experiments, unless

otherwise noted, 250 ng/ml aTc, 1 mM IPTG and 0.2% Ara were used. All concentrations are final concentrations.

Experimental procedure

Bacterial Cell Experiments

Different plasmids expressing gRNAs and targets (listed in Table S1) were transformed into the reporter cells (MG1655 PRO) harboring aTc-inducible CDA-nCas9-ugi (for bacterial experiments, APOBEC1 CDA (*95*) was used as the writing module). Single transformant colonies were grown in LB + Carb + Cam for 6-8 hours to obtain seed cultures. Seed cultures were diluted (1:100) in fresh media containing different combinations of inducers and grown in 96-well plates for multiple days with serial dilution as indicated in induction patterns in corresponding figures. Samples for various analyses including HTS, Sequalizer, and flow cytometry were taken at indicated time points.

Cell Cultures and Mammalian Cell Experiments

Cell culture and transfections were performed as described previously (3). HEK 293T cells were grown in DMEM supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. Lentiviruses were packaged using the FUGW backbone (Addgene #25870) and psPAX2 and pVSV-G helper plasmids in HEK 293T cells. Filtered lentiviruses were used to infect respective cell lines in the presence of polybrene (8 μg/mL). Successful lentiviral integration was confirmed by using lentiviral plasmid constructs constitutively expressing fluorescent proteins or antibiotic resistance genes to serve as infection markers.

A lentiviral plasmid construct was made by placing the nCas9-CDA-ugi-VP64 fusion protein with nuclear localization signals linked to the Puromycin resistance gene with the P2A sequence under the control of constitutive CMV promoter (for mammalian experiments, PmCDA (*102*) was used as the writing module). In addition, repeat arrays (4xOp_1xOp* or 1xOp*) were placed upstream of the minimal pMLV promoter driving EGFP and the resultant reporter constructs were cloned into the same lentiviral construct. The clonal cell lines harboring the two transcriptional units were constructed by infecting early passage HEK 293T cells with high titer lentiviral particles, selecting for pooled populations grown in the presence of Puromycin (7 μg/mL) and picking up clonal populations after seeding pooled population with the density of 0.5 cells per well in a 96-well plate.

On day 0, 440,000 clonal reporter cells were infected with high titer lentiviral particles encoding the sgRNAs driven by the U6 promoter in a 6-well plate with triplicates. Infection efficiency was more than 90% in every sample. The cells were harvested every 3 days until day 15 after the infection. Half of the harvested cells were seeded in a 6-well plate for further culture and a quarter of cells were collected for next-generation sequencing. Microscopic images were obtained just before the harvests.

Microscopy Image Analysis

Fluorescence microscopy images of cells in tissue culture plates were obtained by using the ZEISS ZEN microscope software. For each sample, the total number of EGFP-positive cells and signal intensities were measured from microscopic images of 5 random fields using CellProfiler image analysis software by using the 'ColorToGray', 'IdentifyPrimaryObjects', MeasureObjectIntensity' and 'ExportToSpreadsheet' modules.

Flow Cytometry

An LSR Fortessa II flow cytometer (Becton Dickinson, NJ) was used for all the experiments. GFP expression was measured using 488/FITC laser/filter set. All samples were uniformly gated and flow cytometry data were analyzed by FACSDiva and FlowJo (Becton Dickinson, NJ). For each gated sample, the mean fluorescence and percent of GFP-positive cells were calculated.

High-throughput Sequencing

For each sample, 5 µl of culture was resuspended in 15 µl of QuickExtract DNA Extraction Solution (Epicentre, WI) and lysed by a two-step protocol (15 minutes incubation at 65°C followed by 2 minutes incubation at 98°C). Target sites were PCR amplified using 2 µl of lysed cultures as template and the appropriate primers listed in Table S3. The obtained amplicons were directly used as templates in a second round of PCR to add Illumina barcodes and adaptors. The amplicons were then multiplexed and analyzed by Illumina MiSeq. The obtained sequencing reads were demultiplexed and allele frequencies were calculated using a custom MATLAB script.

Sanger Sequencing and Sequalizer Analysis

For each sample, target sites were PCR amplified by target-specific primers and Sanger sequenced by Quintara Biosciences. The obtained Sanger chromatograms were then analyzed by Sequalizer using seed cultures as reference as described above.

Sequalizer MATLAB code

```
%Sequalizer V1.0
%Authors: Fahim Farzadfard & Nava Gharaei
%This script uses a modified version of SeqDoc (Crowe, M. L.(2005). BMC Bioinformatics, 6(1), 133).
%The path to this script should be specified in the input parameters.
% Perl and SeqDoc dependencies need to be installed.

%input parameters
% 1. Specify path to the modified SeqDoc script
SeqDoc_path='C:\Sequalizer_SeqDoc.pl';
% 2. Select the reference Sanger chromatogram. The reference
%chromatogram will be compared with all the other Sanger chromatograms in the
%same directory (i.e., test samples). The output of Sequalizer will be saved in the
%same directory.
[FileName,PathName] = uigetfile('*.ab1','Select the reference file within directory that contains test files');
% 3. Specify sequence of the WT taget. It is used to align Sanger chromatograms to the target.
WT='CCCAATGGCGATTAAATGGCCTAAATAGAGCCTATGGG';
% 4. Specify positions of the targeted dC or dG residues
target=[18:21];
% end of input parameters

%Analyze the reference chromatogram by SeqDoc
fileID = fopen([PathName 'input_parameters.txt'],'w');
fprintf(fileID,'%s\n',['WT=' WT char(9) num2str(target)]);
fclose(fileID);
y1=-5000;y2=5000;
fileID = fopen([PathName 'output.txt'],'w');
tm=0;
ref=[PathName FileName];
system([SeqDoc_path ' ' ref ' ' ref ' ' PathName]);
imp_ref=textread([PathName 'ref.txt'],'%s','delimiter','\n');
t = ~cellfun(@isempty, regexp(imp_ref,'=>'));
a=find(t==1);
labels_ref=strrep(split(imp_ref(a),' =>'),"","");
st='';
st=['"" char(labels_ref(1,1)) '",[]'];
val=[];
for i=2:size(labels_ref,1);
   st=[st ',"' char(labels_ref(i,1)) '",[]'];
end
REF=eval(['struct(' st ')']);
id_seq=find(ismember(labels_ref, 'sequence')); REF.sequence=char(replace(labels_ref(id_seq,2),{''',',',""},''));
lab={'a_trace','c_trace','g_trace','t_trace','base_pos'};
```

```
for u=1:size(lab,2)
id_pos(u)=find(ismember(labels_ref, lab(u)));
if id_pos(u)<size(labels_ref,1)
O=str2num(char(replace(imp_ref(a(id_pos(u))+1:a(id_pos(u)+1)-2),',',')));
else
O=str2num(char(replace(imp_ref(a(id_pos(u))+1:end-2),',',')));
end
if u==1
   REF.a_trace=O;
elseif u==2
   REF.c_trace=O;
elseif u==3
   REF.g_trace=O;
elseif u==4
   REF.t_trace=O;
elseif u==5
   REF.base_pos=O;
end
  O=" ;
end
xp=localalign(WT,REF.sequence);
start=REF.base_pos(xp.Start(2));stop=REF.base_pos(xp.Stop(2)-tm);
k=0;
for k=0:xp.Stop(2)-tm-xp.Start(2)
   u=REF.sequence(xp.Start(2)+k);
   lb{k+1}=u;
end
tout_labels='filename';
for kk=1:size(target,2)
   tout_labels=[tout_labels char(9) char(lb(target(kk))) num2str(target(kk))];
end
fprintf(fileID,'%s\n',tout_labels);
%End of reference chromatogram analysis

%Analyze each test chromatogram and compare with reference
%estimate position-specific mutation frequencies in the targeted positions
%Plot the differences between normalized chromaogram of ref and test
%samples and export position specific mutation frequencies to an output file.
files=dir([PathName '*.ab1']);
for z=1:size(files,1)
filename=[PathName files(z).name];
system([SeqDoc_path ' ' ref ' ' filename ' ' PathName]);
imp=textread([PathName 'diff.txt'],'%s','delimiter','\n');
imp=replace(imp,"","");imp=replace(imp,',','');
labels={'a_trace => [', 'c_trace => [', 'g_trace => [', 't_trace => ['};
ids={'A', 'C', 'G', 'T'};
```

```
for i=1:size(labels,2)
ind(i)= find(ismember(imp, labels(i)));
end
[ord d]=sort(ind);
leng=ord(2)-ord(1)-2;
diff=struct(ids{d(1)},str2num(char(imp(ord(1)+1:ord(1)+leng))),ids{d(2)},str2num(char(imp(ord(2)+
1:ord(2)+leng))),ids{d(3)},str2num(char(imp(ord(3)+1:ord(3)+leng))),ids{d(4)},str2num(char(imp(or
d(4)+1:ord(4)+leng))));
pl=traceplot(diff);
pl.A.LineWidth=2;pl.C.LineWidth=2;pl.G.LineWidth=2;pl.T.LineWidth=2;pl.C.Color=[0 0
0];pl.G.Color=[0.9 .65 1];
xlim([start stop]);
ylim([y1 y2]);
title(strrep(files(z).name,'_','\_'));
xticks(REF.base_pos(xp.Start(2):xp.Stop(2)-tm));
xticklabels(lb);
set(gca, 'fontsize', 10, 'fontweight', 'b');
for p=1:size(target,2)
   if WT(target(p))=='G'
      xl=REF.base_pos(xp.Start(2)+target(p)-1);
      yl=diff.G(REF.base_pos(xp.Start(2)+target(p)-1));
   val(p)=round(real(sqrt((max(diff.G(REF.base_pos(xp.Start(2)+target(p)-1)-
2:REF.base_pos(xp.Start(2)+target(p)-1)+2))-min(diff.A(REF.base_pos(xp.Start(2)+target(p)-1)-
2:REF.base_pos(xp.Start(2)+target(p)-
1)+2)))/(8*max(REF.g_trace(REF.base_pos(xp.Start(2)+target(p)-1)-
2:REF.base_pos(xp.Start(2)+target(p)-1)+2)))),3);
   elseif WT(target(p))=='C'
      xl=REF.base_pos(xp.Start(2)+target(p)-1);
      yl=diff.C(REF.base_pos(xp.Start(2)+target(p)-1));
   val(p)=round(real(sqrt((max(diff.C(REF.base_pos(xp.Start(2)+target(p)-1)-
2:REF.base_pos(xp.Start(2)+target(p)-1)+2))-min(diff.T(REF.base_pos(xp.Start(2)+target(p)-1)-
2:REF.base_pos(xp.Start(2)+target(p)-
1)+2)))/(8*max(REF.c_trace(REF.base_pos(xp.Start(2)+target(p)-1)-
2:REF.base_pos(xp.Start(2)+target(p)-1)+2)))),3);
   else
      val(p)=0;
   end
   h=text(xl,yl+200,num2str(val(p)));
   set(h, 'rotation', 90)
end
fprintf(fileID,'%s\n',[files(z).name char(9) replace(regexprep(num2str(val),' +',' '),' ',char(9))]);
saveas(gcf,[PathName files(z).name '.jpg']);
end
fclose(fileID);
```

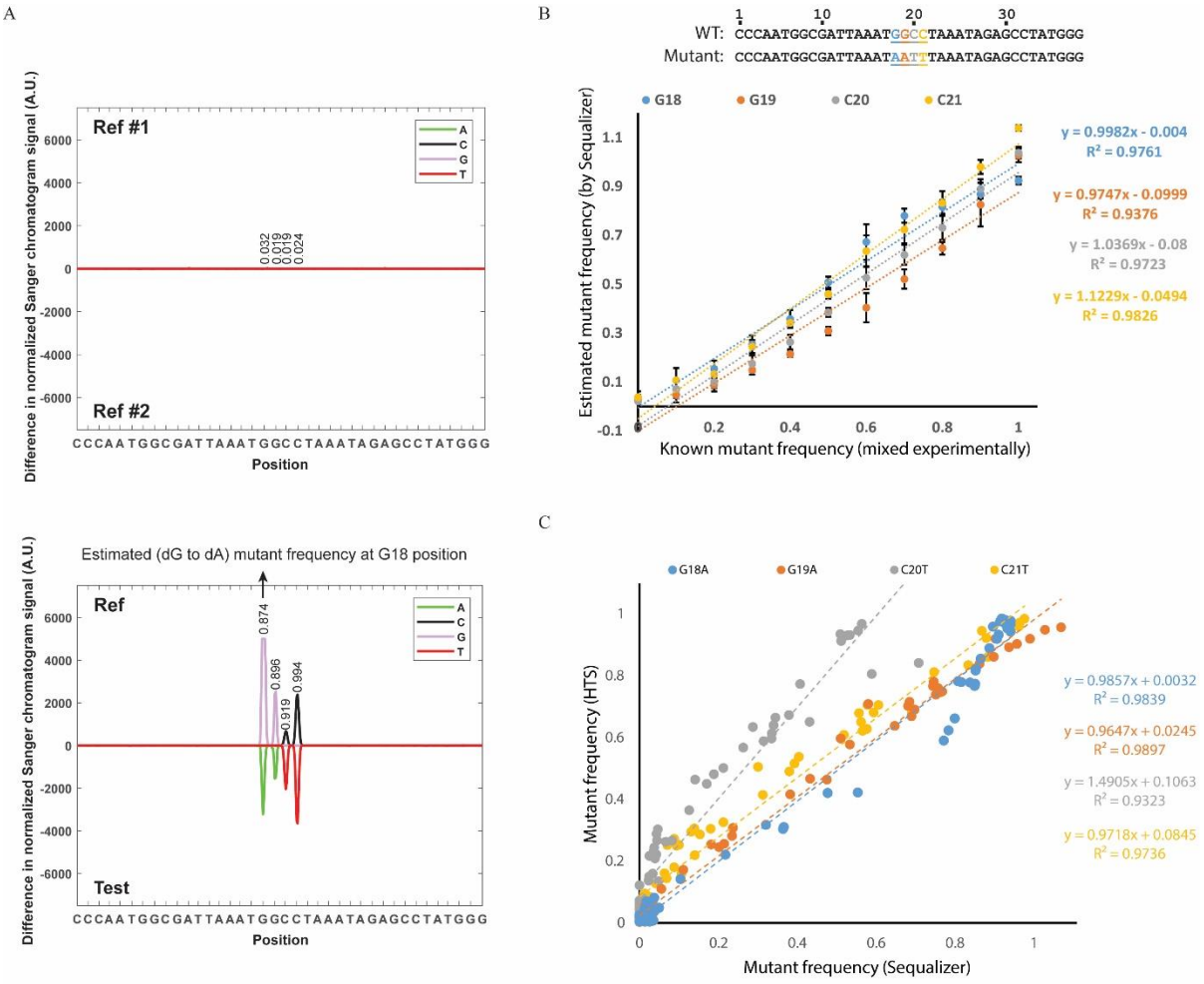


Figure S4.1 | Using Sequalizer to estimate position-specific mutant frequencies from Sanger chromatograms. (A) Sequalizer analysis comparing two instances of WT unmutated (i.e., Ref samples) sequences (top) and a WT unmutated (Ref) sequence vs. Test sample containing a mixture of mutated and unmutated sequences (bottom). The y-axis shows differences between normalized Sanger chromatograms for the samples being compared (Ref #1 vs. Ref #2 or Ref vs. Test). Peaks in these plots indicate differences in the normalized chromatograms and thus mutations in corresponding positions. For example, the peak marked by a black arrow in the bottom plot indicates mutations of dG at position 18 in the Ref to dA in the Test sample. The numbers above target positions (i.e., positions 18-21), show the estimated mutant frequency in that position based on the Sequalizer algorithm, which takes into account the height of Sanger chromatograms in a given position to normalize the calculated difference values. **(B)** Standard curves obtained by analyzing samples containing known mutant ratios by Sequalizer. Two plasmids encoding the pure WT and mutant sequences (as indicated) were mixed at the following mutant:WT ratios: 0:10, 1:9, 2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2, 9:1, and 10:0. The mixtures were Sanger-sequenced and the obtained chromatograms were analyzed by Sequalizer. The estimated mutant frequencies at the four target positions were plotted against the known (i.e., experimentally mixed) mutant ratios. Error bars indicate standard deviation for six independent replicates. **(C)** The position-specific mutant frequencies measured by Sequalizer vs. HTS at four target positions for samples from the experiment described in Fig. 4.1B.

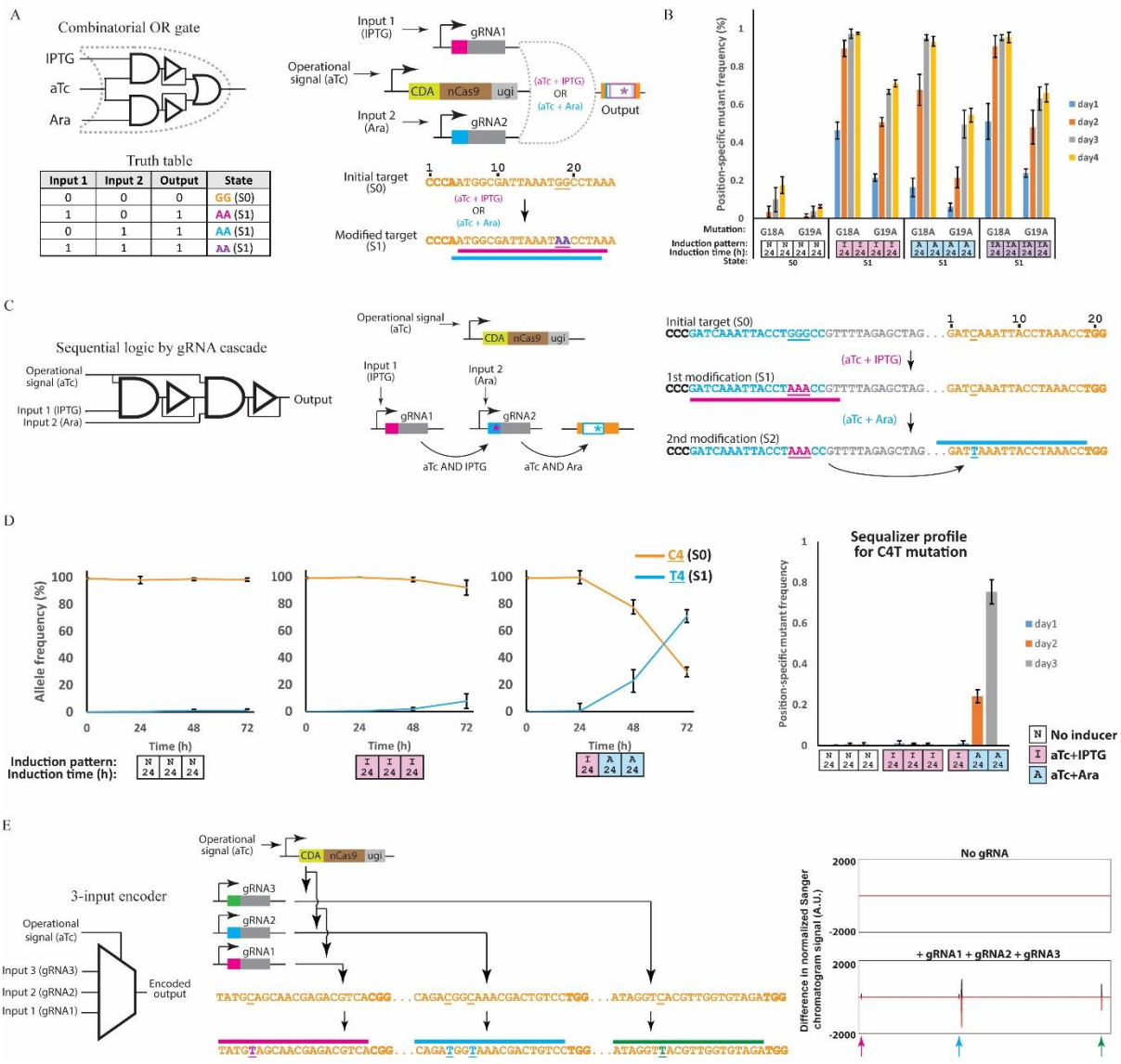


Figure S4.2 | Examples of additional circuits built using DOMINO operators. (A) Schematic representation and truth table for a DOMINO OR gate. **(B)** Sequalizer results for the circuit shown in (A). *E. coli* cells were induced for four days using the indicated patterns and position-specific mutant frequencies were assessed by Sequalizer analysis of Sanger chromatograms. Error bars indicate standard deviation for three biological replicates. **(C)** A sequential AND gate built by a cascade of gRNAs, where the first (IPTG-inducible) gRNA edits and activates a downstream gRNA, which can then edit a downstream target. As demonstrated in this example, gRNA outputs of a DOMINO cascade can be independently regulated by using inducible promoters, such as an Ara-inducible promoter. This offers greater flexibility compared to using mutations as DOMINO outputs (e.g., designs shown in Fig. 2 and 3). **(D)** Dynamics of allele frequencies (i.e., memory states) for the circuit shown in (C) assessed by HTS (left panel) and population-averaged C4T mutation frequency assessed by Sequalizer (right panel). Error bars indicate standard deviation for three biological replicates. **(E)** A multi-input encoder circuit, where the presence of three input gRNAs is converted into *cis*-encoded mutations in the same DNA target locus (*lacZ* ORF in *E. coli* in this case). The circuit can be used to encode multiple transcriptional signals from various loci across a genome into DNA memory within a confined region. The multiplexed and *cis*-encoded signals can then be read and decoded by HTS or Sanger sequencing to reveal information about the original signals.

The plots on the right show the Sequalizer output for cells containing no gRNA (top) and those containing three constitutively expressed input gRNAs (bottom). Mutations in gRNA target sites are reflected as peaks in the bottom Sequalizer plot. This circuit is an example of a DOMINO circuit with more than two inputs, which we envision can be readily extended to additional inputs for *in vivo* memory applications and storing information (spatial, temporal, or artificial) across a genome.

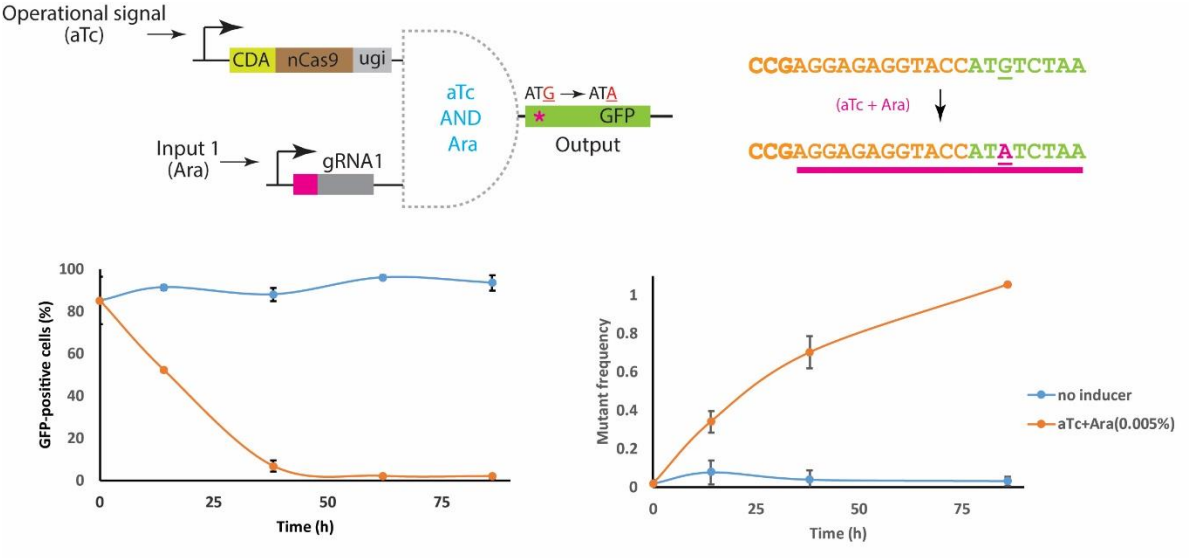


Figure S4.3 | Regulation of gene expression by manipulating functional elements by DOMINO. Conditional conversion of a canonical, efficient initiation codon (ATG) to ATA (which is a non-efficient initiation codon) by an Ara-inducible DOMINO operator was used to down-regulate GFP expression in *E. coli*. Over time, the number of GFP-positive cells decreased and the frequency of mutants increased in induced samples while these quantities minimally changed in non-induced samples. For GFP measurements, samples were grown for six hours in LB with no inducers before flow cytometry to ensure removal of any repression (i.e., CRISPRi) effect enacted by bound CDA-nCas9-ugi. Error bars indicate the standard deviation of three biological replicates.

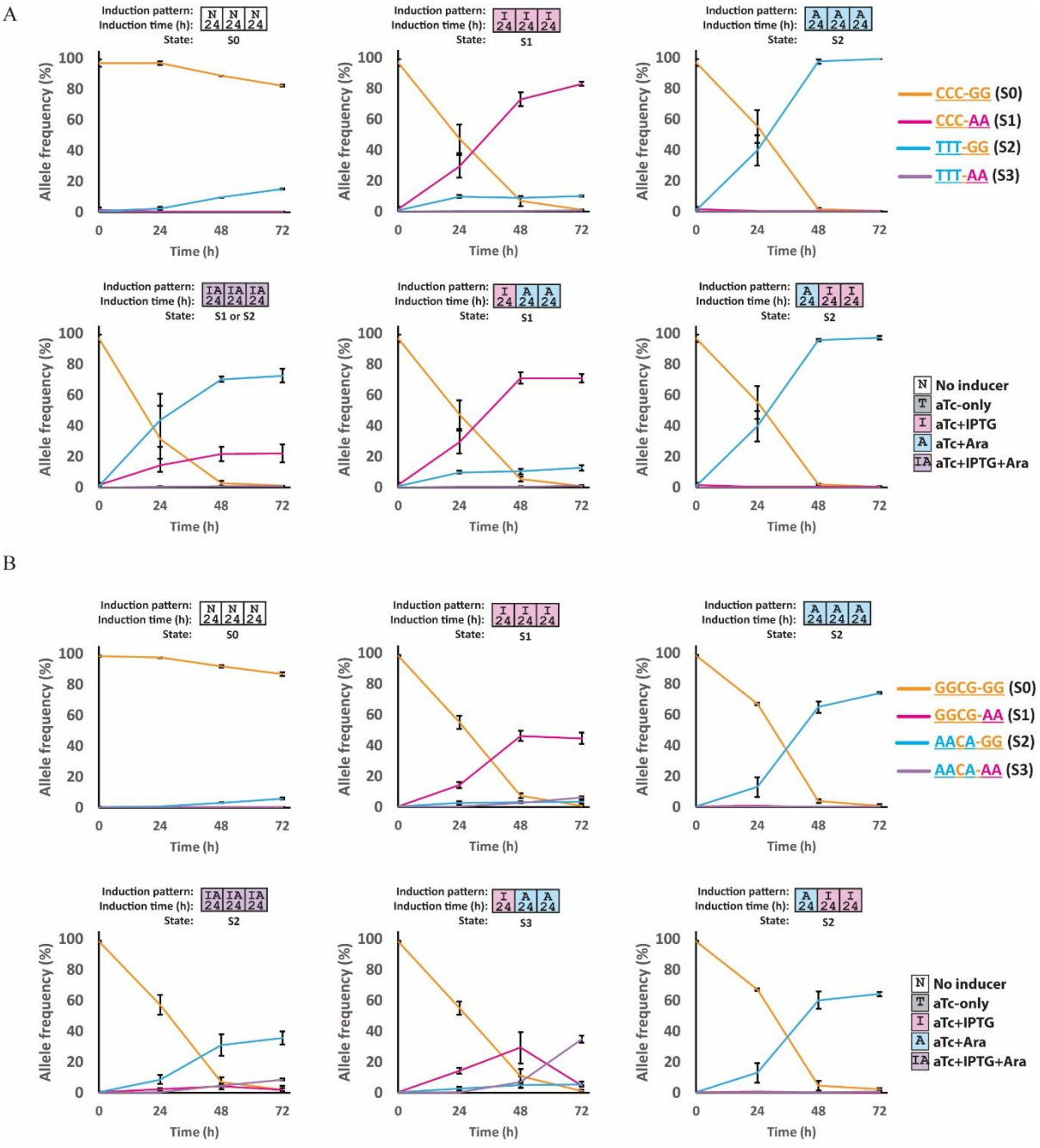


Figure S4.4 | Dynamics of allele frequencies (memory states) for (A) the race-detecting circuit (Fig. 4.2D) and (B) the sequential logic circuit shown in Fig. 4.2E. In each subplot, the dominant allele in the last time point has been used to determine the memory state. Error bars indicate standard deviation for three biological replicates.

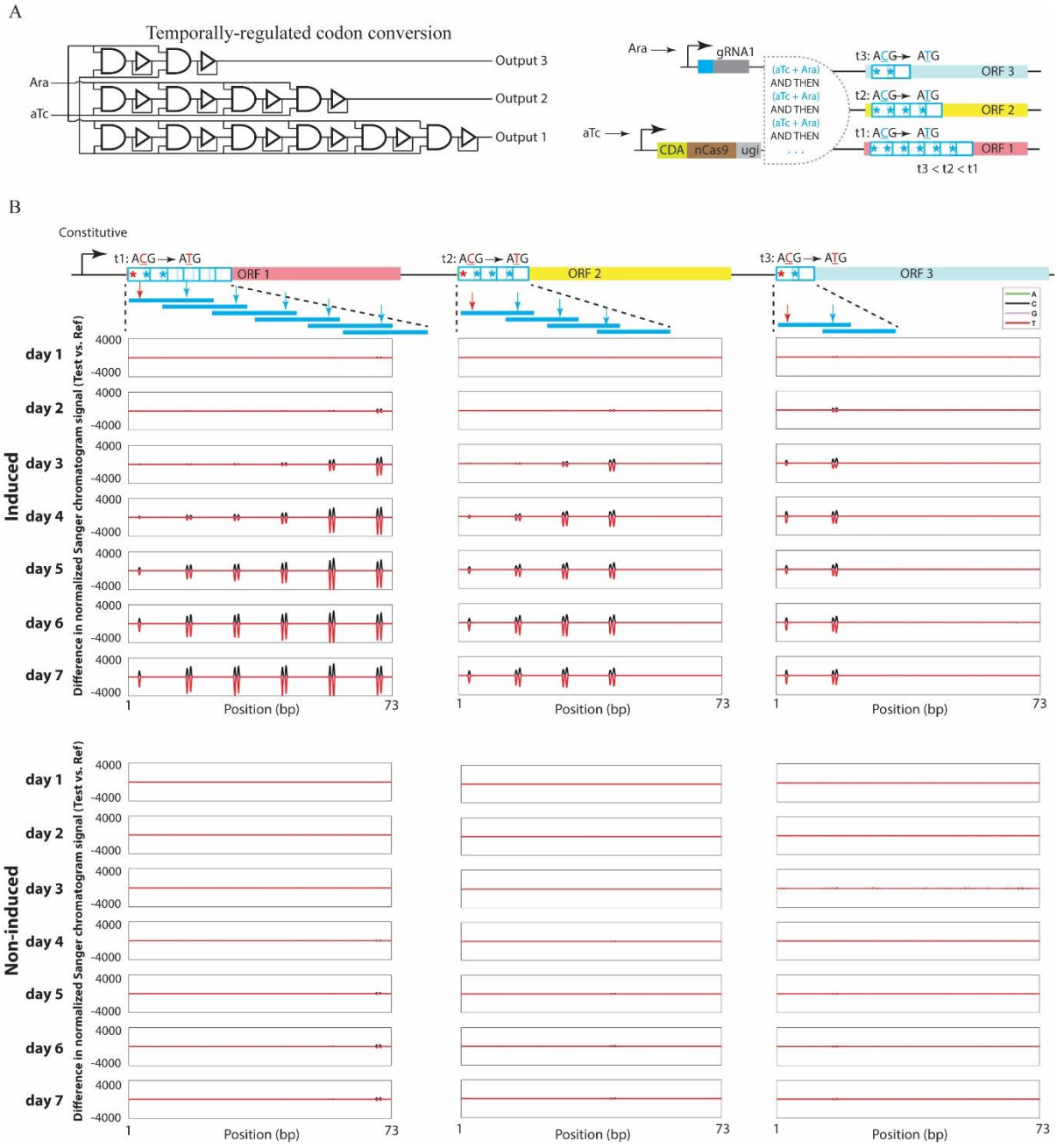


Figure S4.5 | Using DOMINO delay elements to temporally control the conversion of cryptic start codons into canonical start codons in three ORFs. (A) The schematic representation of the time-dependent codon conversion experiment. Three different ORFs with non-canonical (ACG) start codons and different number of delay elements (i.e., overlapping repeats) in their N-termini were placed in a synthetic operon. A gRNA was designed so that it could bind to the 3'-distal repeat element in each array. Sequential recruitment and editing of the repeat elements by this gRNA led to progressive mutation accumulation within the repeat elements toward the 5'-end and eventually editing of the upstream ACG codons to ATG. In this circuit, due to the presence of different number of delay elements in each array, different delay times and thus temporal regulation is achieved. The time required for start codon conversion for ORF 1 (t1) is expected to be longer than the time required for ORF 2 (t2) which itself is expected to be longer than the time required for the conversion in ORF 3 (t3). **(B)** *E. coli* cells harboring the indicated circuit in (A) were induced and then mutation accumulation in the

arrays was monitored by Sanger sequencing and Sequalizer over time. Upon induction of the circuit, time-dependent accumulation of mutations was observed in all the three repeat arrays. The position corresponding to the start codon (shown by red arrow) in the third ORF, which possessed only two repeats in its N-terminus array, was the first that accumulated significant levels of mutations. This was followed by the second ORF, which contained four delay elements and thus experienced a longer delay compared to ORF 3. The first ORF, which possessed six repeats and was thus subject to the longest delay, was the last ORF in which mutations in the position corresponding to the cryptic start codon were accumulated. On the other hand, in non-induced cells, only low levels of mutations accumulated in the downstream repeat of each array and only at the later time points of the experiment, likely due to the background activity of the promoters. Nevertheless, no mutations were detected in positions corresponding to cryptic start codons in non-induced cells.

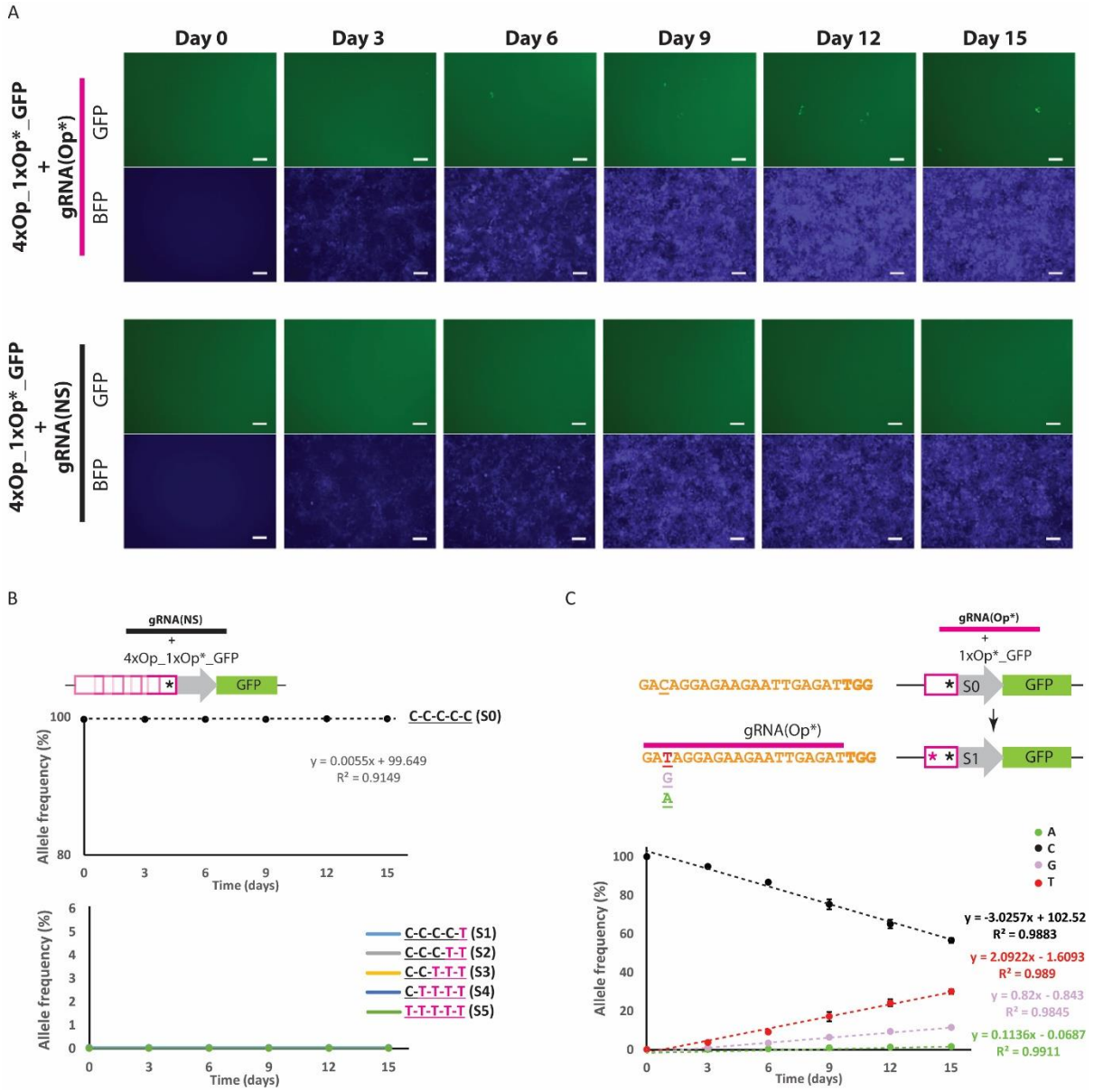


Figure S4.6 | Representative microscopy images and additional data for the experiment shown in Fig. 4. (A) Representative microscopy images for cells harboring the 4xOp_1xOp*_GFP reporter and the Op*-specific gRNA (gRNA(Op*)) or a non-specific gRNA (gRNA(NS)). Scale bars indicate 100 µm. **(B)** Dynamics of allele frequencies (memory states) for cells harboring the 4xOp_1xOp*_GFP reporter and gRNA(NS) (negative control). **(C)** Dynamics of allele frequencies (memory states) for cells harboring the 1xOp*_GFP reporter and gRNA(Op*). The mutable dC residue within the gRNA target site was mutated with a constant rate into dT and constant but lower rates into dG and dA, reflecting the promiscuous repair of deaminated cytidine lesions in mammalian cells. The linear decrease in dC allele frequency, as well as the linear increases in dT, dG, and dA allele frequencies, can be used as an analog readout of gRNA expression duration or intensity.

Table S4.1 | List of the plasmids used in this study.

Name	Plasmid Code	Marker	Used in
P_{tetO}_CDA-nCas9-ugi	pFF1454	Cam	Figs. 4.1-3 & 4.5 Figs. S4.1-S4.5
Comb_AND_gate	pFF1581	Carb	Fig. 4.1B-D
Comb_AND_gate_gRNA_output	pFF1590	Carb	Fig. 4.1E
Seq_AND_gate	pFF1610	Carb	Fig. 4.2A-C
Race_detecting	pFF1684	Cam	Fig. 4.2D Fig. S4.4A
Mixed_seq_logic	pFF1685	Carb	Fig. 4.2E Fig. S4.4B
3x_propagation_delay_seq_AND	pFF1588	Carb	Fig. 4.3
gRNA(Op*)	pYH383	Carb Hygro	Fig. 4.4 Fig. S4.6
gRNA(NS)	pYH384	Carb Hygro	Fig. 4.4 Fig. S4.6
4xOp*_1xOp_GFP_pCMV_nCas9_CDA_ugi_VP64	pYH396	Carb Puro	Fig. 4.4 Fig. S4.6
1xOp*_GFP_pCMV_nCas9_CDA_ugi_VP64	pYH404	Carb Puro	Fig. 4.4 Fig. S4.6
OR_gate	pFF1583	Carb	Fig. S4.2A-B
gRNA_cascade	pFF1586	Carb	Fig. S4.2C-D
Multiplexer	pFF1572	Carb	Fig. S4.2E
Temporal_start_codon_conversion	pFF1573	Carb	Fig. S4.5
ATG_conversion	PFF1604	Carb	Fig. S4.3

Table S4.2 | List of the synthetic parts and their corresponding sequences used in this study.

Part name	Type	Sequence	Source
P_{lacO} ($P_{LlacO-1}$)	IPTG-inducible promoter	AATTGTGAGCGGATAACAATTGACATTGTGAGCGGATAACAAGATACTGAGCACATCAGCAGGACGCACTGACC	(*30*)
P_{tetO}	aTc-inducible promoter	TCCCTATCAGTGATAGAGAAAAGAATTCAAAAGATCTAAAGAGGAGAAAGGATCT	(*30*)
pBAD	Ara-inducible promoter	ACATTGATTATTTGCACGGCGTCACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATA	*E. coli* genome
4xOp_1xOp*	4xOp_1xOp* array upstream of minimal MLP promoter	GACAGGAGAAGAATTGAGACAGGAGAAGAATTGAGACAGGAGAAGAATTGAGACAGGAGAAGAATTGAGACAGGGGGGGCTATAAAAGGGGGTGGGGGCGTTCGTCCTCACTCTAGATCTGCGATCTAAGTAAGCTTGGCATTCCGGTACTGTTGGTAAAGCCACCATGGC	This work
1xOp*	1xOp* upstream of minimal MLP promoter	GACAGGAGAAGAATTGAGATTGGTGGGGGGCTATAAAAGGGGGTGGGGGCGTTCGTCCTCACTCTAGATCTGCGATCTAAGTAAGCTTGGCATTCCGGTACTGTTGGTAAAGCCACCATGGC	This work
pU6	Constitutive RNA Pol III promoter	TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC	
CDA-nCas9-ugi For use in bacterial experiments. The APOBEC1 CDA protein used as the writing module.	read-write head ORF	ATGAGCTCAGAGACTGGCCCAGTGGCTGTGGACCCCACATTGAGACGGCGGATCGAGCCCCATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCCTGCTTTACGAAATTAATTGGGGGGGCCGGCACTCCATTTGGCGACATACATCACAGAACACTAACAAGCACGTCGAAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTCTGTCCGAACACAAGGTGCAGCATTACCTGGTTTCTCAGCTGGAGCCCATGCGGCGAATGTAGTAGGGCCATCACTGAATTCCTGTCAAGGTATCCCCACGTCACTCTGTTTATTTACATCGCAAGGCTGTACCACCACGCTGACCCCCGCAATCGACAAGGCCTGCGGGATTTGATCTCTTCAGGTGTGACTATCCAAATTATGACTGAGCAGGAGTCAGGATACTGCTGGAGAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGGCCTAGGTATCCCCATCTGTGGGTACGACTGTACGTTCTTGAACTGTACTGCATCATACTGGGCCTGCCTCCTTGTCTCCAACATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTCTTGTCATTACCAGCGACTGCCCCCACACATTCTCTGGGCCACCGGGTTGAAAAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTGATAAAAAGTATTCTATTGGTTTAGCCATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCT	(*95*)

		TGTCGAAGAGGACAAGAAACATGAACGGCACCCCAT CTTTGGAAACATAGTAGATGAGGTGGCATATCATGA AAAGTACCCAACGATTTATCACCTCAGAAAAAAGCT AGTTGACTCAACTGATAAAGCGGACCTGAGGTTAAT CTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGG GCACTTTCTCATTGAGGGTGATCTAAATCCGGACAA CTCGGATGTCGACAAACTGTTCATCCAGTTAGTACA AACCTATAATCAGTTGTTTGAAGAGAACCCTATAAA TGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGC CCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATC GCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTC GGTAACCTTATAGCGCTCTCACTAGGCCTGACACCA AATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCC AAATTGCAGCTTAGTAAGGACACGTACGATGACGAT CTCGACAATCTACTGGCACAAATTGGAGATCAGTAT GCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGAT GCAATCCTCCTATCTGACATACTGAGAGTTAATACTG AGATTACCAAGGCGCCGTTATCCGCTTCAATGATCA AAAGGTACGATGAACATCACCAAGACTTGACACTTC TCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAAT ATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGT ACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGG AATTCTACAAGTTTATCAAACCCATATTAGAGAAGA TGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATC GCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACA ACGGTAGCATTCCACATCAAATCCACTTAGGCGAAT TGCATGCTATACTTAGAAGGCAGGAGGATTTTTATC CGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAA TCCTAACCTTTCGCATACCTTACTATGTGGGACCCCT GGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAG AAAGTCCGAAGAAACGATTACTCCATGGAATTTTGA GGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTT CATCGAGAGGATGACCAACTTTGACAAGAATTTACC GAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTA CGAGTATTTCACAGTGTACAATGAACTCACGAAAGT TAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTT TCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCT GTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCA ATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTT CGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATT TAATGCGTCACTTGGTACGTATCATGACCTCCTAAA GATAATTAAAGATAAGGACTTCCTGGATAACGAAGA GAATGAAGATATCTTAGAAGATATAGTGTTGACTCT TACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAG ACTAAAAACATACGCTCACCTGTTCGACGATAAGGT TATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTG GGGACGATTGTCGCGGAAACTTATCAACGGGATAAG AGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCT AAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCA GCTGATCCATGATGACTCTTTAACCTTCAAAGAGGA TATACAAAAGGCACAGGTTTCCGGACAAGGGGACTC ATTGCACGAACATATTGCGAATCTTGCTGGTTCGCC AGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGT AGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAA ACCGGAAAACATTGTAATCGAGATGGCACGCGAAAA TCAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGA GCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACT GGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAA TACCCAATTGCAGAACGAGAAACTTTACCTCTATTAC CTACAAAATGGAAGGGACATGTATGTTGATCAGGAA CTGGACATAAACCGTTTATCTGATTACGACGTCGAT CACATTGTACCCCAATCCTTTTTGAAGGACGATTCAA TCGACAATAAAGTGCTTACACGCTCGGATAAGAACC GAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCG TAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAA ATGCGAAACTGATAACGCAAAGAAAGTTCGATAACT TAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTG	

		ACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAA CCCGCCAAATCACAAAGCATGTTGCACAGATACTAG ATTCCCGAATGAATACGAAATACGACGAGAACGATA AGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGT CAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAAT TCTATAAAGTTAGGGAGATAAATAACTACCACCATG CGCACGACGCTTATCTTAATGCCGTCGTAGGGACCG CACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGT TTGTGTATGGTGATTACAAAGTTTATGACGTCCGTA AGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGG CTACAGCCAAATACTTCTTTTATTCTAACATTATGAA TTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGA GATACGCAAACGACCTTTAATTGAAACCAATGGGGA GACAGGTGAAATCGTATGGGATAAGGGCCGGGACTT CGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGT CAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGG GTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAG TGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCC GAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGC CTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGG AAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATT GGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAA GAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAA GGAAGTAAAAAAGGATCTCATAATTAAACTACCAAA GTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACG GATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAA CGAACTCGCACTACCGTCTAAATACGTGAATTTCCTG TATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCA CCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAG CAGCACAAACATTATCTCGACGAAATCATAGAGCAA ATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGAT GCCAATCTGGACAAAGTATTAAGCGCATACAACAAG CACAGGGATAAACCCATACGTGAGCAGGCGGAAAAT ATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTC CAGCCGCATTCAAGTATTTTGACACAACGATAGATC GCAAACGATACACTTCTACCAAGGAGGTGCTAGACG CGACACTGATTCACCAATCCATCACGGGATTATATG AAACTCGGATAGATTTGTCACAGCTTGGGGGTGACT CTGGTGGTTCTACTAATCTGTCAGATATTATTGAAA AGGAGACCGGTAAGCAACTGGTTATCCAGGAATCCA TCCTCATGCTCCCAGAGGAGGTGGAAGAAGTCATTG GGAACAAGCCGGAAAGCGATATACTCGTGCACACCG CCTACGACGAGAGCACCGACGAGAATGTCATGCTTC TGACTAGCGACGCCCCTGAATACAAGCCTTGGGCTC TGGTCATACAGGATAGCAACGGTGAGAACAAGATTA AGATGCTCTCTGGTGGTTCTCCCAAGAAGAAGAGGA AAGTCTAA	
nCas9-CDA-ugi-VP64 For use in mammalian cell experiments. PmCDA protein (*102*) and minimal VP64 (*103*) domain were used as the write and the transactivation modules, respectively.	read-write-transactivator ORF	ATGGCACCGAAGAAGAAGCGTAAAGTCGGAATCCAC GGAGTTCCTGCGGCAATGGACAAGAAGTACTCCATT GGGCTCGCTATCGGCACAAACAGCGTCGGTTGGGCC GTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAA TTCAAAGTTCTGGGCAATACCGATCGCCACAGCATA AAGAAGAACCTCATTGGCGCCCTCCTGTTCGACTCC GGGGAGACGGCCGAAGCCACGCGGCTCAAAAGAACA GCACGGCGCAGATATACCCGCAGAAAGAATCGGATC TGCTACCTGCAGGAGATCTTTAGTAATGAGATGGCT AAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAG TCCTTTTTGGTGGAGGAGGATAAAAAGCACGAGCGC CACCCAATCTTTGGCAATATCGTGGACGAGGTGGCG TACCATGAAAAGTACCCAACCATATATCATCTGAGG AAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTG CGGTTGATCTATCTCGCGCTGGCGCATATGATCAAA TTTCGGGGACACTTCCTCATCGAGGGGGACCTGAAC CCAGACAACAGCGATGTCGACAAACTCTTTATCCAAC TGGTTCAGACTTACAATCAGCTTTTCGAAGAGAACC CGATCAACGCATCCGGAGTTGACGCCAAAGCAATCC	This work

		TGAGCGCTAGGCTGTCCAAATCCCGGCGGCTCGAAA ACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACG GCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCT GACCCCCAACTTTAAATCTAACTTCGACCTGGCCGAA GATGCCAAGCTTCAACTGAGCAAAGACACCTACGAT GATGATCTCGACAATCTGCTGGCCCAGATCGGCGAC CAGTACGCAGACCTTTTTTTGGCGGCAAAGAACCTG TCAGACGCCATTCTGCTGAGTGATATTCTGCGAGTG AACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGT ATGATCAAGCGCTATGATGAGCACCACCAAGACTTG ACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCT GAGAAGTACAAGGAAATTTTCTTCGATCAGTCTAAA AATGGCTACGCCGGATACATTGATGGCGGAGCAAGC CAGGAGGAATTTTACAAATTTATTAAGCCCATCTTG GAAAAAATGGACGGCACCGAGGAGCTGCTGGTAAAG CTTAACAGAGAAGATCTGTTGCGCAAACAGCGCACT TTCGACAATGGAAGCATCCCCCACCAGATTCACCTG GGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGAT TTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATT GAGAAAATCCTCACATTTCGGATACCCTACTATGTA GGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGG ATGACTCGCAAATCAGAAGAGACCATCACTCCCTGG AACTTCGAGGAAGTCGTGGATAAGGGGGCCTCTGCC CAGTCCTTCATCGAAAGGATGACTAACTTTGATAAA AATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTC TGCTGTACGAGTACTTCACAGTTTATAACGAGCTCA CCAAGGTCAAATACGTCACAGAAGGGATGAGAAAGC CAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCG TGGACCTCCTCTTCAAGACGAACCGGAAAGTTACCG TGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTG AATGTTTCGACTCTGTTGAAATCAGCGGAGTGGAGG ATCGCTTCAACGCATCCCTGGGAACGTATCACGATC TCCTGAAAATCATTAAAGACAAGGACTTCCTGGACA ATGAGGAGAACGAGGACATTCTTGAGGACATTGTCC TCACCCTTACGTTGTTTGAAGATAGGGAGATGATTG AAGAACGCTTGAAAACTTACGCTCATCTCTTCGACG ACAAAGTCATGAAACAGCTCAAGAGGCGCCGATATA CAGGATGGGGGCGGCTGTCAAGAAAACTGATCAATG GGATCCGAGACAAGCAGAGTGGAAAGACAATCCTGG ATTTTCTTAAGTCCGATGGATTTGCCAACCGGAACT TCATGCAGTTGATCCATGATGACTCTCTCACCTTTAA GGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGG GGACAGTCTTCACGAGCACATCGCTAATCTTGCAGG TAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGT TAAGGTCGTGGATGAACTCGTCAAAGTAATGGGAAG GCATAAGCCCGAGAATATCGTTATCGAGATGGCCCG AGAGAACCAAACTACCCAGAAGGGACAGAAGAACAG TAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAA AGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGT TGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCT GTACTACCTGCAGAACGGCAGGGACATGTACGTGGA TCAGGAACTGGACATCAATCGGCTCTCCGACTACGA CGTGGATCATATCGTGCCCCAGTCTTTTCTCAAAGAT GATTCTATTGATAATAAAGTGTTGACAAGATCCGAT AAAAATAGAGGGAAGAGTGATAACGTCCCCTCAGAA GAAGTTGTCAAGAAAATGAAAAATTATTGGCGGCAG CTGCTGAACGCCAAACTGATCACACAACGGAAGTTC GATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCT GAGTTGGATAAAGCCGGCTTCATCAAAAGGCAGCTT GTTGAGACACGCCAGATCACCAAGCACGTGGCCCAA ATTCTCGATTCACGCATGAACACCAAGTACGATGAA AATGACAAACTGATTCGAGAGGTGAAAGTTATTACT CTGAAGTCTAAGCTGGTCTCAGATTTCAGAAAGGAC TTTCAGTTTTATAAGGTGAGAGAGATCAACAATTAC CACCATGCGCATGATGCCTACCTGAATGCAGTGGTA GGCACTGCACTTATCAAAAAATATCCCAAGCTTGAA TCTGAATTTGTTTACGGAGACTATAAAGTGTACGAT	

		GTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATA GGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAAT ATTATGAATTTTTTCAAGACCGAGATTACACTGGCC AATGGAGAGATTCGGAAGCGACCACTTATCGAAACA AACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGT AGGGATTTCGCGACAGTCCGGAAGGTCCTGTCCATG CCGCAGGTGAACATCGTTAAAAAGACCGAAGTACAG ACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAA AGGAACAGCGACAAGCTGATCGCACGCAAAAAAGAT TGGGACCCCAAGAAATACGGCGGATTCGATTCTCCT ACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTG GAGAAAGGGAAGTCTAAAAAACTCAAAAGCGTCAAG GAACTGCTGGGCATCACAATCATGGAGCGATCAAGC TTCGAAAAAAACCCCATCGACTTTCTCGAGGCGAAA GGATATAAAGAGGTCAAAAAAGACCTCATCATTAAG CTTCCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCC GGAAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGA AAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTA ATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAA AGGGTCTCCCGAAGATAATGAGCAGAAGCAGCTGTT CGTGGAACAACACAAACACTACCTTGATGAGATCAT CGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCT CGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTA CAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGC AGAAAACATTATCCACTTGTTTACTCTGACCAACTTG GGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACC ATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTC CTGGACGCCACACTGATTCATCAGTCAATTACGGGG CTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGT GGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAG GTGGGTGGAGGAGGTACCGGCGGTGGAGGCTCAGCA GAATACGTACGAGCTCTGTTTGACTTCAATGGGAAT GACGAGGAGGATCTCCCCTTTAAGAAGGGCGATATT CTCCGCATCAGAGATAAGCCCGAAGAACAATGGTGG AATGCCGAGGATAGCGAAGGGAAAAGGGGCATGATT CTGGTGCCATATGTGGAGAAATATTCCGGTGACTAC AAAGACCATGATGGGGATTACAAAGACCACGACATC GACTACAAAGACGACGACGATAAATCAGGGATGACA GACGCCGAGTACGTGCGCATTCATGAGAAACTGGAT ATTTACACCTTCAAGAAGCAGTTCTTCAACAACAAGA AATCTGTGTCACACCGCTGCTACGTGCTGTTTGAGT TGAAGCGAAGGGGCGAAAGAAGGGCTTGCTTTTGGG GCTATGCCGTCAACAAGCCCCAAAGTGGCACCGAGA GAGGAATACACGCTGAGATATTCAGTATCCGAAAGG TGGAAGAGTATCTTCGGGATAATCCTGGGCAGTTTA CGATCAACTGGTATTCCAGCTGGAGTCCTTGCGCTG ATTGTGCCGAGAAAATTCTGGAATGGTATAATCAGG AACTTCGGGGAAACGGGCACACATTGAAAATCTGGG CCTGCAAGCTGTACTACGAGAAGAATGCCCGGAACC AGATAGGACTCTGGAATCTGAGGGACAATGGTGTAG GCCTGAACGTGATGGTTTCCGAGCACTATCAGTGTT GTCGGAAGATTTTCATCCAAAGCTCTCATAACCAGCT CAATGAAAACCGCTGGTTGGAGAAAACACTGAAACG TGCGGAGAAGTGGAGATCCGAGCTGAGCATCATGAT CCAGGTCAAGATTCTGCATACCACTAAGTCTCCAGCC GTTGGTCCCAAGAAGAAAAGAAAAGTCGGTACCATG ACCAACCTTTCCGACATCATAGAGAAGGAAACAGGC AAACAGTTGGTCATCCAAGAGTCGATACTCATGCTT CCTGAAGAAGTTGAGGAGGTCATTGGGAATAAGCCG GAAAGTGACATTCTCGTACACACTGCGTATGATGAG AGCACCGATGAGAACGTGATGCTGCTCACGTCAGAT GCCCCAGAGTACAAACCCTGGGCTCTGGTGATTCAG GACTCTAATGGAGAGAACAAGATCAAGATGCTATCT GGTGGTTCTCCCAAGAAGAAGAGGAAAGTCGAGGAT CCAAAGAAGAAAAGGAAGGTTGAAGACCCCAAGAAA AAGAGGAAGGTGGATGGGATCGGCTCAGGCAGCAAC GGCGGTGGAGGTTCAGACGCTTTGGACGATTTCGAT	

		CTCGATATGCTCGGTTCTGACGCCCTGGATGATTTC GATCTGGATATGCTCGGCAGCGACGCTCTCGACGAT TTCGACCTCGACATGCTCGGGTCAGATGCCTTGGAT GATTTTGACCTGGATATGCTC	

Table S4.3 | List of HTS primers and their corresponding sequences used in this study.

name	Type	Sequence	Used in
FF_oligo_2525	HTS_Primer_Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTGCTGCCCGACAACCACTA	Figs. 4.1C, 4.3C, S4.1C, S4.2D, S4.4
FF_oligo_2526	HTS_Primer_Reverse	CGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNTGAACAACCACCACTTCAAGTGGG	Figs. 4.1C, 4.3C, S4.1C, S4.2D, S4.4
FF_oligo_2527	HTS_Primer_Forward	CACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGGACAGCAGAGATCCAGTTTGGT	Figs. 4.4 & S4.6
FF_oligo_2528	HTS_Primer_Reverse	GGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNTCGCAGATCTAGAGTGAGGACGAAC	Figs. 4.4 & S4.6

Chapter 5: ENGRAM

High-Capacity DNA Recorder for Concurrent Recording of Analog Information and Chronicle of Molecular Events into DNA

5.1 Abstract

Existing DNA recording technologies are not suitable for long-term and high-capacity recording of biological information. To address these limitations, we developed a compact, high-capacity, and long-term DNA memory architecture that can continuously record the analog properties of desired signals as well as the chronicle of events (lineages) produced by those signal over many generations. We envision that this DNA recording technology, called ENGRAM for ENGineered Random Accumulative Memory, to be especially useful in scenarios that demand extensive memory states such as high-resolution lineage tracing applications.

5.2 Introduction

Existing Cas9-based recording technologies (*115, 117*) rely on stochastic DNA memory states resulting from indels generated by double-strand DNA breaks. These recorders lose their recording capacity after one or a few recording events due to deletions and loss of gRNA target sites and are therefore not ideal for long-term recording of event histories and generating high-resolution cellular lineages. To address some of these problems, our lab previously described the mSCRIBE system (*78*), an engineered self-targeting gRNA (stgRNA) that could recruit Cas9 to its own encoding locus and execute cycles of double-strand break generation and successive indel formation by the Non-Homologous End Joining (NHEJ) pathway. However, due to the prevalence of deletions as a product of NHEJ, these recorders could exhaust their recording capacity due to deletions in the stgRNA handle. Furthermore, new mutations could destroy the previous mutations (i.e., overwrite the previous memory states), which makes deducing lineage histories from these stochastically generated memory states challenging.

DOMINO circuits that rely on deterministic DNA modifications are useful when transitions between a handful of memory states are desired. The autonomous and continuous nature of these DNA writers is especially useful for building long-term DNA

recorders to study signaling dynamics and event histories in their native contexts. However, for some applications, such as lineage tracing, the number of memory states needed to record event histories with high resolution could be orders of magnitude higher than what can be practically achieved by deterministic DNA mutations. Although the memory capacity of DOMINO circuits can be increased by incorporating multiple gRNAs or by increasing the number of repeats in DOMINO arrays, these designs are still not as compact as they could be and may require encoding large numbers of memory registers using dozens of gRNAs and/or hundreds and thousands of bps of DNA.

To address these limitations, we developed a sequential mutation accumulation strategy that can be used to build long-term, autonomous, and minimally disruptive molecular recorders in a compact, and high-capacity memory architectures. In this strategy, named ENGRAM for ENGineered Random Accumulative Memory, the CDA-nCas9-ugi read-write head continuously incorporates position-specific but pseudo-random mutations into a (C-rich) stgRNA locus as a function of time and duration of stgRNA expression (Fig. 5.1A). Mutation accumulation in the stgRNA memory register can be coupled to signals of interest by placing stgRNA expression under the control of the corresponding signal. The degree to which mutations accumulate in this memory register can then be read out by HTS and used to deduce signaling dynamics of the original signal.

5.3 Results

To demonstrate this concept, we placed a C-rich stgRNA (43 bp SDS with 34 dC residues) under the control of an Ara-inducible promoter (Fig. 5.1A) and transformed this construct into *E. coli* cells harboring an aTc-inducible CDA-nCas9-ugi plasmid. We then grew the transformants in the presence or absence of aTc and different concentrations of Ara for multiple cycles with serial dilutions. Mutation accumulation in the stgRNA locus was monitored over the course of the experiment. As shown in Fig. 5.1B, the frequency of mutant alleles in the populations increased in a time- and Ara-dosage-dependent manner, indicating that these recorders are capable of recording analog information in a continuous fashion.

The unidirectional and minimally disruptive nature of CDA-mediated mutations generated by these recorders ensures that previous mutations (i.e., memory states) are preserved after each editing step (Fig. 5.1C). The pseudo-random yet position-specific mutations in locations corresponding to dC residues of the stgRNA memory register can be considered as discrete memory states of the register. Accumulation of mutations

in the stgRNA locus can be thus considered as transitions between memory states. The memory capacity of these recorders is basically the number of memory states, which can be exponentially increased by increasing the number of dC residues within the stgRNA locus. These features make the mutation profiles generated by these recorders especially useful for investigating cellular event histories and lineages in an autonomous and high-resolution fashion. Fig. 5.1D shows an example of a lineage map generated for one of the samples (36 h induction with aTc + Ara (0.2%)) in the experiment described in Fig. 5.1B. We could detect more than 1000 discrete memory states (unique mutations) in the 43 bps stgRNA memory register.

Further analysis of these samples revealed that samples with similar fractions of non-mutated stgRNA (state S0), often had a similar distribution of mutated alleles (states >S0) (Fig. 5.2). This suggests that the average rate of transitions between memory states depends on the allele frequencies in the current state and not the input history. In other words, if a sample that has been induced with a high concentration of the input for a short time and a sample that has been induced with a low concentration of the input for a long time have similar frequencies of the unmutated allele (S0), they are very likely to have similar distributions of mutant allele frequencies. This suggests that while at the single-molecule level any transitions may occur randomly from a lower memory state (less mutation) to a higher memory state (more mutations) with some non-zero probability, at the population level, these transitions are more deterministic and are defined by the frequency of each memory state within the population.

We call this DNA memory architecture, that operates in a distinct probabilistic fashion, ENGRAM (ENGineered Random Accumulative Memory) to distinguish it from the deterministic DOMINO operators. While the memory states and orders of state transitions can be accurately designed and predicted in DOMINO-based memory registers, the exact transitions between memory states in ENGRAM registers are unpredictable and probabilistic. In ENGRAM registers, at the single molecule level each possible transition (i.e., from a lower memory state to a higher memory state) is likely to happen with some probability, however, at the population level, transitions are likely to be statistically predictable (Fig. 5.2) and are thus pseudo-random.

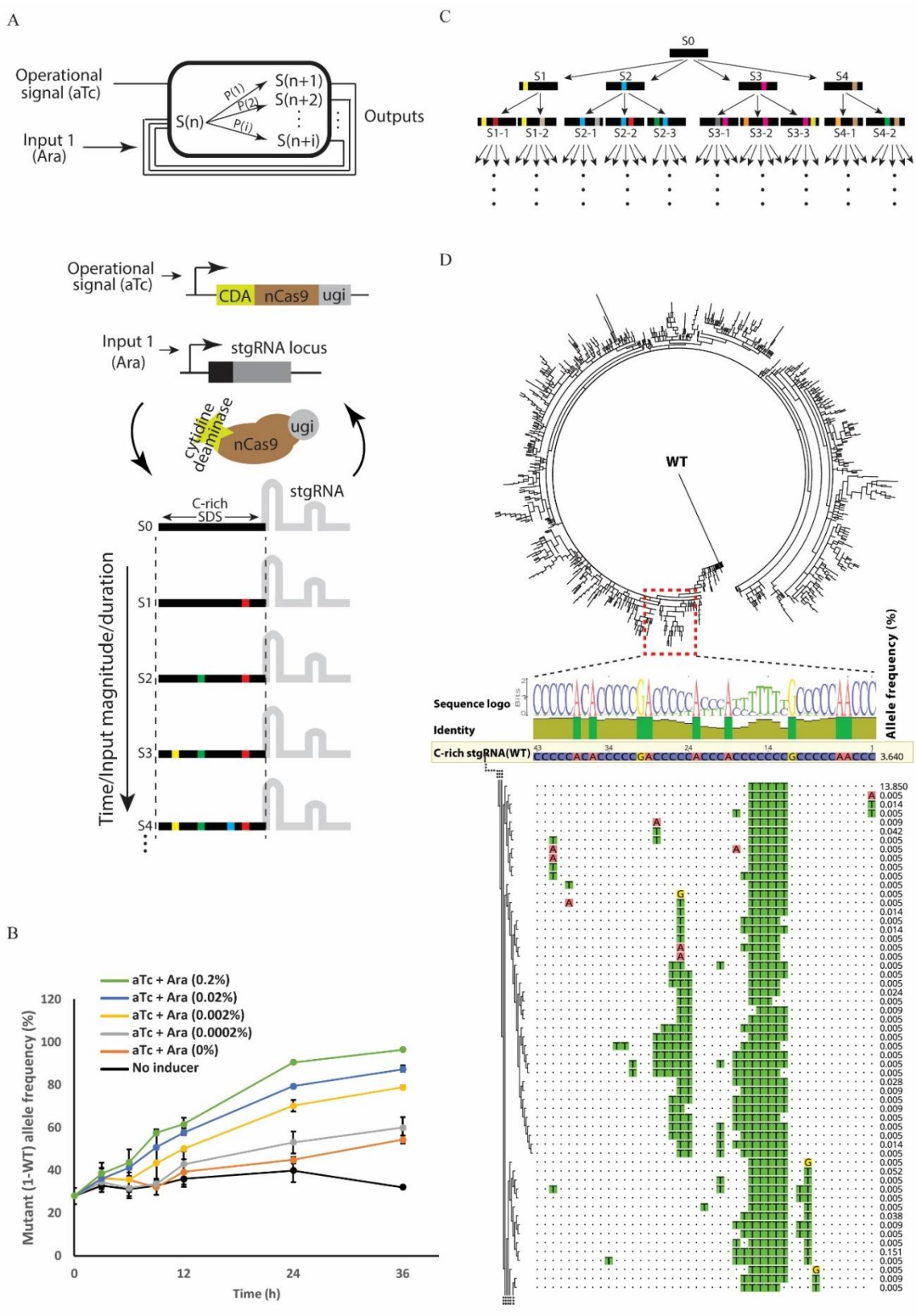


Figure 5.1 | High-capacity, continuous, and long-term ENGRAM recorders for memorizing analog signals and chronicling molecular events. (A) Schematic representation of the ENGRAM high-capacity molecular recording system. A self-targeting gRNA (stgRNA) with a

43-bp C-rich SDS was placed under the control of a desired input. Once expressed, the stgRNA directs the DNA read-write head to its own locus, resulting in dC to dT (and with a lower frequency to dG and dA) mutations that accumulate in the stgRNA locus as a function of duration and magnitude of the signal controlling the gRNA expression. In this design, transitions between memory states are pseudo-random but accumulative, and always occur from a lower memory state (i.e., lower degree of mutations, S(n)) to a higher memory state (i.e., a higher degree of mutations, S(n+i)). **(B)** *E. coli* cells with the circuit shown in (A) were induced with aTc and different concentrations of Ara as indicated, and grown for 36 hours with dilution every 12 hours. Samples were taken at different time points throughout the experiment and assessed for allele frequencies by HTS. The frequency of mutants in the population increased continuously in a time- and Ara dosage-dependent manner, demonstrating that the recorder can continuously record analog information of an incoming signal. **(C)** Unidirectional and pseudo-random mutations that accumulate in the specific positions (i.e., dC residues) within an stgRNA memory register can be considered as non-disruptive and probabilistic transitions between memory states. These mutations (i.e., memory states) can be used to trace back mutation trajectories and cellular lineages. **(D)** An example of a high-resolution cellular lineage generated from the samples shown in (B) (36 h induction, aTc + 0.2% Ara). Positions with the same sequence as the WT stgRNA allele are indicated by dots.

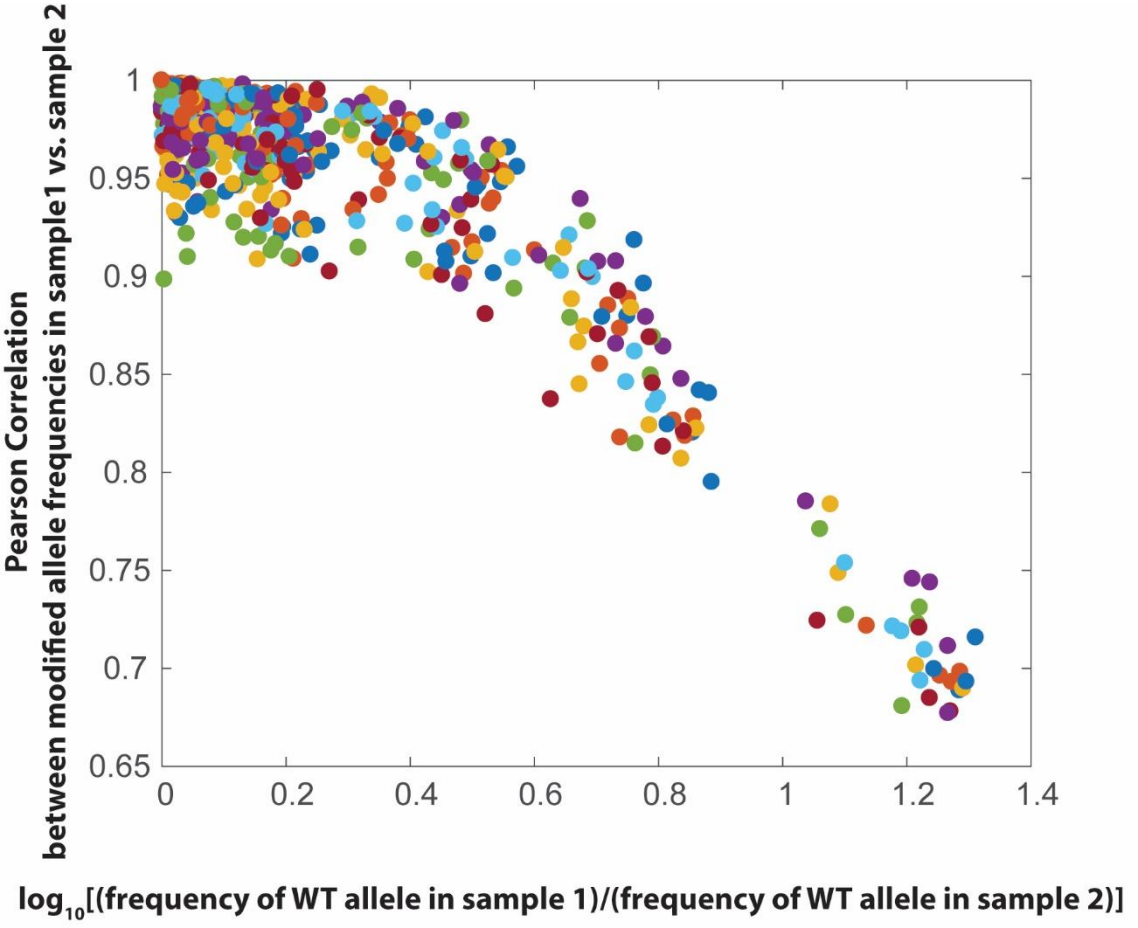


Figure 5.2 | Pearson correlation between frequencies of modified alleles in different samples (obtained from the experiment described in Fig. 5.1), plotted against the ratios of WT (S0) allele frequencies in the corresponding samples. Samples with similar frequencies of the WT allele (x-axis value close to 0) showed high correlation between their frequencies of mutant

alleles as well, independent of their input histories. This was true even for samples that were induced for a long time with a low concentration of the input (Ara) compared with those that were induced for a short time with a high concentration of the input. This suggests that transitions between states are independent of input histories, and depends on the allele frequencies in the current state.

5.4 Discussion

The ENGRAM platform overcomes prior limitations with CRISPR-nuclease-based lineage tracing strategies by providing autonomous, compact, and minimally disruptive DNA writer that continuously and probabilistically writes new, position-specific and pseudo-random mutations on top of a targeted stgRNA barcode. We envision that ENGRAM recorders, which share many features with DOMINO recorders, and at the same time offer higher recording capacities, will be especially useful in generating high-resolution map of cellular lineages. The promiscuous repair of dC lesions in mammalian cells could be especially beneficial for lineage tracking applications, as it can increase the number of potential memory states. Moreover, signal-responsive lineage maps with tunable resolution can be generated with ENGRAM because the activity of the recorder can be modulated by internal or external signals of interest. Combining these recorders with single-cell sequencing and more advanced barcoding schemes, as well as future development of this recording technology in mammalian cells, could pave the way to high-resolution maps of cellular lineages and other applications that require high-density memory storage capacities in living cells.

5.5 Supplementary Information

The methods, DNA constructs and strains described in Chapter 4 were used in the experiments described in this chapter. Additional constructs used in this chapter are described in the following tables.

Table S5.1 | List of the plasmids used in this study.

Name	Plasmid Code	Marker	Used in
P_{BAD}_C-rich_stgRNA	pFF1531	Carb	Figs. 5.1 & 5.2

Table S5.2 | List of the synthetic parts and their corresponding sequences used in this study.

Part name	Sequence	Source
C-rich_stgRNA	CCCCCACACCCCCGACCCCCACCCACCCCCCCGCCCCCAACCC	This study
stgRNA_handle	GGGTTAGAGCTAGAAATAGCAAGTTAACCTAAGGCTAGTCCG TTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTT	(*78*)

Table S5.3 | List of HTS primers and their corresponding sequences used in this study.

name	Type	Sequence	Used in
FF_oligo_2399	HTS_Primer _Forward	ACACTCTTTCCCTACACGACGCTCTTCCGAT CTNNNNNTTT TAT CGCAACTCTCTACTGTTT	Fig. 5.1
FF_oligo_2124	HTS_Primer _Reverse	GGCATTCCTGCTGAACCGCTCTTCCGATCTN NNNNTTCAAGTTGATAACGGACTAGCCTT	Fig. 5.1

Chapter 6: DRIVE

6.1 Abstract

Evolution is a continuous process of genetic diversification and phenotypic selection that tunes the genetic makeup of living organisms and maximizes their fitness in a given environment. Genomic DNA is an evolvable functional memory that records history of these adaptive changes over evolutionary timescales. Efficient DNA writers, such as HiSCRIBE and DOMINO, could enable the autonomous, continuous and targeted diversification of desired loci *in vivo* in a temporally- and spatially-programmable manner, without a concomitant increase in the global mutation rate. HiSCRIBE DNA writers are especially useful to introduce predesigned diversity into cells while DOMINO writers are useful to introduce de novo diversity in target loci of interest. These targeted diversity generation strategies could be coupled with a continuous selection or screening setup to achieve adaptive writing and tune cellular fitness continuously and autonomously with minimal human intervention, in an approach we call DRIVE for Directed and Recurring *In Vivo* Evolution. We demonstrate that the DRIVE platform to be especially useful to optimize cellular traits of interest and engineer cells that can undergo synthetic Lamarckian evolution and tune certain segments of their genome with rates much faster by Darwinian evolution.

6.2 Introduction

Evolutionary engineering of cellular phenotypes, in the form of genetic diversification and phenotypic selection, is a powerful approach for engineering living systems. However, in many cases, natural mutation rates are not high enough to allow desirable genetic changes to be accessible on practical timescales. Simply elevating the global mutation rate is an inefficient strategy to optimize cellular fitness, as infrequent beneficial mutations are often masked by much more frequent deleterious ones, an issue that increases with genome size (*119, 120*). Targeted diversity generation strategies, such as Multiplex Automated Genome Engineering (MAGE) (*62*), address this limitation by localizing diversifications to loci of interest while minimizing the rate of unwanted global mutation. However, MAGE relies on high-efficiency transformation protocols and require human intervention and thus cannot be easily scaled up, applied to organisms with low transformation efficiencies, or coupled to continuous selection or screening. Furthermore, in these methods DNA writing events cannot be linked to cellular regulatory networks, thus they cannot be used for continuous and dynamic

manipulation of cellular phenotypes and tuning of cellular phenotypes. additionally, these methods still rely on *in vitro*-generated diversity, and thus cannot be used to engineer for continuous rounds of evolutions. As evolutionary success depends on the number of evolutionary cycles performed, platforms that enable parallel and continuous rounds of evolution are highly desirable for engineering cellular phenotypes. Continuous *in vivo* evolution techniques like Phage-Assisted Continuous Evolution (PACE) (*49*) enable dozens of rounds of evolution to be performed without the need for human intervention. PACE is powerful but is only applicable to the evolution of phenotypes that can be encoded within a phage genome and connected to protein expression in hosts with well-characterized phages. None of these methods can be used to engineer self-evolvable cells that could autonomously tune their evolvability and increase their fitness with rates higher than possible by natural evolution

Genomic DNA is the ultimate storage medium for life. The information stored in this medium is mainly written, rewritten and scoured by Darwinian evolution forces over evolutionary timescales. However, in certain cases, where the rate of Darwinian evolution is not enough to adapt and cope with treat of ever-changing an environment, living cells have evolved mechanisms to selectively elevate mutation rate in specific segments of their genome, to evolve faster than possible by natural Darwinian evolution. The ability to selectively increase *de novo* mutation rates of specific genomic segments related to a phenotype of interest without a concomitant increase in the global mutation rate could provide selective advantages for an adapting organism. For example, cells may use targeted diversity generation mechanisms to tune their evolvability (*121*) in uncertain environments or in the arms race between fast evolving parasites or viruses. There are examples of targeted diversity-generating mechanisms in living cells, including diversity-generating retroelements in phages and bacteria (*122*), phase and antigenic variation mechanisms in some pathogenic bacteria (*123-125*), CRISPR adaptation in bacteria (*126-129*), and the adaptive immune system in higher vertebrates (*129, 130*). These mechanisms can be all considered as examples of natural Lamarckian evolution that act at the molecular level. Endowing living cells with a synthetic ability to undergo Lamarckian evolution to tune the evolvability of specific segments of their genome could have a great potential for studying and evolutionary engineering of these biological machines. However, the abovementioned systems require *cis*-regulatory elements for targeting, are confined to specific genes and specific organisms, and currently are not amenable to engineering to be redirected to desired targets.

Here, we demonstrate that when combined with a continuous delivery system and appropriate selections or screens, HiSCRIBE DNA writers enable the continuous optimization of a trait of interest. We further show that the system can be used to selectively increase the *de novo* mutation rate of desired genomic loci while minimizing the background mutation rate, as opposed to using a generalized hypermutator phenotype, thus allowing one to tune the evolvability of specific genomic segments.

Finally, we demonstrate that in addition to the rational implementation of logic and memory, DOMINO writers can be used to introduce *de novo* targeted diversity to desired genomic segments. Under a selective pressure, this could result in an increase in fitness and evolution much faster than possible by natural Darwinian evolution. Thus, this type of continuous *de novo* targeted diversity generation and adaptation strategy in the presence of a selective pressure can be considered as a form of synthetic molecular Lamarckian evolution. This strategy, which we refer to as DRIVE, could be especially useful in tuning evolvability of living cells and evolutionary engineering of cellular phenotypes.

6.3 Results

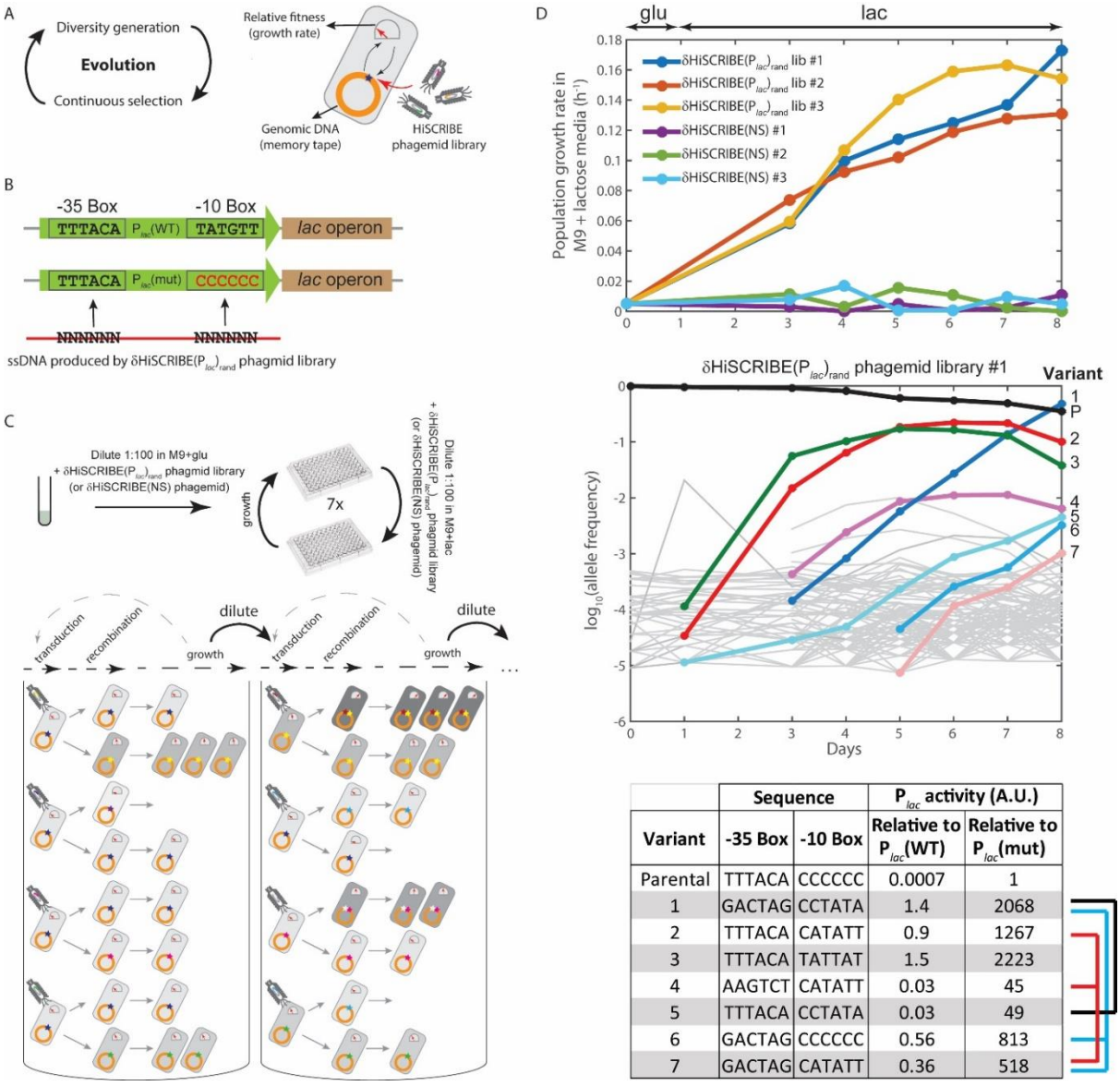
Continuous *in vivo* Evolution

The precise DNA writing enabled by HiSCRIBE offers the capacity to introduce targeted mutations into desired genome loci. This in vivo targeted diversification can be combined with an appropriate continuous selections or screens thus enabling to perform continuous optimization of traits of interest (Fig. 6.1A). To demonstrate this with HiSCRIBE DNA writers, we linked cellular fitness (i.e., growth rate) to a cell's ability to consume lactose (lac) as the sole carbon source. To enable a wide dynamic range in fitness to be explored, we first weakened the activity of the native *lac* operon promoter (P_{lac}) by introducing mutations into its -10 box (P_{lac}(mut), Fig. 6.1B) in the MG1655 *exo*$^-$ strain. Cells with the P_{lac}(mut) promoter (hereafter referred to as the parental strain) grew poorly in minimal media (M9) when lactose was present as the sole carbon source. We then used a randomized δHiSCRIBE phagemid library (δHiSCRIBE$(P_{lac})_{rand}$) to continuously introduce diversity into the -10 and -35 sequences of this promoter (Fig. 6.1B). Starting from an overnight culture, parental cells were diluted into M9 + glucose media and divided into two groups, which were then treated with phagemid particles from either a δHiSCRIBE$(P_{lac})_{rand}$ library or δHiSCRIBE(NS). After this initial growth in glucose, cells were diluted and regrown in M9 + lactose in the presence of phagemid particles for six additional rounds to allow for concomitant diversification, selection, and propagation of beneficial mutations (Fig.

6.1C, see Supplementary Methods). As shown in Fig. 6.1D (top panel), the overall growth rates of cell populations in lactose increased when they were transduced with the δHiSCRIBE(P_{lac})$_{rand}$ phagemid library. In contrast, the growth rates of cell populations exposed to the control δHiSCRIBE(NS) phagemid particles did not change over time. These results demonstrate that the δHiSCRIBE library can introduce targeted diversity into desired loci (-10 and -35 boxes of the P_{lac} promoter) that result in fitness increases of the population under selection over relatively short timescales, and much faster than what can be achieved by natural Darwinian evolution (i.e., in cells transformed with non-targeting δHiSCRIBE(NS)).

To monitor the dynamics of mutants in these cultures, we amplified the P_{lac} region by PCR and performed deep sequencing at different time points over the course of the experiment. The diversity and frequency of P_{lac} alleles in samples that had been exposed to the δHiSCRIBE(NS) phagemid did not change significantly over time and the parental allele comprised ~100% of the population at all analyzed time points (Fig. S6.1A and S6.1B). Further inspection of the rare variants observed in these samples revealed mostly single nucleotide changes compared to the parental allele, suggesting that these arose from sequencing errors. On the other hand, the diversity of P_{lac} alleles greatly increased in cultures that were exposed to the δHiSCRIBE(P_{lac})$_{rand}$ phagemid library when they were initially grown in the M9 + glucose condition (Fig. S6.1A). This initial increase in allele diversity was followed by a significant drop upon dilution of cells in lactose media, likely due to sampling drift and strong selection for alleles that allow for lactose metabolism. Throughout the experiment, however, the number of unique variants remained significantly higher in the δHiSCRIBE(P_{lac})$_{rand}$ cultures than in the negative controls. Moreover, the frequency of P_{lac} alleles from samples that had been exposed to δHiSCRIBE(P_{lac})$_{rand}$ changed dynamically over time (Fig. 6.1D, middle panel). Notably, by the end of the experiment, the frequency of the parental allele dropped to less than 50% and one variant (variant #1) became the dominant allele in the population. Further analysis of frequent variants within the diversified population indicated that multiple mutations occurred in the -10 and -35 boxes in discrete steps, in which secondary mutations arising on top of primary mutations led to an increase in fitness (Fig. 6.1D, bottom panel). For example, based on allele enrichment and P_{lac} activity data (see below), the dominant allele (variant #1) was likely produced from an initial, less active mutant (variant #5) and subsequently took over the population based on increased fitness (i.e., P_{lac} activity). The sequences of successful variants that evolved in our experiments were especially AT-rich (Fig. 6.1D,

bottom panel, and Fig. S6.1C), as is expected from the canonical sequences of these regulatory elements in *E. coli*.



Variant	Sequence		P_{lac} activity (A.U.)	
	-35 Box	-10 Box	Relative to P_{lac}(WT)	Relative to P_{lac}(mut)
Parental	TTTACA	CCCCCC	0.0007	1
1	GACTAG	CCTATA	1.4	2068
2	TTTACA	CATATT	0.9	1267
3	TTTACA	TATTAT	1.5	2223
4	AAGTCT	CATATT	0.03	45
5	TTTACA	CCTATA	0.03	49
6	GACTAG	CCCCCC	0.56	813
7	GACTAG	CATATT	0.36	518

Figure 6.1 | Continuous evolution of a desired genomic locus via HiSCRIBE. (A) diversity generation enabled by HiSCRIBE can be coupled to continuous selection to accelerate the rate of evolution of desired target sites. A randomized δHiSCRIBE library was encoded on phagemids that were continuously delivered into cells. In the presence of a selective pressure, δHiSCRIBE-mediated mutations lead to adaptive genetic changes that increase fitness. An increase in fitness results in faster replication and amplification of the associated genotype, increasing the chance that cells containing the genotype can undergo additional rounds of diversification. **(B)** The sequences of -35 and -10 boxes of the wild-type P_{lac} (P_{lac}(WT)) and mutated P_{lac} (P_{lac}(mut)) targeted by a phagemid-encoded randomized δHiSCRIBE(P_{lac})$_{rand}$ library in the evolution experiment. **(C)** Schematic representation of the evolution experiment. The -35 and -10 boxes of the P_{lac} locus were targeted with an ssDNA library produced *in vivo* from a δHiSCRIBE phagemid library delivered by phagemid transduction. Cells that acquired

beneficial mutations in their P_{lac} locus were expected to metabolize lactose better (indicated by darker gray shading) and be enriched in the population over time. **(D)** Growth rate profiles of cell populations exposed to δHiSCRIBE$(P_{lac})_{rand}$ and δHiSCRIBE(NS) (top) as well as the dynamics of P_{lac} alleles over the course of the experiment are shown as time series for cells exposed δHiSCRIBE$(P_{lac})_{rand}$ phagemid library (middle). The bottom panel shows the identities of the most frequent alleles at the end of the experiment as well as the fold-change in the β-galactosidase activity of those alleles in comparison to the WT and parental alleles. Alleles that are likely ancestors / descendants are linked by brackets. The dynamics of allele enrichment for cells exposed to δHiSCRIBE(NS) and additional parallel evolution experiments are presented in Fig. S6.1.

To validate that the identified variants were indeed responsible for increases in fitness, we reconstructed these variants in the parental strain background and assessed their activity by measuring β-galactosidase activity. As shown in Fig. 6.1D (bottom panel), all the evolved variants showed a significant increase in β-galactosidase activity over the parental variant, indicating successful tuning of the activity of the P_{lac} promoter. For example, the dominant variant at the end of the experiment (variant #1) exhibited a >2000-fold increase in β-galactosidase activity relative to the parental strain, corresponding to a 1.4-fold increase over the wild-type P_{lac} promoter.

These results demonstrate that once coupled to a continuous selection or screen (*131*), HiSCRIBE can be used for adaptive writing and continuous and autonomous diversity generation in desired target loci, enabling easy and flexible continuous evolution experiments requiring minimal human intervention. In the current setup, the continuous diversity generation system relies on the continuous and multiplexed (Fig. S6.2) delivery of phagemid-encoded HiSCRIBE variants that compete for writing on the target locus once inside the cells. In future work, incorporating a conditional origin of replication into phagemids or conjugative plasmids may help to increase the rate of evolution by enforcing writing and curing steps in a more controlled fashion.

Tuning Evolvability and Synthetic Lamarckian Evolution by HiSCRIBE writers

The previous experiment demonstrated that HiSCRIBE can selectively target and mutate genomic loci for evolutionary genome engineering. However, the system still relies on pre-existing diversity that is encoded on the HiSCRIBE plasmids to diversify a target. Since HiSCRIBE DNA writing is mediated through transcription, reverse-transcription, and recurring Beta-mediated ssDNA integration processes, which have a lower fidelity than DNA replication, we investigated if this lower fidelity could be leveraged to increase the *de novo* mutation rate of a target site, without increasing the background mutation rate (Fig. 6.2A).

We used a well-established plating assay (*132-135*) and fluctuation analysis (*136, 137*) to measure locus-specific *de novo* mutation rates induced by HiSCRIBE at targeted and non-targeted loci. Using this assay, mutation rates at two different loci, *rpoB,* and *gyrA*, were estimated based on the frequency of rifampicin-resistant (Rif^R) and nalidixic acid-resistant (Nal^R) cells in the population, respectively. Specifically, we measured locus-specific mutation rates in MG1655 *exo*⁻ cells harboring δHiSCRIBE(*rpoB*)$_{WT}$ (which encodes a 72-bp ssDNA with the same sequence as WT *rpoB*), δHiSCRIBE(*gyrA*)$_{WT}$ (which encodes a 72-bp ssDNA with the same sequence as WT *gyrA*), or δHiSCRIBE(NS). Targeting δHiSCRIBE to *rpoB* increased the mutation rate at this locus (measured by the frequency of Rif^R mutants) while having a minimal effect on the mutation rate at the *gyrA* locus (measured by the frequency of Nal^R mutants) (Figs. 6.2A and S6.3A). Similarly, expressing δHiSCRIBE(*gyrA*)$_{WT}$ resulted in a significant increase in the mutation rate at the *gyrA* locus while having a minimal effect on the mutation rate at the *rpoB* locus. These results suggest that HiSCRIBE can selectively increase the mutation rate of a desired target site without increasing the background mutation rate.

We then investigated whether the rate or spectrum of targeted mutations could be modulated by overexpressing an ssDNA-specific modifying enzyme such as human activation-induced cytidine deaminase (AID). AID is an ssDNA-specific cytidine deaminase that is involved in the diversification of the immunoglobin locus in vertebrates and was previously shown to retain its functionality to deaminate cytidine in *E. coli* (*135*). We surmised that AID could act on ssDNA substrates produced by HiSCRIBE and/or on unwound ssDNA segments generated during passage of the replication fork (*138*) that are likely to be more accessible due to the presence of recombineering factors. As shown in Fig. 6.2A, overexpression of AID alongside δHiSCRIBE(*rpoB*)$_{WT}$ from a synthetic operon (hereafter referred to as δHiSCRIBE_AID(*rpoB*)$_{WT}$) increased the targeted mutation rate of the *rpoB* locus even further. However, it also slightly increased the background mutation rate as measured by the Nal^R phenotype at the *gyrA* locus, likely due to the non-specific action of AID on genomic DNA.

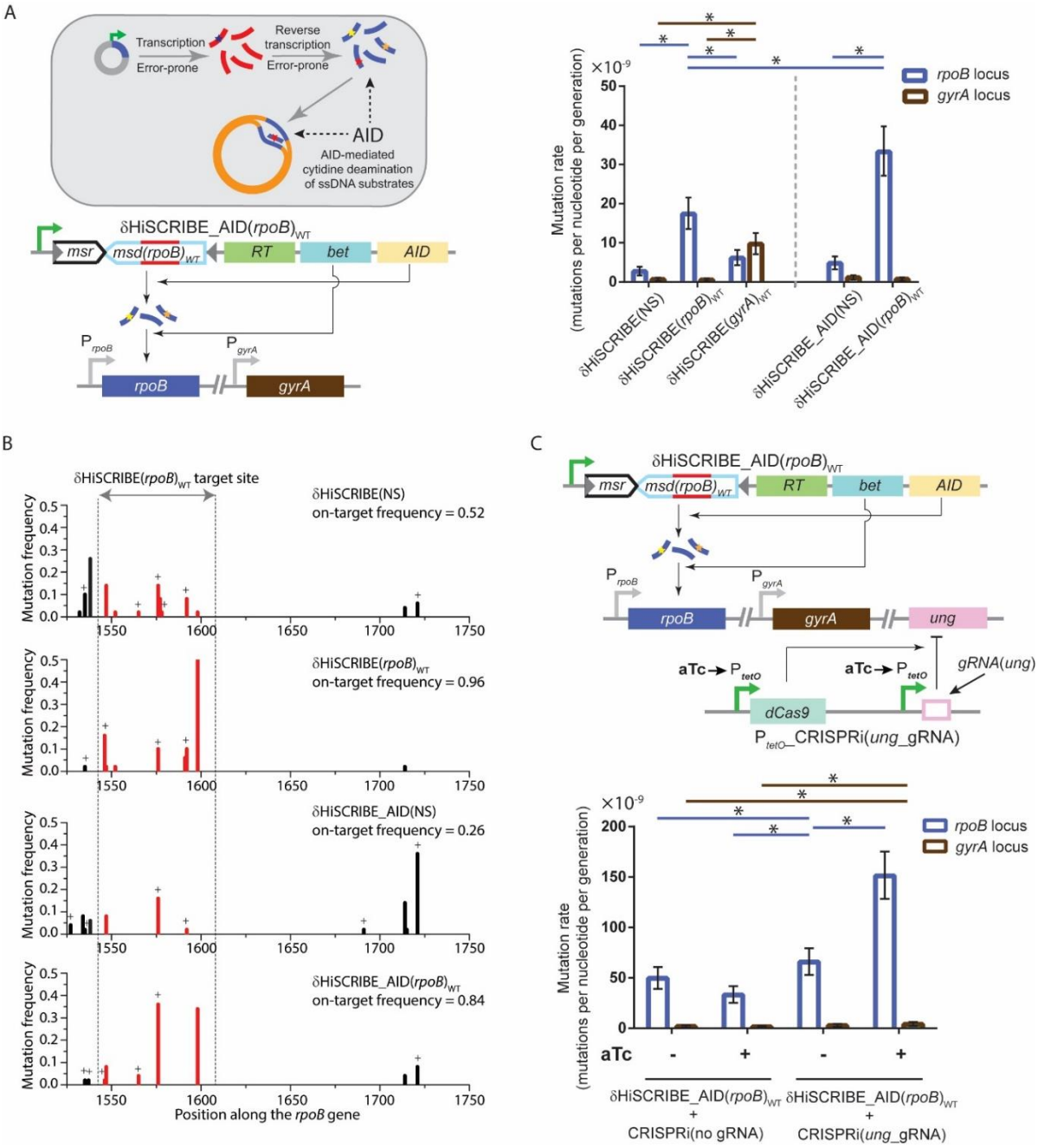


Figure 6.2 | *De novo* targeted mutagenesis via HiSCRIBE. (A) Instead of encoding a library of predefined mutations into HiSCRIBE, we hypothesized that *de novo* mutations could be introduced into HiSCRIBE-expressed ssDNAs during transcription and reverse transcription since these processes are inherently more error-prone than replication. Incorporation of these mutated ssDNAs into target loci should result in targeted *de novo* diversity generation. To enhance the rate of ssDNA mutagenesis, we co-expressed AID with δHiSCRIBE. AID can deaminate cytidine in intracellularly expressed ssDNAs as well as ssDNA regions exposed during the passage of replication forks, thus modulating mutation frequency and spectra. The δHiSCRIBE_AID operon was constructed by placing the AID gene into the δHiSCRIBE operon. Observed frequencies of Rif^R and Nal^R mutants (Fig. S6.3) were used to estimate locus-specific mutation rates of strains expressing different δHiSCRIBE plasmids at *rpoB* and *gyrA*

loci, respectively, using the Maximum Likelihood Estimator (MSS-MLE) method. Error bars indicate 95% confidence intervals for each sample calculated based on 24 parallel cultures. Significant differences in mutation rates ($p < 0.01$) are marked by asterisks. **(B)** The frequency of mutations observed in different positions along the *rpoB* locus. The red columns indicate on-target mutations (i.e., mutations that occurred within δHiSCRIBE($rpoB$)$_{WT}$ target site). Mutations in dC/dG positions are marked by plus signs. 50 colonies were sequenced for each sample. **(C)** Mutation rates of *rpoB* and *gyrA* loci, estimated using MSS-MLE, in strains expressing the δHiSCRIBE_AID($rpoB$)$_{WT}$ plasmid and the aTc-inducible CRISPRi plasmid targeting *E. coli* Uracil-DNA glycosylase (*ung*). Error bars indicate 95% confidence intervals for each sample calculated based on 18 parallel cultures. Significant differences in mutation rates ($p < 0.01$) are marked by asterisks.

To identify the nature of the identified mutants, we Sanger-sequenced the *rpoB* locus of fifty Rif^R colonies from each strain and plotted the observed frequency of each mutation versus its position along the *rpoB* gene (Fig. 6.2B). In cells expressing δHiSCRIBE($rpoB$)$_{WT}$, Rif^R mutations were almost exclusively observed in the 72 bp target region. However, in cells expressing δHiSCRIBE(NS), Rif^R mutations occurred both inside and outside of this region. This suggests that δHiSCRIBE($rpoB$)$_{WT}$ not only increased the mutation rate of the *rpoB* locus, but more specifically did so by elevating the mutation rate within the target region defined by the δHiSCRIBE template. Consistent with the previous reports (*135*), overexpression of AID increased the frequency of mutations at dC/dG positions (Fig. 6.2B and Fig. S6.3B). In cells expressing δHiSCRIBE_AID($rpoB$)$_{WT}$, most mutations in dC/dG positions were observed within the 72 bp target window. This observation was in contrast to cells expressing δHiSCRIBE_AID(NS), where such mutations were observed mostly outside of the targeted region. These results demonstrate that HiSCRIBE can selectively increase the mutation rate at a desired target locus, and that the spectrum of mutations can be tuned by using ssDNA-modifying enzymes.

In order to increase the targeted mutation rate even further, we conditionally knocked down the uracil DNA glycosylase gene (*ung*) of *E. coli*, which is responsible for the repair of deaminated cytidines (*135*), with an aTc-inducible CRISPRi system. As shown in Fig. 6.2C, a significant increase in the mutation rate of the targeted locus (*rpoB*) was observed in cells expressing both δHiSCRIBE_AID($rpoB$)$_{WT}$ and CRISPRi(*ung*_gRNA) upon induction of the CRISPRi system. The background mutation rate in the non-targeted locus (*gyrA*), measured by the Nal^R phenotype, was not significantly affected. These results suggest that by conditionally knocking down systems that repair introduced lesions, one can increase the rate of targeted mutations without affecting the global mutation rate. We anticipate that targeted diversity

generation could be further augmented by additional strategies, including using error-prone RNA polymerases (*46*) and reverse-transcriptases (*48, 139*), RNA and ssDNA modifying enzymes (*140*), and/or conditionally suppressing machinery involved in the repair of corresponding lesions (e.g., MMR) using CRISPRi.

Tuning Evolvability and Synthetic Lamarckian Evolution by DOMINO writers

While HiSCRIBE writers provide an efficient tool to introduce in vitro-generated diversity into desired genomic loci, they have limited capacity to generate genetic *de novo* diversity in compare to platforms that are based on base editors such as DOMINO. We sought to demonstrate the concept by coupling targeted diversity generation achieved by DOMINO with a selective pressure. Specifically, we showed that *E. coli* cells with an initially weak *lac* operon promoter (P_{lac}) can be engineered to autonomously evolve a stronger promoter at the presence of lactose as the sole carbon source, with a rate much faster than possible by natural evolution, and without a requirement for in vitro-generated diversity. As mentioned previously, lactose utilization in *E. coli* relies on the activity of *lac* operon, and at the presence of lactose as the sole carbon source, cells fitness (i.e. growth rate) correlates with their ability to metabolize lactose (i.e. P_{lac} operon activity). In order to increase the fitness range, we weakened the wild-type P_{lac} (P_{lac}(WT)) by replacing the -35 and -10 boxes of this promoter with dC residues. This mutant promoter (P_{lac}(mut)) has a very low activity and cells harboring this promoter (which hereafter are referred to as parental cells) grow very poorly at the presence of lactose (see the first time point in Figs. 6.3D and 6.3E). We then introduced the CDA-nCas9-ugi writer with or without two gRNAs targeting the -35 and -10 boxes of the P_{lac}(mut) into these cells and grew the cells at the presence of glucose (glu) and lactose (lac) for multiple days (Figs. 6.3B and 6.3C). The *lac* operon in *E. coli* is repressed at the presence of glucose, thus, glucose-containing media acts as a non-selective media for these cells. However, in media containing lactose as the sole carbon source, the diversified P_{lac} alleles would compete for consumption of lactose, and those with higher P_{lac} activity are expected to enrich the population over time.

We monitored the growth rate and P_{lac} activity of cultures throughout this experiment. As shown in Fig. 6.3D, the growth rate (in lactose) of cultures that did not express gRNAs only slightly increased toward the end of the experiment (after 72 hours). On the other hand, the growth rate (in lactose) of cultures harboring the P_{lac} containing promoters significantly increased over time, indicating a significant increase in the fitness and that these cells had evolved the ability to metabolize much faster than cells that did not express the gRNAs. These results were further confirmed by measuring

the P_{lac} activity, where a significant increase in the activity of P_{lac} was observed in cultures that express P_{lac} targeting gRNAs, while the activity of P_{lac} in cells that did not express the gRNAs did not increase over time.

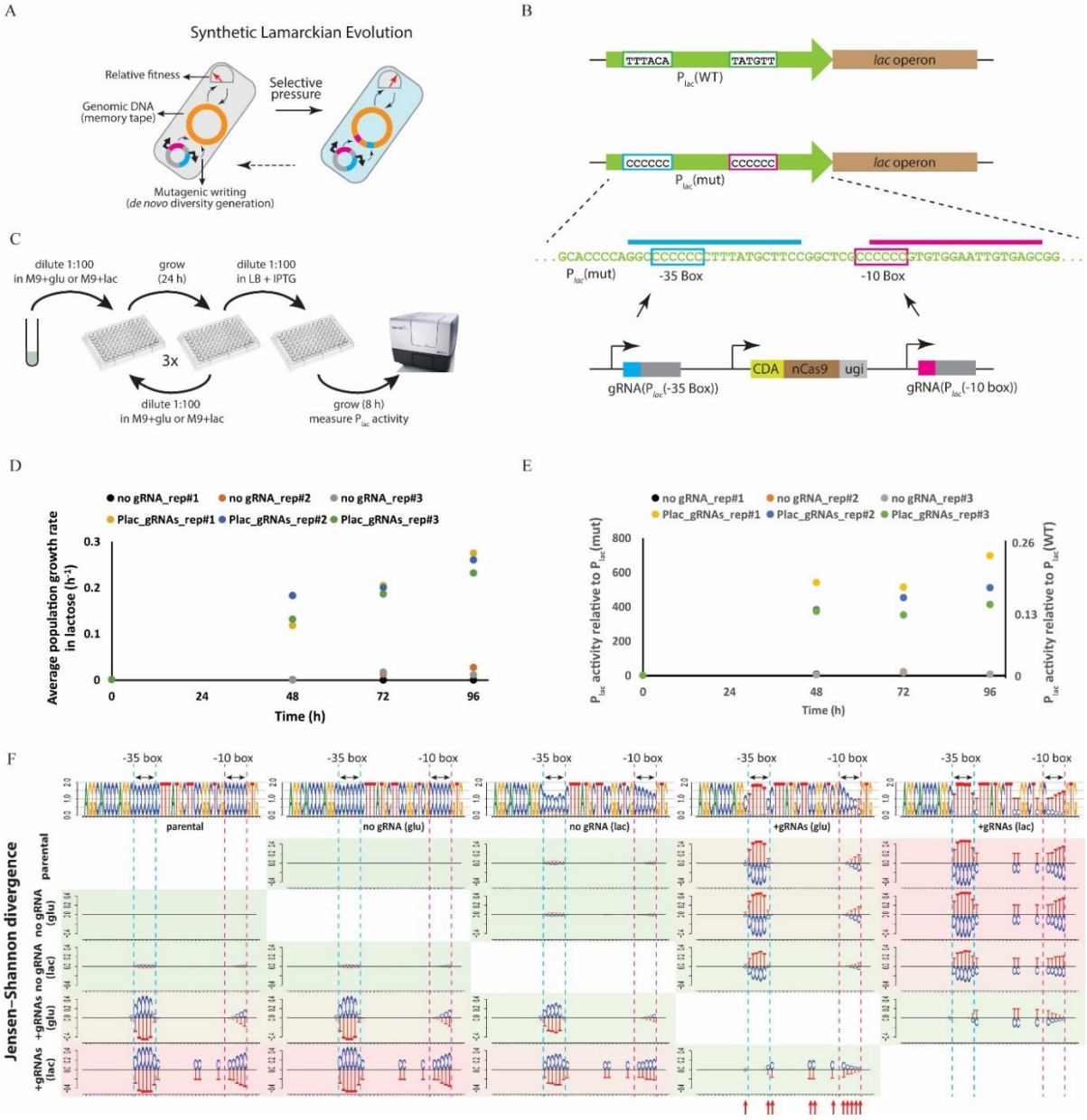


Figure 6.3 | Continuous synthetic Lamarckian evolution of cellular phenotypes enabled by coupling *de novo* diversity generation with continuous selection by DRIVE. (A) Continuous *de novo* targeted diversity generation can be coupled with a selective pressure (or screening) to allow optimizing phenotype of interest without a concomitant increase in the global mutation rate. **(B)** To achieve a large dynamic span in fitness, we weakened P_{lac} promoter of *E. coli*, which controls fitness (i.e., growth rate) of cells at the presence of lactose as the sole carbon source, by introducing 6-bp poly-dC into -35 and -10 regulatory boxes of this promoter to make a mutant P_{lac} promoter (P_{lac}(mut)). We then introduced complementary gRNAs targeting these two regulatory regions to endow cells with the ability to site-specifically

increase their de-novo mutation rate. **(C)** Cells harboring the DNA writer with or without the P_{lac}-targeting gRNAs were grown either in selective media (containing lactose as the sole carbon source) or non-selective media (containing glucose as the sole carbon source) for three successive grow and dilutions cycles. The growth rate of cells in lactose, as well as activity of P_{lac} promoter, was monitored throughout the experiment. **(D)** Average population growth rate of parallel cultures with or without P_{lac}-targeting gRNAs in lactose. **(E)** P_{lac} activity for parallel cultures with or without P_{lac}-targeting gRNAs grown in lactose. **(F)** The sequence logo of position weight matrixes for the parental strain, as well as cells with or without P_{lac}-targeting gRNAs grown in either glucose or lactose are shown (top panel). Jensen-Shannon divergence values for pair-wise comparison of these samples are shown in the bottom panel. For each subplot, positions that harbor different nucleotide distributions are indicated by the letters corresponding to each nucleotide. The letters in the upper section of each subplot correspond to the nucleotides over-represented in the sample in the corresponding column, while the letter in the lower section corresponds to the sample in the corresponding row. Comparing the mutant distribution in cells harboring P_{lac}-targeting gRNAs that were grown in the selective media (lactose) and non-selective media (glucose, reveals adaptive mutations (marked by red arrows) in the vicinity of gRNA target sites on the P_{lac}).

To investigate the evolution of P_{lac} alleles at the molecular level, we PCR amplified the P_{lac} locus and sequenced the amplicons by high-throughput sequencing. As shown in Fig. 6.3F, dC to dT mutations accumulated in the vicinity of the P_{lac} promoter in gRNA expressing cells, indicating targeted *de novo* diversity generation in this locus. Analysis of the enriched variants between gRNA-expressing cells grown in and glucose revealed a series of positions (marked by red arrows in Fig. 6.3F) in which mutations were more strongly enriched in the selective medium (lac) than non-selective medium (glu). The differential enrichment of mutation in these positions suggests that these positions were under positive selection and thus their corresponding mutations can be considered as adaptive mutations.

We also observed some level of mutations in cells with no gRNA that were grown in lactose, but these mutations were only detectable in the later time-points and were significantly lower than the level of mutations in cells expressing the gRNAs. These mutations were likely generated non-specifically as a result of an increase in global mutation rate due to overexpression of the cytidine deaminase, which is further supported by that fact that these mutations only enriched in cells that were under selection (grown in lactose) and not those that were grown in non-selective media (glucose).

6.4 Discussion

Targeted mutagenesis strategies, which as opposed to using a generalized hypermutator genetic background or mutagen chemicals could elevate mutation rate of desired loci without increasing global mutation rate, could have broad utility in evolutionary engineering applications. We used phagemid-encoded HiSCRIBE writers to

continuously and autonomously tune a genomic segment (P_{lac} promoter) and its connected phenotype (ability to metabolize lactose). We demonstrated that HiSCRIBE phagemid libraries and cells comprise a self-contained and rapidly evolving synthetic ecosystem that can continuously and autonomously traverse evolutionary paths imposed by the diversity of the HiSCRIBE library and the applied selective pressure. This platform could facilitate evolutionary genome engineering and gene resurrection studies, which have been traditionally limited due to the lack of suitable tools for *in vivo* targeted genome mutagenesis, and provide new insights into accessible evolutionary trajectories (*141-145*). In addition to phagemid delivery, inducible writing or conjugative delivery of HiSCRIBE libraries (as shown in chapter 3) could be linked to selection or screening strategies to enable temporally- or spatially-restricted diversification and continuous evolution applications. Unlike recombineering-based targeted mutagenesis strategies like Multiplexed Automated Genome Engineering (MAGE), where the library size is limited by the capacity to electroporate synthetic oligos into a limited number of cells, HiSCRIBE diversity generation can be readily scaled-up using alternative delivery methods such as transduction and conjugation. This feature could greatly expand the practical diversity that can be experimentally introduced into a population and the breadth of organisms that can be targeted. Furthermore, unlike MAGE, the diversity generation step for HiSCRIBE can be regulated both spatially and temporally, coupled to cellular regulatory circuits, and performed in a completely autonomous fashion, all of which provide greater ease and flexibility in adaptive writing and evolution experiments.

We further demonstrated that in addition to writing pre-existing diversity on a desired target, the HiSCRIBE system can be used to generate targeted diversity *de novo* without increasing the global mutation rate. Additionally, we demonstrated strategies to augment mutation rate of this *de novo* targeted diversity generation approach. Functional genetic elements (e.g., RNA aptamers, peptides, and proteins) could be potentially encoded within HiSCRIBE templates to build self-targeting HiSCRIBE cassettes that autonomously undergo continuous and accelerated rounds of evolution under a suitable selective pressure.

Although potentially more generalizable, the *de novo* diversity generation achieved by HiSCRIBE writers is currently less efficient than DOMINO. We combined *de novo* targeted diversity generation achieved by DOMINO (i.e., an addressable DNA writer) with suitable selective pressure and engineered cells that can autonomously increase the mutation rate of specific segments of their genomes and undergo (synthetic Lamarckian) evolution with a rate much faster than possible by Darwinian evolution.

The outcome of the DRIVE platform is reminiscent of natural diversity generation mechanism by the DGR system (*122*) in phages and bacteria, but instead of dA residues in the DGR system, here dC residues are targeted for mutation, and the system can be easily retargeted to desired sequences. Incorporation of the recently developed adenosine deaminase (*113*) into DRIVE would make it possible to introduce all the possible transition mutations into desired target sites. This less-explored but powerful approach that converts genetic DNA into a targetable substrate for continuous and autonomous evolution in the laboratory, could open up new avenues to study and engineer biological systems. We envision that the synthetic Lamarckian evolution strategy achieved by DRIVE could have broad applicability in studying and evolutionary engineering of living systems, from engineering smart, fast-adaptable cells that can tune their response and find new solution in response to internal or external cues, to engineering adaptable therapeutics and biomolecules to devising continuous in vivo evolution strategies, to optimizing cellular traits and metabolic pathways, to engineering bacteriophages that can autonomously mutagenize their tail fiber and expand their host-range with a rate much faster than possible by natural evolution under specific user-specified condition.

6.5 Supplementary Information

Unless otherwise specified, the methods, DNA constructs, and strains described in Chapters 3 and 4 were used in the experiments described in this chapter.

Continuous evolution of the P_{lac} promoter

The efficient genome editing achieved by HiSCRIBE can be coupled to continuous selection or screening to enable continuous evolution of desired target loci. In order to demonstrate this adaptive writing strategy, we chose to evolve P_{lac} in *E. coli* (Fig. 6.1). To achieve a wider dynamic range of fitness, we started with a weakened P_{lac} promoter, created by mutating the -10 sequence of P_{lac} promoter from "TATGTT" to "CCCCCC". This mutation leads to the poor growth of cells in M9 media when lactose is the sole carbon source. An overnight culture of the parental strain harboring the mutated P_{lac} promoter (MG1655 Δ*recJ* Δ*xonA* F^+ $P_{lac("TATGTT"\rightarrow"CCCCCC")}$) was diluted (1:100) into M9 + glu (0.2%) and divided into two groups, each with three parallel cultures. Samples in each group were treated with phagemid particles (MOI = 100) from either a δHiSCRIBE(P_{lac}) phagemid library or the non-specific δHiSCRIBE(NS) control, and incubated in a microplate reader at 37°C with continuous shaking (250 RPM). The cultures were grown for 1 hour before antibiotic selection (Carb and Cam for phagemid delivery and F-plasmid maintenance, respectively). Cells were then grown for 23

additional hours, diluted (1:100) into M9 + lactose (0.2%) + phagemid + antibiotics, and grown for 48 hours at 37°C in a microplate reader as above. The dilution and regrowth (24 h) cycles were repeated five additional times to permit selection and propagation of beneficial mutations. OD_{600} was monitored and samples were taken for Illumina sequencing throughout the experiment. Population growth rates based on OD_{600} were calculated using the GrowthRates tool (*146*).

To verify the activity of the identified variants in the P_{lac} evolution experiments, we reconstructed these variants in the parental background using oligo-mediated recombineering (*147*). The reconstructed variants were grown overnight in LB, diluted (1:100) in fresh media supplemented with IPTG (1 mM), and grown for 8 hours (37°C, 700 RPM). The activities of reconstructed P_{lac} promoter variants were measured by Miller assay using Fluorescein di-β-D-galactopyranoside (FDG) as substrate. 50 µL of each culture was mixed with 50 µL of B-PER II reagent (Pierce Biotechnology) containing FDG (0.005 mg/mL final concentration). The fluorescent signal (absorption/emission: 485/515 nm) was monitored in a plate reader with continuous shaking for 2 hours at 37°C. β-galactosidase activity was calculated by normalizing the rate of FDG hydrolysis (obtained from fluorescence signal) to the initial OD_{600}. For each sample, β-galactosidase activity was reported as the mean of three independent biological replicates.

δHiSCRIBE library construction

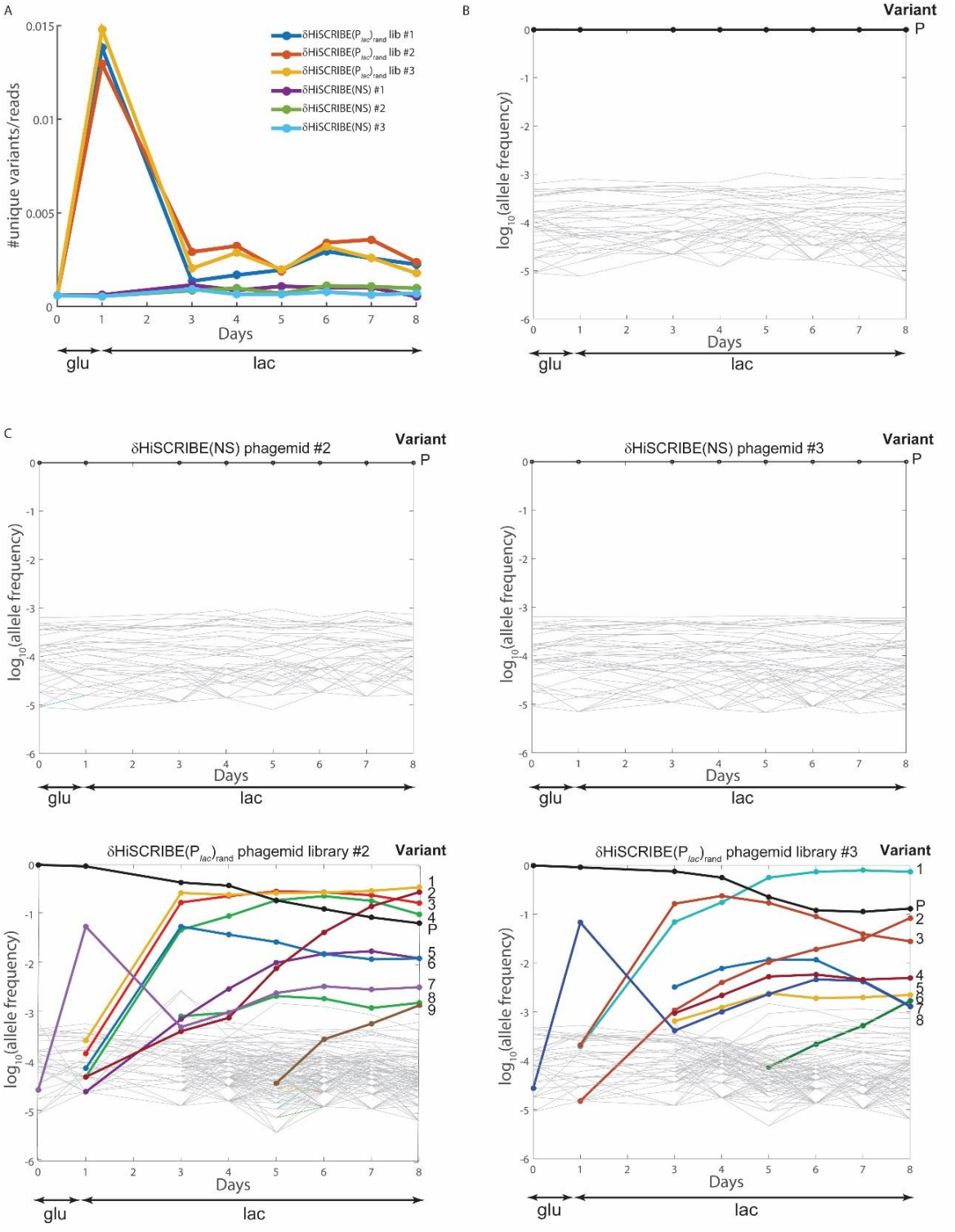
Randomized δHiSCRIBE phagemid and mobilizable libraries (for experiments shown in Figs. 6.1 and S6.1, respectively) were constructed by a modified Quik-Change (Agilent) protocol. Briefly, δHiSCRIBE plasmids were PCR amplified using primers containing the randomized regions within the desired target site in the overhangs. The primers also contained compatible sites for the type IIS enzyme Esp3I. PCR products were used in a Golden Gate assembly (*90*) using this cut site to circularize the linear vector. Circularized vector libraries were amplified by transformation into Electro-ten Blue electrocompetent cells (Agilent). Amplified libraries were then packaged into phagemid particles for transduction experiments (as described in Chapter 3).

Calculating mutation rate

Different δHiSCRIBE or δHiSCRIBE_AID plasmids (as shown in Fig. 6.2) were transformed into the MG1655 Δ*recJ* Δ*xonA* strain. Six single colonies from each transformation plate were inoculated in 1 mL LB + Kan in 24-deep-well plates and

incubated (37°C, 700 RPM) for 24 hours. The number of Rif^R and Nal^R mutants in each sample was determined by plating 400 µL (or 200 µL in experiments shown in Fig. 6.2C) of saturated culture from each sample on LB + Kan + Rif and LB + Kan + Nal plates. The experiments were repeated 4 times (total 24 parallel cultures for each strain). The mutation rates were calculated using FALCOR (*136*) based on the Maximum Likelihood Estimator (MSS-MLE) method (*137*).

To investigate the nature and spectrum of Rif^R mutations, the *rpoB* locus from 50 Rif^R colonies from each sample were PCR amplified. After column purification, they were analyzed by Sanger sequencing. More than 98% of the samples contained mutations within the sequenced region.



Variant	Sequence		P_{lac} activity (A.U.)	
	-35 Box	-10 Box	Relative to P_{lac}(WT)	Relative to P_{lac}(mut)
Parental	TTTACA	CCCCCC	0.0007	1
1	TTTACA	ATAAAG	0.2	339
2	TTTACA	ATTATA	0.9	1428
3	ATGATA	CCCCCC	0.2	365
4	TTTACA	AACATT	0.5	701
5	TTTACA	AAAGTA	NA	NA
6	TTTACA	AGAGTT	NA	NA
7	TTTACA	TATGTT	NA	NA
8	ATGATA	ATAAAG	NA	NA
9	ATGATA	ATTATA	NA	NA

Variant	Sequence		P_{lac} activity (A.U.)	
	-35 Box	-10 Box	Relative to P_{lac}(WT)	Relative to P_{lac}(mut)
Parental	TTTACA	CCCCCC	0.0007	1
1	TTATTC	TGTAAA	1.3	1962
2	TTTACA	TATCAT	1.5	2155
3	TTTACA	TAGGTT	0.5	675
4	TTTACA	TGTAAA	0.02	35
5	TTATTC	CCCCCC	NA	NA
6	TTATTC	TATCAT	NA	NA
7	TTTACA	TATGTT	NA	NA
8	TTTACA	CCAAAA	NA	NA

Figure S6.1 | Dynamics of P_{lac} alleles in the P_{lac} evolution experiment. (A) The diversity of P_{lac} alleles observed in the evolution experiment shown in Fig. 6.1 as well as two additional parallel cultures, reported as the number of unique variants per sequencing read. The diversity of the P_{lac} locus in cultures exposed to the HiSCRIBE$(P_{lac})_{rand}$ phagemid library was significantly higher than those exposed to δHiSCRIBE(NS) phagemids. **(B)** Dynamics of P_{lac} alleles for cultures that were exposed to δHiSCRIBE(NS) phagemids in the experiment shown in Fig. 6.1. **(C)** Changes in P_{lac} alleles frequencies over the course of the experiment shown as time series for cells exposed to the δHiSCRIBE(NS) (top) or the δHiSCRIBE$(P_{lac})_{rand}$ library phagemid particles (middle) for two additional parallel cultures of the experiment shown in Fig. 6.1. The identities of the most frequent alleles at the end of the experiment, as well as fold-change in the β-galactosidase activity of the corresponding allele compared to the WT and parental alleles, are shown in the bottom tables. Alleles that are likely ancestors/descendants are linked by brackets.

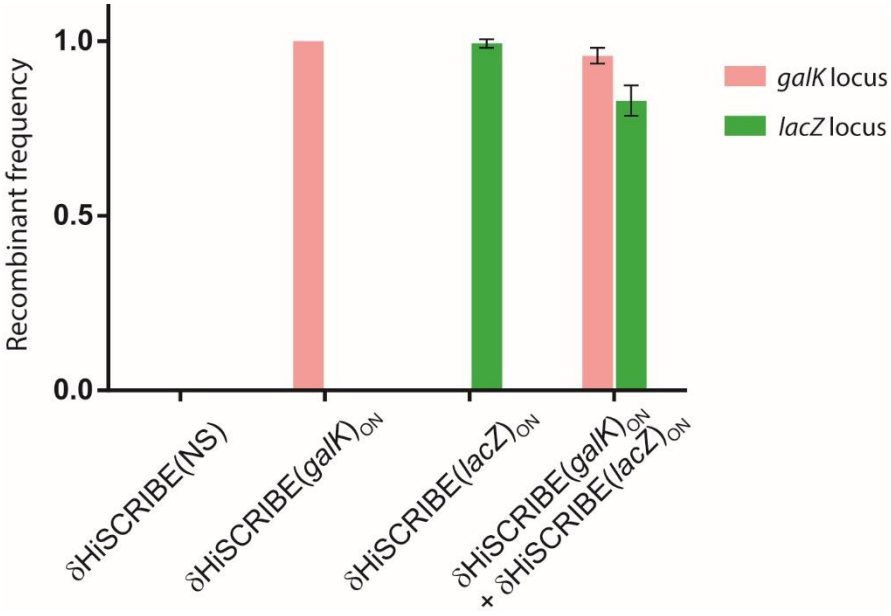
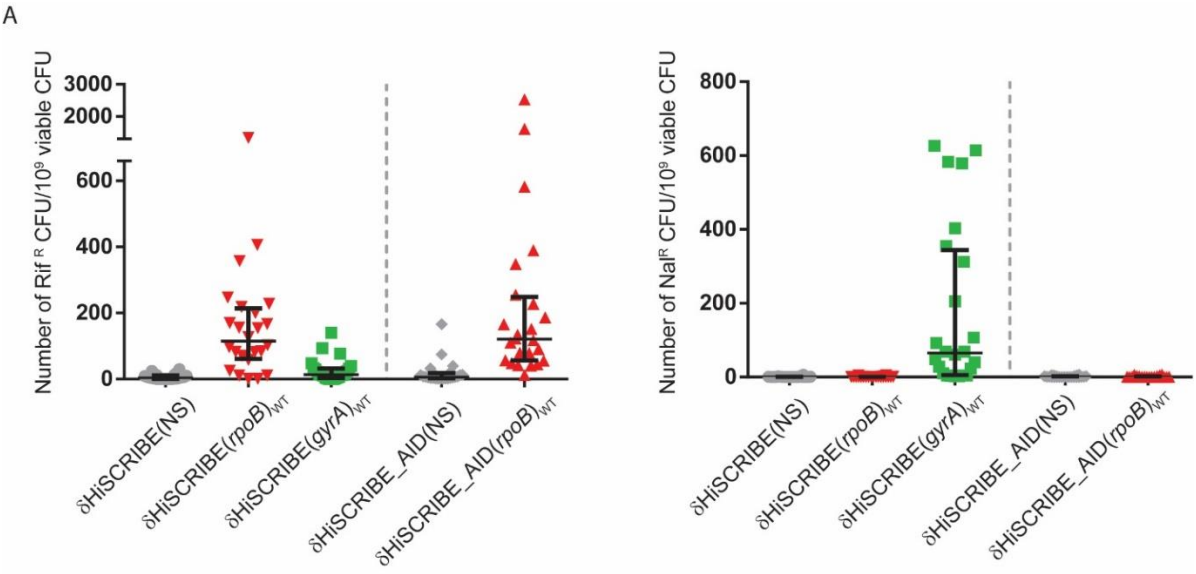


Figure S6.2 | Multiplexed writing in different loci using δHiSCRIBE. An MG1655 F$^+$ Δ*recJ* Δ*xonA* $galK_{OFF}$ $lacZ_{OFF}$ reporter strain (harboring two premature stop codons within *galK* and *lacZ* ORFs) was transduced with either δHiSCRIBE($galK$)$_{ON}$ or δHiSCRIBE($lacZ$)$_{ON}$ (MOI = 50) phagemid particles, or both (MOI = 100 each). Serial dilutions of the samples were spotted on LB + X-gal + IPTG + Carb or MacConkey + gal + Carb plates to measure the frequency of recombinants in the *lacZ* locus (blue colonies) and *galK* locus (pink colonies), respectively. Error bars indicate standard errors for three biological replicates.

A



B

		Frequency of mutations in dC/dG positions	
	Strain	**Total**	**On-target:total**
- AID	δHiSCRIBE(NS)	0.42	0.62
	δHiSCRIBE(*rpoB*)$_{WT}$	0.38	0.95
+ AID	δHiSCRIBE_AID(NS)	0.62	0.29
	δHiSCRIBE_AID(*rpoB*)$_{WT}$	0.54	0.78

Figure S6.3 | D*e novo* targeted mutagenesis by HiSCRIBE. (A) Frequencies of Rif^R and Nal^R mutants, which harbor mutations in the *rpoB* and *gyrA*, respectively, observed in MG1655 Δ*recJ* Δ*xonA* expressing different δHiSCRIBE plasmids. Bars indicate median and interquartile of each sample set. For each strain, the mutant frequencies in 24 parallel cultures were measured and the data was used to calculate the mutation rates shown in Fig. 6.3A. **(B)** The frequency of mutations at dC/dG positions based on the data shown in Fig. 6.3B. AID expression increases the total frequency of mutations at dC/dG positions. However, in cells expressing δHiSCRIBE˙AID(NS), dC/dG mutations mostly occur outside of the target sites. Expression of δHiSCRIBE˙AID(*rpoB*)$_{WT}$ directs dC/dG mutations towards the target site (*rpoB*) and increases the frequency of on-target:total dC/dG mutations.

Table S6.1 | List of the reporter strains used in this study.

Name	Strain Code	Genotype	Used in
MG1655 *exσ* P_{lac}(-10mut)	FFF1032	FFF964 P_{lac}(mut) where -10 Box of P_{lac} promoter in FFF964 is mutated from TATGTT to CCCCC (For transduction experiments, the F-plasmid (from CJ236 (NEB)) was introduced to this strain via conjugation)	Fig. 6.1 Fig. S6.1
MG1655 $galK_{OFF}$	FFF1086	MG1655 $galK_{L187TAA, L188TGA}$	Fig. 6.2 Fig. S6.3
MG1655 P_{lac}(-35, -10mut)	FFF1290	P_{lac}(mut) where -35 and -10 Box of P_{lac} promoter in MG1655 is mutated from WT sequences to CCCCC	Fig. 6.3

Table S6.2 | List of the plasmids used in this study.

Name	Plasmid Code	Maker	Used in	Ref
P_{tetO}_CRISPRi(*ung* gRNA)	pFF1369	Cam	Fig. 6.2C	This work
δHiSCRIBE(*rpoB*)$_{WT}$ (Strong RBS)	pFF1328	Kan	Fig. 6.2 Fig. S6.3	This work
δHiSCRIBE(*gyrA*)$_{WT}$ (Strong RBS)	pFF1336	Kan	Fig. 6.2 Fig. S6.3	This work
δHiSCRIBE_AID(*rpoB*)$_{WT}$ (Strong RBS)	pFF1329	Kan	Fig. 6.2 Fig. S6.3	This work
δHiSCRIBE_AID(*gyrA*)$_{WT}$ (Strong RBS)	pFF1329	Kan	Fig. 6.2 Fig. S6.3	This work
δHiSCRIBE(*lacZ*)$_{ON}$ (Strong RBS)	pFF1299	Carb	Fig. S6.2	This work
P_{tetO}_CDA-nCas9-ugi	pFF1454	Cam	Fig. 6.3	This work (same as chapter 4)
P_{tetO}_CDA-nCas9-ugi (gRNA(P_{lac}(-35 box) & gRNA(P_{lac}(-10 box)	pFF1455	Cam	Fig. 6.3	This work

Table S3 | List of the synthetic parts and their corresponding sequences used in this study (to avoid duplications, only sequences that were not listed in chapters 3 and 4 are listed here).

Part name	Type	Sequence	Ref
msd(*lacZ*)$_{ON}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCACC CAACTTAATCGCCTTGCAGCACATCCCCCTT TCGCCAGCTGGCGTAATAGCGAAGAGGCCC GCACCGATCGCCCTGAATTCAGGAAAACAG ACAGTAACTCAGA	(*67*)
msd(*rpoB*)$_{WT}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCACC GCCTGGGCCGAGTGCGGAGATACGACGTTT GTGCGTAATCTCAGACAGCGGGTTGTTCTG GTCCATAAAGAATTCAGGAAAACAGACAGT AACTCAGA	This work
msd(*gyrA*)$_{WT}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCGAA TGGCTGCGCCATGCGGACGATCGTGTCATA GACCGCCGAGTCACCATGGGGATGGTATTT ACCGATTACGAATTCAGGAAAACAGACAGT AACTCAGA	This work
msd(P_{lac}) (highlighted regions indicate positions in the msd corresponding to the randomized -10 and -35 boxes of P_{lac}	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCAAT GTGAGTTAGCTCACTCATTAGGCACCCCAG GCNNNNNNCTTTATGCTTCCGGCTCGNNNN NNGTGTGGAATTGTGAGCGGATAACAATTT CACACAGGAATTCAGGAAAACAGACAGTAA CTCAGA	This work
AID	Activation-induced Cytidine Deaminase	ATGGACAGCCTCTTGATGAACCGGAGGAAG TTTCTTTACCAATTCAAAAATGTCCGCTGG GCTAAGGGTCGGCGTGAGACCTACCTGTGC TACGTAGTGAAGAGGCGTGACAGTGCTACA TCCTTTTCACTGGACTTTGGTTATCTTCGCA ATAAGAACGGCTGCCACGTGGAATTGCTCT TCCTCCGCTACATCTCGGACTGGGACCTAG ACCCTGGCCGCTGCTACCGCGTCACCTGGT TCACCTCCTGGAGCCCCTGCTACGACTGTG CCCGACATGTGGCCGACTTTCTGCGAGGGA ACCCCAACCTCAGTCTGAGGATCTTCACCG CGCGCCTCTACTTCTGTGAGGACCGCAAGG CTGAGCCCGAGGGGCTGCGGCGGCTGCACC GCGCCGGGGTGCAAATAGCCATCATGACCT TCAAAGATTATTTTTACTGCTGGAATACTTT TGTAGAAAACCATGAAAGAACTTTCAAAGC CTGGGAAGGGCTGCATGAAAATTCAGTTCG TCTCTCCAGACAGCTTCGGCGCATCCTTTTG CCCCTGTATGAGGTTGATGACTTACGAGAC GCATTTCGTACTTTGGGACTTTGA	This work
-10 box gRNA	gRNA protospacer	CCCCCGTGTGGAATTGTGAG	This work
-35 box gRNA	gRNA protospacer	GGCCCCCCCCTTTATGCTTC	This work
*ung*_gRNA	gRNA protospacer	GGACTGCCGCTCGCTGGCGA	This work

Table S5 | List of the sequencing primers used in this study

Primer code	Name	Sequence
FF_oligo_1831	lacZ(+)	ACACGACGCTCTTCCGATCTNNNNNCTG GAA AGC GGG CAG TGA GC
FF_oligo_1833	lacZ(-)	CGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNN CCCAGTCACGACGTTGTAAAACGAC

Chapter 7: crisprTF

Tunable and Multi-Functional Eukaryotic Transcription Factors Based on CRISPR/Cas

This chapter is adapted from

Farzadfard, F., Perli, S.D. and Lu, T.K., 2013. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. ACS synthetic biology, 2(10), pp.604-613. (http://dx.doi.org/10.1021/sb400081r).

to fit the format of this thesis (with permission from American Chemical Society).

7.1 Abstract

Transcriptional regulation is central to the complex behavior of natural biological systems and synthetic gene circuits. Platforms for the scalable, tunable, and simple modulation of transcription would enable new abilities to study natural systems and implement artificial capabilities in living cells. Previous approaches to synthetic transcriptional regulation have relied on engineering DNA-binding proteins, which necessitate multi-step processes for construction and optimization of function. Here, we show that the CRISPR/Cas system of *Streptococcus pyogenes* can be programmed to direct both activation and repression to natural and artificial eukaryotic promoters through the simple engineering of guide RNAs with base-pairing complementarity to target DNA sites. We demonstrate that the activity of CRISPR-based transcription factors (crisprTFs) can be tuned by directing multiple crisprTFs to different positions in natural promoters and by arraying multiple crisprTF-binding sites in the context of synthetic promoters in yeast and human cells. Furthermore, externally controllable regulatory modules can be engineered by layering gRNAs with small molecule-responsive proteins. Additionally, single nucleotide substitutions within promoters are sufficient to render them orthogonal with respect to the same gRNA-guided crisprTF. We envision that CRISPR-based eukaryotic gene regulation will enable the facile construction of scalable synthetic gene circuits and open up new approaches for mapping natural gene networks and their effects on complex cellular phenotypes.

7.2 Introduction

Complex and sophisticated phenotypes in eukaryotic cells manifest from layered regulatory networks and specific expression programs involving the regulated

transcription of many genes (*148*). As major players in these networks, eukaryotic transcriptional factors (TFs) can integrate multiple signals and perform complex, combinatorial functions on promoters, where regulatory information is encoded in the form of binding sites for TFs and interactions between TFs, to modulate gene expression patterns (*148-150*).

Rewiring endogenous transcriptional networks by natural or synthetic TFs is a powerful strategy for interrogating cellular functions and controlling cellular phenotypes (*40, 151-160*). Previously, natural DNA-binding domains (DBDs, mainly from bacterial sources, such as TetR, LacI, and LexA) have been used to recruit effector (e.g. activator and repressor) domains to the regulatory regions of eukaryotic genes in order to modulate their transcription (*161-163*). This necessitates the placement of DBD-specific operator site(s) in the *cis*-regulatory region of the promoters for specific genes which is a labor- and time-intensive process, especially if the regulation of multiple genes is desired. Moreover, engineering and modulating complex transcriptional networks requires tunable, extensible, and orthogonal transcription factors. However, only a few orthogonal variants of natural DBDs are well-characterized and changing their specificity has proven to be challenging (*164*). As such, the use of natural DBD-based TFs for wiring complex transcriptional networks and synthetic gene circuits has been limited.

To address these limitations, synthetic TFs based on Zinc Fingers (ZFs) and Transcriptional Activator-Like Effectors (TALEs) have been developed (*165-171*). The ability to program the specificity of ZFs and TALEs to potentially target any sequence makes these DBDs appealing for designing libraries of orthogonal transcription factors. Synthetic ZF- and TALE-based TFs have been shown to work in a wide range of eukaryotes (*168, 172, 173*); however, obtaining a TF for a given target site requires tedious selection processes or multi-stage DNA assembly protocols (*174, 175*). Furthermore, the scale of regulation that can be achieved by these TFs is potentially limited by the metabolic burden imposed on the cells and the number of TFs that can be simultaneously encoded in a given cell (*176*).

Here, we present a strategy for modulating eukaryotic transcription at natural and synthetic promoters using programmable and tunable synthetic transcription factors based on a bacterial CRISPR (clustered regularly interspaced short palindromic repeats)/Cas system. Many bacteria use CRISPR-based immune systems to degrade genetic materials of invading phages (*177, 178*). In these systems, short RNAs expressed from CRISPR loci are used to target an endonuclease protein (Cas9) against

invading genetic material. Recently, it has been shown that Cas9 can be used as a programmable tool for genome editing across various organisms (*63, 179-183*). In this context, small customizable guide RNAs (gRNAs) can be used to program and target Cas9 endonuclease to specific loci in living cells to induce double (or single)-stranded breaks in DNA. Upon cleavage, error-prone or template-directed repair pathways are triggered, generating variants of the original target loci. Recently, Qi et al. (*69*) showed that an endonuclease-deficient Cas9 (dCas9, with D10A H841A mutations relative to the wild-type Cas9) can be used as a programmable "CRISPRi" tool for gene silencing in *Escherichia coli*. When targeted to a promoter or ORF of a gene of interest, dCas9 can block progression of RNA polymerase and hence silence expression of the targeted gene. They also provide evidence that CRISPRi is functional in human cells, albeit with much lower efficiency compared with *E. coli*. In addition, Bikard et al. demonstrated that along with programmed transcriptional repression, transcriptional activation can be achieved in *E. coli* by fusing the omega subunit of RNA polymerase to the endonuclease-deficient Cas9 (*184*).

7.3 Results

Here, we achieved versatile, programmable, and multiplexable tools for gene regulation in eukaryotes by functionalizing dCas9 with effector domains and targeting both natural and synthetic promoters. As a proof of concept, we made an RNA-guidable transcription factor by fusing dCas9 to an activator domain. Using this CRISPR-based transcription factor (crisprTF), we teased apart the regulatory maps of several natural eukaryotic promoters (in *Saccharomyces cerevisiae* and HEK 293T cells) without the need to modify promoter architectures. Unlike previous generations of customizable DBDs (i.e., ZFs and TALEs) that require multi-stage design and cloning strategies, crisprTFs can be readily customized and retargeted to different loci and regulatory regions *in vivo* using specific gRNAs with homology to target sites (Fig. 7.1). dCas9 thus offers a powerful tool for targeting functions of interest to specific genomic loci in living cells, which can potentially be used to regulate gene expression at will, construct scalable synthetic gene circuits, or rewire endogenous regulatory networks.

To implement crisprTFs in *Saccharomyces cerevisiae*, we fused the SV40 nuclear localization sequence (NLS) and four tandem copies of Herpes Simplex Viral Protein 16 (VP64, a commonly used eukaryotic transcription activator domain) to a codon-optimized *S. pyogenes* dCas9 (Fig. 7.1A) (*185*). The crisprTF cassette was then cloned under the control of pTPGI, a synthetic promoter which can be induced by growing cells in galactose + anhydrotetracycline (aTc) media (*169*), and integrated into the

yeast genome. To assess the activity of crisprTF, *gfp* was placed under the control of a minimal CYC1 promoter (pCYC1m) and also integrated into the yeast genome. pCYC1m retains one of the two endogenous TATA boxes of the wild-type CYC1 promoter and lacks binding sites for endogenous regulatory factors in the upstream activating sequence (UAS) (*186, 187*). gRNAs were expressed constitutively from the RNA polymerase III-dependent pRPR1 promoter and the 3'-ends of the gRNAs were defined by the pRPR1 terminator (*188*).

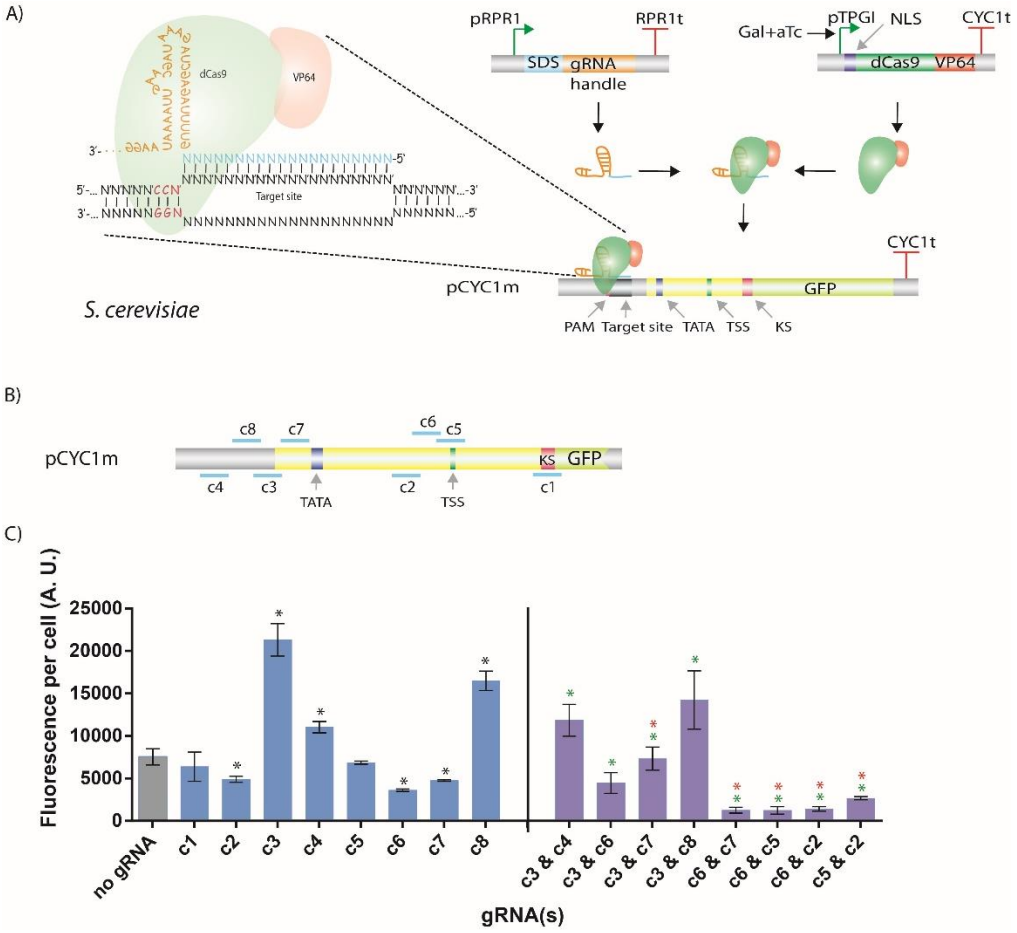


Figure 7.1 | A schematic view of the programmable CRISPR/Cas-based eukaryotic transcriptional regulation system implemented in *S. cerevisiae*. **(A)** crisprTF (dCas9_VP64) expression is induced by growing cells in galactose + aTc media. crisprTFs are guided to target sites by guide RNAs (gRNAs), which are constitutively expressed from the pRPR1 promoter and bind to the respective target sites. Specificity of crisprTFs is primarily determined by the 20 bp Specificity Determinant Sequence (SDS) at the 5'-end of the gRNA along with the presence of a PAM motif (NGG) at the target site. **(B)** Map of pCYC1m illustrating the relative positions of known regulatory elements. TATA: TATA box, TSS: Transcription Start Site, KS: Kozak Sequence. Blue lines indicate target sites for each gRNA (c1-c8). **(C)** Left panel: Regulation of *gfp* expression from pCYC1m by crisprTFs based on the individual gRNAs shown in (B). Yeast cells expressing crisprTFs and containing the reporter construct were transformed with plasmids expressing gRNAs labeled as shown in the x-axis. Targeting crisprTFs to sequences upstream of the TATA boxes (by c3, c4, and c8 gRNAs) resulted in

higher *gfp* expression than the no gRNA control. On the other hand, targeting crisprTFs to sequences spanning the TATA box and the Kozak sequence (by c1, c6, and c7 gRNAs) resulted in reduced *gfp* expression relative to the no gRNA control. Error bars indicate the standard error of the mean for three independent biological replicates. Asterisks (*) on each bar indicate statistically significant changes in *gfp* expression relative to the no gRNA control (based on the two-sided Welch's t-test, p-value < 0.05). Right panel: Co-expression of multiple gRNAs resulted in synergistic gene regulation. Pairwise combinations of non-neutral gRNAs were expressed from pRPR1 promoters on pRS423 and pRS425 backbones. Green and red asterisks (*) indicate statistically significant changes in *gfp* expression in samples with co-expressed gRNAs relative to the 1st gRNA only and the 2nd gRNA only, respectively (two-sided Welch's t-test, p-value < 0.05).

The expression of gRNAs targeting different regions in the pCYC1m (as shown in Fig. 7.1B) resulted in various statistically significant levels of reporter fluorescence compared to the no gRNA control (Fig. 7.1C, left panel). Targeting crisprTFs to the sequences upstream of the TATA boxes (by c3, c4, and c8 gRNAs) led to the activation of the reporter. However, targeting crisprTFs to the sequences spanning the TATA box and the Kozak sequence (KS) resulted in the repression of *gfp* expression to various degrees. Stronger repression was achieved when crisprTFs were targeted to the proximity of TATA box (using c7 gRNA) and to the vicinity of the TATA box and the transcription start site (using c2 and c6 gRNAs), likely due to the interference of crisprTFs with the formation of the transcriptional initiation complex (*189*). No activation was observed with any of the eight tested gRNAs when dCas9, without a fused activator domain, was targeted to pCYC1m (Fig. S7.1). All the tested gRNAs in this strain repressed *gfp* expression to some extent (Fig. S7.1), with the highest repression observed with c6 and c7. These results demonstrate that dCas9 is able to repress transcription but requires an activation domain (VP64) to activate transcription of a target locus and further supports the hypothesis that dCas9 (or as a fusion to VP64) can act as a repressor by interfering with the formation of the transcriptional initiation complex. Similar results were achieved with the GAL1 promoter (pGAL1) and its variants, where targeting crisprTFs to sequences upstream and downstream of TATA box led to activation and repression of the GFP reporter, respectively (Fig. S7.2). These results indicate that a single crisprTF can be programmed to act as both an activator and a repressor by targeting it to different positions across endogenous promoters.

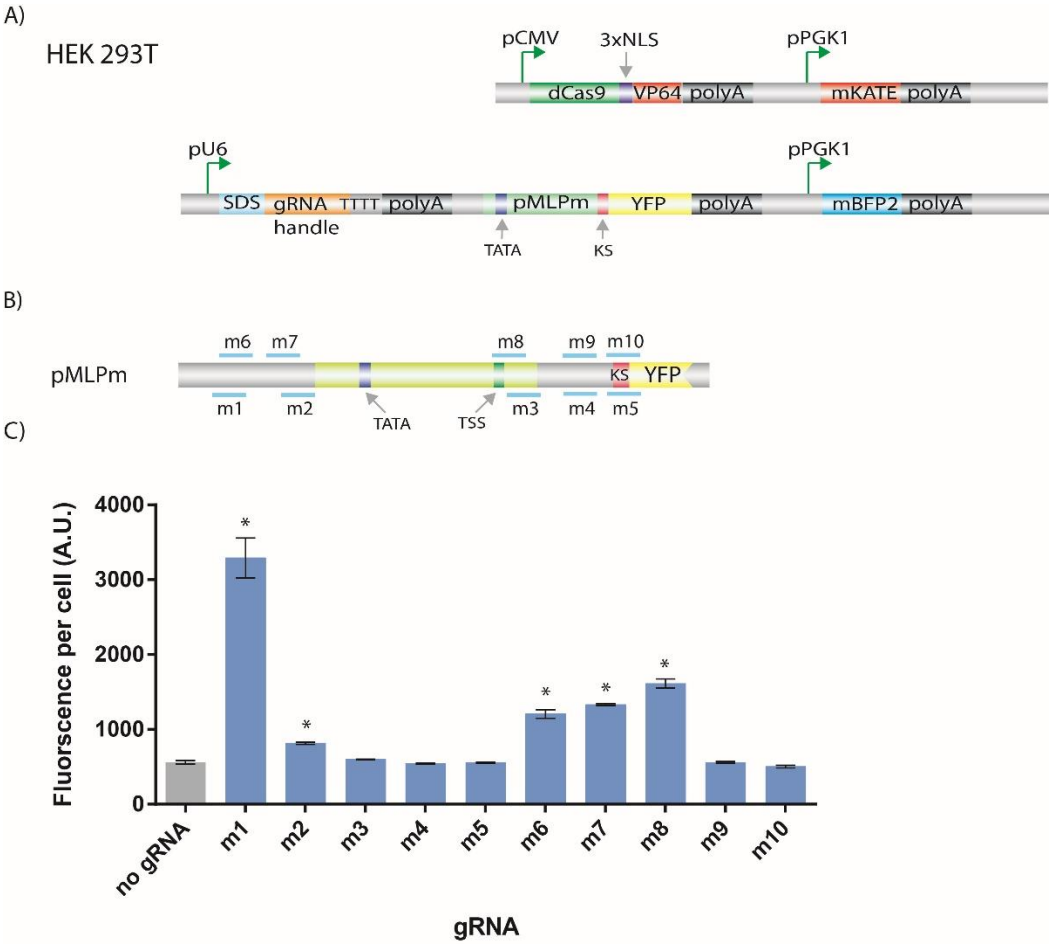


Figure 7.2 | Regulation of *yfp* expression from a minimal MLP promoter (pMLPm) by crisprTFs in HEK 293T cells. (A) dCas9_VP64 is expressed in HEK 293T cells by the pCMV promoter and directed to target sequences in pMLPm. The mKATE (red) and mBFP2 (blue) fluorophores act as flow-cytometry gating controls for successful plasmid transfections. **(B)** Map of pMLPm illustrating the relative positions of known regulatory elements. Blue lines indicate target sites for each gRNA. **(C)** Regulation of *yfp* expression from pMLPm by crisprTFs based on the gRNAs shown in (B). HEK 293T cells were co-transfected with the plasmids shown in (A), with specific gRNAs labeled as shown in the x-axis. Targeting crisprTFs to sequences upstream of the TATA box (by m1, m2, m6 and m7 gRNAs) resulted in higher *yfp* expression compared with the no gRNA control. Error bars indicate the standard error of the mean for three independent biological replicates. Asterisks (*) on each bar indicate statistically significant changes in *yfp* expression relative to the no gRNA control (based on the two-sided Welch's t-test, p-value < 0.05).

The activity of a promoter is determined by combinatorial interactions between transcriptional regulatory factors bound to that promoter. We thus investigated the effects of binding of multiple crisprTFs targeted to the same promoter. To this end, pairwise combinations of gRNAs with non-neutral effects (those that showed either activation or repression in the left panel of Fig. 7.1C) were co-expressed. As shown in the right panel of Fig. 7.1C, co-expression of repressor gRNAs resulted in synergistic repression of the reporter (up to 7x repression was achieved with co-expression of the

c5 & c6 pair as well as the c6 & c7 pair). On the other hand, when a repressor gRNA was co-expressed with an activator gRNA (e.g., the c3 & c6 pair), an intermediate level of GFP expression was achieved, indicating an antagonistic interaction between the two gRNAs. Moreover, the effects of repressor gRNAs were dominant over activator gRNAs, suggesting that interruption of the formation of the transcription initiation complex has a stronger effect than activation (*190*). Co-expression of two activator gRNAs (e.g. c3 & c4) did not result in synergistic activation of the reporter, which suggests that the relative positions and interactions of bound activators are important for determining synergistic activation (*190*). Consistent with our results, it has been shown that synergistic activation from synthetic promoters with multiple GAL4 operator sites depends on the distance and helical phase of the operator sites (*191*). Furthermore, in another study, it has been shown that not all of the combinations of TALE-activators targeted to the same promoter result in synergistic activation (*171*).

We next sought to investigate the activity of crisprTFs in human cells. To this end, a human-codon-optimized crisprTF cassette was placed on a plasmid under the control of the constitutive cytomegalovirus immediate-early promoter (pCMV). The gRNAs were expressed constitutively from a separate plasmid by the RNA polymerase III-dependent U6 promoter (pU6), as previously described (*182, 192*) (Fig. 7.2A). After transfection of these plasmids into HEK 293T cells, we investigated the regulatory architecture of the minimal adenovirus major late promoter (pMLPm) (*193*) by targeting crisprTFs to different positions across this promoter (Fig. 7.3A). YFP was used as the readout for pMLPm promoter activity.

Consistent with the results obtained in *S. cerevisiae*, crisprTFs activated gene expression when targeted to sequences upstream of the pMLPm TATA box (using m1, m2, m6, or m7 gRNA) or downstream of the transcription start site (m8 gRNA) (Fig. 7.2B and 7.2C). Since the basal expression level of the pMLPm promoter is low, it was challenging to detect significant repression from this promoter. Thus, to demonstrate that crisprTFs can function as transcriptional repressors in mammalian cells, we placed mKATE under the control of a constitutive mammalian promoter, phosphoglycerate kinase 1 (pPGK1) (*194*), and targeted crisprTFs to this promoter (Fig. 7.3A). pPGK1 is a strong, constitutive, TATA-less promoter that contains a CCAAT box (*195*) and five GC-boxes (*196*). These sites are the binding sites for the endogenous human transcription factors CBP and SP1, respectively. Targeting dCas9 alone, dCas9 fused to the VP64 domain, or dCas9 fused to KRAB domain to the CCAAT box or the GC-boxes resulted in significant repression of the reporter gene

(Fig. 7.3B), presumably by preventing endogenous transcription factors from binding to specific DNA recognition elements (*195, 196*) within the pPGK1 promoter.

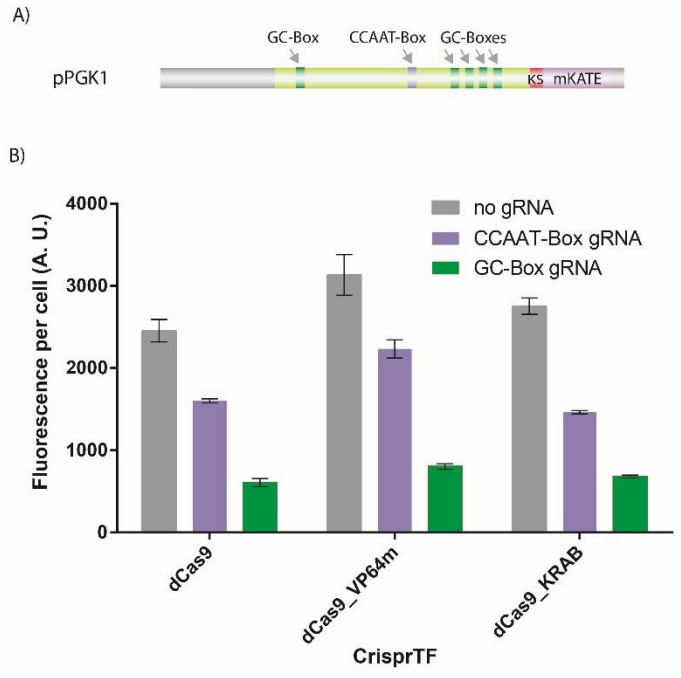


Figure 7.3 | CrisprTF-mediated repression of the constitutive pPGK1 promoter in HEK 293T cells. (A) Map of the pPGK1 promoter illustrating the relative positions of known regulatory elements. **(B)** CrisprTF-based targeted repression of the constitutive pPGK1 promoter. Constructs expressing different dCas9-based proteins (dCas9, dCas9_VP64, and dCas9_KRAB) were co-transfected with plasmids containing pPGK1_mKATE and constructs expressing no gRNAs or gRNAs targeting the CCAAT box or the GC-box gRNA. Significant repression of the pPGK1 promoter relative to the no gRNA control was observed with all of the three different dCas9 constructs (dCas9, dCas9-VP64, and dCas9-KRAB). Error bars indicate the standard error of the mean for three independent biological replicates.

We further sought to explore the tunability of crisprTFs in the context of synthetic promoters. In order to do so, we engineered multiple artificial binding sites (operators), separated by twenty base-pair sequences, upstream of the pCYC1m in *S. cerevisiae* (Fig. 7.4A). Expression of a gRNA which targeted these arrayed operator sites resulted in synergistic activation of the *bfp* reporter (Fig. 7.4B). Higher levels of *bfp* expression (up to 70-fold activation with 12x gRNA operator sites) were achieved by increasing the number of gRNA binding sites upstream of the engineered pCYC1m. This level of activation in yeast is comparable to the activation reported for commonly used endogenous yeast promoters (e.g. pGAL1 (*197*) and pCUP1 (*198*)) and synthetic promoters that are modulated by engineered bacterial DNA binding domains (e.g. TetON/TetOFF promoters (*162*)). Our results are consistent with previous observations that arraying multiple binding sites for a transcriptional activator

upstream of a promoter results in longer transcriptional bursts from the promoter and thus leads to higher levels of expression from the targeted promoter (*199*). We saw similar synergistic activation in HEK 293T cells when multiple gRNA operator sites were placed upstream of pMLPm (Figure 7.4C and 7.4D), with up to 56x activation attained with 3x gRNA operator sites. The level of activation that is achieved by crisprTFs in human cells is comparable to the levels of activation reported for ZF- and TALE-activators (*165, 170, 171*), where higher activation levels can be achieved by increasing the number of operator sites or by targeting multiple synthetic transcription factors to the same locus (*171, 173*). These results demonstrate that crisprTFs can be used to build synthetic promoters with tunable strengths by the straightforward engineering of gRNA-binding sites.

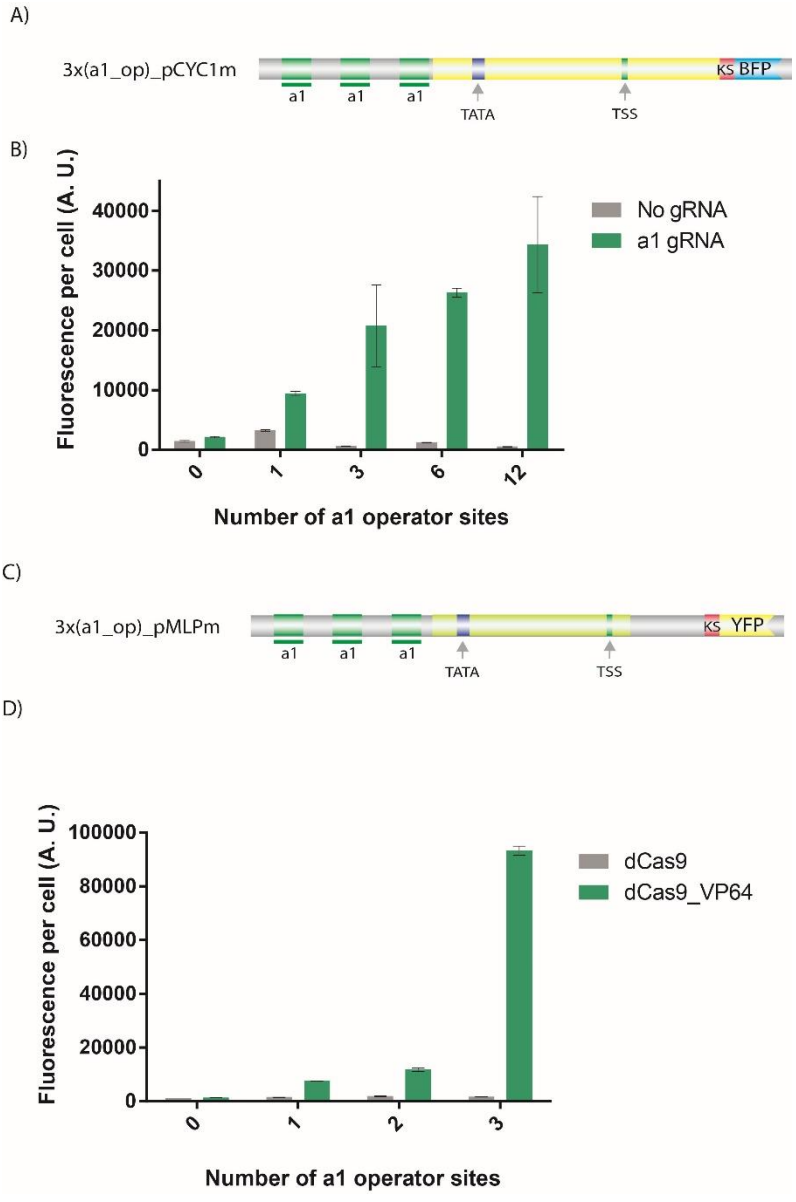


Figure 7.4 | Synergistic and tunable activation of synthetic promoters with arrayed operator sites upstream of pCYC1m in *S. cerevisiae* and pMLPm in HEK 293T cells using crisprTFs. **(A)** A schematic view of the pCYC1m synthetic promoter with three a1_gRNA operator sites (3x DNA sequences recognized by the a1_gRNA) arrayed upstream of pCYC1m, thus named 3x(a1_op)_pCYC1m. **(B)** Increasing the number of arrayed a1_gRNA operator sites upstream of pCYC1m resulted in higher *bfp* expression in *S. cerevisiae* cells expressing the a1_gRNA compared to the no gRNA controls. Error bars indicate the standard error of the mean for three independent biological replicates. **(B)** A schematic view of the pMLPm synthetic promoter with three a1_gRNA operator sites arrayed upstream of pMLPm, thus named 3x(a1_op) _pMLPm. **(D)** Increasing the number of arrayed a1_gRNA operator sites upstream of pMLPm resulted in higher *yfp* expression in HEK 293T cells when co-transfected with a1_gRNA and dCas9_VP64 versus when co-transfected with a1_gRNA and dCas9. Error bars indicate the standard error of the mean for three independent biological replicates.

In many applications, control of the activity of a transcription factor by an inducer (e.g., a small molecule) is desired. With crisprTFs, one viable strategy is to constitutively express the protein component of the system (i.e., dCas9) and then modulate the amount of gRNA available for binding to dCas9 and thus the activity achieved at the target DNA. To test this strategy, we constructed an anhydrotetracycline (aTc)-inducible pRPR1 promoter by placing a TetR operator site (1xTetO) in the pRPR1 promoter, as previously described (*200*), and constitutively expressing Tet repressor (TetR) (Fig. 7.5A). Furthermore, to make the expression of dCas9_VP64 independent of aTc, we placed it under the control of pGAL1. We tested this system in cells containing a pCYC1m promoter with six a1_gRNA operator sites, named 6x(a1_op) _pCYC1m, controlling the expression of *bfp*. As shown in Fig. 7.5B, *bfp* expression increased about 20-fold when *S. cerevisiae* cells were induced with galactose and aTc, compared with galactose only, thus demonstrating external control of the crisprTF activity.

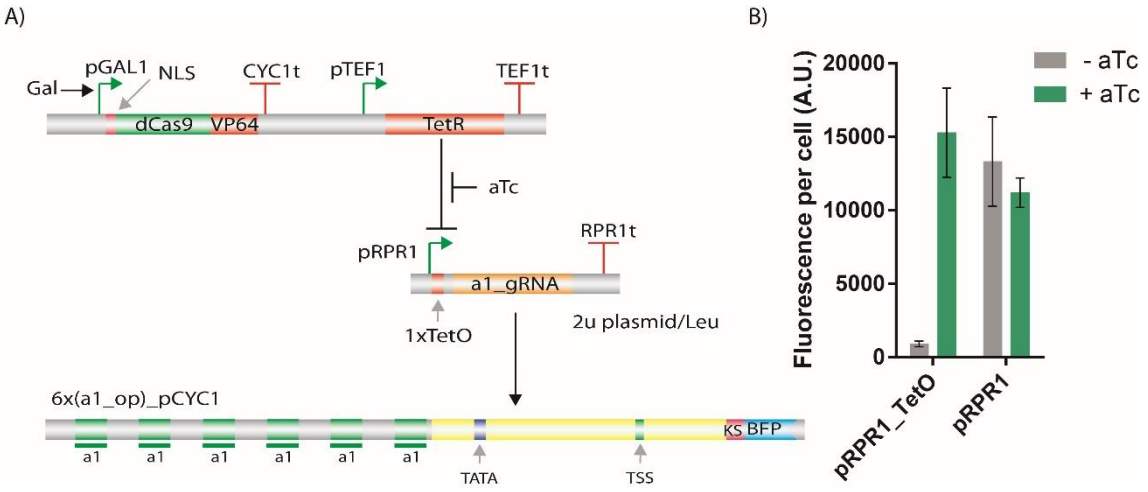


Figure 7.5 | Inducible crisprTF-guided activation of synthetic promoters. (A) Schematic of the aTc-inducible pRPR1_TetO promoter. Expression of dCas9_VP64 is driven by the galactose-inducible pGAL1 promoter. A TetR operator site (1xTetO) was placed in the pRPR1 promoter to make an aTc-responsive pRPR1_TetO promoter. Addition of aTc releases TetR-mediated repression on the pRPR1_TetO promoter and results in a1_gRNA expression. **(B)** aTc-dependent *bfp* expression from a synthetic 6x(a1_op) _pCYC1m promoter. *S. cerevisiae* cells containing the circuit shown in (A) were grown in galactose media with either 250 ng/mL aTc or no aTc. Error bars indicate the standard error of mean for three biological replicates.

Although the exact parameters that determine the sequence specificity of Cas9 are not yet well-characterized, it has been shown that the specificity of Cas9 for target loci is mainly determined by the PAM motif (NGG) and the 12 base-pairs preceding this motif (seed sequence). Mutations in any of these fifteen positions can severely affect the targeting efficiency and binding specificity of Cas9 (*181, 182, 201, 202*). To test whether this property can be used to create orthogonal synthetic promoters, new PAM motifs (marked with asterisks, Fig. 7.6A) or single or multiple base-pair mismatches (marked with dashes, Fig. 7.6A) were introduced into the wild-type pCYC1m sequence to design a modified pCYC1m promoter (pCYC1m(modified), Table S7.1)). As shown in the left panel of Fig. 7.6B, the modified promoter did not respond to the wild-type gRNAs (except for c4, which still perfectly matched target sites within this modified promoter). However, this modified promoter responded to a new set of gRNAs (cm1, cm2, and cm5) that were designed to match the modified target sites. The wild-type pCYC1m promoter did not respond to the new set of gRNAs (cm1-cm6) (Fig. 7.6B, left panel). These results demonstrate that as little as a single base-pair mismatch is sufficient to direct the crisprTF to one locus while preventing activity at another locus.

To further demonstrate the potential of crisprTFs towards constructing synthetic promoters and gRNAs that are orthogonal with respect to each other, we tested three randomly designed gRNAs (a1, a2 and a3 gRNAs) for their ability to activate each other's target sequences. As shown in Fig. 7.6C, each of the gRNAs exhibited high activity at their cognate target sequences but low activity at non-cognate sequences. These results suggest that one can construct synthetic promoters and gRNAs that are orthogonal with respect to each other and to the host genome, especially within eukaryotes with smaller genomes, such as yeasts.

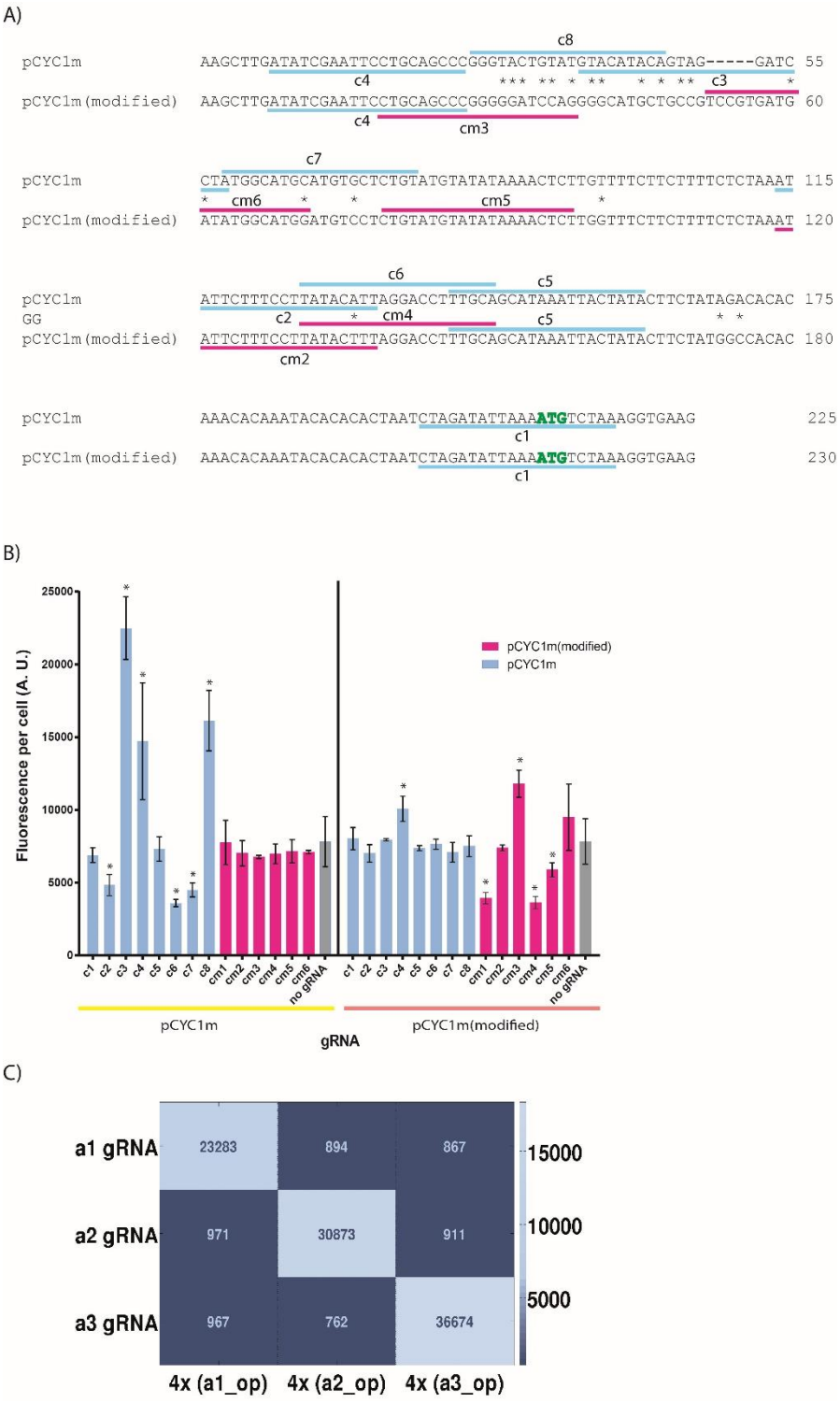


Figure 7.6 | Constructing orthogonal crisprTF-responsive promoters. (A) A schematic view of gRNAs targeting the wild-type and modified pCYC1m promoters in yeast. Only the c1-c8 gRNAs have perfect homology to target sequences in pCYC1m. On the other hand, c1, c4, c5, and cm1-cm6 gRNAs have perfect homology to the sequences in the pCYC1m(modified) promoter. Mismatches between pCYC1m and pCYC1m(modified) are marked by asterisks (*). **(B)** pCYC1m only responds to gRNAs that are perfectly matching gRNAs (c1-c8 gRNAs) and not to those that contain mismatches (cm1-cm6 gRNAs). The pCYC1m(modified) promoter

responds to the cm1-cm6 gRNAs. Those gRNAs that bind to the sequences upstream of the TATA boxes activate *gfp* expression and those that target sequences downstream of the TATA boxes repress *gfp* expression. The c1 and c5 gRNAs have similarly neutral effects on both the wild-type and modified promoters. Error bars indicate standard error of the mean for three independent biological replicates. Asterisks (*) on each bar indicate statistically significant activation or repression relative to no gRNA controls (based on the two-sided Welch's t-test, p-value < 0.05). (C) Heat map illustrating the orthogonality of crisprTFs in human cells. Plasmids encoding three orthogonal gRNAs (a1, a2 and a3 gRNAs) were co-transfected into HEK293T cells along with one of the three reporter plasmids (each encoding 4x operator sites for a given gRNA) upstream of the pMLPm promoter driving *yfp* expression. Only cognate interactions between gRNAs and target binding sites resulted in significant activation. The standard error of the mean of YFP fluorescence for three independent biological replicates is indicated in each cell of the heat-map plot.

During the course of the peer-review for this work, similar systems for transcriptional control in eukaryotic cells were described. Gilbert et al. (*203*) demonstrated that CRISPR-mediated gene repression and activation can be achieved in both yeast and mammalian cells by using fusions of dCas9 with repressor and activator domains respectively. Furthermore, Maeder et al. (*204*) and Perez-Pinera et al. (*205*) showed synergistic CRISPR/Cas-based gene activation in human cells with multiple gRNAs. In this paper, we additionally show that both activation and repression functions can be achieved with only one transcription factor, by targeting dCas9_VP64 fusions to different regulatory sequences along a promoter. Our finding that one can activate or repress the expression of a gene of interest by directing a single protein to different positions of a promoter is advantageous for the efficient design of synthetic transcriptional networks or rewiring natural ones. This property obviates the need for using separate orthogonal Cas9 protein fusions as activators and repressors. Furthermore, we demonstrate that more sophisticated regulatory motifs, such as small-molecule responsive modules can be built for crisprTFs, thus enabling external control of crisprTF-based transcriptional circuits. Such synthetic modules can be interfaced with other regulatory elements to achieve more complex regulation for synthetic biology.

7.4 Discussion

Our results show that dCas9 can be used as a customizable RNA-guided DNA-binding platform for the regulation of gene expression at natural and synthetic promoters in eukaryotic cells. The ease of design and expression of customized gRNAs in comparison to ZFs and TALEs make CRISPR-based transcription factors appealing as synthetic TFs for modulating endogenous gene expression as well as for synthetic biology. The ability to customize the target site of dCas9 via the expression of short gRNAs obviates the need to engineer multiple orthogonal DBDs in order to construct complex

transcriptional circuits. This could potentially reduce the overall metabolic burden on cells and enable the integration of more complex synthetic computation and logic within living cells (*41, 206*). More complex regulatory and logic circuits, such as cascades and complex digital logic gates can be built by layering crisprTFs. The possibility of integrating multiple inputs at a single promoter expands the regulatory potential and provides us with increased flexibility that can be leveraged while designing synthetic transcriptional networks or rewiring endogenous pathways.

Furthermore, since both activation and repression functions can be achieved with crisprTFs, the crisprTF platform may be advantageous compared to non-coding RNA-based gene regulatory platforms where only repression can be achieved. In a way, crisprTFs combine the multiplexability of RNA-based regulatory approaches with the flexibility and rich functionality repertoire of protein-based gene regulatory approaches: Cas9 can be functionalized with regulatory domains of interest (e.g., activation, repression, or epigenetic effector) and then be targeted to multiple loci using different gRNAs.

Future work is needed to define the range of effector domains that can be used with dCas9 for a variety of regulatory functions, including transcriptional regulation and epigenetic modifications. In addition, the identification, characterization, and optimization of Cas9 homologs or evolved variants may enable enhanced activity and specificity of this system. Moreover, the ability to synthesize random libraries of gRNAs opens the possibility for high-throughput perturbations of transcriptional networks and screening for desirable phenotypes. Ultimately, we envision that crisprTFs will enable the regulation and perturbation of natural transcriptional networks as well as the construction of complex synthetic circuits at an unprecedented speed and scale.

7.5 Supplementary Information

Materials and Methods

Strain and Plasmid Construction

Saccharomyces cerevisiae

dCas9 (endonuclease-deficient Cas9, with D10A and H841A mutations relative to the wild-type sequence of *S. pyogenes* Cas9 (*69*)) with an N-terminal SV40 nuclear localization signal (NLS) was codon-optimized for expression in *S. cerevisiae* and cloned into a pRS314 backbone under control of the pTPGI promoter (*169*). The RNA-guided transcription factors (crisprTFs) were built by fusing four repeats of the minimal domain of the herpes simplex viral protein 16 (VP16) to the C-terminus of dCas9

(dCas9_VP64). The crisprTF-expressing plasmid was then integrated into the TRP1 locus of *S. cerevisiae* W303.

The reporter plasmids were built by cloning yeast-enhanced *gfp* under the control of the wild-type or modified pCYC1m promoter into pRS406 using one-step Gibson assembly. The reporters for the multiple-gRNA-binding-site experiment (Fig. 7.4A) were built by cloning the corresponding number of binding sites upstream of the pCYC1m promoter driving production of EBFP2. All reporters were integrated into the bla1 locus of the integrated crisprTF plasmid.

To build gRNA-expressing plasmids, empty gRNA expressing vectors were first made by cloning the pRPR1 promoter (an RNA-polymerase-III-dependent promoter (*188*)), the gRNA handle (flanked by HindIII and Xho1 sites), and the RPR terminator into the SacI and KpnI sites of either the pRS423 or pRS425 plasmid using one-step Gibson assembly. The specificity determinant sequence (SDS) for each gRNA was then cloned into the HindIII site of these vectors by one-step Gibson assembly. Sequences of the constructs used in this study are listed in Table S7.1.

HEK 293T Cells

To construct the mammalian dCas9_VP64 expressing plasmid, we first introduced D10A and H841A mutations into hCas9 (*182*) (Addgene, Plasmid #41815). Then, three repeats of SV40 NLS (3xNLS) were fused to the C-terminus of the mutated hCas9 using a PCR-based assembly protocol. Using a multi-part Gibson assembly protocol, the immediate-early promoter of cytomegalovirus (pCMV), dCas9_3xNLS, VP64, and SV40 polyA terminator were cloned into the NotI site of the pG5-Luc plasmid (Promega). To monitor successfully transfected cells by flow cytometry, we replaced the original luciferase gene in pG5-Luc with mKATE (Evrogen). The resulting pPGK1_mKATE cassette served as a constitutive fluorescent protein control that was used to gate for the presence of the crisprTF-expressing plasmid with flow cytometry.

The gRNA expression plasmids were constructed by cloning the 138 bp human U6 promoter (an RNA-polymerase-III-dependent promoter (*192*)), along with the gRNA handle and terminator into a plasmid containing pPGK1-eBFP2 flanked by the SV40 polyA terminator (a gift from Lior Nissim). A SacI site was placed at the 3'-end of the U6 promoter to enable the cloning of different specificity determining sequences for each gRNA. The reporters were assembled into the gRNA-expressing plasmid through a one-step Gibson assembly reaction, where the upstream polyadenylation signal and transcriptional pause site from pG5-Luc, along with a 41 bp, minimal adenovirus type

2 major late promoter (pMLPm), mYFP, and HSV polyA signal were cloned into the AatII site of the gRNA-expressing plasmids.

For the synthetic promoter experiments, additional gRNA operator sites were cloned in the NheI site upstream of the pMLPm promoter (see Supplementary Information). For the repression experiments, dCas9_KRAB was constructed by cloning a 366 bp KRAB domain to the C-terminus of dCas9. GCCACC was used as the Kozak sequence for the expression of dCas9_VP64, mYFP, eBFP2, and mKATE.

Unless directly targeted by gRNAs for repression assays, the mKATE fluorescent protein on the crisprTF-expression plasmid and the eBFP2 fluorescent protein on the reporter/gRNA plasmid served as our gating controls for flow cytometry analysis.

Fluorescence Assays

To assess expression of the reporter constructs, yeast cells expressing different gRNAs (or no gRNA as control) were grown overnight (900 rpm, 30°C) in 96-deep-well plates in yeast minimal media supplemented with glucose with appropriate selection (three independent cultures for each sample). 10 μL of these cultures were then transferred into fresh media supplemented with galactose + 250 ng/ml anhydrotetracycline (aTc) and grown for 20 hours (900 rpm, 30°C) before analysis by flow cytometry.

For the human cell culture experiments, HEK 293T kidney epithelial cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS), 1% glutamine, and 1% penicillin/streptomycin. Cells were grown under 5% CO2 at 37°C. HEK 293T cells were transfected with Fugene-HD transfection reagent (Promega) and assayed for gene expression with flow cytometry at 48 hours post-transfection.

An LSR Fortessa II flow cytometer equipped with 405 nm, 488 nm, and 561 nm lasers was used for all the experiments. GFP/YFP, BFP, and mKATE levels were detected using 488/FITC, 405/Pacific-Blue, and 561/TX-red laser/filter sets, respectively. All samples were uniformly gated by forward and side scatter. Additional gating for the presence of red and blue fluorophores was applied to the HEK 293T samples to ensure only cells successfully transfected with both the crisprTF and the reporter/gRNA plasmids are analyzed. For each gated sample, the mean fluorescence per cell was calculated. Three independent biological samples were used to calculate the mean and standard error of the mean for each data point.

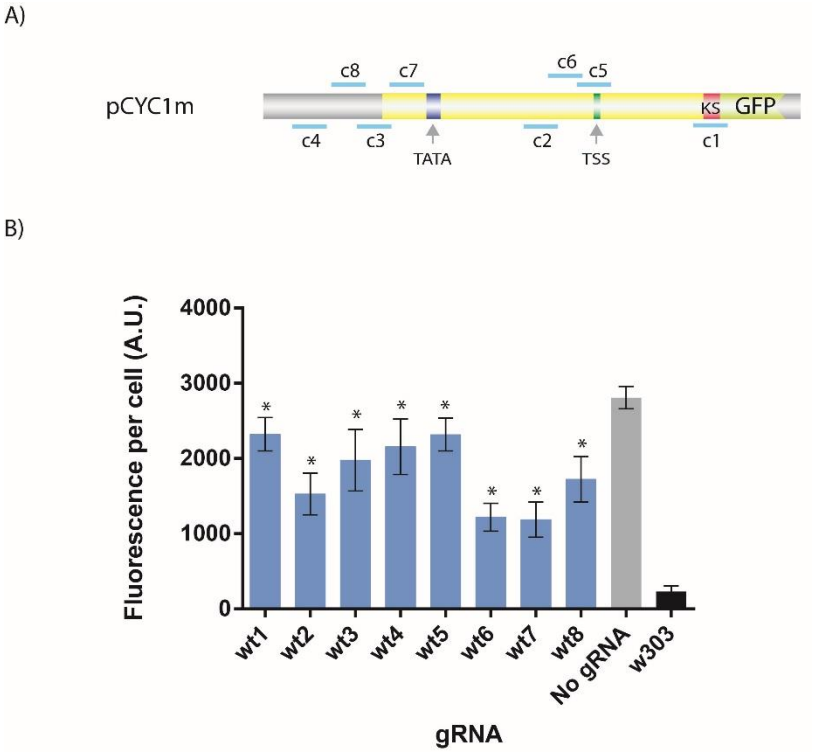


Figure S7.1 | Effect of dCas9 alone (without an attached VP64 domain) on *gfp* expression when targeted to the pCYC1m promoter. (A) A map of gRNA target sites on pCYC1m used in this experiment. **(B)** Cells expressing dCas9 and containing pCYC1m driving *gfp* expression were transformed with vectors expressing different gRNAs whose target sites are shown in (A). All the tested gRNAs showed various levels of repressive or neutral effects, suggesting that dCas9 by itself cannot activate the pCYC1m promoter. The greatest repression was observed with c6 and c7 gRNAs, as was seen with dCas9_VP64 (Fig. 7.1). Error bars indicate standard error of the mean for three independent biological replicates. Asterisks (*) on each bar indicate statistically significant changes in GFP expression relative to the no gRNA control (based on the two-sided Welch's t-test, p-value < 0.05).

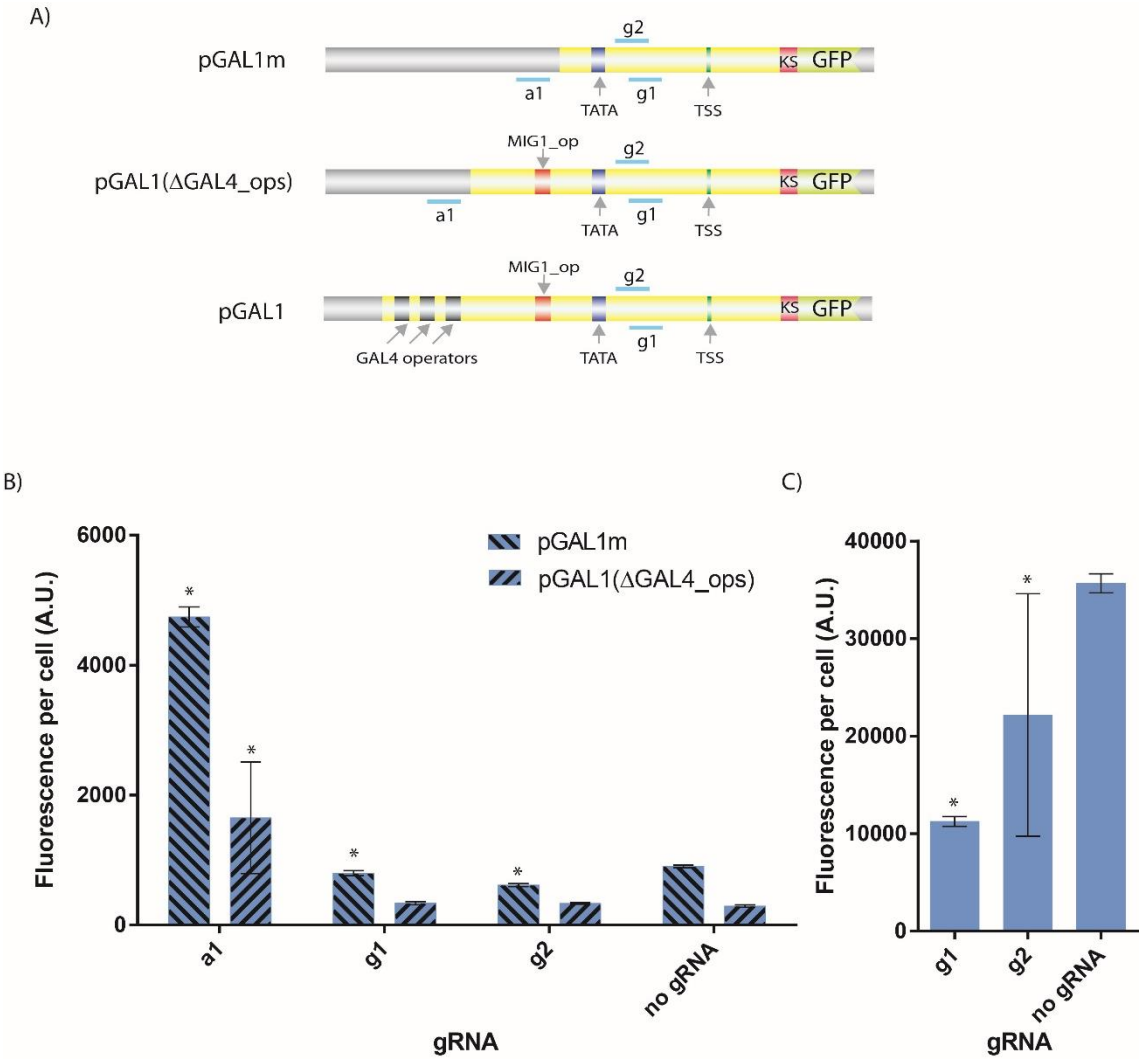


Figure S7.2 | Regulation of *gfp* expression from pGAL1, pGAL1 lacking GAL4 operator sites (pGAL1(ΔGAL4_ops)) and minimal pGAL1 (pGAL1m) promoters by crisprTFs. (A) Maps of pGAL1m, pGAL1(ΔGAL1_ops) and pGAL1 promoters illustrating the relative positions of binding sites for endogenous pGAL1 regulatory factors (GAL4 and MIG1) along the promoters. Blue lines indicate target sites for tested gRNAs. **(B)** *S. cerevisiae* cells expressing crisprTFs and containing the pGAL1(ΔGAL4_ops) or pGAL1m driving *gfp* as reporter constructs (as indicated in (A)) were transformed with plasmids expressing gRNAs labeled as shown in the x-axis. Targeting crisprTFs to sequences upstream of the TATA boxes (by a1_gRNA) resulted in higher *gfp* expression than the no gRNA control. On the other hand, targeting crisprTFs to sequences downstream of the TATA boxes (by g1 and g2 gRNAs) resulted in reduced *gfp* expression relative to the no gRNA control. Error bars indicate the standard error of the mean for three independent biological replicates. Asterisks (*) on each bar indicate statistically significant changes in GFP expression relative to the no gRNA control (based on the two-sided Welch's t-test, p-value < 0.05). **(C)** Targeting crisprTF to sequences downstream of TATA box in pGAL1 (with g1 and g2 gRNAs) result in the repression of *gfp* expression from pGAL1. Error bars indicate the standard error of the mean for three independent biological replicates. Asterisks (*) on each bar indicate statistically significant changes in GFP expression relative to the no gRNA control (based on the two-sided Welch's t-test, p-value < 0.05).

Table S7.1: Sequences of the constructs used in *S. cerevisiae* experiments.

Name	Sequence
NLS_dCas9_VP64	ATGTCTAGAGCCCCAAAGAAGAAGAGAAAAGTTAGACCCGGGGATAA GAAATACTCTATTGGTTTGGCTATCGGTACAAACTCTGTTGGTTGGG CTGTTATTACTGATGAATACAAGGTTCCATCCAAGAAGTTCAAGGTTT TGGGTAACACTGATAGACACTCCATCAAAAAGAACTTGATTGGTGCC TTGTTGTTCGATTCTGGTGAAACTGCTGAAGCTACTAGATTGAAAAG AACCGCTAGAAGAAGATACACCAGAAGAAAGAACAGAATCTGCTACT TGCAAGAAATCTTCTCCAACGAAATGGCCAAGGTTGATGATTCATTCT TCCACAGATTGGAAGAATCCTTCTTGGTCGAAGAAGATAAGAAGCAC GAAAGACATCCAATCTTCGGTAACATCGTTGATGAAGTTGCTTACCAC GAAAAGTACCCAACTATCTACCATTTGAGAAAGAAGTTGGTTGACTC TACCGATAAGGCTGATTTGAGATTGATCTATTTGGCTTTGGCCCACAT GATTAAGTTCAGAGGTCATTTCTTGATCGAAGGTGATTTGAACCCAG ATAACTCCGATGTTGATAAGTTGTTCATCCAATTAGTCCAAACCTACA ATCAATTATTCGAAGAAAACCCAATCAACGCCTCTGGTGTTGATGCTA AAGCTATTTTGTCTGCCAGATTGTCCAAGTCCAGAAGATTAGAAAAT TTGATCGCCCAATTACCAGGTGAAAAGAAGAATGGTTTGTTCGGTAA TTTGATTGCCTTGTCTTTGGGTTTGACTCCAAACTTCAAGTCCAATTT CGATTTGGCTGAAGATGCCAAGTTGCAATTATCTAAGGATACCTACG ATGACGATTTGGATAACTTGTTGGCTCAAATCGGTGATCAATACGCT GATTTGTTTTTGGCTGCTAAGAACTTGTCCGATGCCATTTTGTTGTCC GATATTTTGAGAGTCAACACCGAAATTACTAAGGCTCCATTGTCTGCC TCTATGATCAAAAGATACGATGAACACCACCAAGACTTGACTTTGTTG AAGGCTTTGGTCAGACAACAATTACCTGAAAAGTACAAAGAAATTTT CTTCGATCAATCCAAGAACGGTTACGCCGGTTATATTGATGGTGGTG CTTCTCAAGAAGAATTTTACAAGTTCATCAAGCCAATCTTGGAAAAGA TGGACGGTACTGAAGAATTATTGGTCAAGTTGAACAGAGAAGATTTG TTGAGAAAGCAAAGAACCTTCGACAACGGTTCTATTCCACATCAAATT CACTTGGGTGAATTGCACGCAATTTTGAGAAGACAAGAAGATTTTTA TCCATTCTTGAAGGACAACAGAGAAAAGATCGAAAAGATTCTGACCT TCAGAATCCCTTACTACGTTGGTCCATTGGCTAGAGGTAATTCAAGAT TTGCCTGGATGACTAGAAAGTCCGAAGAAACTATTACTCCTTGGAAC TTCGAAGAAGTTGTAGATAAGGGTGCTTCTGCCCAATCCTTTATTGA AAGAATGACCAACTTCGACAAGAACTTGCCAAACGAAAAGGTTTTGC CAAAGCACTCTTTGTTGTACGAATACTTCACCGTCTACAACGAATTGA CTAAGGTTAAGTACGTCACCGAAGGTATGAGAAAACCAGCTTTTTTA TCCGGTGAACAAAAGAAGGCTATCGTCGATTTGTTGTTCAAGACCAA CAGAAAGGTTACTGTCAAGCAATTAAAAGAAGATTACTTCAAGAAAA TCGAATGCTTCGACTCCGTTGAAATTTCTGGTGTCGAAGATAGATTC AATGCCTCTTTAGGTACTTACCATGACTTGTTGAAAATCATCAAGGAC AAGGATTTCTTGGACAACGAAGAAAACGAAGATATTTTGGAAGATAT TGTCTTGACATTGACCTTGTTTGAAGATAGAGAAATGATTGAAGAAA GATTGAAAACCTACGCCCACTTGTTCGATGATAAGGTTATGAAGCAA TTAAAGAGAAGAAGATACACTGGTTGGGGTAGATTGTCCAGAAAATT GATTAACGGTATCAGAGACAAGCAATCCGGTAAGACCATTTTGGACT TTTTGAAGTCTGATGGTTTCGCTAACAGAAACTTCATGCAATTAATCC ACGACGATTCCTTGACTTTCAAAGAAGATATACAAAAGGCCCAAGTCT CTGGTCAAGGTGATTCTTTACATGAACATATCGCTAACTTGGCTGGTT CTCCAGCTATTAAGAAGGGTATTTTACAAACCGTTAAGGTCGTTGAC GAATTGGTCAAAGTTATGGGTAGACATAAGCCAGAAAACATCGTTAT CGAAATGGCTAGAGAAAATCAAACCACCCAAAAGGGTCAAAAGAACT CCAGAGAAAGAATGAAGAGAATCGAAGAAGGTATCAAAGAATTGGGT TCCCAAATTTTGAAAGAACACCCAGTTGAAAACACCCAATTACAAAAC GAAAAGTTGTACTTGTACTACTTGCAAAACGGTAGAGATATGTACGT

	TGACCAAGAATTGGACATCAACAGATTGTCTGATTACGATGTTGACG CTATCGTTCCACAATCTTTTTTGAAGGATGACTCCATTGACAACAAGG TCTTGACTAGATCCGATAAGAATAGAGGTAAGTCCGATAACGTTCCA TCTGAAGAAGTCGTTAAGAAAATGAAGAACTATTGGAGACAATTATT GAACGCCAAGTTGATCACCCAAAGAAAGTTTGACAATTTGACCAAGG CTGAAAGAGGTGGTTTGTCTGAATTGGATAAGGCAGGTTTCATCAAA AGACAATTAGTAGAAACCAGACAAATCACCAAGCACGTTGCTCAAAT TTTGGATAGTAGAATGAACACTAAGTACGACGAAAACGACAAATTGA TCAGAGAAGTTAAGGTCATTACCTTGAAGTCCAAGTTGGTTTCCGAT TTCAGAAAGGACTTCCAATTCTACAAGGTCAGAGAAATCAACAACTAC CATCATGCACATGATGCTTACTTGAATGCTGTTGTTGGTACTGCCTTG ATTAAGAAGTATCCAAAGTTGGAATCCGAATTTGTCTACGGTGATTA CAAGGTTTACGACGTTAGAAAGATGATCGCCAAGTCCGAACAAGAAA TTGGTAAAGCTACTGCCAAATACTTCTTCTACTCCAATATTATGAATT TCTTTAAGACCGAAATCACTTTGGCCAACGGTGAAATTAGAAAAAGA CCATTGATTGAAACTAATGGTGAAACAGGTGAAATCGTTTGGGATAA GGGTAGAGATTTTGCCACTGTTAGAAAGGTATTGTCCATGCCACAAG TAAACATCGTCAAAAAGACCGAAGTTCAAACTGGTGGTTTCTCCAAA GAATCCATTTTGCCTAAGAGAAACTCCGATAAGTTGATCGCTAGAAA AAAAGACTGGGACCCAAAAAAGTACGGTGGTTTTGATTCTCCAACTG TTGCTTACTCTGTTTTGGTTGTTGCTAAGGTCGAAAAGGGTAAGAGT AAGAAGTTGAAGTCCGTCAAAGAATTATTAGGTATCACTATCATGGA AAGATCCTCATTCGAAAAGAATCCTATCGACTTTTTGGAAGCCAAGG GTTACAAAGAAGTCAAGAAGGACTTGATCATTAAGTTGCCAAAGTAC AGTTTGTTCGAATTGGAAAATGGTAGAAAGAGAATGTTGGCTTCTGC CGGTGAATTACAAAAGGGTAATGAATTGGCTTTGCCATCCAAGTACG TTAATTTCTTATACTTGGCCTCCCACTACGAAAAATTGAAAGGTTCTC CTGAAGATAACGAACAAAAGCAATTATTTGTCGAACAACACAAGCAC TACTTGGACGAAATCATTGAACAAATTTCCGAATTTTCCAAAAGAGTC ATTTTGGCTGACGCCAATTTGGACAAAGTTTTGTCAGCTTACAACAA GCACAGAGATAAGCCAATTAGAGAACAAGCTGAAAACATCATTCACT TGTTCACTTTGACTAACTTGGGTGCTCCAGCTGCTTTTAAGTATTTCG ATACCACTATCGACAGAAAGAGATACACCTCTACCAAAGAAGTTTTG GACGCTACTTTGATCCACCAATCTATTACTGGTTTGTACGAAACTAGA ATCGACTTGTCTCAATTAGGTGGTGATGGTTCTGGTAGATCTGGAGT CGACGGTGGAGGTTCTGACGCTTTGGACGACTTCGACTTGGATATGC TGGGTTCTGATGCGCTAGATGACTTTGACCTCGACATGCTTGGAAGT GACGCCTTAGATGATTTTGACCTGGATATGCTTGGATCAGACGCTCT GGACGATTTCGACTTAGACATGCTTTCCTAG
pCYC1m	AAGCTTGATATCGAATTCCTGCAGCCCGGGTACTGTATGTACATACA GTAGGATCCTATGGCATGCATGTGCTCTGTATGTATATAAAACTCTT GTTTTCTTCTTTTCTCTAAATATTCTTTCCTTATACATTAGGACCTTT GCAGCATAAATTACTATACTTCTATAGACACACAAACACAAATACACA CACTAATCTAGATATTAAAATGTCTAAAGGTGAAG
pCYC1m(modified)	AAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCAGGGGCATGCT GCCGTCCGTGATGATATGGCATGGATGTCCTCTGTATGTATATAAAA CTCTTGGTTTCTTCTTTTCTCTAAATATTCTTTCCTTATACTTTAGGA CCTTTGCAGCATAAATTACTATACTTCTATGGCCACACAAACACAAAT ACACACACTAATCTAGATATTAAAATGTCTAAAGGTGAAG
0xa1_pCYC1m	AAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCTATGGCATGCA TGTGCTCTGTATGTATATAAAACTCTTGTTTTCTTCTTTTCTCTAAAT ATTCTTTCCTTATACATTAGGACCTTTGCAGCATAAATTACTATACTT CTATAGACACACAAACACAAATACACACACTAATCTAGATAAAAAATG GTTTCCAAG

1xa1_pCYC1m	AAGCTTGATATCGAATTCCTGCAGCCCGGGTACTGTATGTACATACA GTAGGATCCTATGGCATGCATGTGCTCTGTATGTATATAAAACTCTT GTTTTCTTCTTTTCTCTAAATATTCTTTCCTTATACATTAGGACCTTT GCAGCATAAATTACTATACTTCTATAGACACACAAACACAAATACACA CACTAATCTAGATAAAAAATGGTTTCCAAG
3xa1_pCYC1m	AAGCTTGATATCGAATTCCTGCAGCCCGGGTACTGTATGTACATACA GTAGCTCGAAAATATTAATTACTGTATGTACATACAGTAGATCCTCGA AAATATTAATTACTGTATGTACATACAGTACTATGGCATGCATGTGCT CTGTATGTATATAAAACTCTTGTTTTCTTCTTTTCTCTAAATATTCTT TCCTTATACATTAGGACCTTTGCAGCATAAATTACTATACTTCTATAG ACACACAAACACAAATACACACACTAATCTAGATAAAAAATGGTTTCC AAGGGTGAAG
6xa1_pCYC1	AAGCTTGATATCGAATTCCTGCAGCCCGGGTACTGTATGTACATACA GTAGGATCCTCGAAAATATTAATTACTGTATGTACATACAGTAGGAT CCTCGAAAATATTAATTACTGTATGTACATACAGTAGGATCCTACTGT ATGTACATACAGTAATTAATATTTTCGAGGATCCTACTGTATGTACAT ACAGTAATTAATATTTTCGAGGATCCTACTGTATGTACATACAGTAAT TAATATTTTCGAGGATCCTATGGCATGCATGTGCTCTGTATGTATATA AAACTCTTGTTTTCTTCTTTTCTCTAAATATTCTTTCCTTATACATTA GGACCTTTGCAGCATAAATTACTATACTTCTATAGACACACAAACACA AATACACACACTAATCTAGATAAAAAATGGTTTCCAAG
12xa1_pCYC1	AAGCTTGATATCGAATTCCTGCAGCCCGGGTACTGTATGTACATACA GTAGGATCCTCGAAAATATTAATTACTGTATGTACATACAGTAGGAT CCTCGAAAATATTAATTACTGTATGTACATACAGTAGGATCCTCGAAA ATATTAATTACTGTATGTACATACAGTAGGATCCTCGAAAATATTAAT TACTGTATGTACATACAGTAGGATCCTCGAAAATATTAATTACTGTAT GTACATACAGTAGGATCCTACTGTATGTACATACAGTAATTAATATTT TCGAGGATCCTACTGTATGTACATACAGTAATTAATATTTTCGAGGAT CCTACTGTATGTACATACAGTAATTAATATTTTCGAGGATCCTACTGT ATGTACATACAGTAATTAATATTTTCGAGGATCCTACTGTATGTACAT ACAGTAATTAATATTTTCGAGGATCCTACTGTATGTACATACAGTAAT TAATATTTTCGAGGATCCTATGGCATGCATGTGCTCTGTATGTATATA AAACTCTTGTTTTCTTCTTTTCTCTAAATATTCTTTCCTTATACATTA GGACCTTTGCAGCATAAATTACTATACTTCTATAGACACACAAACACA AATACACACACTAATCTAGATAAAAAATGGTTTCCAAG
pRPR_TetO	GGGGGATCTGCCAATTGAACATAACATGGTAGTTACATATACTAGTA ATATGGTTCGGCACACATTAAAAGTATAAAAACTATCTGAATTACGA ATTACATATATTGGTCATAAAAATCAATCAATCATCGTGTGTTTTATA TGTCTCTTATCTAAGTATAAGAATATCCATAGTTAATATTCACTTACG CTACCTTTTAACCTGTAATCATTGTCAACAGGATATGTTAACGACCCA CATTGATAAACGCTAGTATTTCTTTTTCCTCTTCTTATTGGCCGGCTG TCTCTATACTCCCCTATAGTCTGTTTCTTTTCGTTTCGATTGTCCCTA TCAGTGATAGAGATGGCGCACATGGTACGCTGTGGTGCTCGCGGCTG GGAACGAAACTCTGGGAGCTGCGATTGGCAG

gRNA expression cassette (SacI_pRPR1_HindIII_SDS_gRNA backbone_XhoI_RPR1t_KpnI)	GAGCTCGGGGGATCTGCCAATTGAACATAACATGGTAGTTACATATA CTAGTAATATGGTTCGGCACACATTAAAAGTATAAAAACTATCTGAAT TACGAATTACATATATTGGTCATAAAAATCAATCAATCATCGTGTGTT TTATATGTCTCTTATCTAAGTATAAGAATATCCATAGTTAATATTCAC TTACGCTACCTTTTAACCTGTAATCATTGTCAACAGGATATGTTAACG ACCCACATTGATAAACGCTAGTATTTCTTTTTCCTCTTCTTATTGGCC GGCTGTCTCTATACTCCCCTATAGTCTGTTTCTTTTCGTTTCGATTGT TTTACGTTTGAGGCCTCGTGGCGCACATGGTACGCTGTGGTGCTCGC GGCTGGGAACGAAACTCTGGGAGCTGCGATTGGCAGAAGCTTNNNNN NNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT TTTCTCGAGCCATATCCAACTTCCAATTTAATCTTTCTTTTTTAATTT TCACTTATTTGCGATACAGAAAGAAAAAAGCGATAGTAACTATTGAA TTTTGTTTGGATTTGGTTAGATTAGATATGGTTTCTCTTTATATTTAC ATGCTAAAAATGGGCTACACCAGAGATACATAATTAGATATATATAC GCCAGTACACCTTATCGGCCCAAGCCTTGTCCCAAGGCAGCGTTTTGT TCTTGGAAACGCTGCCCTACACGTTCGCTATGCTTCAAGAACTTTTCT GAGCACTTCATGATGCATGTTTGTTCCTTATTGGTTAGCTTTGATGTT GTGAAGTCATTGACACAGTCTGTGAAACATCTTTCTACCAGATTAGA GTACAAACGCATGAAATCCTTCATTTGCTTTTGTTCCACTACTTTTTG GAACTCTTGTTGTTCTTTGGTACC
	Where NNNNNNNNNNNNNNNNNNNN is one of the following:
c1 gRNA	CTAGATATTAAAATGTCTAA
c2 gRNA	ATATTCTTTCCTTATACATT
c3 gRNA	GTACATACAGTAGGATCCTA
c4 gRNA	ATATCGAATTCCTGCAGCCC
c5 gRNA	TATAGTAATTTATGCTGCAA
c6 gRNA	TGCAAAGGTCCTAATGTATA
c7 gRNA	ACAGAGCACATGCATGCCAT
c8 gRNA	TGTATGTACATACAGTACCC
cm1 gRNA	ATATTCTTTCCTTATACTTT
cm2 gRNA	CTGTATGTATATAAAACTCT
cm3 gRNA	CTGCAGCCCGGGGGATCCAG
cm4 gRNA	TGCAAAGGTCCTAAAGTATA
cm5 gRNA	AGAGTTTTATATACATACAG
cm6 gRNA	CCATGCCATATCATCACGGA
a1 gRNA	TTACTGTATGTACATACAGT
g1 gRNA	ACTAATACTTTCAACATTTT
g2 gRNA	GTTGAAAGTATTAGTTAAAG

Table S7.2: Sequences of the constructs used in HEK 293T experiments.

Name	Sequence
dCas9_3xNLS_VP64	GCCACCATGGACAAGAAGTACTCCATTGGGCTCGCCATCGGCACAAAC AGCGTCGGCTGGGCCGTCATTACGGACGAGTACAAGGTGCCGAGCAA AAAATTCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAAGAAGA ACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCA CGCGGCTCAAAAGAACAGCACGGCGCAGATATACCCGCAGAAAGAAT CGGATCTGCTACCTGCAGGAGATCTTTAGTAATGAGATGGCTAAGGT GGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGGTGGAGG AGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGAC GAGGTGGCGTACCATGAAAAGTACCCAACCATATATCATCTGAGGAA GAAGCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTTGATCTATC TCGCGCTGGCGCATATGATCAAATTTCGGGGACACTTCCTCATCGAGG GGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAAC TGGTTCAGACTTACAATCAGCTTTTCGAAGAGAACCCGATCAACGCAT CCGGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAAATCCC GGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAAC GGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCCAAC TTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGC AAAGACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGATCGGC GACCAGTACGCAGACCTTTTTTTGGCGGCAAAGAACCTGTCAGACGCC ATTCTGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGC TCCGCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCACCAAG ACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAAGT ACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACA TTGACGGCGGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCC ATCTTGGAAAAAATGGACGGCACCGAGGAGCTGCTGGTAAAGCTTAA CAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACAATGGAAGCA TCCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCCTCAGGCGGC AAGAGGATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAG AAAATCCTCACATTTCGGATACCCTACTATGTAGGCCCCCTCGCCCGG GGAAATTCCAGATTCGCGTGGATGACTCGCAAATCAGAAGAGACCAT CACTCCCTGGAACTTCGAGGAAGTCGTGGATAAGGGGGCCTCTGCCC AGTCCTTCATCGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACG AAAAGGTGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTT ATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAG CCAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTC TTCAAGACGAACCGGAAAGTTACCGTGAAACAGCTCAAAGAAGACTA TTTCAAAAAGATTGAATGTTTCGACTCTGTTGAAATCAGCGGAGTGG AGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAAA TCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACATT CTTGAGGACATTGTCCTCACCCTTACGTTGTTTGAAGATAGGGAGAT GATTGAAGAACGCTTGAAAACTTACGCTCATCTCTTCGACGACAAAGT CATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGT CAAGAAAACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACA ATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAACTTCATG CAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAGAAA GCACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAA TCTTGCAGGTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTA AGGTCGTGGATGAACTCGTCAAAGTAATGGGAAGGCATAAGCCCGAG AATATCGTTATCGAGATGGCCCGAGAGAACCAAACTACCCAGAAGGG ACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAA AAGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCC AGCTTCAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGG GACATGTACGTGGATCAGGAACTGGACATCAATCGGCTCTCCGACTA

	CGACGTGGATGCCATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTAT TGATAATAAAGTGTTGACAAGATCCGATAAAAATAGAGGGAAGAGTG ATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTATTGG CGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAA TCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCG GCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCACG TGGCCCAAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATG ACAAACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTG GTCTCAGATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGAT CAACAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGG CACTGCACTTATCAAAAAATATCCCAAGCTTGAATCTGAATTTGTTTA CGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTG AGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATA TTATGAATTTTTTCAAGACCGAGATTACACTGGCCAATGGAGAGATTC GGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAATCGTG TGGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTGTCCAT GCCGCAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCGGAGGCT TCTCCAAGGAAAGTATCCTCCCGAAAAGGAACAGCGACAAGCTGATC GCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTCGATTC TCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAG GGAAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGCTGGGCATCACA ATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAG GCGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTTCC CAAGTACTCTCTCTTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGC TAGTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTA AATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAG GGTCTCCCGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACAC AAACACTACCTTGATGAGATCATCGAGCAAATAAGCGAATTCTCCAAA AGAGTGATCCTCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTAC AATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAAACATTAT CCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGTA CTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGT CCTGGACGCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAAC AAGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGGCTGACGGGC CCTCACTGGGTTCAGGGTCACCCAAGAAGAAACGCAAAGTCGAGGAT CCAAAGAAGAAAAGGAAGGTTGAAGACCCCAAGAAAAAGAGGAAGGT GGATGGGATCGGCTCAGGCAGCAACGGCGGTGGAGGTTCAGACGCTT TGGACGATTTCGATCTCGATATGCTCGGTTCTGACGCCCTGGATGATT TCGATCTGGATATGCTCGGCAGCGACGCTCTCGACGATTTCGACCTCG ACATGCTCGGGTCAGATGCCTTGGATGATTTTGACCTGGATATGCTCT CATGATGA
KRAB	GATGCTAAGTCACTAACTGCCTGGTCCCGGACACTGGTGACCTTCAAG GATGTATTTGTGGACTTCACCAGGGAGGAGTGGAAGCTGCTGGACAC TGCTCAGCAGATCGTGTACAGAAATGTGATGCTGGAGAACTATAAGA ACCTGGTTTCCTTGGGTTATCAGCTTACTAAGCCAGATGTGATCCTCC GGTTGGAGAAGGGAGAAGAGCCCTGGCTGGTGGAGAGAGAAATTCAC CAAGAGACCCATCCTGATTCAGAGACTGCATTTGAAATCAAATCATCA GTTTCCAGCAGGAGCATTTTTAAAGATAAGCAATCCTGTGACATTAAA ATGGAAGGAATGGCAAGGAATGATCTCTGG

pPGK1	AATTCTACCGGGTAGGGGAGGCGCTTTTCCCAAGGCAGTCTGGAGCA TGCGCTTTAGCAGCCCCGCTGGGCACTTGGCGCTACACAAGTGGCCTC TGGCCTCGCACACATTCCACATCCACCGGTAGGCGCCAACCGGCTCCG TTCTTTGGTGGCCCCTTCGCGCCACCTTCTACTCCTCCCCTAGTCAGG AAGTTCCCCCCCCGCCCCGCAGCTCGCGTCGTGCAGGACGTGACAAATG GAAGTAGCACGTCTCACTAGTCTCGTGCAGATGGACAGCACCGCTGA GCAATGGAAGCGGGTAGGCCTTTGGGGCAGCGGCCAATAGCAGCTTT GCTCCTTCGCTTTCTGGGCTCAGAGGCTGGGAAGGGGTGGGTCCGGG GGCGGGCTCAGGGGCGGGCTCAGGGGCGGGGCGGGCGCCCGAAGGTC CTCCGGAGGCCCGGCATTCTGCACGCTTCAAAAGCGCACGTCTGCCGC GCTGTTCTCCTCTTCCTCATCTCCGGGCCTTTCGACCTGCATCCCAAG CTTGCCACCATG
pMLPm	GGGGGGCTATAAAAGGGGGTGGGGGCGTTCGTCCTCACTCTAGATCT GCGATCTAAGTAAGCTTGGCATTCCGGTACTGTTGGTAAAGCCACCAT GGCGCCACCATG
0xa1_pMLPm	ATCGATAGCTAGCGGGGGGCTATAAAAGGGGGTGGGGGCGTTCGTCC TCACTCTAGATCTGCGATCTAAGTAAGCTTGGCATTCCGGTACTGTTG GTAAAGCCACCATGGCGCCACCATG
1xa1_pMLPm	ATCGATAGCTAGCCTGAAAATATTAATTACTGTATGTACATACAGTAG GCTAGCGGGGGGCTATAAAAGGGGGTGGGGGCGTTCGTCCTCACTCT AGATCTGCGATCTAAGTAAGCTTGGCATTCCGGTACTGTTGGTAAAG CCACCATGGCGCCACCATG
2xa1_pMLPm	ATCGATAGCTAGCCTGAAAATATTAATTACTGTATGTACATACAGTAG GCTAGCCTGAAAATATTAATTACTGTATGTACATACAGTAGGCTAGCG GGGGGCTATAAAAGGGGGTGGGGGCGTTCGTCCTCACTCTAGATCTG CGATCTAAGTAAGCTTGGCATTCCGGTACTGTTGGTAAAGCCACCATG GCGCCACCATG
3xa1_pMLPm	ATCGATAGCTAGCCTGAAAATATTAATTACTGTATGTACATACAGTAG GCTAGCCTGAAAATATTAATTACTGTATGTACATACAGTAGGCTAGCC TGAAAATATTAATTACTGTATGTACATACAGTAGGCTAGCGGGGGGC TATAAAAGGGGGTGGGGGCGTTCGTCCTCACTCTAGATCTGCGATCT AAGTAAGCTTGGCATTCCGGTACTGTTGGTAAAGCCACCATGGCGCC ACCATG
pU6_gRNA	TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATC CGGTACCAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTT CATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTA ATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGA AAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAA ATGGACTATCATTGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGG CTTTATATATCTTGTGGAAAGGACGAAACACCGNNNNNNNNNNNNNNN NNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG TTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT
	where NNNNNNNNNNNNNNNNNNNN is one of the following:
m1 gRNA	CTGTCCCCAGTGCAAGTGC
m2 gRNA	TTCTCTATCGATAGCTAGC
m3 gRNA	TCTGCGATCTAAGTAAGCT
m4 gRNA	ACTGTTGGTAAAGCCACCA
m5 gRNA	ATGGCGCCACCATGAGCAG
m6 gRNA	GGCACCTGCACTTGCACTG
m7 gRNA	TATCGATAGAGAAATGTTC

m8 gRNA	AGATCGCAGATCTAGAGTG
m9 gRNA	GTGGCTTTACCAACAGTAC
m10 gRNA	GCTCATGGTGGCGCCATGG
CCAAT gRNA	GAAGGAGCAAAGCTGCTAT
GC-box gRNA	GGCGGGCTCAGGGGCGGGG
a1 gRNA	TACTGTATGTACATACAGT
a2 gRNA	AGTCGCGTGTAGCGAAGCAT
a3 gRNA	CAACGCGACGCTAGATAGCA

Chapter 8: PRISM

Randomized CRISPR-Cas Transcriptional Perturbation Screening Reveals Protective Genes Against Alpha-Synuclein Toxicity

This chapter is adapted from

Chen, Y.C.*, Farzadfard, F.*, Gharaei, N., Chen, W.C., Cao, J. and Lu, T.K., 2017. Randomized CRISPR-Cas Transcriptional Perturbation Screening Reveals Protective Genes against Alpha-Synuclein Toxicity. Molecular Cell, 68(1), pp.247-257. (http://dx.doi.org/10.1016/j.molcel.2017.09.014)

to fit the format of this thesis (with permission from Elsevier Inc.).

8.1 Abstract

The genome-wide perturbation of transcriptional networks with CRISPR-Cas technology has primarily involved systematic and targeted gene modulation. Here, we developed PRISM (Perturbing Regulatory Interactions by Synthetic Modulators), a screening platform that uses randomized CRISPR-Cas transcription factors (crisprTFs) to globally perturb transcriptional networks. By applying PRISM to a yeast model of Parkinson's disease (PD), we identified guide RNAs (gRNAs) that modulate transcriptional networks and protect cells from alpha-synuclein (αSyn) toxicity. One gRNA identified in this screen outperformed the most protective suppressors of αSyn toxicity reported previously, highlighting PRISM's ability to identify modulators of important phenotypes. Gene expression profiling revealed genes differentially modulated by this strong protective gRNA that rescued yeast from αSyn toxicity when overexpressed. Human homologs of top-ranked hits protected against αSyn-induced cell death in a human neuronal PD model. Thus, high-throughput and unbiased perturbation of transcriptional networks via randomized crisprTFs can reveal complex biological phenotypes and effective disease modulators.

8.2 Introduction

The systematic perturbation of transcriptional networks enables the elucidation of gene functions and regulatory networks that underlie biological processes. Current methods of modulating transcriptional networks mainly rely on targeted single-gene

overexpression, knockout, and knockdown (*207-210*). With the advent of artificial transcription factors, such as zinc-finger-, Transcriptional Activator-Like Effector (TALE)-, and CRISPR-Cas9-based transcription factors (crisprTFs), customized transcriptional perturbations are possible (*211-214*). For example, crisprTF-based platforms enable bi-directional gene activation and repression in eukaryotic systems (*69, 110, 215-218*) and have been used for genome-wide targeted screens owing to the ease of designing and synthesizing guide RNAs (gRNAs) (*219-221*). In addition, strategies for higher-order perturbations using barcoded combinatorial genetic screens in human cells have been adapted to be compatible with CRISPR-Cas9 screens (*222, 223*). Existing CRISPR-Cas9-based screening strategies rely on gRNAs designed to target individual genes while minimizing off-target effects (*219, 224-230*). Although these technologies provide powerful strategies for perturbing individual genes, they may not be suitable for global or combinatorial perturbation of transcriptional networks. Many complex diseases, as well as treatments required to counteract those conditions, may involve simultaneous or dynamic changes in the expression levels of many genes, which are not accessible by screens that target genes one at a time (*231, 232*).

To address this limitation, we explored the use of randomized gRNAs and crisprTFs in an approach called PRISM (Perturbing Regulatory Interactions by Synthetic Modulators) in order to effect global transcriptional perturbations conferring enhanced cellular resistance to alpha-synuclein (αSyn). The aggregation of misfolded αSyn in intraneuronal Lewy bodies is one of the pathological hallmarks of Parkinson's disease (PD) (*233, 234*). The overexpression of αSyn in various eukaryotic model organisms has been used to elucidate the complex cellular processes associated with PD (*235-237*). Because of its conserved molecular mechanisms and the availability of genetic tools, *Saccharomyces cerevisiae* has been extensively used as a model to systematically study and identify genes involved in neurodegenerative diseases such as PD and Alzheimer's disease (*238, 239*).

Here, we demonstrate that one of the strongest protective gRNAs identified in our PRISM screens outperformed any individual overexpressed gene that we tested in suppressing αSyn toxicity, including the strongest protective genes found in previous genome-wide screens (*231, 240-242*). These results highlight that randomized gRNA/crisprTF perturbations can achieve powerful phenotypic modulation compared with other targeted gene perturbation methods.

8.3 Results

Randomized gRNA Screening Design

We cloned a dCas9-VP64 expression cassette under the control of a doxycycline (Dox)-inducible (Tet-ON) promoter. To build the yeast screen strain, this construct was integrated into the genome of an αSyn-expressing *S. cerevisiae* strain (hereafter referred to as the yeast parental strain) in which two copies of the human wild-type αSyn (*SNCA*) gene fused to yellow fluorescent protein (YFP) are overexpressed under the control of a galactose (Gal)-inducible promoter (*242*) (Fig. 8.1A). Both the yeast parental strain and the screen strains showed significant cellular growth defects in the presence of Gal due to overexpression of αSyn. The expression of dCas9-VP64 with no gRNA in the screen strain did not interfere with normal cellular growth or αSyn-associated toxicity (Fig. S8.1A).

A randomized gRNA-expressing plasmid library was built by co-transforming the screen strain with linearized high-copy *2μ* plasmids, flanked by the *RPR1* promoter (*RPR1p*) and gRNA handle at the ends, and a randomized oligo library encoding 20-mer randomized nucleotides flanked by homology arms to the ends of the vector (Fig. 8.1A). We observed approximately 100 million colony forming units (CFUs) per library transformation, which is comparable with the theoretical diversity of the seed sequence, the protospacer adjacent motif (PAM)-proximal 12 nucleotides (4^{12} = ~1.67 x10^7). Library representation and sequence distributions were determined by deep sequencing (Figs. S8.2A-2G). After cells were transformed with the library, they were recovered in liquid culture with Dox (1 μg/mL) for 12 hours to amplify the library and induce crisprTF expression. The cultures were then plated on synthetic complete media (Scm)−Uracil (Ura)+Gal+Dox plates, and gRNAs from surviving colonies were characterized by colony PCR followed by Sanger sequencing.

gRNA Suppressors of αSyn Toxicity were Identified by a Randomized Screen in *Saccharomyces cerevisiae*

To validate the activity of the identified gRNAs, they were re-cloned in both high-copy *2μ* and low-copy *ARS/CEN* plasmids, and transformed back into both parental and screen strains. We confirmed that two gRNAs (designated as gRNA 6-3 and gRNA 9-1), expressed from either high-copy and low-copy plasmids, could rescue the screen strain from αSyn toxicity (Fig. 8.1B). gRNA 6-3 moderately suppressed αSyn toxicity whereas gRNA 9-1 strongly suppressed it; gRNA 9-1 was thus chosen for further

characterization. Although no perfect match between the identified gRNAs and the yeast genome was found, relaxing the search criteria (to find up to two mismatches inside the seed region) revealed the presence of a few dozen sites that could potentially serve as off-target binding sites of these gRNAs, including one in the *GAL4* gene (Table S8.1). As additional controls, we confirmed that gRNA 9-1-mediated suppression of αSyn toxicity relied on the presence of dCas9-VP64 (Fig. S8.1B) and that *GAL4* and αSyn expression levels were not directly affected by gRNA 9-1/crisprTF (Figs. S8.1E-1F). Furthermore, we re-encoded the putative gRNA 9-1 off-target binding site in *GAL4* so that there were five matches in the seed sequence and found that gRNA 9-1 preserved its ability to rescue the screen yeast strain from αSyn toxicity (Figs. S8.1G-1I).

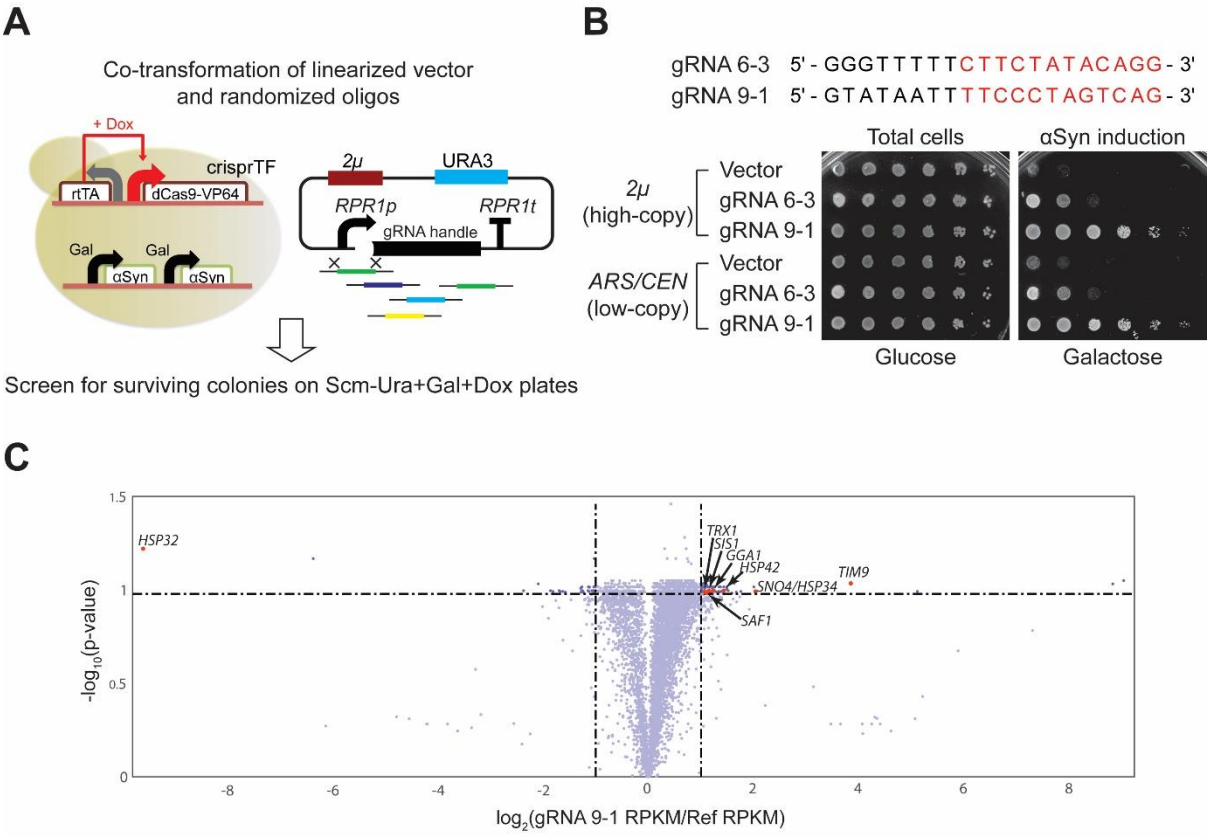


Figure 8.1 | Randomized gRNA/crisprTF screens identify genetic modifiers of αSyn toxicity in *S. cerevisiae*. **(A)** Schematic illustration of the engineered yeast screen strain expressing αSyn and crisprTF (left) and the strategy used for building the randomized gRNA library (right). See Methods section and Figs. S8.1-2 for details. **(B)** Sequences of the two identified gRNAs (designated as gRNA 6-3 and 9-1) that could suppress αSyn-mediated toxicity. 5-fold serial dilutions of saturated cultures were spotted on Scm (Synthetic complete media)−Uracil (Ura)+Glucose+Doxycycline (Dox) plates to quantify the total number of viable cells and Scm−Ura+Galactose (Gal)+Dox plates to score cell viability upon αSyn induction. gRNA 9-1

is a strong suppressor of αSyn toxicity while gRNA 6-3 is a moderate suppressor. Both gRNAs performed better than the negative control (empty vector), and suppression levels were independent of gRNA plasmid copy number. See also Fig. S8.1. **(C)** The transcriptome analysis of the screen strain harboring gRNA 9-1 in comparison with the reference strain (screen strain with no gRNA) represented as a volcano plot (x-axis: fold change versus y-axis: statistical significance). A list of differentially expressed genes is provided in Table S8.2.

Gene Expression Profiling of αSyn-Resistant Cells by gRNA 9-1/crisprTF Revealed Suppressors of αSyn Toxicity

We compared the transcriptome of screen cells expressing gRNA 9-1 and dCas9-VP64 to that of cells expressing dCas9-VP64 but no gRNA by using RNA sequencing to map transcriptional perturbations enacted by the αSyn-protective crisprTF (Fig. 8.1C and Figs. S8.2H-2I). We identified 114 differentially expressed genes with at least two-fold changes in mRNA expression levels compared with the non-gRNA control (false discovery rate (FDR)-adjusted p-value ≤ 0.1) (Table 8.1 and Table S8.2). Most of these genes (93%) have not been previously identified in single gene knockout or overexpression screens as suppressors of αSyn toxicity (*231, 240, 242*). Intriguingly, they were enriched in Gene Ontology (GO) categories including protein quality control, ER/Golgi trafficking, lipid metabolism, mitochondrial function, and stress responses (Table S8.3). Almost all of the newly identified genes exhibited only modest changes in gene expression (109 out of 114 genes had fold-changes <5).

We systematically tested the effects of our differentially expressed genes on αSyn toxicity in the screen strain by overexpressing 95 of them that were found in the Yeast ORF Collection (Open Biosystems). Overexpression of 57 out of 95 (60%) genes (13 down-regulated and 44 up-regulated by gRNA 9-1/crisprTF) significantly suppressed αSyn toxicity (Fig. S8.3A, summarized in Table S8.2; representative candidates are shown in Fig. 8.2A), whereas only 5 out of 34 (14.7%) genes randomly chosen from the Yeast ORF Collection suppressed αSyn toxicity (Figure S8.3B and Table S8.4). Thus, our randomized gRNA/crisprTF screening approach enriched the search for αSyn-toxicity suppressors.

Furthermore, there was no significant correlation between observed αSyn expression levels and toxicity (Figs. S8.1E-1F). *UBP3* (ubiquitin-specific protease) was used as a positive control. *UBP3*, previously shown to be a strong suppressor of αSyn toxicity, is known to participate in the degradation of misfolded proteins in the vesicular trafficking processes (*242-244*). We found that 29 genes whose expression was

modulated by gRNA 9-1 protected against αSyn-toxicity similarly to or better than *UBP3*. Notably, gRNA 9-1 alone outperformed the overexpression of any single gene in abrogating αSyn-associated phenotypes based on cell viability assay results, shown in Fig. 8.2A and microscopy (see below and Figs. S8.4A-4B), suggesting that gRNA 9-1 plays a master role in mitigating αSyn stress.

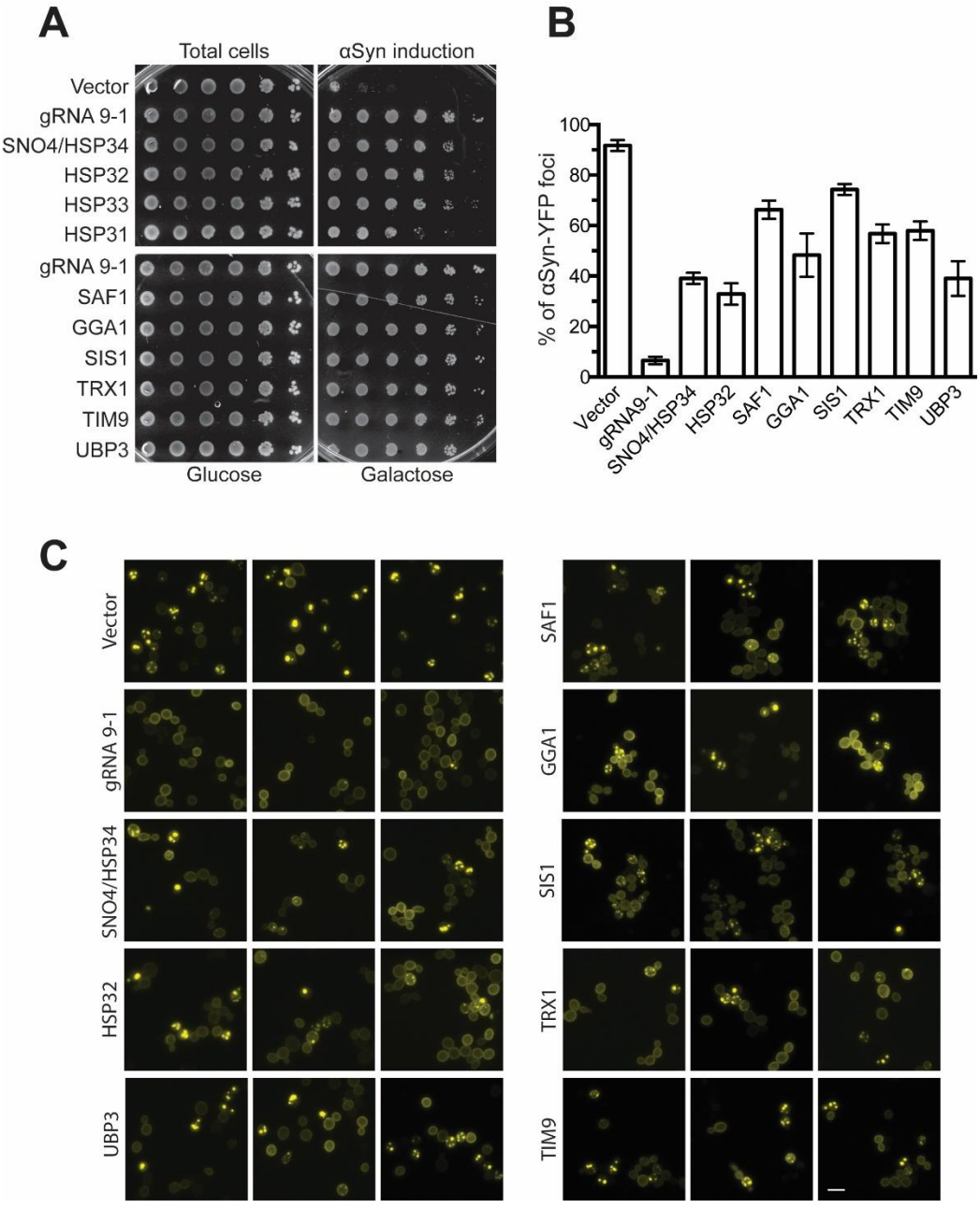


Figure 8.2 | Overexpressing genes identified from the gRNA 9-1/crisprTF screen rescues αSyn-associated cellular defects in yeast. (A) Survival of the screen strain harboring gRNA 9-1 ('gRNA 9-1') compared to cells expressing the empty vector ('Vector') and those overexpressing *HSP31-34* (heat shock proteins) (top) as well as top-ranked αSyn suppressors identified in the

screen (bottom). *UBP3*, a known strong αSyn suppressor, was used as a positive control. **(B)** Quantification of αSyn-YFP foci in the screen strain harboring either no gRNA or gRNA 9-1, or cells harboring plasmids that overexpress the indicated genes. Cytoplasmic YFP foci represent αSyn aggregates produced as a result of defects in vesicular trafficking. Cells expressing crisprTF and gRNA 9-1 robustly inhibited αSyn aggregates, evidenced by the absence of cytoplasmic YFP foci in these samples. Cells overexpressing *UBP3* were used as a positive control in this assay. Data were presented as mean ± SEM of three biological replicates. **(C)** Representative images of αSyn-expressing cells mentioned in **b**. Bar = 10 μm. See also Figs. S8.3-4.

Alterations in membrane trafficking and localization of αSyn from the plasma membrane into cytoplasmic foci are well-established hallmarks of PD (*245*). Owing to highly conserved mechanisms involved in membrane trafficking, yeast cells have been used to study αSyn-coupled vesicular trafficking defects, which has led to mechanistic insights into modifiers of αSyn toxicity, such as *UBP3* and the Rab family GTPase *YPT1* and their human homologs (*242-244*). We quantitatively measured the effect of gRNA 9-1 on the localization of αSyn-YFP by microscopy. In this assay, aggregated αSyn-YFP is detected as cytoplasmic foci, which are distinguishable from the membrane-localized, non-aggregated form of the protein. As shown in Figs. 8.2B and 8.2C, upon 6 hours of αSyn induction, 92% of yeast cells with dCas9-VP64 but no gRNA (negative control) contained aggregated αSyn-YFP foci. Overexpression of dCas9-VP64 along with gRNA 9-1 resulted in localization of αSyn-YFP to the plasma membrane such that aggregated αSyn-YFP foci were observed in only ~7% of cells. This was significantly lower than the percentage of cells overexpressing *UBP3* (~39% cells with αSyn-YFP foci), which we used as a positive control.

Human *DJ-1/PARK7*, *ALS2*, *GGA1*, and *DNAJB1* Homologs were Identified as Robust Protectors Against αSyn Toxicity

One of the interesting functional categories identified in our screen involves the heat shock chaperones. Specifically, *HSP31-34* heat shock proteins are homologs of the human *DJ-1/PARK7* gene, in which autosomal recessive mutations are associated with early onset of familial PD (*246*). *DJ-1* is thought to protect neurons from mitochondrial oxidative stress by acting as a redox-dependent chaperone to inhibit αSyn aggregates (*246, 247*). The roles of *HSP31-34* in protecting yeast cells from αSyn toxicity have been previously investigated (*248*); however, these genes have not been identified in previous genome-wide screens for modifiers of αSyn toxicity. We identified

SNO4/HSP34 and *HSP32* as two of the differentially expressed genes in our screen. As shown in Fig. 8.2, either *SNO4/HSP34* or *HSP32,* when overexpressed, significantly rescued αSyn-induced growth defects and membrane-trafficking abnormalities. Interestingly, *SNO4/HSP34* was moderately up-regulated by gRNA 9-1, whereas *HSP32* was extremely down-regulated (Fig. 8.1C and Table 8.1), which could reflect evolutionarily conserved functions of these paralog proteins, despite their being under the control of different gene regulatory programs. Furthermore, overexpression either of the other two yeast *DJ-1* homologs (*HSP31* and *HSP33*) also significantly suppressed αSyn toxicity (Fig. 8.2A), even though they were not significantly modulated by gRNA 9-1. This further supports the involvement of this class of paralog heat-shock proteins in suppressing αSyn toxicity. Consistently, *HSP31* (which among *HSP31-34* shows the least homology with *DJ-1*) was recently shown to be a chaperone involved in mitigating various protein misfolding stresses, including that of αSyn (*249*).

Among other top αSyn-toxicity suppressors (Table 8.1 and Fig. 8.2), yeast *SAF1* encodes an F-Box protein that selectively targets unprocessed vacuolar/lysosomal proteins for proteasome-dependent degradation (*250, 251*). The homolog of this protein in mice and humans, *ALS2/alsin*, functions as a guanine nucleotide exchange factor (GEF) that activates the small GTPase *Rab5*, an evolutionarily conserved protein involved in membrane trafficking in endocytic pathways (*252*). Mutations in human *ALS2* have been shown to cause autosomal recessive motor neuron diseases (*253*). In addition, we found that *GGA1* and its paralog *GGA2* could both alleviate αSyn toxicity (Figs. 8.2 and S8.3), which was interesting because neither of them had been reported previously to have this activity. The yeast *GGA1* protein has been implicated in binding ubiquitin, thus facilitating the sorting of cargo proteins from the trans-Golgi network to endosomal compartments (*254, 255*). Human *GGA1* overexpression attenuates amyloidogenic processing of amyloid precursor proteins (APP) in Alzheimer's disease and a rare inherited lipid-storage disease, Niemann-Pick type C (NPC) (*256, 257*). Finally, *SIS1,* the yeast Hsp40 homolog of human *DNAJ/HSP40* family proteins, was identified as αSyn suppressor via PRISM. *DNAJ* family proteins play roles in priming the specificity of *HSP70* chaperoning complexes. It has been shown that mammalian *DNAJ* and *HSP70* are up-regulated in response to αSyn overexpression (*258*). In addition, the *DNAJB* subfamily has been shown to suppress polyglutamine (polyQ) aggregates (*259*). These results demonstrate that randomized transcriptional perturbations with PRISM enable the discovery of modulators of disease-relevant phenotypes.

Table 8.1 | A summary of top-ranked genes that are differentially regulated by gRNA 9-1 and that suppressed αSyn toxicity in yeast when overexpressed. A complete list of genes differentially modulated by gRNA 9-1 is provided in Table S8.2.

Yeast Gene	Human Homologs	Log_2(fold change)	Survival Score	Fluorescent Foci Score	Biological Function
SNO4/HSP34	*DJ-1/PARK7*	2.035	6	7	Chaperone and cysteine protease
HSP32	*DJ-1/PARK7*	-9.593	6	7	Chaperone and cysteine protease
HSP42	*HSPB1, HSPB3, HSPB6, HSPB7, HSPB8, HSPB9*	1.434	5	6	Chaperone
SIS1	*DNAJB1-B9*	1.154	6	3	Chaperone
GGA1	*GGA1, GGA2, GGA3*	1.241	6	6	ER to Golgi vesicular trafficking
SRN2		1.031	6	4	Ubiquitin-dependent protein sorting
SAF1	*ALS2, RCC1*	1.180	6	4	Proteasome-dependent degradation
TRX1	*TXN, TXNDC2, TXNDC8*	1.072	6	5	Thioredoxin
TIM9	*TIMM9*	3.846	6	5	Mitochondrial intermembrane protein
OXR1	*OXR1, NCOA7, TLDC2*	1.003	5	3	Oxidative damage resistance
STF2	*SERBP1, HABP4*	2.004	6	3	mRNA stabilization
gRNA 9-1			6	10	
UBP3			6	7	
Vector			1	1	

Verification of Human Homologs of the Identified Hits in a Neuronal PD Model

To investigate the neuroprotective effects of the human homologs of the protective yeast genes described above, we overexpressed *DJ-1, ALS2, GGA1,* and *DNAJB1* in an αSyn-overexpressing human neuroblastoma cell line (SH-SY5Y), an established neuronal model of PD (*260*). SH-SY5Y cells were differentiated into cells with dopaminergic neuron-like phenotypes upon retinoic acid (RA) treatment. When ß-galactosidase (ß-gal) was expressed in these cells, no toxicity was observed. In contrast, αSyn-expressing cells gradually exhibited neurite retraction and only 40-50% viability at 6 days of differentiation (Figs. S8.5A-5B). Expressing *DJ-1* or *ALS2* alone did not alter cell survival in the absence of αSyn, but strongly suppressed αSyn-inducible cell

death (Fig. 8.3B). αSyn-expressing cells that were transfected with *GGA1* or *DNAJB1* exhibited about 60% viability, which was similar to the effect of expressing the known anti-apoptotic gene *Bcl-xL* (positive control). Consistent with these results, overexpression of *DJ-1* or *ALS2* resulted in reductions in the dead cell populations, as did treatment with the apoptotic inhibitor zVAD (Fig. 8.3C).

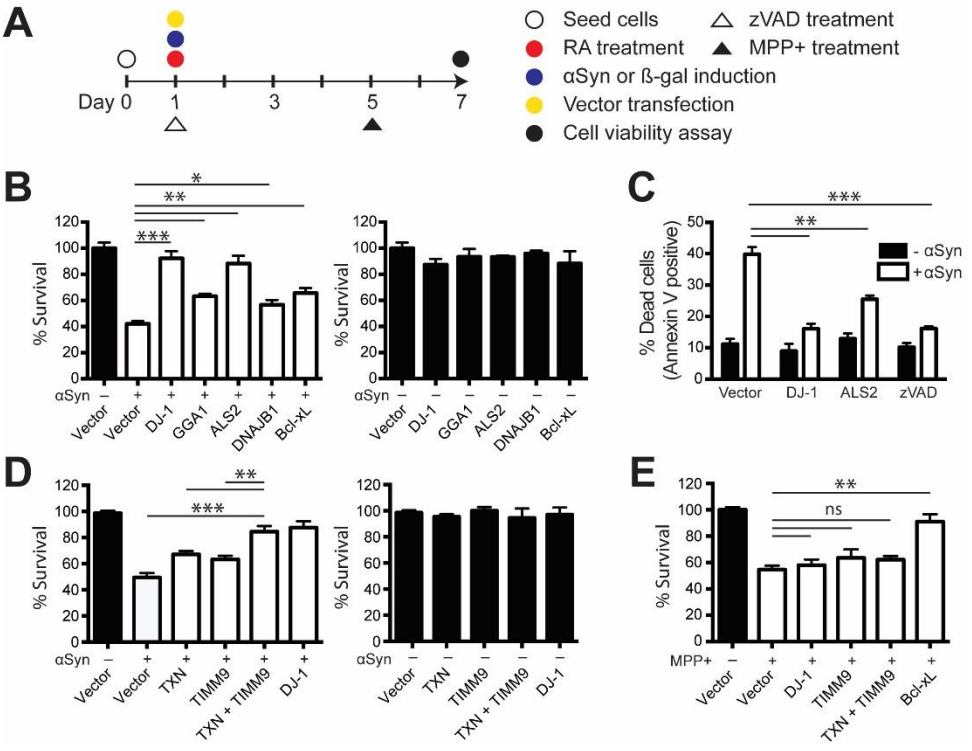


Figure 8.3 | Human homologs of yeast αSyn-toxicity suppressors in a human neuronal PD model. (A) A schematic representation of the experimental procedure used to test the human homologs of yeast αSyn-toxicity suppressors in differentiated neuronal cell lines. Different constructs expressing individual genes were transfected into the SH-SY5Y neuroblastoma cell line via transient transfection to examine their ability to protect against αSyn toxicity. αSyn expression was induced by removal of Dox from the media and retinoic acid (RA) treatment was used for neuronal differentiation over the course of a six-day period. The anti-cell-death drug zVAD and the toxin MPP+ were applied in control experiments. See also Fig. S8.5. **(B)** Cell viability of differentiated cell lines overexpressing αSyn and the indicated constructs (white bars) were determined by the CellTiter-Glo luminescent assay. The expression of individual genes did not significantly affect cell survival of differentiated cells in the absence of αSyn induction (black bars). Constructs expressing human *DJ-1* (homolog of yeast *SNO4/HSP34* and *HSP32*), *GGA1* (*GGA1*), *ALS2* (*SAF1*), or *DNAJB1* (*SIS1*) were tested. *Bcl-xL*, which protects against apoptotic neuronal death, was used a positive control (*261*). **(C)** The percentage of dead cells upon αSyn induction was quantitated by FITC-Annexin V staining followed by flow cytometry. Effects of overexpressing *DJ-1* or *ALS2* via plasmid transfection were compared with effects of zVAD. **(D)** Constructs expressing human *TXN* (homolog of yeast *TRX1*) or *TIMM9* (homolog of yeast *TIM9*) were transfected individually or co-transfected together to test for synergistic effects on protection from αSyn toxicity. *TXN*

+ *TIMM9* synergistically rescued these cells from αSyn toxicity when co-transfected together. **(E)** Overexpression of *DJ-1*, *TIMM9,* or *TXN* + *TIMM9* did not protect against MPP+ toxicity, in contrast with *Bcl-xL* overexpression. Transfected and differentiated cells were treated with 6 mM MPP+ and then tested for cell viability 48 hours afterward. All data were presented as mean ± SEM of triplicate sets. *p < 0.05, **p < 0.01, ***p < 0.001; ns, not significant.

Human *TNX* and *TIMM9* Synergistically Protect Cells against αSyn Toxicity

Increased oxidative stresses and defective mitochondrial function are pathological mechanisms involved in sporadic PD (*262*). We identified that yeast thioredoxin *TRX1*, an oxidoreductase involved in maintaining the cellular redox potential, and *TIM9*, a mitochondrial chaperone involved in the transport of hydrophobic proteins across mitochondrial intermembrane space (*263*), participate in the suppression of αSyn toxicity in yeast cells (Fig. 8.2 and Figs. S8.4C-4E). Neuronal cells transfected with the human homologs of either of these genes, *TXN* or *TIMM9*, respectively, exhibited about ~60% survival upon αSyn induction compared with <50% with the vector control expressing no transgene. Intriguingly, co-expression of the two human genes *TXN* and *TIMM9* led to enhanced survival in the presence of αSyn toxicity (~88 % survival) (Fig. 8.3D). Furthermore, the neuroprotective effects of expressing *DJ-1*, *TXN*, and *TIMM9* were specific to αSyn-associated toxicity, as these genes did not protect against 1-methyl-4-phenyl pyridinium (MPP+)-induced neurodegeneration (*261*) (Fig. 8.3E and Figs. S85C-5D).

To further investigate these genes as potential therapeutic targets for neuroprotection in PD, we engineered lentiviral vectors expressing *DJ-1*, *TXN*, or *TIMM9*, or co-expressing *TXN* and *TIMM9*. We then used these vectors to stably infect cells prior to inducing αSyn stress. Consistent with our transient transfection experiments, *DJ-1* reliably protected differentiated SH-SY5Y cells from αSyn-induced cell death and neuronal abnormalities, as did co-expression of *TXN* and *TIMM9* (Fig. 8.4). These results also suggest that activation of these endogenous genes could be explored as a potential therapeutic direction for neuroprotection in PD.

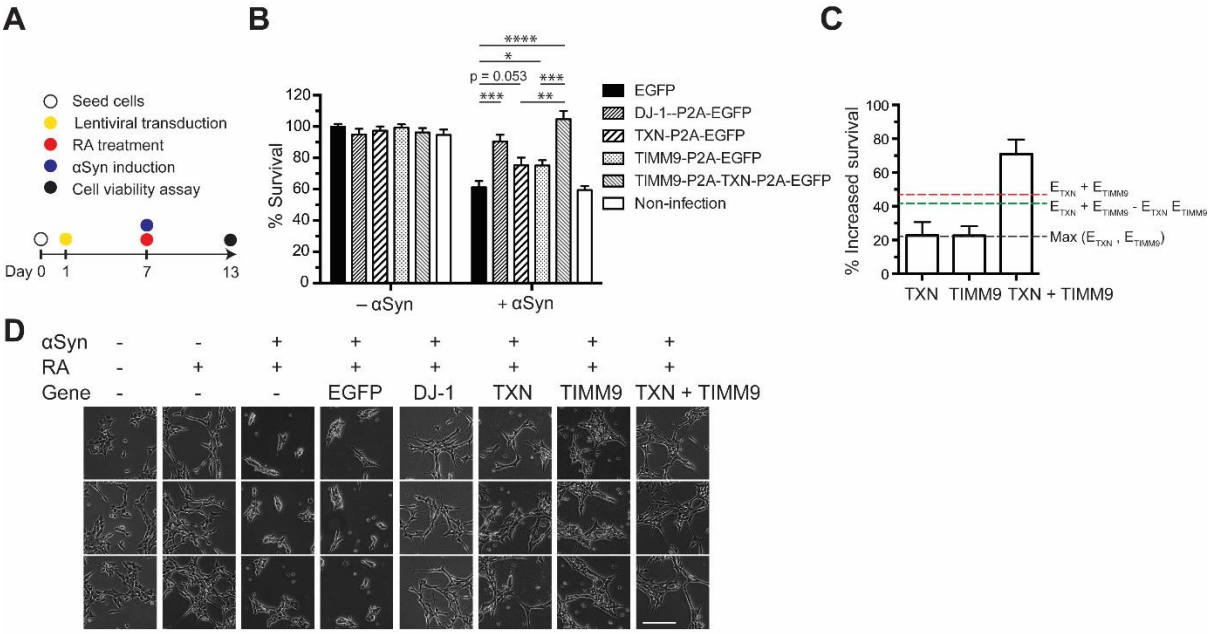


Figure 8.4 | Lentiviral expression of human *DJ-1*, *TXN*, and *TIMM9* protects against αSyn-associated toxicity in a neuronal PD model. (A) The human homologs of yeast αSyn-toxicity suppressors were stably expressed via lentiviral vectors six days before RA treatment and αSyn induction, as indicated in the experimental procedure diagram. **(B)** Overexpression of *DJ-1* or *TXN* + *TIMM9* significantly increased neuronal viability in the presence of αSyn. The 2A peptide sequence (P2A) was used to achieve the simultaneous expression of multiple genes from a single promoter. **(C)** *TXN* and *TIMM9* work synergistically protect neuronal cells from αSyn toxicity based on Highest Single Agent [Max(E_{TXN}, E_{TIMM9})] (*264*), Linear Interaction Effect (E_{TXN} + E_{TIMM9}) (*265*), and Bliss Independence (E_{TXN} + E_{TIMM9} - E_{TXN} E_{TIMM9}) (*266*) models (dashed lines). The effect of *TXN* + *TIMM9* was greater than the threshold values obtained from these models. **(D)** Representative images of neuronal morphology and cell density of cells transfected with lentiviral vectors overexpressing the indicated human genes. Bar = 400 μm. All data were presented as mean ± SEM, n = 6. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.

8.4 Discussion

In this study, we introduce the PRISM screening platform to probe mechanisms underlying cellular responses to αSyn stress. This platform takes advantage of the promiscuity of crisprTF activity with randomized gRNAs for global transcriptional network perturbations. In contrast to the typical targeted CRISPR screens, in which gRNAs are designed to modulate individual genes, PRISM does not require any assumptions regarding potential targets and enables unbiased and high-level perturbations of cellular networks. Thus, randomized gRNA screening with PRISM is complementary to targeted screening strategies that have been used with CRISPR-Cas nucleases, crisprTFs, and RNA interference (*211, 212, 219, 221-225, 230, 267-272*). Randomized gRNA screening with crisprTFs involves a simple library construction

procedure and enables global perturbations of transcriptional networks that might not be accessible by traditional single- or multiple-gene perturbations. Such high-order perturbations may be especially important when studying sophisticated phenotypes involving multi-layered regulatory networks, such as those associated with complex human diseases or stress tolerance.

As a proof of concept, we applied this system to a yeast model of PD and identified a diverse set of differentially expressed genes that could individually and collectively rescue αSyn-associated phenotypes. Transcriptomic analysis of the top αSyn-toxicity-protecting gRNA candidate (gRNA 9-1) revealed modest changes in the expression of multiple genes involved in various pathways associated with PD. Intriguingly, genes perturbed by gRNA 9-1 were enriched for αSyn toxicity suppressors versus random selection, most of which have not been reported in previous unbiased genome-wide screens. We verified that over half of these newly identified genes, when overexpressed, rescued yeast cells from αSyn-mediated toxicity. Moreover, gRNA 9-1/crisprTF ameliorated αSyn-associated phenotypes more than overexpression of any individual gene, highlighting the power of global transcription factor screening and suggesting that combinatorial or global effects can enhance desired phenotypes.

To verify the physiological relevance of our hits, we overexpressed the human homologs of validated yeast hits in a human neuronal model of αSyn toxicity. Interestingly, human genes *TXN* and *TIMM9*, homologous to yeast genes *TRX1* and *TIM9,* respectively*,* worked synergistically to suppress αSyn toxicity in human cells (Fig. 8.4). Thioredoxin is known to facilitate the mitochondrial import of *TIM9* in yeast (*273*) and to act as a neuroprotective agent against oxidative stress in neuronal cells (*274*). The observed synergistic effect of *TXN* and *TIMM9*, as well as the protective effect of redox-dependent chaperones, in suppressing αSyn toxicity further points to the potential importance of mitochondrial maintenance and oxidative stress in PD. Future efforts will be needed to determine whether combinatorial modulation of mitochondrial function pathways and cellular redox may help treat αSyn-associated dysfunction in animal models and clinical settings. In addition, future work should investigate the underlying mechanisms of neuroprotection from the hits identified in this study. These insights could help to develop neuroprotective strategies or engineer αSyn-resistant neuronal cells that could help prevent progressive neurodegeneration in PD patients diagnosed early.

In this screening effort, we used a first-generation crisprTF that results in modest levels of gene activation or repression (*215, 219*). Stronger activation and repression could be achieved by recently improved variants of crisprTFs (*110, 221, 275*) while other types of perturbations could be introduced via dCas9 fused to epigenetic regulatory domains (*112, 276, 277*).

We performed Sanger sequencing on the surviving colonies to identify gRNA 9-1 and gRNA 6-3, which were then tested in validation experiments. In several cases, we identified multiple different gRNAs within a single yeast cell. Performing high-throughput sequencing on amplicons obtained from surviving yeast colonies could reveal additional gRNAs that could rescue cells from αSyn toxicity when combinatorially expressed. However, for the purpose of this study, we chose to focus only on individual gRNAs that had suppressive effects against αSyn toxicity on their own. In future work, yeast single-copy centromeric (*ARS/CEN*) plasmids could be used for gRNA expression to ensure that each transformant receives only one gRNA variant.

Even though many genes were up- or down-regulated by gRNA 9-1, we validated them only through overexpression using the readily available Yeast ORF Collection. Thus, our current data suggests that genes modulated by gRNA hits from PRISM screens are enriched for effects on the desired phenotype, but does not indicate the impact of the directionality of gene modulation on the phenotype. In future work, genes identified via PRISM screens can be tested via knockdown as well as overexpression in order to determine whether directionality makes a difference. Furthermore, we chose to overexpress genes from the *GAL1* promoter in the Yeast ORF Collection rather than using CRISPR activation because efficient targeted CRISPR activation still requires tuning and optimization for each gene of interest. However, overexpression of ORFs by the strong *GAL1* promoter can result in expression levels much higher than those achievable by our first-generation crisprTFs. In future work, assessing how different expression levels of identified genes modulate the phenotypes identified through PRISM screening will also be of interest. Finally, we envision that genetic interactions between genes identified by PRISM screening can be further mapped through combinatorial CRISPR technologies (*223, 278, 279*) and both gene activation and inhibition technologies. Thus, high-throughput randomized crisprTF screens should provide access to a broader range of biological phenotypes across a wide range of organisms in the future.

We predicted 51 potential binding sites for gRNA 9-1 in the yeast genome (Table S8.1). Although direct crisprTF binding should be identifiable by Chromatin

Immunoprecipitation Sequencing (ChIP-Seq) experiments, we were unable to achieve this despite trying multiple different approaches (*228, 229, 280*). We hypothesize that this may be due to weak binding of dCas9 in the absence of perfect-match binding sites for gRNA 9-1 in the genome. Furthermore, it remains challenging to infer transcription regulatory networks solely based on predicted crisprTF binding sites and changes in RNA levels without mapping the transient and indirect cascades involved in transcriptional perturbations (*281*). Therefore, drawing direct connections between crisprTFs and regulated genes in PRISM screens remains a challenge that needs to be addressed in future work.

8.5 Supplementary Information

Experimental Model and Subject Details

Yeast Strains and Growth Condition

Strains used in this study are all derivatives of W303 (MATa *ade2-1 trp1-1 can1-100 leu2-3, 112 his3-11, 15 ura3*). The ITox2C yeast strain (*242*) harboring two copies of wild-type αSyn (*SNCA*)-YFP under control of the Gal-inducible *GAL1* promoter (hereafter referred to as the parental strain, a generous gift from Dr. Susan Lindquist, Whitehead Institute, USA) was used for the construction of the crisprTF-expressing screen strain. The Dox-inducible (Tet-ON) promoter was constructed by cloning the pTRE promoter and reverse tetracycline-controlled transactivator (rtTA, from Addgene plasmid #31797) upstream of a minimal *CYC1* promoter in the pRS405 backbone. The dCas9-VP64 expression cassette (Addgene plasmid #49013) was then cloned into this vector using Gibson assembly. A sense mutation was introduced within the *LEU2* ORF by using the QuikChange system (Stratagene) in order to generate a unique PstI site in the vector. The pRS405-pTetON-dCas9-VP64-PstI plasmid was linearized by PstI and transformed into the ITox2C parental strain to build the screen strain. Leucine-positive integrants were verified by genomic PCRs as well as testing for the presence of αSyn-mediated defects by the survival assay and microscopy after Gal induction.

To build the *GAL4** strain, a sequence containing full endogenous *GAL4* promoter (-257 to 214) was first PCR amplified by oligos (forward: 5'-CCCAGTATTTTTTTTATTCTACAAACC -3'; reversed: 5'-AAATCAGTAGAAATAGCTGTTCCAGTCTTTCTAGCCTTGATTCCACTTCT GTCAGgTGaGCtCggGTtaaCGGAGACCTTTTGGTTTTGG -3'). This fragment was then assembled (by Gibson assembly) with a kanMX6 expression cassette amplified from pFA6a-kanMX (Addgene plasmid #39296) using oligos (forward: 5'-

GGGGCGATTGGTTTGGGTGCGTGAGCGGCAAGAAGTTTCAAAACGTCCG CGTCCTTTGAGACAGCATTCGGAATTCGAGCTCGTTTAAAC -3'; reversed: 5'-GAAGGTTTGTAGAATAAAAAAAATACTGGGCGGATCCCCGGGTTAATTAA -3'). The assembled kanMX-*GAL4** cassette was then purified and transformed into yeast cells and transformants were selected in presence of 200 mg/L G418 (Thermo Fisher Scientific). Integrants were confirmed by yeast colony PCR and Sanger sequencing.

Yeast cells were cultured in either YPD (1% yeast extract, 2% Bacto-peptone and 2% glucose) or Synthetic complete medium (Scm) supplemented with 2% glucose, raffinose, or galactose. Doxycycline (Sigma) was added directly to culture media or plates immediately before pouring (final concentration of 1 µg/mL).

Neuroblastoma Cell Culture and Gene Expression

Parental and engineered SH-SY5Y cell lines (*260*) (kindly provided by Dr. Leonidas Stefanis, Biomedical Research Foundation Academy Of Athens, Greece) were grown in Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F-12) base medium plus 1% GlutaMAX™ supplemented with 15% heat-inactivated FBS (Fetal Bovine Serum) and 1X antibiotic-antimycotic (Invitrogen) at 37°C with 5% CO_2. Cells were seeded at an initial density of 10^4 cells/cm^2 in culture dishes coated with 0.05 mg/mL collagen (Invitrogen). Cells were maintained with 2 µg/mL Dox as previously described (*260*), in order to repress expression of αSyn and ß-galactosidase (ß-gal), which are driven by the Tet-OFF promoter (*260, 282*). The expression of αSyn and ß-gal was induced by removing Dox from the media. Cells were differentiated by treating the cells with 10 µM all-trans Retinal (RA; Sigma) for 6 days. For transient expression of human genes, cells were transfected by adding 1 µg plasmid DNA/ 4 µL FuGENE® HD Transfection Reagent (Promega).

Lentivirus production and transduction were performed as previously described (*283*). Viral supernatants from HEK 293T fibroblasts were collected at 48-hr after transfection and filtered through a 0.45 µm polyethersulfone membrane. For transduction with individual vector constructs, 2 ml filtered viral supernatant was used to infect 2 x 10^6 cells in the presence of 8 µg/mL polybrene (Sigma) overnight. Cells

were washed with fresh culture medium 1 day after infection, and cultured for following 6 days before RA treatment and αSyn induction.

Randomized gRNA Library Construction and Screening

To build the randomized gRNA library, random oligos containing 20 bp random nucleotide flanked by homology arms to the vector were co-transformed into yeast with a linearized *2`* vector flanked by *RPR1* promoter and gRNA handle at the ends into the screen yeast strain. Once inside the cells, a gRNA-expressing library was reconstituted by the yeast homologous recombination machinery. The randomized oligo library was synthesized by the IDT hand-mixed protocol for randomized oligos using the following template: 5'-GCTGGGAACGAAACTCTGGGAGCTGCGATTGGCAG(N1:32181832)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC-3', where N1 indicates the hand-mixed nucleotide with the following ratio: A:C:G:T = 32:18:18:32. The GC content of the randomized portion of the oligo pool was set to 64% to match with the average GC content of yeast promoters (http://rulai.cshl.edu/SCPD/). The libraries were screened in the presence of both galactose and Dox, and the gRNA content of surviving colonies was characterized by colony PCR followed by Sanger sequencing. Individual gRNAs were verified by cloning each gRNA sequence into the empty gRNA vector and transforming these vectors back into the screen strain to validate gRNA activity in a clean background.

Yeast Growth and Viability Assays

The yeast screen strain was transformed with gRNAs or individual genes obtained from yeast ORF library (Open Biosystems). Single transformant colonies were grown overnight in Scm-Uracil (Ura)+raffinose media in the presence of Dox (1 μg/mL) to induce crisprTF expression. Saturated cultures were diluted to $OD_{600} = 0.1$ in Scm-Ura+Glucose+Dox and Scm-Ura+Galactose+Dox media and grown at 30°C in a Synergy H1 Microplate Reader (BioTek). OD_{600} and fluorescence (excitation and emission spectrum at 508 and 534 nm, respectively) were monitored over the course of the experiments. For measuring cell viability by spotting assays (*284*), cultures were serially diluted (5-fold dilutions) and spotted on Scm-Ura+Glucose+Dox plates for visualizing total viable cells and on Scm-Ura+Galactose+Dox plates for measuring survival. Plates were incubated at 30°C for 2 days.

RNA Preparation and Sequencing

The screen strain was transformed with either a vector expressing gRNA 9-1 or the empty gRNA vector. Two single-colony transformants from each sample were grown overnight in Scm-Ura+Glucose+Dox. These cultures were diluted into the same fresh media to OD_{600} = 0.1 and were incubated at 30°C, 300 RPM. Samples were collected in mid-logarithmic phase (OD_{600} = 0.8) and flash-frozen in liquid nitrogen. Samples were kept in -80°C until further processing. Total RNA samples were prepared using the MasterPure Yeast RNA Purification kit (Epicentre) following the manufacturer's protocol. mRNA libraries were prepared using the Illumina TruSeq library preparation kit, barcoded, multiplexed and sequenced by Illumina HiSeq. The reads were processed by the MIT BioMicroCenter facility pipeline and mapped to the *S. cerevisiae* reference genome (sacCer3). RPKM values were calculated using ArrayStar and differentially expressed genes were identified by t-test (p-value ≤ 0.1, FDR correction(*285*)). Genes that exhibited at least twofold changes in expression in cells containing the gRNA 9-1 compared with the reference (empty gRNA vector) were considered as differentially expressed. Functional classification of the identified genes was performed using the FunSpec webserver (*286*).

Western Blotting and Fluorescence Imaging

Yeast protein extracts were prepared for Western blotting by trichloroacetic acid extraction (*287*). Blots were probed in phosphate-buffered saline containing 0.1% Tween containing 1% (w/v) dried milk. Overexpression constructs containing a 6xHis tag were detected using anti-His monoclonal antibody (1:2000; R93025, Life Technologies) followed by anti-mouse-HRP secondary antibody. αSyn (*SNCA*) was detected with mouse monoclonal anti-αSyn antibodies (1:1000; Syn-1/Clone 42, BD Biosciences).

αSyn-YFP expressing cells were directly visualized under an inverted fluorescence microscope (Zeiss) after 6 days of αSyn induction. The phenotypes were quantified by counting αSyn foci in at least 100 individual cells in multiple randomly chosen fields of view for three independent sets of experiments.

Neuroblastoma Cell Viability and Death Assays

Viable SH-SY5Y cells were quantified by using the CellTiter-Glo Luminescent Cell Viability Assay (Promega). Images were captured using the EVOS™ FL Cell Imaging System directly from culture plates under 10x magnification. Cell death was measured by the FITC Annexin V Apoptosis Detection Kit I (BD Biosciences) followed by LSR

Fortessa II flow cytometry analysis. At least 10,000 cells were recorded per sample in each data set. In the cell death assay (Fig. 8.3C), caspase inhibitor zVAD (Z-VAD-FMK; BD Biosciences) was added to the media upon αSyn induction (100 μM final concentration). For the cell survival assay (Fig. 8.3E), MPP+ iodide (1-methyl-4-phenylpyridinium iodide; Sigma) was added to media of transfected cells 48 hours before processing for cell viability assay.

Quantification and Statistical Analysis

Potential Target Site Analysis

Potential target sites for gRNAs 6-3 and 9-1 in the *S. cerevisiae* genome were identified using CasOT CRISPR off-target search tool (*288*). All potential target sites with up to two mismatches inside the seed region are presented in Table S8.1.

Scoring Strategy for αSyn-toxicity Suppression in Yeast Survival Assays

A defined scoring system, which quantified the numbers of total and full spots in the spotting assays with serial dilutions, was used to score yeast survival upon αSyn induction: cells expressing the empty vector (which showed the least survival upon αSyn induction; sick colonies in the first spot) were scored as 1, and samples overexpressing gRNA 9-1 and *UBP3* (the positive control for αSyn-toxicity suppression) were scored as 6 (five full spots and healthy colonies in the sixth spot). Other samples were scored by visual inspection and comparing the spotting assay survival results with the two abovementioned reference points.

Score	Number of total spots	Number of full spots	Score reference
6	6	5	*UBP3* and gRNA 9-1
5	5	4	
4	4	3	
3	3	2	
2	2	1	
1	1	0	Vector
0	0	0	

Scoring Strategy for αSyn Aggregate Suppression in Fluorescence Microscopy Assays

A defined scoring system, which distinguished the percentage of cell-containing αSyn-YFP foci, was used to score αSyn-aggregate suppression: cells expressing the empty vector were scored as 1 (91.7%), and the samples overexpressing gRNA 9-1 were scored as 10 (6.5%).

Score	% cells with αSyn aggregates	Score reference
10	0 - 10%	gRNA 9-1
9	10 - 20%	
8	20 - 30%	
7	30 - 40%	
6	40 - 50%	
5	50 - 60%	
4	60 -70%	
3	70 - 80%	
2	80 - 90%	
1	90 -100%	Vector

Synergy Quantification

The increased survival against αSyn toxicity by overexpression of *TXN*, *TIMM9*, and *TXN* + *TIMM9* was normalized to the vector control (Fig. 8.3D) or the *EGFP* control (Fig. 8.4B). We considered co-expression of *TXN* + *TIMM9* to be interacting synergistically if the observed combination effect was greater than the expected effect given by Highest Single Agent (*264*), Linear Interaction Effect (*265*), and Bliss Independence (*266*) models. Synergy was calculated based on data presented in Fig. 8.4B and tested by three models respectively, as illustrated in Fig. 8.4C.

Data and Software Availability

The accession number for the RNA-Seq data reported in this paper is GEO: GSE87547. The sequences of recombinant DNA reported in this study have been deposited at Mendeley database (http://dx.doi.org/10.17632/wfskh3hjj5.1).

Randomized gRNA Library Construction and Screening Protocol

1. Prepare yeast competent cells
   a. Grow 100 mL overnight culture of the yeast screen strain in YPD media.
   b. Collect cells by spinning down for 3 minutes at 5000 RPM. Wash twice with 40 mL distilled water by centrifugation for 3 minutes at 5000 RPM and collect cells again.
   c. Resuspend cells in 800 µL distilled water and transfer to a fresh 2 mL microcentrifuge tube. Spin down for 1 minute at 5000 RPM. Remove water and resuspend cells in 1 mL filtered 0.1 M Lithium Acetate (Sigma) by pipetting.
   d. Incubate cells for 30 minutes at 30°C.
   e. Prepare the transformation master mix. For each reaction:

Reagents	Volume (µL)
1 M Lithium Acetate (Sigma)	40
10 mg/mL Denatured Salmon Sperm DNA (Life Technologies)	5
50% Polyethylene Glycol 3350 (Sigma)	240
Distilled Water	65
Total	**350**

2. Construct and screen randomized gRNA library
   a. Linearize the gRNA backbone vector (pRS426-gRNA-HindIII-EcoRI) by HindIII-HF and EcoRI-HF double digestion for 2 hours at 37°C.

Reagents	Volume (µL)
Vector DNA (1 µg/µl)	2
HindIII-HF (New England Biolabs)	1
EcoRI-HF (New England Biolabs)	1
10X CutSmart Buffer (New England Biolabs)	5
Distilled Water	41
Total	**50**

   b. The linearized DNA (5761 bp) was purified and cleaned by QIAquick Gel Extraction Kit (Qiagen). Elute DNA in 40 µL distilled water in the final step, followed by DNA concentration.
   c. The randomized oligo library was synthesized by IDT using the following template: 5'-GCTGGGAACGAAACTCTGGGAGCTGCGATTGGCAG(N1:32181832)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)(N1)GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC-3', where N1 indicates the hand-mixed nucleotide with the following ratio: A:C:G:T = 32:18:18:32.
   d. Mix linearized vector DNA, randomized oligos, 50 µL yeast competent cells, and 350µL transformation mix. Vortex briefly to mix thoroughly.

Reagents	Volume (µL)
Linearized Vector DNA (1 µg/µl)	1
Randomized Oligo Library (100 µM)	3

Yeast Competent Cells	50
Transformation Master Mix	350

e. Incubate at 30°C for 20 minutes. Invert tubes several times at 10 minutes after the start of incubation to mix contents well. Incubate an addition 20 minutes at 42°C.
f. Collect cells by spinning down for 3 minutes at 5000 RPM. Wash with 1 mL distilled water by centrifugation for 3 minutes at 5000 RPM and collect cells again.
g. Resuspend and grow the transformed cells in 100 mL YPD media for 2 hours.
h. Plate 100 µL of YDP culture on Scm−Ura+Glucose plates to determine the library transformation efficiency (with triplicates).
i. Collect remaining cells by spinning down for 3 minutes at 5000 RPM. Wash twice with 40 mL distilled water by centrifugation for 3 minutes at 5000 RPM and collect cells again.
j. Grow cells in 200 mL Scm−Ura+Glucose media in the presence of 1 µg/mL Doxycycline overnight (~12 hours) to induce crisprTF expression.
k. Spin down cells for 3 minutes at 5000 RPM. Resuspend cells in 5 mL distilled water and plate on Scm−Ura+Galactose+Doxycycline plates (100 µL per 10-cm agar plate). Spread cells using sterile glass beads.
l. Plates were incubated at 30 °C for 2-3 days.

3. Characterization and validation of protective gRNAs
   a. Pick up surviving colonies from Scm−Ura+Galactose+Doxycycline plates.
   b. gRNA sequences of surviving colonies were directly PCR-amplified by using KAPA2G Robust PCR Kit (Kapa Biosystems) with T3-promoter (5'-AATTAACCCTCACTAAAGG-3') and T7-promoter (5'-TAATACGACTCACTATAGG-3') primers.
   c. PCR-amplified gRNA fragments (1034 bp) were purified by QIAquick Gel Extraction Kit (Qiagen), and then were cloned back in the linearized pRS426-gRNA-HindIII-EcoRI vector by using Gibson assembly. Independent clones were randomly chosen for Sanger sequencing.
   d. Verified gRNA constructs were individually transformed in both the parental and screen yeast strains to test the suppression of αSyn toxicity.

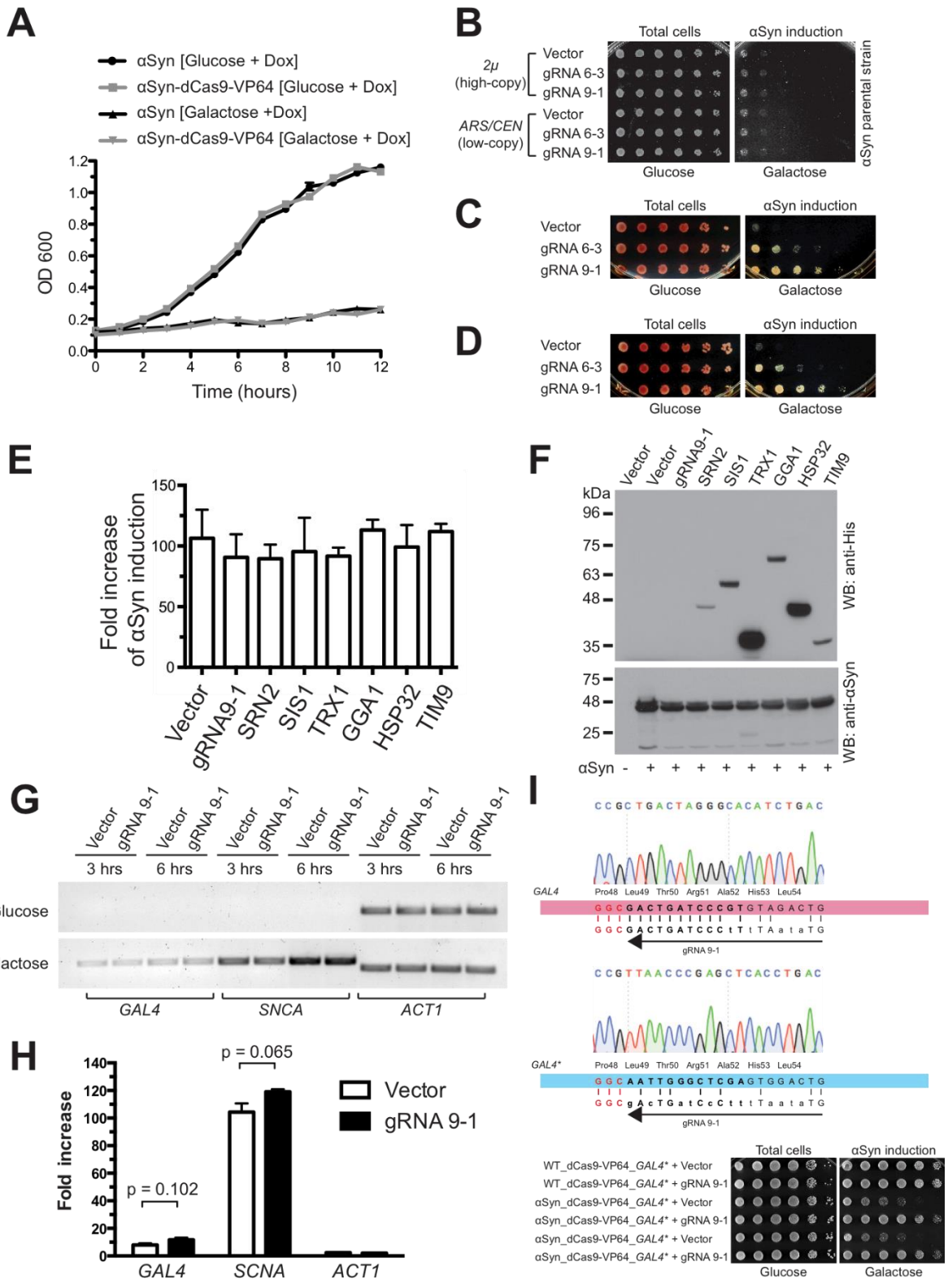
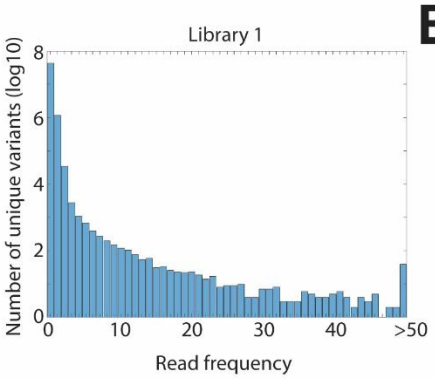
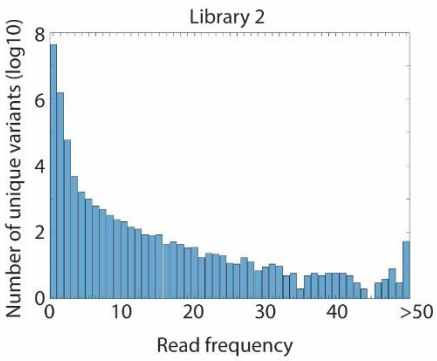

A
αSyn [Glucose + Dox]
αSyn-dCas9-VP64 [Glucose + Dox]
αSyn [Galactose +Dox]
αSyn-dCas9-VP64 [Galactose + Dox]
OD 600
Time (hours)
B
Total cells
αSyn induction
2μ (high-copy)
ARS/CEN (low-copy)
Vector
gRNA 6-3
gRNA 9-1
αSyn parental strain
Glucose
Galactose
C
D
E
Fold increase of αSyn induction
Vector
gRNA9-1
SRN2
SIS1
TRX1
GGA1
HSP32
TIM9
F
kDa
WB: anti-His
WB: anti-αSyn
αSyn
G
3 hrs
6 hrs
Glucose
Galactose
GAL4
SNCA
ACT1
H
p = 0.065
p = 0.102
Fold increase
Vector
gRNA 9-1
I
GAL4
GAL4*
gRNA 9-1
WT_dCas9-VP64_GAL4* + Vector
WT_dCas9-VP64_GAL4* + gRNA 9-1
αSyn_dCas9-VP64_GAL4* + Vector
αSyn_dCas9-VP64_GAL4* + gRNA 9-1
αSyn_dCas9-VP64_GAL4* + Vector
αSyn_dCas9-VP64_GAL4* + gRNA 9-1

Figure S8.1 | gRNA-mediated suppression of αSyn toxicity. (A) Growth profiles of the αSyn-expressing parental yeast strain (black lines) as well as the αSyn and crisprTF (dCas9-VP64)-expressing screen strain (gray lines) were determined in glucose and galactose media (which induces αSyn expression), and in the presence of Dox for dCas9-VP64 induction. These cells did not contain any gRNAs. Cell density was measured by OD_{600} at indicated time points. Parental and screen strains exhibited similar growth profiles in glucose media and both showed severe growth defects upon αSyn induction by galactose, suggesting that expression of dCas9-VP64 by itself does not affect αSyn-mediated toxicity. Error bars represent the standard error of three independent biological replicates. **(B)** gRNA 6-3 and gRNA 9-1 were tested individually in the αSyn-expressing parental yeast strain, which does not express dCas9-VP64 (crisprTF). Neither of the two gRNAs was able to suppress αSyn toxicity in the absence of dCas9-VP64 (crisprTF). These results, along with the data presented in Fig. 8.1B, demonstrate that the protective effect of gRNA 6-3 and gRNA 9-1 in αSyn suppression depends on the expression of dCas9-VP64. **(C-D)** With integrated crisprTF, both gRNAs also rescued yeast viability in the other two high αSyn-expressing strains, HiTox (middle) and 4xαSynTox (bottom), which were published by Dr. Lindquist's group (*231, 240, 242, 245*). **(E)** Overnight cultures of the screen strain overexpressing the indicated genes were induced in Scm-Ura+Gal+Dox for 18 hours. The expression level of αSyn-YFP was quantified by flow cytometry (using LSR Fortessa II flow cytometer equipped with 488/FITC laser/filter set) and normalized to the non-induced control (Scm-Ura+Glucose+Dox). Data are presented as mean ± SEM of three biological replicates. **(F)** The expression of αSyn-YFP and individual genes were further validated by Western blotting of yeast whole cell lysates. **(G)** Overnight cultures of screen yeast cells harboring no gRNA ('Vector') or gRNA 9-1 ('gRNA 9-1') were grown in glucose and galactose media for 3 and 6 hours. Total RNA was extracted from these samples, and the expression level of *GAL4*, *SNCA* (αSyn) and *ACT1* were analyzed by RT-PCR using gene-specific primers. Representative data from two independent experiments are shown. **(H)** Quantitative real-time PCR was performed with the same primers in (G) and normalized to the gene expression in glucose cultures (6 hours, n =4). Primers are listed in Table S8.5. **(I)** One of the potential binding sites of gRNA 9-1 was located within the *GAL4* ORF (Table S8.1). To investigate the effect of gRNA 9-1/crisprTF on *GAL4* expression and exclude the possibility that gRNA 9-1's protective effect was mediated by repressing *GAL4* expression (which acts as the activator of *GAL1* promoter, which drives the expression of αSyn), the potential gRNA 9-1 binding site in the *GAL4* ORF in the screen strain was removed by replacing six synonymous codons from Leu49 to Leu54 (this modified *GAL4* is designated as *GAL4**). Compared with the vector control, gRNA 9-1 consistently achieved αSyn-toxicity suppression in two independent *GAL4** screen strains, indicating that the rescued growth phenotype by gRNA 9-1/crisprTF was independent of the interaction between *GAL4* and gRNA 9-1.

A



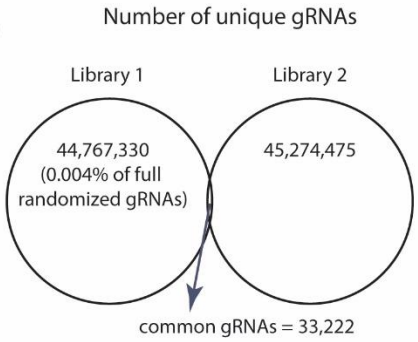
B



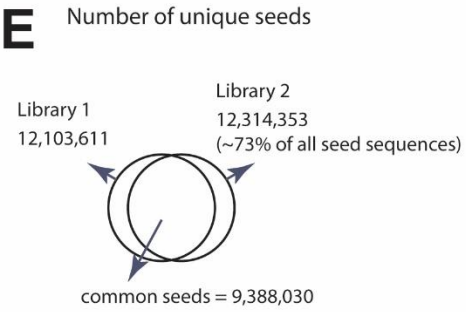
C

Replicate	Number of total reads	Number of variants	Variants per total reads
Library 1	46,053,460	44,767,330	0.982
Library 2	46,993,730	45,274,475	0.979

D



E



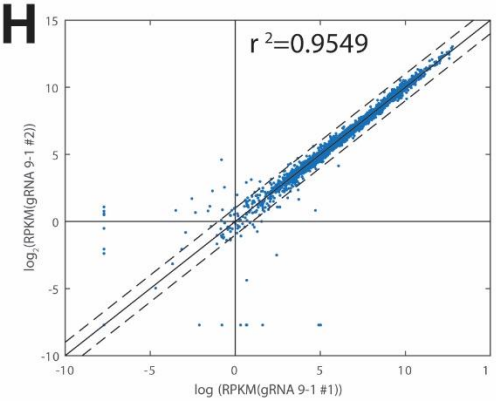
F

Library 1	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10
G	20.11	20.30	21.11	21.51	20.45	19.17	18.74	18.37	18.27	18.73
A	33.62	31.76	30.18	29.78	29.82	30.42	30.49	30.65	30.76	30.64
T	38.17	32.83	33.71	33.80	34.58	34.91	35.28	35.48	35.48	35.00
C	8.11	15.11	15.00	14.91	15.15	15.50	15.50	15.50	15.49	15.64
Library 1	**N11**	**N12**	**N13**	**N14**	**N15**	**N16**	**N17**	**N18**	**N19**	**N20**
G	19.53	19.93	20.08	20.25	19.63	18.70	18.43	18.34	17.70	23.09
A	29.83	28.91	28.30	28.65	29.62	30.99	31.30	28.27	26.96	24.93
T	34.54	34.38	34.20	33.63	33.64	34.03	34.83	36.57	37.49	36.86
C	16.10	16.77	17.43	17.47	17.11	16.27	15.44	16.82	17.84	15.12

G

Library 2	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10
G	20.34	21.02	21.99	22.48	21.02	19.35	18.52	17.94	17.82	18.55
A	33.65	31.17	29.64	29.50	29.80	30.74	31.13	31.35	31.19	30.76
T	38.12	32.00	32.58	32.58	33.63	34.22	35.10	35.62	35.81	35.10
C	7.88	15.81	15.78	15.44	15.54	15.68	15.25	15.09	15.17	15.59
Library 2	**N11**	**N12**	**N13**	**N14**	**N15**	**N16**	**N17**	**N18**	**N19**	**N20**
G	19.81	20.72	21.18	21.42	20.55	19.29	18.54	17.90	17.09	22.23
A	29.55	28.35	27.52	27.87	29.10	30.96	31.43	28.78	27.61	25.55
T	34.19	33.40	32.85	32.20	32.46	33.09	34.47	36.70	37.81	37.18
C	16.45	17.54	18.45	18.51	17.90	16.66	15.55	16.62	17.49	15.04

H



I

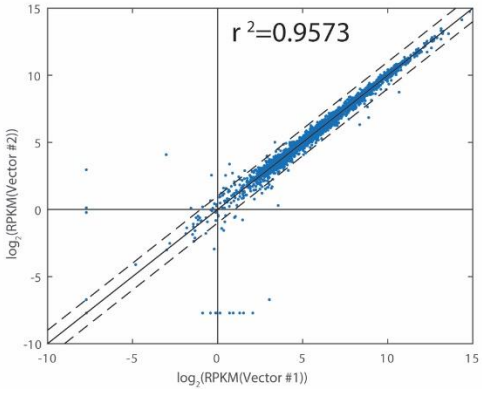


Figure S8.2 | Randomized gRNA libraries for crisprTF-based screening. (A-B) The numbers of unique gRNAs were quantitated by pooled Illumina sequencing after the introduction of the library into the yeast screen strain. The two biological replicates (Library 1 and 2) were transformed and built from the same DNA library described in Detailed Protocols. **(C)** About 98% of reads obtained from Illumina sequencing were unique variants when sequenced at a depth of ~45 million reads. **(D)** Due to the large oligo library (4^{12} = ~ 1 trillion) used to build the libraries prior to yeast transformation, only about 0.004% of the theoretical diversity of the 20-nucleotide randomized gRNAs was covered in each transformation reaction. Therefore, gRNAs that overlapped between two libraries (that were transformed from the same oligo stock) were rare (<0.1%). **(E)** The experimental coverage for the seed sequences was ~73% of the theoretical diversity (~12 million unique seed sequences were detected in either library out of a theoretical diversity of 4^{12}). About 56% of all possible seed sequences were present in both of our transformed libraries (~9.4 million out of 4^{12}). **(F-G)** The DNA library used to encode randomized gRNAs was synthesized to be AT-rich (A = 32%, T = 32%, C =18%, and G =18% at each position) to match the average GC-content of yeast promoters. The frequency of the randomized 20-mer nucleotides was analyzed by two deep sequencing replicates. The normalized values match the designed GC-content of the randomized synthetic oligos, indicating that the library was accurately reconstructed from the randomized oligos *in vivo*. **(H)** Scatter plot of the RNA-Seq data demonstrates that most of the gRNA 9-1 modulated genes had a concordant change in the levels of expression between two biological replicates, $r^2 = 0.9549$. **(I)** Linear regression between two sequencing replicates carrying the vector control, $r^2 = 0.9573$. Mean of RPKM values of the biological replicates for gRNA 9-1 and vector control were compared and plotted as a Volcano plot (Fig. 8.1C); there were numerous outliers (FDR-adjusted p-value ≤ 0.1, fold-change ≥ 2, indicating that a specific group of genes (detailed in Table S8.2) was differentially modulated by gRNA 9-1.

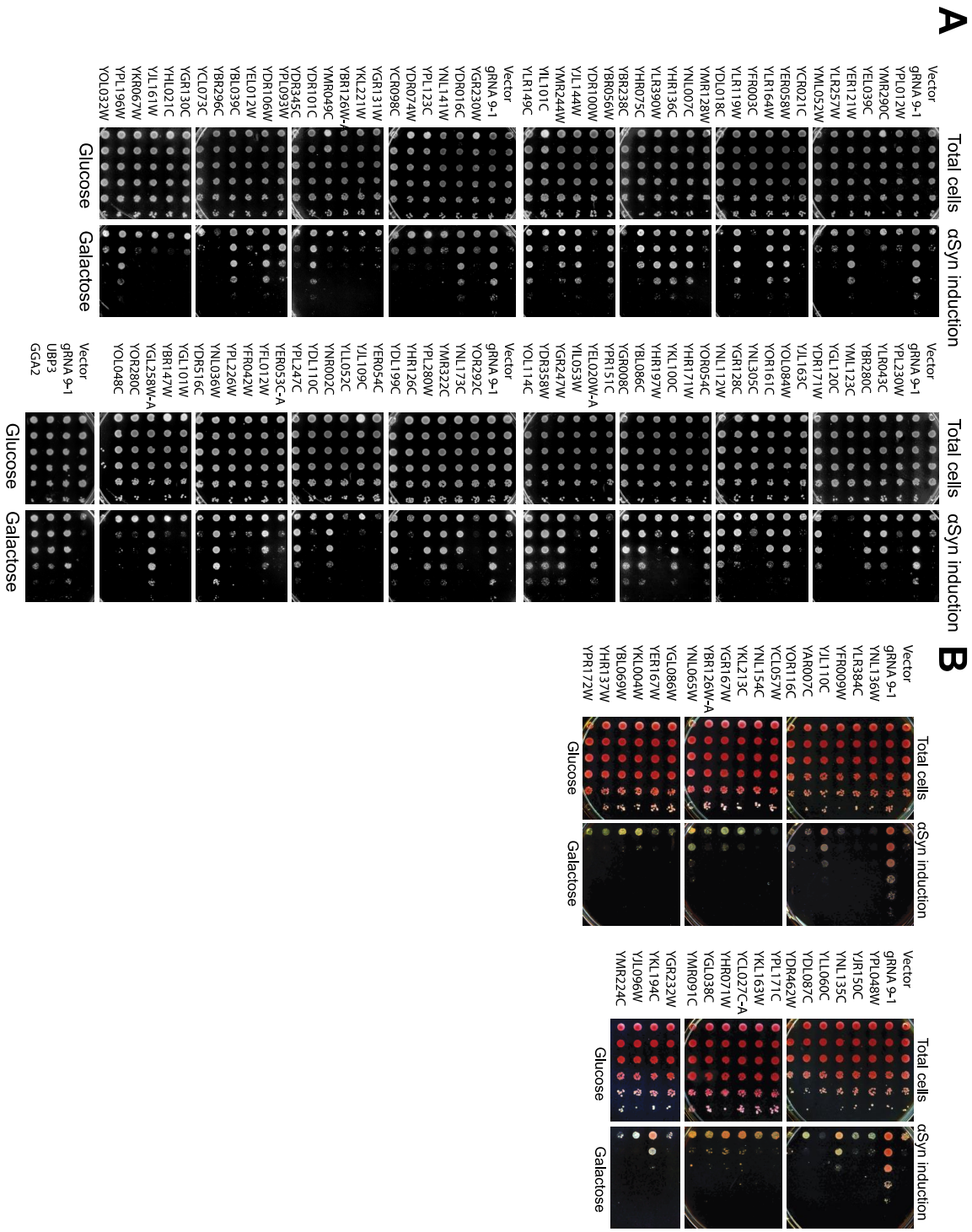


Fig. S8.3 | Systematic overexpression of genes modulated by gRNA 9-1 and evaluation of their effects on αSyn toxicity in yeast. (A) Plasmids containing genes that are modulated by gRNA 9-1 were obtained from the yeast ORF library (Open Biosystems Yeast ORF Collection) and transformed into the screen strain. Cells expressing individual genes were spotted onto galactose plates and scored for the suppression of αSyn toxicity in comparison to cells expressing dCas9-VP64 and gRNA 9-1 as well as those expressing dCas9-VP64 and vector control. *UBP3* (a known suppressor of αSyn toxicity) was used as a positive control. A complete list of differentially expressed genes and annotations, as well as associated scores, are

presented in Table S8.2. **(B)** Examination of αSyn-toxicity suppression by a randomly selected set of overexpressed genes from the yeast ORF library. Thirty-four yeast genes were randomly chosen from the yeast ORF library (Open Biosystems Yeast ORF Collection) and transformed into the screen strain. Cell survival in the presence of αSyn induction was measured by spotting assays and compared to survival of cells expressing dCas9-VP64 and gRNA 9-1 ('gRNA 9-1'; scored as 6) as well as those expressing dCas9-VP64 and vector control ('Vector'; scored as 1). Only five genes (*YJL110C*, *YOR116C*, *YNL065W*, *YNL135C*, and *YKL194C*) out of 34 genes scored greater than or equal to 3. A complete list of genes and annotations as well as associated scores are presented in Table S8.4.

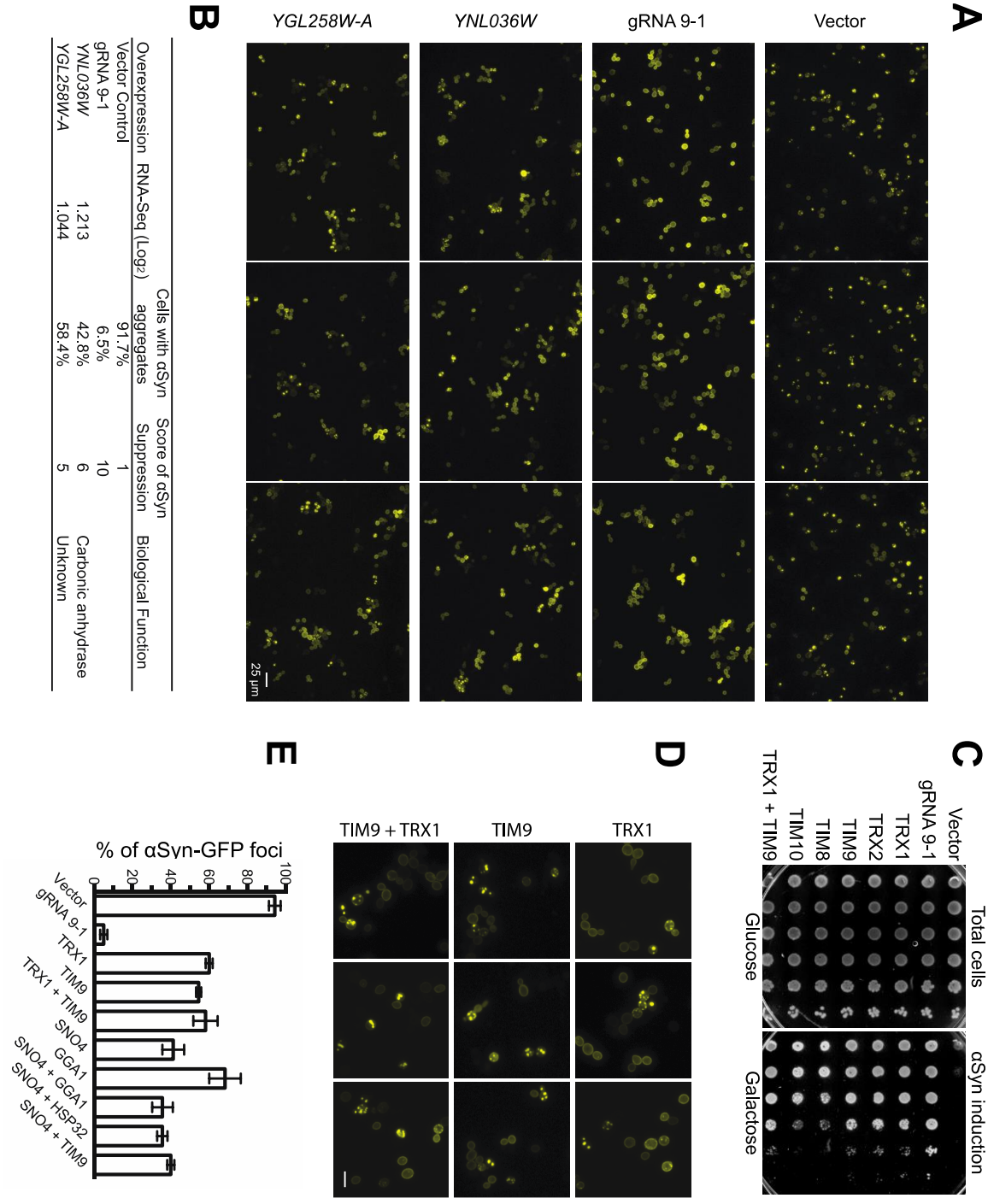


Figure S8.4 | gRNA 9-1/crisprTF is the strongest modulator of αSyn-induced phenotypes. (A-B) We systematically tested the protective effects of the individual genes that were differentially controlled by gRNA 9-1/crisprTF. Fluorescent microscopy clearly showed that gRNA 9-1/crisprTF was the strongest suppressor of αSyn aggregates compared with any individual genes found in this study or previous genome-wide screens. αSyn-YFP foci were observed in only 6.5% of cells that overexpressed gRNA 9-1/crisprTF (scored as 10). This significantly outperformed cells overexpressing *YNL036W* (42.8% cells with αSyn-YFP foci; scored as 6), *YGL258W-A* (58.4% cells with αSyn-YFP foci; scored as 5), and other individual genes in Figure 2 (summarized in Table 8.1 and Table S8.2). **(C)** Yeast *TRX* and *TIM* family

proteins function together to protect mitochondria from oxidative stresses (*273*). Genes in both families were highlighted in gRNA 9-1 expression profiling. Cells harboring individual genes from *TRX* (*TRX1* and *TRX2*) and *TIM* (*TIM8*, *TIM9*, and *TIM10*) families were overexpressed in the yeast screen strain to test for αSyn-toxicity protection. All these proteins strongly suppressed αSyn toxicity when overexpressed. We did not observe synergistic protective effects in yeast assays when *TRX1* and *TIM9* were co-expressed, in contrast to our results in human cells. **(D)** Representative images of αSyn-YFP foci in screen yeast cells overexpressing *TRX1*, *TIM9* or both. Bar = 10 μm. **(E)** Other co-expressed gene pairs (*SNO4 + GGA1, SNO4 + HSP32,* and *SNO4 + TIM9*) were tested in αSyn-YFP foci assays. None of them demonstrated synergistic αSyn protection compared to single gene expression.

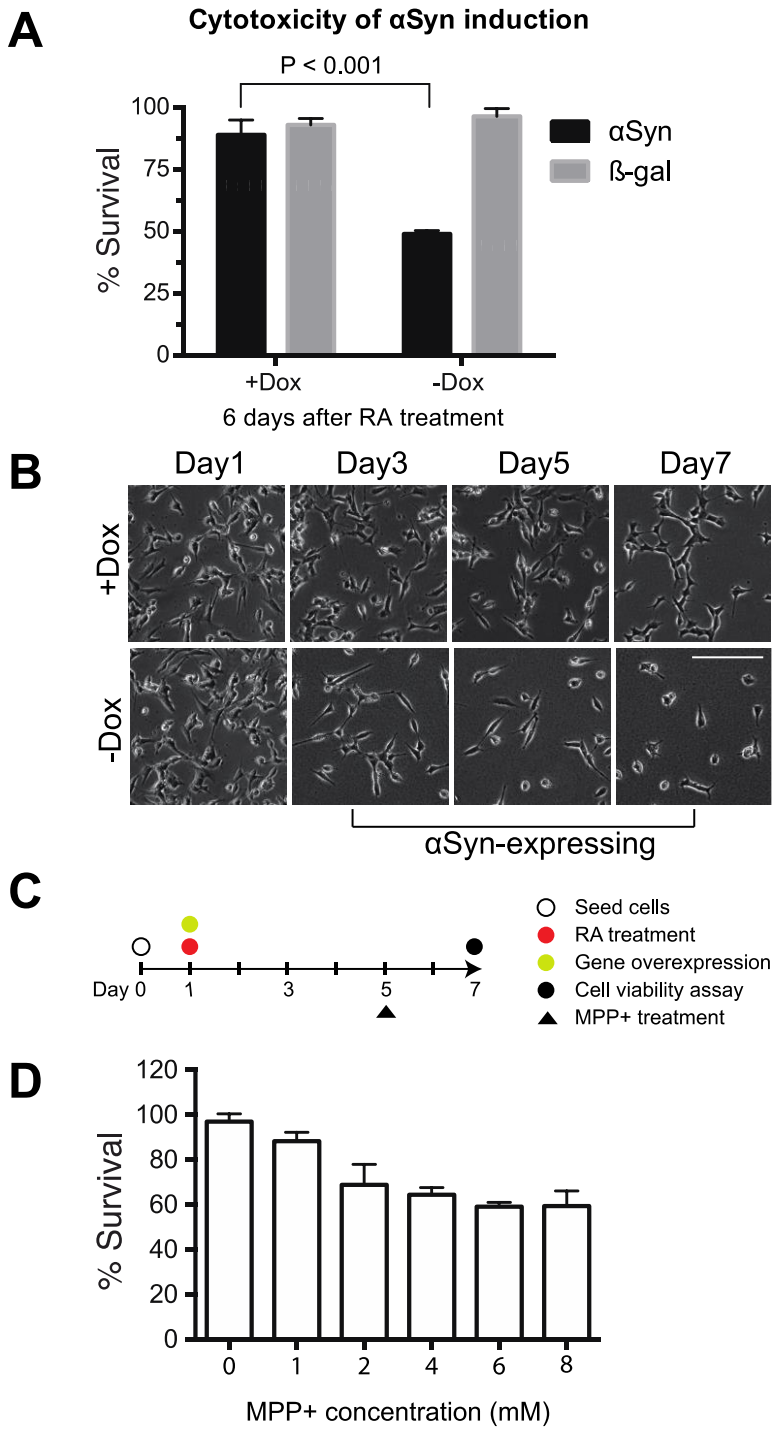


Figure S8.5 | Inducible expression of αSyn in the human neuronal model of PD. (A) αSyn and ß-gal (non-toxic negative control) expression were induced in human SH-SY5Y neuroblastoma cells by removal of Dox from media. αSyn-expressing cells significantly lost viability at the 6th-day post-differentiation (retinoic acid treatment). **(B)** Representative images showing retraction of neuritic processes, membrane blebbing, and cell death in αSyn-expressing cells (-Dox condition). Bar = 10 µm. **(C)** Schematic of the experimental procedure used to study the effect of MPP+, a known inducer of neuronal cell death, on differentiated SH-SY5Y cells. **(D)** A series of titration treatments were performed to identify the minimal concentration of MPP+

that result in maximal toxicity. Cells were treated with different concentrations of MPP+ for 48 hours and cell viability was measured by CellTiter-Glo luminescent assay and normalized to the non-MPP+ treatment (n = 3). 6 mM MPP+ was found to be the optimal concentration for maximal toxicity and therefore was used in the survival assay.

Table S8.1 | Potential binding sites for gRNA 6-3 and gRNA 9-1 in the yeast genome.

gRNA	Target Site	Target Sequence	PAM	Number of Mismatches in Seed Region	Total Number of Mismatches	Gene Name
gRNA 6-3	I:115314-115337:-	GGtaaTga˙CTTCTtgACAGG-TGGC	NGG	2	7	YAL019W-A
gRNA 6-3	II:141581-141604:+	ttcaacaT˙CTTCTgTACgGG-AAGA	NAG	2	9	PRE7
gRNA 6-3	II:212896-212919:-	atGaaTac˙CTTCaATACtGG-TGGT	NGG	2	8	SLA1
gRNA 6-3	II:447874-447897:+	aGGggaaT˙CTTgTATACAGa-AAGT	NAG	2	7	SIF2
gRNA 6-3	II:524119-524142:-	GatTaTTg˙CTTCTATAttGG-ATGG	NNGG	2	6	IRA1
gRNA 6-3	III:164910-164933:-	tttaaaga˙CTTCTATAgAtG-AAGA	NAG	2	10	NPP1
gRNA 6-3	III:210476-210499:+	ttcTTcTT˙CTTCTtTACtGG-GAGT	NAG	2	6	IMG1
gRNA 6-3	IV:1723-1746:-	GctTTTcg˙CTTtTATACAGc-AGGA	NGG	2	6	COS7
gRNA 6-3	IV:17896-17919:-	ttcgggTa˙CTTCTAaACAGa-CGGA	NGG	2	9	AAD4
gRNA 6-3	IV:112128-112151:+	tatcggaa˙aTTCTtTACAGG-TAGG	NAG	2	10	SNF3
gRNA 6-3	IV:145433-145456:-	tGtagcac˙CTTCTATAgAGG-AAGT	NAG	1	8	AIR2
gRNA 6-3	IV:309773-309796:+	tGcaaTTT˙CTTCTAaACAGt-GCGG	NNGG	2	6	RPL13A
gRNA 6-3	IV:354026-354049:-	GatTggag˙CaTCTATAtAGG-GAGC	NAG	2	8	MBP1
gRNA 6-3	IV:579322-579345:+	GatTTcca˙CTTCTgTACAGa-TGGA	NGG	2	7	AIM7
gRNA 6-3	IV:588552-588575:+	tcccTTag˙aTTCTgTACAGG-AAGA	NAG	2	8	FMP16
gRNA 6-3	IV:700228-700251:+	tttcTcag˙CTTaTATAaAGG-ATGG	NNGG	2	9	YDR124W
gRNA 6-3	IV:754827-754850:+	cGGaggaa˙CTTCaATAgAGG-TAGA	NAG	2	8	KGD2
gRNA 6-3	IV:780586-780609:-	GttTgTcT˙CTTaTATACAGc-CGGA	NGG	2	6	YDR161W
gRNA 6-3	IV:825659-825682:-	ataTccgg˙CTTCTtTgCAGG-GAGT	NAG	2	9	SCC2
gRNA 6-3	IV:868491-868514:-	GttTaacT˙CTTCaATAgAGG-TCGG	NNGG	2	7	MSS4
gRNA 6-3	IV:1032570-1032593:+	cGtgggTT˙CTTCTATAgAGG-GAGA	NAG	1	6	ZIP1
gRNA 6-3	IV:1174819-1174842:+	ctcaTaTa˙tTTgTATACAGG-AAGA	NAG	2	8	
gRNA 6-3	IV:1188425-1188448:+	tGaaaTca˙CTTCTtTACAaG-AGGA	NGG	2	8	SPC110
gRNA 6-3	IV:1195737-1195760:+	aacgTaaT˙CTTgTATAaAGG-TGGA	NGG	2	8	BCP1
gRNA 6-3	IV:1197663-1197686:+	tttgaaaa˙tTTgTATACAGG-AGGA	NGG	2	10	TFC6
gRNA 6-3	IV:1266351-1266374:+	tttagaag˙CTTCTATtCAaG-AAGA	NAG	2	10	SXM1
gRNA 6-3	IV:1320005-1320028:+	taaaTaga˙CaTCTATACAcG-TAGT	NAG	2	9	DYN2
gRNA 6-3	IV:1385972-1385995:-	tatgcTTc˙CcTCcATACAGG-CAGG	NAG	2	8	YDR461C-A
gRNA 6-3	IX:102306-102329:+	cctTcaaT˙CTTCTATAgAGc-CGGT	NGG	2	8	FKH1

gRNA 6-3	IX:187934-187957:+	tGtTTTTa˙aTTaTATACAG G-TTGG	NNG G	2	5	LYS12
gRNA 6-3	IX:237025-237048:+	acGagTca˙CTTCTATAagG G-TAGG	NAG	2	8	YIL067C
gRNA 6-3	V:20258-20281:-	aaccTTgT˙CTTCaATcCAG G-CGGC	NGG	2	7	DSF1
gRNA 6-3	V:133587-133610:+	taaggaaa˙CTTCTAaACAGt-TCGG	NNG G	2	10	GLC3
gRNA 6-3	V:192917-192940:+	tttcTgac˙CTTCaATACAtG-GGGG	NGG	2	9	SPC25
gRNA 6-3	V:263157-263180:+	cctgacTT˙tTTaTATACAGG -TGGC	NGG	2	8	GIP2
gRNA 6-3	V:282867-282890:+	tcGaggcT˙CTTCTtTACcGG -GAGT	NAG	2	8	YER064C
gRNA 6-3	V:391439-391462:+	cctTaaac˙CTTCTATAaAtG-CAGA	NAG	2	9	BOI2
gRNA 6-3	V:393983-394006:+	tttcaacg˙CTTCTAaAtAGG-GAGA	NAG	2	10	BOI2
gRNA 6-3	VI:79298-79321:+	atacaTaT˙tTTCTATACAG G-GGGT	NGG	1	7	CAK1
gRNA 6-3	VI:232258-232281:+	atGaTTcT˙CTTCTATAtAG G-CAGG	NAG	1	5	IRC5
gRNA 6-3	VI:243249-243272:-	aacgTggT˙CTaCTATACAG G-AGGA	NGG	1	7	YFR045W
gRNA 6-3	VII:15025-15048:-	taaTaacc˙CTTtTATACAtG-TTGG	NNG G	2	9	ADH4
gRNA 6-3	VII:15750-15773:-	aacaTaag˙CTTCgATACAGt -GAGT	NAG	2	9	ADH4
gRNA 6-3	VII:158407-158430:+	tctcaTTc˙tTTCTATAaAGG -GGGC	NGG	2	8	GTS1
gRNA 6-3	VII:385070-385093:-	cataTaTa˙CTTaTATACAGc -GAGA	NAG	2	8	PUS2
gRNA 6-3	VII:780897-780920:-	aaccaaaT˙CTTCaAcACAGG -TAGC	NAG	2	9	THI4
gRNA 6-3	VII:1023603-1023626:+	tGtcagTg˙CTTCTAaACAaG -ATGG	NNG G	2	8	YGR266W
gRNA 6-3	VII:1030947-1030970:-	GaaTcTcc˙tTTtTATACAG G-TTGG	NNG G	2	7	YTA7
gRNA 6-3	VII:1045389-1045412:-	atccaaga˙CTTCTgTACAaG-AGGA	NGG	2	10	RNH70
gRNA 6-3	VIII:19909-19932:-	atccaaTT˙CTTCcATAtAGG -CTGG	NNG G	2	8	ARN1
gRNA 6-3	VIII:76355-76378:-	ttaaTTag˙CTTCTtTACAtG-CGGC	NGG	2	8	YLF2
gRNA 6-3	VIII:145978-146001:-	tGaccTTc˙aTTCaATACAG G-TTGG	NNG G	2	7	YHR020W
gRNA 6-3	VIII:231050-231073:-	ttcgTTgc˙CTTtgATACAGG -GAGT	NAG	2	8	SSF1
gRNA 6-3	VIII:300415-300438:+	ccGgaagT˙CaTCTcTACAG G-ATGG	NNG G	2	8	SFB3
gRNA 6-3	VIII:397328-397351:+	aGtgTTgg˙aTTCTATACtG G-AGGC	NGG	2	7	PEX28
gRNA 6-3	X:36038-36061:-	cGacaggc˙aTTCcATACAGG -AGGA	NGG	2	9	OPT1
gRNA 6-3	X:53270-53293:+	ttGaaaca˙CTTaaATACAGG -AAGA	NAG	2	9	RCY1
gRNA 6-3	X:98173-98196:+	aGcTgggT˙CTTCTATACAc a-TGGG	NGG	2	7	CPS1
gRNA 6-3	X:146252-146275:-	tGcgTTgT˙CTTtTATAtAG G-CGGA	NGG	2	6	SFH5
gRNA 6-3	X:544977-545000:-	cGacgaaa˙CTcaTATACAGG -AGGT	NGG	2	9	CDC8
gRNA 6-3	X:632972-632995:-	taaaaTTa˙gTTCTATAaAG G-AAGA	NAG	2	8	CPA2
gRNA 6-3	XI:36621-36644:+	tctaTagg˙CTaCTATACAtG-AAGG	NAG	2	9	SAC1

gRNA 6-3	XI:208050-208073:-	caaaaTaT˙CTTtTATACAaG-GAGA	NAG	2	8	YPK1
gRNA 6-3	XI:247133-247156:-	GtGgTcgc˙CTTCTtTACAaG-AAGA	NAG	2	7	LAP4
gRNA 6-3	XI:304194-304217:-	aaGgTgca˙CTTtTATACAaG-CTGG	NNG	2	8	
gRNA 6-3	XI:339579-339602:-	tatTTTTT˙CTTCgATAtAGG-GAGA	NAG	2	5	MDM35
gRNA 6-3	XI:528075-528098:+	tGcTggag˙CgTCTAcACAGG-GCGG	NNG	2	8	TRK2
gRNA 6-3	XI:582181-582204:+	ttccaTTT˙CTTgTATAaAGG-TAGT	NAG	2	7	ECM4
gRNA 6-3	XII:248557-248580:-	atGTgcag˙CTTCTAaACAGc-ACGG	NNG	2	8	YLR053C
gRNA 6-3	XII:260469-260492:+	aaacggaT˙CTTCTgTACAGc-GAGA	NAG	2	9	REX2
gRNA 6-3	XII:322130-322153:-	tataTaTa˙CaTaTATACAGG-TAGG	NAG	2	8	XDJ1
gRNA 6-3	XII:490870-490893:+	atGaTaaa˙CTTCTAcACtGG-AAGG	NAG	2	8	RRT15
gRNA 6-3	XII:531094-531117:-	GatcagTT˙CTTtTATgCAGG-TAGA	NAG	2	7	ATG26
gRNA 6-3	XII:546408-546431:+	cttaggTc˙CTTCTATtaAGG-AAGA	NAG	2	9	NOP56
gRNA 6-3	XII:708601-708624:+	aatTTcac˙CTTCagTACAGG-TAGA	NAG	2	8	NNT1
gRNA 6-3	XII:892092-892115:-	tGcTgTTT˙CTTCTgTAgAGG-AGGT	NGG	2	5	IKI3
gRNA 6-3	XIII:150756-150779:-	cacaacaT˙tTTtTATACAGG-GAGT	NAG	2	9	PIF1
gRNA 6-3	XIII:559528-559551:+	cacaagTg˙CTgCaATACAGG-AGGA	NGG	2	9	YMR147W
gRNA 6-3	XIII:647508-647531:-	tGcagaTT˙CTTCTATgCAGt-CAGC	NAG	2	7	GYL1
gRNA 6-3	XIII:647736-647759:-	taaaggaT˙CTTCTATACgGc-GTGG	NNG	2	9	GYL1
gRNA 6-3	XIII:804718-804741:+	GtagcTca˙CcTCTATACAGG-TGGT	NGG	1	7	PRP24
gRNA 6-3	XIII:919017-919040:-	cacggTcc˙CTTCTATAaAGa-TGGT	NGG	2	9	SNO4
gRNA 6-3	XIV:30102-30125:-	GaGcaggg˙CTTCTAaACAaG-ATGG	NNG	2	8	FIG4
gRNA 6-3	XIV:291671-291694:+	cataTTaT˙CTcCTATACAcG-AGGC	NGG	2	7	UBP10
gRNA 6-3	XIV:318849-318872:+	tcagaaTa˙tTTCTATAtAGG-AAGT	NAG	2	9	FMP41
gRNA 6-3	XIV:476045-476068:+	acaaaTac˙aTTaTATACAGG-GAGT	NAG	2	9	PMS1
gRNA 6-3	XIV:504894-504917:+	atGggTTc˙CTTCTtcACAGG-TAGA	NAG	2	7	AQR1
gRNA 6-3	XIV:676546-676569:-	ttccgTTT˙CTTCaATAgAGG-AGGA	NGG	2	7	CPR8
gRNA 6-3	XIV:775608-775631:+	aaccTTgT˙CTTCaATcCAGG-CGGC	NGG	2	7	YNR073C
gRNA 6-3	XV:9645-9668:+	aaGagTTc˙tTTCTATACAtG-TAGA	NAG	2	7	YOL163W
gRNA 6-3	XV:124102-124125:+	tGaTcaaT˙CTTCTAcgCAGG-GAGA	NAG	2	7	ITR2
gRNA 6-3	XV:251169-251192:-	tGtagcTT˙CcTCTATACAtG-CTGG	NNG	2	7	NGL1
gRNA 6-3	XV:390765-390788:+	GtGTTgTT˙CTTaTATACAGG-AGGC	NGG	1	3	HMS1
gRNA 6-3	XV:787651-787674:+	attTcaca˙CTTtTATACAaG-AGGA	NGG	2	9	MET7
gRNA 6-3	XV:1052491-1052514:-	aacgaagT˙CTTCTATACAaa-GAGA	NAG	2	9	RDR1

gRNA 6-3	XVI:11926-11949:+	cacggTcc˙CTTCTATAaAGa-TGGT	NGG	2	9	HSP32
gRNA 6-3	XVI:11782 0-117843:+	atcggTTT˙CTTCTATtCAtG-TAGT	NAG	2	7	YPL229 W
gRNA 6-3	XVI:17537 7-175400:-	GaGcaacg˙tTaCTATACAG G-GAGT	NAG	2	8	OXR1
gRNA 6-3	XVI:41544 3-415466:-	ataTaTaT˙tTTCTATAaAG G-TAGT	NAG	2	7	GCR1
gRNA 6-3	XVI:55223 8-552261:+	GccaTTgg˙CTTCTAaACAG c-TAGA	NAG	2	7	ULA1
gRNA 6-3	XVI:60912 8-609151:-	GtaaTaTg˙CTTtTATAtAG G-TTGG	NNG G	2	7	EAF3
gRNA 6-3	XVI:77198 5-772008:+	tGcgcTaa˙CTTCTATAaAG G-AGGG	NGG	1	7	CLB2
gRNA 9-1	II:124775-124798:+	aacacgcT˙TTCCCTAGTCtG-TAGC	NAG	1	8	PIN4
gRNA 9-1	II:141015-141038:-	GccTAtgc˙TTCaCTAGTCA c-AGGC	NGG	2	7	PRE7
gRNA 9-1	II:190562-190585:+	tcActtcT˙TTCCCTAcTCAt-GGGC	NGG	2	8	PEP1
gRNA 9-1	III:174520-174543:+	agggcAaT˙TTCCCcAaTCA G-TTGG	NNG G	2	8	SYP1
gRNA 9-1	IV:187150-187173:+	GTtTtcgT˙TgCCCTcGTCA G-CCGG	NNG G	2	6	ATG9
gRNA 9-1	IV:203591-203614:-	tTgaccTT˙TTCCtgAGTCA G-AAGA	NAG	2	7	BPL1
gRNA 9-1	IV:361768-361791:+	tgATAcag˙TaCCCcAGTCA G-TGGC	NGG	2	7	MCH1
gRNA 9-1	IV:373436-373459:+	aatgAtaT˙TTCCCcAtTCAG -TGGA	NGG	2	8	FAD1
gRNA 9-1	IV:512002-512025:-	tctccAgc˙TTCaCTAGaCAG-TGGT	NGG	2	9	LYS14
gRNA 9-1	IV:1216099 -1216122:+	agcagcTg˙TTtCaTAGTCAG -CGGA	NGG	2	9	XRS2
gRNA 9-1	IV:1222881 -1222904:-	acAaAATc˙TTCCCTAGctA G-TTGG	NNG G	2	6	FRQ1
gRNA 9-1	IV:1453832 -1453855:+	aagTtgag˙TTCtCaAGTCAG -CGGT	NGG	2	9	SAM2
gRNA 9-1	IX:341811-341834:-	Gctatgca˙TTCCCaAtTCAG-AAGA	NAG	2	9	FAA3
gRNA 9-1	V:77444-77467:+	tccTgcgT˙TTCCgTAGTCAa -GGGT	NGG	2	8	YEF1
gRNA 9-1	V:339086-339109:+	catctggT˙TTCtCaAGTCAG-CGGT	NGG	2	9	TRP2
gRNA 9-1	V:394153-394176:-	aaAggtga˙TTCtCTAGTCAc-GTGG	NNG G	2	9	BOI2
gRNA 9-1	VI:190133-190156:-	catTttaT˙TTCtaTAGTCAG-AAGT	NAG	2	8	FAB1
gRNA 9-1	VI:191084-191107:+	aTggAtTa˙TTCCtTAGTCA t-TGGT	NGG	2	7	FAB1
gRNA 9-1	VIII:89453-89476:-	cacattTg˙TTCCaTtGTCAG-TTGG	NNG G	2	9	YHL009 W-B
gRNA 9-1	VIII:23966 8-239691:-	ccAaAtTT˙TTCCCcAGTgA G-GGGA	NGG	2	6	YHR071 C-A
gRNA 9-1	VIII:29417 2-294195:+	tcAgttcT˙TTCCCTAGTatG-TAGT	NAG	2	8	HXT5
gRNA 9-1	X:201460-201483:-	cacattTg˙TTCCaTtGTCAG-TTGG	NNG G	2	9	YJL113W
gRNA 9-1	X:361346-361369:-	actcggaT˙TTCCCTgGTCtG-GAGC	NAG	2	9	YJL043W
gRNA 9-1	X:377567-377590:-	tagTAATa˙TTtaCTAGTCA G-TGGG	NGG	2	6	IRC18
gRNA 9-1	X:389949-389972:+	tctTtgaa˙TTCCCTttTCAG-AAGT	NAG	2	9	VPS53
gRNA 9-1	X:716216-716239:+	GatcAcTT˙TTtCCcAGTCA G-TAGA	NAG	2	6	DAN4

gRNA 9-1	XI:101766-101789:+	GTtgAAaT˙TTCtCTAGTCAa-TGGT	NGG	2	5	FAS1
gRNA 9-1	XI:290574-290597:-	cTtgAcTT˙TTCCCTAGTtcG-TAGA	NAG	2	6	DHR2
gRNA 9-1	XI:451589-451612:-	aacTctTg˙TTCCCTgGcCAG-CAGT	NAG	2	8	MRPL13
gRNA 9-1	XI:609043-609066:+	GTAatccg˙TTCagTAGTCAG-AGGA	NGG	2	7	PXL1
gRNA 9-1	XII:329488-329511:+	tgccgtcT˙gaCCCTAGTCAG-GAGC	NAG	2	9	GIS3
gRNA 9-1	XII:689622-689645:+	ccAgtgcT˙TTCCCTAGTCcG-TGGT	NGG	1	7	PIG1
gRNA 9-1	XII:780325-780348:-	aTATAtaa˙aTCCCTcGTCAG-GGGA	NGG	2	6	PEX30
gRNA 9-1	XII:820724-820747:-	aaATAAaT˙TgCCCgAGTCAG-TGGA	NGG	2	5	YLR345W
gRNA 9-1	XIII:471225-471248:+	agtgtATT˙TTCCCTccTCAG-GGGA	NGG	2	7	YMR102C
gRNA 9-1	XIII:794455-794478:+	GagggAga˙TgCCCTgGTCAG-GAGC	NAG	2	8	YMR262W
gRNA 9-1	XIV:48908-48931:+	aaATgtca˙TTCCaTAGcCAG-TGGA	NGG	2	8	RFA2
gRNA 9-1	XIV:176427-176450:+	cTtTctaa˙TTCCCTcaTCAG-GCGG	NNGG	2	8	RAD50
gRNA 9-1	XIV:197156-197179:+	accggtcT˙TTCCaTAGTCAa-GAGA	NAG	2	9	ZWF1
gRNA 9-1	XIV:660306-660329:+	cgActAgT˙TTCCCcAGTCtG-ACGG	NNGG	2	7	ACC1
gRNA 9-1	XV:660238-660261:-	tTcTcATT˙TTtCCTAtTCAG-AGGA	NGG	2	5	ALE1
gRNA 9-1	XV:729770-729793:+	tccaAgcT˙TTCtCTtGTCAG-CTGG	NNGG	2	8	NOC2
gRNA 9-1	XVI:82193-82216:+	GTcagATg˙TgCCCTAGTCAG-CGGA	NGG	1	5	GAL4
gRNA 9-1	XVI:142896-142919:-	tTgattcg˙gTCCCTcGTCAG-GAGA	NAG	2	9	BMS1
gRNA 9-1	XVI:237718-237741:-	tccTgtcT˙TTCCgTgGTCAG-TGGG	NGG	2	8	ATG29
gRNA 9-1	XVI:337889-337912:-	tcATtcTa˙TTCCtTtGTCAG-TAGA	NAG	2	7	PEX25
gRNA 9-1	XVI:439175-439198:+	cacattTg˙TTCCaTtGTCAG-TTGG	NNGG	2	9	YPL060C-A
gRNA 9-1	XVI:544973-544996:-	GacaAAcc˙TTCCtTgGTCAG-CAGC	NAG	2	7	NCR1
gRNA 9-1	XVI:587480-587503:+	GTActcTa˙cTCCCaAGTCAG-CGGA	NGG	2	6	YPR014C
gRNA 9-1	XVI:619512-619535:+	aTtggcTc˙TTCtCTcGTCAG-TAGG	NAG	2	8	ATH1
gRNA 9-1	XVI:883548-883571:+	tTccAAgT˙TTaCCTAGcCAG-AAGA	NAG	2	6	YPR170C

Table S8.2 | gRNA 9-1 regulated genes that suppress αSyn toxicity when overexpressed.

Systematic Name	Standard Name	Fold change (log2(gRNA9 RPKM/Ref RPKM))	p-value	Survival Score	Human Homolog and Ortholog
YPL280W	HSP32	-9.59341	0.06	6	PARK7
YEL020W-A	TIM9	3.84556	0.09	6	TIMM9
YLR164W	SHH4	1.63076	0.09	6	SDHD
YMR322C	SNO4	2.03475	0.1	6	PARK7
YGR008C	STF2	2.00423	0.1	6	HABP4, SERBP1
YDR358W	GGA1	1.24093	0.1	6	GGA1, GGA2, GGA3, TOM1, TOM1L1, TOM1L2, WDFY1, WDFY2
YER058W	PET117	1.23464	0.1	6	
YLR390W	ECM19	1.2306	0.1	6	
YBR280C	SAF1	1.18006	0.1	6	ALS2, RCC1
YDR106W	ARP10	1.16556	0.1	6	
YNL007C	SIS1	1.15439	0.1	6	DNAJ (B1-B9), DNAJC5, DNAJC5B, DNAJC5G
YJL144W		1.14196	0.1	6	
YDR016C	DAD1	1.10875	0.1	6	
YBL086C		1.1009	0.1	6	
YLR043C	TRX1	1.07168	0.1	6	TXN, TXNDC2, TXNDC8
YOR054C	VHS3	1.07053	0.1	6	PPCDC
YBR056W		1.06549	0.1	6	
YLR119W	SRN2	1.03094	0.1	6	
YDR101C	ARX1	-1.04375	0.1	6	PA2G4, XPNPEP1, XPNPEP2
YHR136C	SPL2	-1.26719	0.1	6	
YNL112W	DBP2	-1.69614	0.1	6	DDX17, DDX41, DDX43, DDX5, DDX53, DDX59
YDL199C		1.14338	0.09	5	
YOL084W	PHM7	1.58682	0.1	5	TMEM63 (A-C)
YOL114C		1.49153	0.1	5	ICT1
YPL247C		1.49053	0.1	5	DCAF7
YER121W		1.43396	0.1	5	
YDR171W	HSP42	1.43394	0.1	5	CRYAA, CRYAB, HSPB1, HSPB3, HSPB6, HSPB7, HSPB8, HSPB9
YNL036W	NCE103	1.21279	0.1	5	
YLR149C		1.1335	0.1	5	
YNR002C	ATO2	1.08266	0.1	5	
YGR247W	CPD1	1.06796	0.1	5	
YGL258W-A		1.04446	0.1	5	
YOR161C	PNS1	1.0314	0.1	5	SLC44 (A1-A5)
YKL100C		1.01091	0.1	5	HM13, SPPL2A, SPPL2C, SPPL3

YPL196W	OXR1	1.00292	0.1	5	NCOA7, OXR1, TLDC2
YBL039C	URA7	-1.0346	0.1	5	CTPS1, CTPS2
YMR244W		-1.8109	0.1	5	
YHR075C	PPE1	1.05839	0.09	4	PPME1
YFL012W		1.49125	0.1	4	
YNL173C	MDG1	1.19461	0.1	4	
YGR128C	UTP8	-1.00839	0.1	4	
YPL093W	NOG1	-1.14212	0.1	4	GTPBP4
YDR074W	TPS2	1.10559	0.09	3	
YGR230W	BNS1	1.08088	0.09	3	
YPL123C*	RNY1	1.3355	0.1	3	RNASET2
YPL230W	USV1	1.19881	0.1	3	KLF (1-17), SP (5-7)
YGL101W		1.16132	0.1	3	HDDC2
YBR147W	RTC2	1.07812	0.1	3	C3orf55, PQLC2, TMEM44
YNL305C	BXI1	1.07108	0.1	3	FAIM2, GRINA, TMBIM1, TMBIM4
YKR067W	GPT2	1.05353	0.1	3	
YIL101C	XBP1	1.02357	0.1	3	
YHL021C	AIM17	1.00773	0.1	3	BBOX1, TMLHE
YCR098C	GIT1	-1.01065	0.1	3	
YMR049C	ERB1	-1.03363	0.1	3	BOP1
YMR290C	HAS1	-1.23137	0.1	3	DDX18
YJL109C	UTP10	-1.26035	0.1	3	HEATR1
YDR345C	HXT3	-1.5735	0.1	3	
YCL073C	GEX1	5.11103	0.1	2	
YDR516C	EMI2	1.40186	0.1	2	GCK, HK1, HK2, HK3, HKDC1
YDL110C	TMA17	1.27099	0.1	2	
YGR130C*		1.177	0.1	2	
YOL048C	RRT8	1.17182	0.1	2	
YJL161W	FMP33	1.16756	0.1	2	
YEL012W*	UBC8	1.07261	0.1	2	UBE2H
YFR042W	KEG1	1.02711	0.1	2	
YER053C-A		1.01554	0.1	2	
YOR280C	FSH3	1.01108	0.1	2	OVCA2
YLR257W		1.00554	0.1	2	
YML052W	SUR7	1.00544	0.1	2	
YHR197W	RIX1	-1.00962	0.1	2	
YPL226W	NEW1	-1.07183	0.1	2	
YKL221W	MCH2	1.07998	0.09	1	SLC16 (A1-A14)
YOL032W	OPI10	1.67029	0.1	1	C11orf73
YJL163C		1.51658	0.1	1	SLC46 (A1-A3)
YCR021C*	HSP30	1.3464	0.1	1	
YBR126W-A		1.27059	0.1	1	

YHR171W	ATG7	1.26014	0.1	1	ATG7
YGR131W	FHN1	1.11873	0.1	1	
YMR128W	ECM16	-1.03422	0.1	1	DHX37
YPL012W	RRP12	-1.25061	0.1	1	RRP12
YNL141W	AAH1	-1.67229	0.1	1	ADA, ADAL
YML123C	PHO84	-2.08859	0.09	0	
YDR100W	TVP15	1.38301	0.1	0	
YOR292C		1.25053	0.1	0	MPV17
YEL039C	CYC7	1.2486	0.1	0	CYC5
YDL018C	ERP3	1.21249	0.1	0	TMED1, TMED2, TMED3, TMED4, TMED5, TMED6, TMED7, TMED-TICAM2
YER054C	GIP2	1.13455	0.1	0	PPP1R3 (A-G)
YFR003C	YPI1	1.05663	0.1	0	PPP1R11
YPR151C	SUE1	1.05337	0.1	0	
YGL120C	PRP43	-1.12706	0.1	0	DHX15, DHX32, DQX1
YIL053W	GPP1	-1.13681	0.1	0	
YBR238C		-1.27286	0.1	0	
YLL052C	AQY2	-1.6217	0.1	0	AQP(1-10), MIP
YHR126C	ANS1	-1.85683	0.1	0	
YBR296C	PHO89	-2.36723	0.1	0	SLC20A1, SLC20A2
YAL028W	FRT2	1.09066	0.09	N/A	
YBR230W-A		1.1175	0.1	N/A	
YBR285W		1.34955	0.1	N/A	
YBR302C	COS2	1.10159	0.1	N/A	
YDR169C-A		8.81987	0.09	N/A	
YDR258C	HSP78	1.305	0.1	N/A	CLPB
YDR342C	HXT7	1.10241	0.09	N/A	
YGR027W-B		-6.3627	0.07	N/A	
YHR086W-A		1.75662	0.1	N/A	
YHR087W	RTC3	1.27268	0.1	N/A	
YJR005C-A		1.37524	0.1	N/A	CCDC124
YLR401C	DUS3	-1.02504	0.07	N/A	DUS3L
YML132W	COS3	1.10159	0.1	N/A	
YMR247W-A		1.41759	0.1	N/A	
YMR262W		1.37104	0.1	N/A	TATDN3
YOL161C	PAU20	9.02934	0.09	N/A	
YOL164W-A		1.0321	0.09	N/A	
YOR341W	RPA190	-1.11917	0.1	N/A	POLR1A
YPR010C	RPA135	-1.28417	0.1	N/A	POLR1B

Table S8.3 | Functional categories of gRNA9-1 regulated genes.

MIPS Functional Classification+A2:C33

Category	p-value	In Category from Cluster
enzyme inhibitor [18.02.01.02]	0.0005366	YPI1 SPL2 VHS3
unfolded protein response (e.g. ER quality control) [32.01.07]	0.0010971	HSP42 HSP78 TIM9 SNO4 SIS1 HSP32
UNCLASSIFIED PROTEINS [99]	0.0020048	YBR126W-A RTC2 YBR230W-A SAF1 YBR285W COS2 TMA17 TVP15 ARP10 YDR169C-A YER053C-A YER121W YFL012W KEG1 YGL101W YGL258W-A YGR130C AIM17 YHR086W-A RTC3 ANS1 FMP33 YJL163C YJR005C-A YKL100C YLR149C YLR257W YMR247W-A BXI1 OPI10 RRT8 PHM7 YOL114C YOL164W-A PNS1 YOR292C YPL247C
regulation of phosphate metabolism [01.04.04]	0.0033019	GIP2 YPI1 VHS3
ribosome biogenesis [12.01]	0.0045886	ARX1 RIX1 ECM16 RRP12 NOG1
stress response [32.01]	0.0063805	HSP30 TPS2 CYC7 STF2 XBP1 YJL144W PAU20 VHS3
PROTEIN FATE (folding, modification, destination) [14]	0.007599	SNO4 HSP32
rRNA processing [11.04.01]	0.0081719	UTP8 RIX1 UTP10 ERB1 ECM16 HAS1 DBP2 RRP12
homeostasis of phosphate [34.01.03.03]	0.0096615	PHO89 PHO84

GO Molecular Function

Category	p-value	In Category from Cluster
molecular˙function [GO:0003674]	6.14E-08	FRT2 YBL086C YBR056W RTC2 YBR230W-A YBR238C YBR285W COS2 HSP30 ERP3 TMA17 TVP15 ARX1 YDR169C-A EMI2 YER053C-A PET117 YER121W YFL012W KEG1 YGL101W YGL258W-A STF2 YGR130C FHN1 BNS1 AIM17 YHR086W-A RTC3 ANS1 RIX1 YJL144W FMP33 YJL163C YJR005C-A YKL100C YLR149C YLR164W YLR257W ECM19 SUR7 COS3 ERB1 YMR244W YMR247W-A YMR262W MDG1 BXI1 OPI10 RRT8 PHM7 YOL114C PAU20 YOL164W-A PNS1 FSH3 YOR292C RRP12 OXR1 YPL247C SUE1
transporter activity [GO:0005215]	0.00082742	GIT1 YDL199C HXT7 HXT3 MCH2 AQY2 PHO84
protein phosphatase inhibitor activity [GO:0004864]	0.00170312	YPI1 VHS3
inorganic phosphate transmembrane transporter activity [GO:0005315]	0.00280679	PHO89 PHO84
symporter activity [GO:0015293]	0.00759902	PHO89 MCH2

GO Cellular Component

Category	p-value	In Category from Cluster
membrane raft [GO:0045121]	0.00025849	YGR130C FHN1 SUR7 MDG1
90S preribosome [GO:0030686]	0.00158222	PRP43 UTP8 UTP10 ECM16 HAS1 RRP12

plasma membrane [GO:0005886]	0.00244257	PHO89 GEX1 HSP30 GIT1 HXT7 HXT3 FHN1 ANS1 AQY2 SUR7 MDG1 ATO2 PHM7 PNS1
integral to membrane [GO:0016021]	0.00277969	FRT2 RTC2 PHO89 COS2 GEX1 HSP30 GIT1 ERP3 YDL199C TVP15 HXT7 HXT3 TIM9 YER053C-A KEG1 FHN1 UTP10 FMP33 YJL163C YKL100C MCH2 GPT2 YLR164W ECM19 SUR7 PHO84 COS3 YMR244W BXI1 ATO2 RRT8 PHM7 PAU20 PNS1 YOR292C
t-UTP complex [GO:0034455]	0.00576345	UTP8 UTP10
fungal-type vacuole [GO:0000324]	0.00584756	COS2 TRX1 COS3 BXI1 PHM7 YOR292C
membrane [GO:0016020]	0.00591148	FRT2 RTC2 YBR238C PHO89 COS2 GEX1 HSP30 GIT1 ERP3 YDL199C TVP15 HXT7 HXT3 TIM9 YER053C-A KEG1 FHN1 ANS1 ATG7 FMP33 YJL163C YKL100C MCH2 GPT2 AQY2 TRX1 SRN2 YLR164W ECM19 SUR7 PHO84 COS3 YMR244W MDG1 BXI1 ATO2 RRT8 PHM7 PAU20 PNS1 YOR292C
cellular˙component [GO:0005575]	0.00757535	YBL086C YBR230W-A YBR285W YDR169C-A YER121W YFL012W YGL258W-A BNS1 YHR086W-A ANS1 YJR005C-A YLR149C YMR244W YMR247W-A YMR262W SNO4 YOL114C PAU20 YOL164W-A FSH3 HSP32
rDNA heterochromatin [GO:0033553]	0.00759902	UTP8 UTP10
eisosome [GO:0032126]	0.00966151	YGR130C SUR7

Table S8.4 | List of the genes randomly chosen from the yeast ORF library and their αSyn suppression ability when overexpressed.

SystematicA2:C32	Standard Name	Survival Score
YNL136W	EAF7	0
YLR384C	IKI3	0
YFR009W	GCN20	1
YJL110C	GZF3	3
YAR007C	RFA1	1
YOR116C	RPO31	3
YCL057W	PRD1	0
YNL154C	YCK2	0
YKL213C	DOA1	1
YGR167W	CLC1	1
YBR126W-A		1
YNL065W	AQR1	3
YGL086W	MAD1	1
YER167W	BCK2	1
YKL004W	AUR1	1
YBL069W	AST1	1
YHR137W	ARO9	1
YPR172W		1
YPL048W	CAM1	1
YJR150C	DAN1	1
YNL135C	FPR1	3
YLL060C	GTT2	0
YDL087C	LUC7	1
YDR462W	MRPL28	1
YPL171C	OYE3	1
YKL163W	PIR3	1
YCL027C-A	HBN1	2
YHR071W	PCL5	2
YGL038C	OCH1	2
YMR091C	NPL6	2
YGR232W	NAS6	1
YKL194C	MST1	2
YJL096W	MRPL49	1
YMR224C	MRE11	1

Table S8.5 | Yeast strains, primer sets, and plasmids used in this study.

Yeast Stain	Genotype	Source
W303-1A	MATa *ade2-1, ura3-1 his3-11, -15 trp1-1 leu2-3, -112 can1-100*	Thomas and Rothstein, 1989
W303-crisprTF	W303-1A *TetON-dCas9-VP64-LEU2*	This study
ITox2C	W303-1A *GAL1p-SNCA-YFP-TRP1, GAL1p-SNCA-YFP-HIS3*	Cooper et al., 2006
ITox2C-crisprTF	ITox2C *TetON-dCas9-VP64-LEU2*	This study
W303-crisprTF-GAL4*	W303-crisprTF *kanMX-GAL4**	This study
ITox2C-crisprTF-GAL4*	ITox2C-crisprTF *kanMX-GAL4**	This study

Primer Name	Oligo Sequence	Length (nt)
GAL4_qF1	5'- GGTCTTCGAGTCAGGTTCCA -3'	20
GAL4_qR1	5'- CGGCGTCTTTGTTCCAGAAT -3'	20
SNCA_qF1	5'- CAAACAGGGTGTGGCAGAAG -3'	20
SNCA_qR1	5'- CTCCCTCCACTGTCTTCTGG -3'	20
ACT1_qF1	5'- CGAATTGAGAGTTGCCCCAG -3'	20
ACT1_qR1	5'- CAAGGACAAAACGGCTTGGA -3'	20

Plasmid	Source
pRS405	Sikorski and Hieter, 1989
pLenti CMV rtTA3G Blast	Addgene #31797
pTPGI˙dCas9˙VP64	Addgene #49013
pRS405-TetON-dCas9-VP64	This study
pRS405-TetON-dCas9-VP64-PstI	This study
pRPR1_gRNA_handle_RPR1t	Addgene #49014
pRS426	Sikorski and Hieter, 1989
pRS426-gRNA 6-3	This study
pRS426-gRNA 9-1	This study
pRS426-gRNA-HindIII-EcoRI	This study
pRS416	Sikorski and Hieter, 1989
pRS416-gRNA 6-3	This study
pRS416-gRNA 9-1	This study
pRS416-gRNA-HindIII-EcoRI	This study
pBG1805	Open Biosystems
pBG1805-SNO4	Open Biosystems
pBG1805-HSP32	Open Biosystems
pBG1805-HSP31	Open Biosystems
pBG1805-HSP33	Open Biosystems
pBG1805-SIS1	Open Biosystems
pBG1805-SAF1	Open Biosystems
pBG1805-GGA1	Open Biosystems

pBG1805-GGA2	Open Biosystems
pBG1805-TRX1	Open Biosystems
pBG1805-TRX2	Open Biosystems
pBG1805-TIM8	Open Biosystems
pBG1805-TIM9	Open Biosystems
pBG1805-TIM10	Open Biosystems
pBG1805-HSP42	Open Biosystems
pBG1805-UBP3	Open Biosystems
pFA6a-kanMX	Addgene #39296
pCMV-mKate2	This study
pCMV-DJ-1-mKate2	This study
pCMV-ALS2-mKate2	This study
pCMV-GGA1-mKate2	This study
pCMV-DNAJB1-mKate2	This study
pCMV-TIMM9-mKate2	This study
pCMV-TXN-mKate2	This study
pCMV-Bcl-xL-mKate2	This study
FuGW-EGFP	Lois et al., 2002
FuGW-DJ-1-P2A-EGFP	This study
FuGW-TXN-P2A-EGFP	This study
FuGW-TIMM9-P2A-EGFP	This study
FuGW-TIMM9-P2A-TXN-P2A-EGFP	This study
Delta8.9	Lois et al., 2002
VSV-G	Lois et al., 2002

Chapter 9: Conclusion

In this thesis, I described several platforms and strategies for performing various types of computation and memory operations in living cells. This section outlines some more general aspects of this research. The readers are referred to the corresponding chapters for specific discussions on features and applications of the corresponding platforms.

The common feature of the platforms described in this thesis is the use of genome DNA as a readable and writable medium for computing and memory. SCRIBE, HiSCRIBE, DOMINO and ENGRAM DNA writing platforms enable to record information about biological cues – such as the concentration of small molecules and metabolite or transcription factors – in the form of precise mutations in the genomic DNA. I demonstrated that these DNA writers can be used for analog (i.e., continuous) recording or signaling dynamics over time, by using the distributed genomic content of cell populations. I further showed that by using efficient DNA writers, such as DOMINO, one can perform various types of dynamic digital logic and memory operations in a robust and scalable fashion.

It remains to be determined which type of computation – analog vs. digital – provides a more suitable strategy for performing more sophisticated forms of information processing in living cells. There may not be a universal strategy to address all the needs, and as is the case of natural gene circuits, different strategies may be used for different purposes. Digital computation is more robust to biological noise, but at the same time more resource-intensive. The genomic memory platforms – such as the one exemplified by dynamic DOMINO logic - offer a strategy to reduce the metabolic burden on cells by minimizing protein expression and permanently storing the result of previous steps of computations in DNA. On the other hand, analog computation where memory and computation operations are distributed within cell populations, could reduce the overall metabolic load on individual cells and could be used for continuous recording.

As mentioned in previous chapters, molecular recording platforms that enable continuous and autonomous memorization of biological cues over time and space, such as the one described in this thesis, would have broad utility across various disciplines in biology, from recording spatiotemporal signaling dynamics to lineage tracing and

etc. However, the recording frequency of DNA recorders, for which recording is mediated through mutation and repair, is limited by the rate of the associated DNA repair mechanism in a given organism, and thus may not be suitable for recording events that occur in faster timescales. It remains to be determined whether alternative media and recording mechanisms, such as epigenetic marks (methylation, acetylation etc.), RNA and protein editing/modifications can be used for information storage, either as the final storage medium, or as an intermediate (i.e., working) form of memory, before converting that information to more permanent form of genetic memory. Alternative strategies, based on DNA/RNA synthesis using error-prone polymerases, or other variants of chimeric enzymes that do not depend on DNA repair for recording could be explored. Such platforms, if developed, can be especially useful for recording events that occur in faster timescales and may someday help to decipher and map the activity of neural circuits in the brain in a high-throughput and resolution fashion. Moreover, analogous strategies to what was previously described (chapter 4) for recording bacterial connectome, by using mobilizable DNA writers that can pass through synapses may be used to map neural connectomes. There certainly are plenty that needs to be done and technical challenges that need to be overcome to achieve these goals and to be able to map and decipher inner working of the brain, the holy grails of biological computation. Nevertheless, there are plenty of interesting biological questions and applications – such as high-resolution mapping of cellular lineages – that can be pursued and investigated by current DNA writing-based molecular recorders that perform in slower timescales.

As discussed previously, genomic DNA is an evolvable functional memory that records the history of adaptive changes over course of evolution. The ability to precisely target and mutate specific genomic sites by the described DNA writers in response to desired biological cues enables to reverse the flow of information that is expected from central dogma. While this does not necessarily negate the central dogma – specifically, that the sequence information encoded in proteins cannot be transferred back to protein, RNA or DNA – it is tempting to speculate whether such information can at least partially be transferred back to other forms by using some form of intermediate but sophisticated synthetic gene circuits. Nevertheless, the precise and conditional DNA writing and diversity generation described in this work, as described in chapter 6, foreshadow the dawn of synthetic Lamarckian evolution strategies and dynamic evolutionary engineering era. While we could perhaps see only a few steps ahead, there is a tremendous potential for using dynamic genome engineering strategies to develop

fast-adaptive biological systems ecosystem based on evolutionary designs, which could be especially useful for many biomedical and biotechnological applications.

We have just begun to decipher the astonishing complexity of biological machines at the molecular level and realize the sophisticated strategies these machines use to perform various forms of computation and memory operations in different timescales. Devised over billion years of evolution to address various challenges that living cells face with, these strategies share striking similarities and differences with man-made machines. Both bottom-up and top-down approaches are needed to tease apart various layers of complexity of these molecular machines. Understanding the molecular architecture that gives the living cells the unique capacity to perform robust computation and find efficient solutions in uncertain environments with minimal energy could have great impacts in designing man-made computers and developing artificial intelligence algorithms. Studying various mechanisms that nature has devised (many of which yet to be discovered) could be especially insightful toward this goal. For example, careful generalization reveals a recurring theme of memory and learning at various timescales and molecular layers in living systems; from evolutionary forces acting on genomic materials over course of evolution to learning mediated by re-enforcing synaptic connections in the brain occurring within much shorter time scales, living cells seem to be subjected to learning and adaptation at various levels. Turing's effort to describe the human brain cortex by postulating an "unorganized machine" is in many ways reminiscent of what we know today about how biological systems learn and adapt at different molecular levels and timescales. It remains to be determined whether such recurring design principles can be identified, generalized, and applied to man-made machines, and whether evolutionary principles can be effectively implemented in such machines.

At the same time, applying design principles developed in other domains, such as electrical circuits, into living cells could be a valuable strategy to explore the extent to which biological machines are amenable to our known engineering principles. Such approach could help charting the unknowns and determining differences between biological and man-made machines at the molecular level. Towards this goal and to fully control and decipher cellular programs and design principles of biological machines, various enabling platforms for precise control of genetic, epigenetic, transcriptional and post-transcriptional programs over space and time are needed. The systems and approaches described in this work is a step forward in that direction.

Bibliography

1. E. Schrodinger, *What is life?* , (University Press: Cambridge, 1943).
2. O. T. Avery, C. M. Macleod, M. McCarty, Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *The Journal of experimental medicine* **79**, 137-158 (1944); published online EpubFeb 01 (
3. J. D. Watson, F. H. Crick, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953); published online EpubApr 25 (
4. M. H. Wilkins, A. R. Stokes, H. R. Wilson, Molecular structure of deoxypentose nucleic acids. *Nature* **171**, 738-740 (1953); published online EpubApr 25 (
5. F. Sanger, S. Nicklen, A. R. Coulson, DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463-5467 (1977); published online EpubDec (
6. A. M. Maxam, W. Gilbert, A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 560-564 (1977); published online EpubFeb (
7. A. M. Turing, Computing machinery and intelligence. *Mind* **59**, 433-460 (1950).
8. A. W. Burks, H. H. Goldstine, J. Von Neumann, Preliminary discussion of the logical design of an electronic computer instrument. (1946).
9. N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipos, E. Birney, Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **494**, 77-80 (2013); published online EpubFeb 7 (10.1038/nature11875 nature11875 [pii]).
10. G. M. Church, Y. Gao, S. Kosuri, Next-generation digital information storage in DNA. *Science* **337**, 1628 (2012); published online EpubSep 28 (science.1226355 [pii] 10.1126/science.1226355).
11. M. C. Inniss, P. A. Silver, Building synthetic memory. *Curr Biol* **23**, R812-816 (2013); published online EpubSep 9 (10.1016/j.cub.2013.06.047 S0960-9822(13)00772-0 [pii]).
12. T. S. Gardner, C. R. Cantor, J. J. Collins, Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339-342 (2000); published online EpubJan 20 (10.1038/35002131).
13. J. W. Kotula, S. J. Kerns, L. A. Shaket, L. Siraj, J. J. Collins, J. C. Way, P. A. Silver, Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc Natl Acad Sci U S A* **111**, 4838-4843 (2014); published online EpubApr 1 (10.1073/pnas.1321321111 1321321111 [pii]).
14. D. R. Burrill, M. C. Inniss, P. M. Boyle, P. A. Silver, Synthetic memory circuits for tracking human cell fate. *Genes Dev* **26**, 1486-1497 (2012); published online EpubJul 1 (10.1101/gad.189035.112 26/13/1486 [pii]).
15. B. P. Kramer, A. U. Viretta, M. Daoud-El-Baba, D. Aubel, W. Weber, M. Fussenegger, An engineered epigenetic transgene switch in mammalian cells. *Nat Biotechnol* **22**, 867-870 (2004); published online EpubJul (10.1038/nbt980 nbt980 [pii]).
16. C. M. Ajo-Franklin, D. A. Drubin, J. A. Eskin, E. P. Gee, D. Landgraf, I. Phillips, P. A. Silver, Rational design of memory in eukaryotic cells. *Genes Dev* **21**, 2271-2276 (2007); published online EpubSep 15 (21/18/2271 [pii]

10.1101/gad.1586107).
17. P. Siuti, J. Yazbek, T. K. Lu, Synthetic circuits integrating logic and memory in living cells. *Nat Biotechnol*, (2013); published online EpubFeb 10 (10.1038/nbt.2510
nbt.2510 [pii]).
18. J. Bonnet, P. Subsoontorn, D. Endy, Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 8884-8889 (2012); published online EpubJun 05 (10.1073/pnas.1202344109).
19. J. Bonnet, P. Yin, M. E. Ortiz, P. Subsoontorn, D. Endy, Amplifying Genetic Logic Gates. *Science*, (2013); published online EpubMar 28 (science.1232758 [pii]
10.1126/science.1232758).
20. A. E. Friedland, T. K. Lu, X. Wang, D. Shi, G. Church, J. J. Collins, Synthetic gene networks that count. *Science* **324**, 1199-1202 (2009); published online EpubMay 29 (10.1126/science.1172005).
21. H. M. Ellis, D. Yu, T. DiTizio, D. L. Court, High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A* **98**, 6742-6746 (2001); published online EpubJun 5 (10.1073/pnas.121164898
121164898 [pii]).
22. N. Costantino, D. L. Court, Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 15748-15753 (2003); published online EpubDec 23 (10.1073/pnas.2434959100).
23. D. Yu, H. M. Ellis, E. C. Lee, N. A. Jenkins, N. G. Copeland, D. L. Court, An efficient recombination system for chromosome engineering in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5978-5983 (2000); published online EpubMay 23 (10.1073/pnas.100127597).
24. J. A. Sawitzke, N. Costantino, X. T. Li, L. C. Thomason, M. Bubunenko, C. Court, D. L. Court, Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *Journal of molecular biology* **407**, 45-59 (2011); published online EpubMar 18 (10.1016/j.jmb.2011.01.030).
25. B. Swingle, E. Markel, N. Costantino, M. G. Bubunenko, S. Cartinhour, D. L. Court, Oligonucleotide recombination in Gram-negative bacteria. *Molecular microbiology* **75**, 138-148 (2010); published online EpubJan (10.1111/j.1365-2958.2009.06976.x).
26. S. Datta, N. Costantino, X. Zhou, D. L. Court, Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 1626-1631 (2008); published online EpubFeb 5 (10.1073/pnas.0709089105).
27. T. Yee, T. Furuichi, S. Inouye, M. Inouye, Multicopy single-stranded DNA isolated from a gram-negative bacterium, Myxococcus xanthus. *Cell* **38**, 203-209 (1984); published online EpubAug (0092-8674(84)90541-5 [pii]).
28. B. C. Lampson, M. Inouye, S. Inouye, Retrons, msDNA, and the bacterial genome. *Cytogenet Genome Res* **110**, 491-499 (2005)84982 [pii]
10.1159/000084982).
29. D. Lim, W. K. Maas, Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in E. coli B. *Cell* **56**, 891-904 (1989); published online EpubMar 10 (0092-8674(89)90693-4 [pii]).
30. R. Lutz, 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* **25**, 1203-1210 (1997); published online EpubMar 15 (
31. P. L. Sharma, V. Nurpeisov, R. F. Schinazi, Retrovirus reverse transcriptases containing a modified YXDD motif. *Antivir Chem Chemother* **16**, 169-182 (2005).
32. J. R. Mao, M. Shimada, S. Inouye, M. Inouye, Gene regulation by antisense DNA produced in vivo. *J Biol Chem* **270**, 19684-19687 (1995); published online EpubAug 25 (
33. S. Warming, N. Costantino, D. L. Court, N. A. Jenkins, N. G. Copeland, Simple and highly efficient BAC recombineering using galK selection. *Nucleic acids research* **33**, e36 (2005); published online EpubFeb 24 (10.1093/nar/gni035).
34. R. Ohlendorf, R. R. Vidavski, A. Eldar, K. Moffat, A. Moglich, From dusk till dawn: one-plasmid systems for light-regulated gene expression. *J Mol Biol* **416**, 534-542 (2012); published online EpubMar 2 (10.1016/j.jmb.2012.01.001 S0022-2836(12)00011-3 [pii]).
35. M. S. Huen, X. T. Li, L. Y. Lu, R. M. Watt, D. P. Liu, J. D. Huang, The involvement of replication in single stranded oligonucleotide-mediated gene repair. *Nucleic acids research* **34**, 6183-6194 (2006)10.1093/nar/gkl852).
36. A. R. Poteete, Involvement of DNA replication in phage lambda Red-mediated homologous recombination. *Mol Microbiol* **68**, 66-74 (2008); published online EpubApr (10.1111/j.1365-2958.2008.06133.x MMI6133 [pii]).
37. A. R. Poteete, Involvement of DNA Replication Proteins in Phage Lambda Red-Mediated Homologous Recombination. *PLoS One* **8**, e67440 (2013)10.1371/journal.pone.0067440 PONE-D-13-05904 [pii]).
38. B. E. Dutra, V. A. Sutera, Jr., S. T. Lovett, RecA-independent recombination is efficient but limited by exonucleases. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 216-221 (2007); published online EpubJan 2 (10.1073/pnas.0608293104).
39. S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, R. Weiss, A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130-1134 (2005); published online EpubApr 28 (nature03461 [pii] 10.1038/nature03461).
40. S. Auslander, D. Auslander, M. Muller, M. Wieland, M. Fussenegger, Programmable single-cell mammalian biocomputers. *Nature* **487**, 123-127 (2012); published online EpubJul 5 (10.1038/nature11149 nature11149 [pii]).
41. R. Daniel, J. R. Rubens, R. Sarpeshkar, T. K. Lu, Synthetic analog computation in living cells. *Nature* **497**, 619-623 (2013); published online EpubMay 30 (10.1038/nature12148 nature12148 [pii]).
42. E. J. Olson, L. A. Hartsough, B. P. Landry, R. Shroff, J. J. Tabor, Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nat Methods* **11**, 449-455 (2014); published online EpubApr (10.1038/nmeth.2884 nmeth.2884 [pii]).
43. J. A. Mosberg, C. J. Gregg, M. J. Lajoie, H. H. Wang, G. M. Church, Improving lambda red genome engineering in Escherichia coli via rational removal of endogenous nucleases. *PloS one* **7**, e44638 (2012)10.1371/journal.pone.0044638).
44. S. Miyata, A. Ohshima, S. Inouye, M. Inouye, In vivo production of a stable single-stranded cDNA in Saccharomyces cerevisiae by means of a bacterial

retron. *Proc Natl Acad Sci U S A* **89**, 5735-5739 (1992); published online EpubJul 1 (
45. O. Mirochnitchenko, S. Inouye, M. Inouye, Production of single-stranded DNA in mammalian cells by means of a bacterial retron. *J Biol Chem* **269**, 2380-2383 (1994); published online EpubJan 28 (
46. S. Brakmann, S. Grzeszik, An error-prone T7 RNA polymerase mutant generated by directed evolution. *Chembiochem : a European journal of chemical biology* **2**, 212-219 (2001); published online EpubMar 2 (
47. J. D. Roberts, K. Bebenek, T. A. Kunkel, The accuracy of reverse transcriptase from HIV-1. *Science* **242**, 1171-1173 (1988); published online EpubNov 25 (
48. K. Bebenek, J. Abbotts, S. H. Wilson, T. A. Kunkel, Error-prone polymerization by HIV-1 reverse transcriptase. Contribution of template-primer misalignment, miscoding, and termination probability to mutational hot spots. *The Journal of biological chemistry* **268**, 10324-10334 (1993); published online EpubMay 15 (
49. K. M. Esvelt, J. C. Carlson, D. R. Liu, A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503 (2011); published online EpubApr 28 (10.1038/nature09929).
50. Y. Amir, E. Ben-Ishay, D. Levner, S. Ittah, A. Abu-Horowitz, I. Bachelet, Universal computing by DNA origami robots in a living animal. *Nat Nanotechnol* **9**, 353-357 (2014); published online EpubMay (10.1038/nnano.2014.58
nnano.2014.58 [pii]).
51. L. Qian, E. Winfree, J. Bruck, Neural network computation with DNA strand displacement cascades. *Nature* **475**, 368-372 (2011); published online EpubJul 21 (10.1038/nature10262
nature10262 [pii]).
52. G. Seelig, D. Soloveichik, D. Y. Zhang, E. Winfree, Enzyme-free nucleic acid logic circuits. *Science* **314**, 1585-1588 (2006); published online EpubDec 8 (314/5805/1585 [pii]
10.1126/science.1132493).
53. P. W. Rothemund, Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297-302 (2006); published online EpubMar 16 (nature04586 [pii]
10.1038/nature04586).
54. S. M. Douglas, H. Dietz, T. Liedl, B. Hogberg, F. Graf, W. M. Shih, Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414-418 (2009); published online EpubMay 21 (10.1038/nature08016
nature08016 [pii]).
55. S. M. Douglas, I. Bachelet, G. M. Church, A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335**, 831-834 (2012); published online EpubFeb 17 (10.1126/science.1214081
335/6070/831 [pii]).
56. S. M. Chirieleison, P. B. Allen, Z. B. Simpson, A. D. Ellington, X. Chen, Pattern transformation with DNA circuits. *Nat Chem* **5**, 1000-1005 (2013); published online EpubDec (10.1038/nchem.1764
nchem.1764 [pii]).
57. C. A. Brosey, C. Yan, S. E. Tsutakawa, W. T. Heller, R. P. Rambo, J. A. Tainer, I. Ivanov, W. J. Chazin, A new structural framework for integrating replication protein A into DNA processing machinery. *Nucleic Acids Res* **41**, 2313-2327 (2013); published online EpubFeb 1 (10.1093/nar/gks1332
gks1332 [pii]).
58. S. Kosuri, G. M. Church, Large-scale de novo DNA synthesis: technologies and applications. *Nat Methods* **11**, 499-507 (2014); published online EpubMay (10.1038/nmeth.2918

nmeth.2918 [pii]).
59. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682 (2012); published online EpubJul (10.1038/nmeth.2019 nmeth.2019 [pii]).
60. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-6645 (2000); published online EpubJun 6 (10.1073/pnas.120163297 120163297 [pii]).
61. G. Pines, E. F. Freed, J. D. Winkler, R. T. Gill, Bacterial Recombineering: Genome Engineering via Phage-Based Homologous Recombination. *ACS synthetic biology* **4**, 1176-1185 (2015); published online EpubNov 20 (10.1021/acssynbio.5b00009).
62. H. H. Wang, F. J. Isaacs, P. A. Carr, Z. Z. Sun, G. Xu, C. R. Forest, G. M. Church, Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898 (2009); published online EpubAug 13 (10.1038/nature08187).
63. W. Jiang, D. Bikard, D. Cox, F. Zhang, L. A. Marraffini, RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* **31**, 233-239 (2013); published online EpubMar (10.1038/nbt.2508).
64. C. Ronda, L. E. Pedersen, M. O. Sommer, A. T. Nielsen, CRMAGE: CRISPR Optimized MAGE Recombineering. *Scientific reports* **6**, 19452 (2016); published online EpubJan 22 (10.1038/srep19452).
65. R. J. Citorik, M. Mimee, T. K. Lu, Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. *Nature biotechnology* **32**, 1141-1145 (2014); published online EpubNov (10.1038/nbt.3011).
66. L. Cui, D. Bikard, Consequences of Cas9 cleavage in the chromosome of Escherichia coli. *Nucleic acids research* **44**, 4243-4251 (2016); published online EpubMay 19 (10.1093/nar/gkw223).
67. F. Farzadfard, T. K. Lu, Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* **346**, 1256272 (2014); published online EpubNov 14 (10.1126/science.1256272).
68. J. W. Chase, C. C. Richardson, Exonuclease VII of Escherichia coli. Mechanism of action. *The Journal of biological chemistry* **249**, 4553-4561 (1974); published online EpubJul 25 (
69. L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, W. A. Lim, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183 (2013); published online EpubFeb 28 (10.1016/j.cell.2013.02.022).
70. R. M. Schaaper, R. L. Dunn, Spectra of spontaneous mutations in Escherichia coli strains defective in mismatch correction: the nature of in vivo DNA replication errors. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 6220-6224 (1987); published online EpubSep (
71. B. J. Caliando, C. A. Voigt, Targeted DNA degradation using a CRISPR device stably carried in the host genome. *Nature communications* **6**, 6989 (2015); published online EpubMay 19 (10.1038/ncomms7989).
72. Y. Gao, Y. Zhao, Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *Journal of integrative plant biology* **56**, 343-349 (2014); published online EpubApr (10.1111/jipb.12152).
73. M. G. Ross, C. Russ, M. Costello, A. Hollinger, N. J. Lennon, R. Hegarty, C. Nusbaum, D. B. Jaffe, Characterizing and measuring bias in sequence data. *Genome biology* **14**, R51 (2013)10.1186/gb-2013-14-5-r51).

74. D. I. Lou, J. A. Hussmann, R. M. McBee, A. Acevedo, R. Andino, W. H. Press, S. L. Sawyer, High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 19872-19877 (2013); published online EpubDec 3 (10.1073/pnas.1319590110).
75. M. W. Schmitt, S. R. Kennedy, J. J. Salk, E. J. Fox, J. B. Hiatt, L. A. Loeb, Detection of ultra-rare mutations by next-generation sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 14508-14513 (2012); published online EpubSep 4 (10.1073/pnas.1208715109).
76. N. Roquet, A. P. Soleimany, A. C. Ferris, S. Aaronson, T. K. Lu, Synthetic recombinase-based state machines in living cells. *Science* **353**, aad8559 (2016); published online EpubJul 22 (10.1126/science.aad8559).
77. S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016); published online EpubJul 29 (10.1126/science.aaf1175).
78. S. D. Perli, C. H. Cui, T. K. Lu, Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* **353**, (2016); published online EpubSep 09 (10.1126/science.aag0511).
79. R. I. Zeitoun, A. D. Garst, G. D. Degen, G. Pines, T. J. Mansell, T. Y. Glebes, N. R. Boyle, R. T. Gill, Multiplexed tracking of combinatorial genomic mutations in engineered cell populations. *Nature biotechnology* **33**, 631-637 (2015); published online EpubJun (10.1038/nbt.3177).
80. T. Aparicio, S. I. Jensen, A. T. Nielsen, V. de Lorenzo, E. Martinez-Garcia, The Ssr protein (T1E˙1405) from Pseudomonas putida DOT-T1E enables oligonucleotide-based recombineering in platform strain P. putida EM42. *Biotechnology journal* **11**, 1309-1319 (2016); published online EpubOct (10.1002/biot.201600317).
81. C. D. Nadell, K. Drescher, K. R. Foster, Spatial structure, cooperation and competition in biofilms. *Nature reviews. Microbiology* **14**, 589-600 (2016); published online EpubSep (10.1038/nrmicro.2016.84).
82. A. M. Zador, J. Dubnau, H. K. Oyibo, H. Zhan, G. Cao, I. D. Peikon, Sequencing the connectome. *PLoS biology* **10**, e1001411 (2012)10.1371/journal.pbio.1001411).
83. J. I. Glaser, B. M. Zamft, G. M. Church, K. P. Kording, Puzzle Imaging: Using Large-Scale Dimensionality Reduction Algorithms for Localization. *PloS one* **10**, e0131593 (2015)10.1371/journal.pone.0131593).
84. I. D. Peikon, J. M. Kebschull, V. V. Vagin, D. I. Ravens, Y. C. Sun, E. Brouzes, I. R. Correa, Jr., D. Bressan, A. M. Zador, Using high-throughput barcode sequencing to efficiently map connectomes. *Nucleic acids research*, (2017); published online EpubApr 26 (10.1093/nar/gkx292).
85. S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature* **547**, 345-349 (2017); published online EpubJul 20 (10.1038/nature23017).
86. K. C. Murphy, M. G. Marinus, RecA-independent single-stranded DNA oligonucleotide-mediated mutagenesis. *F1000 biology reports* **2**, 56 (2010); published online EpubJul 22 (10.3410/B2-56).
87. H. Jung, J. Liang, Y. Jung, D. Lim, Characterization of cell death in Escherichia coli mediated by XseA, a large subunit of exonuclease VII. *Journal of microbiology* **53**, 820-828 (2015); published online EpubDec (10.1007/s12275-015-5304-0).
88. M. S. Dillingham, S. C. Kowalczykowski, RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiology and molecular biology reviews : MMBR* **72**, 642-671, Table of Contents (2008); published online EpubDec (10.1128/MMBR.00020-08).

89. D. G. Gibson, Enzymatic assembly of overlapping DNA fragments. *Methods in enzymology* **498**, 349-361 (2011)10.1016/B978-0-12-385120-8.00015-2).
90. C. Engler, S. Marillonnet, Golden Gate cloning. *Methods in molecular biology* **1116**, 119-131 (2014)10.1007/978-1-62703-764-8˙9).
91. R. Milo, P. Jorgensen, U. Moran, G. Weber, M. Springer, BioNumbers--the database of key numbers in molecular and cell biology. *Nucleic acids research* **38**, D750-753 (2010); published online EpubJan (10.1093/nar/gkp889).
92. L. Chasteen, J. Ayriss, P. Pavlik, A. R. Bradbury, Eliminating helper phage from phage display. *Nucleic acids research* **34**, e145 (2006)10.1093/nar/gkl772).
93. K. R. Yamamoto, B. M. Alberts, R. Benzinger, L. Lawhorne, G. Treiber, Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. *Virology* **40**, 734-744 (1970); published online EpubMar (
94. L. Ferrieres, G. Hemery, T. Nham, A. M. Guerout, D. Mazel, C. Beloin, J. M. Ghigo, Silent mischief: bacteriophage Mu insertions contaminate products of Escherichia coli random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *Journal of bacteriology* **192**, 6418-6427 (2010); published online EpubDec (10.1128/JB.00621-10).
95. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424 (2016); published online EpubMay 19 (10.1038/nature17946).
96. P. Siuti, J. Yazbek, T. K. Lu, Synthetic circuits integrating logic and memory in living cells. *Nature biotechnology* **31**, 448-452 (2013); published online EpubMay (10.1038/nbt.2510).
97. L. Zelcbuch, N. Antonovsky, A. Bar-Even, A. Levin-Karp, U. Barenholz, M. Dayagi, W. Liebermeister, A. Flamholz, E. Noor, S. Amram, A. Brandis, T. Bareia, I. Yofe, H. Jubran, R. Milo, Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic acids research* **41**, e98 (2013); published online EpubMay (10.1093/nar/gkt151).
98. P. Siuti, J. Yazbek, T. K. Lu, Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnology* **31**, 448-452 (2013); published online EpubMay (10.1038/nbt.2510).
99. F. Farzadfard, T. K. Lu, Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* **346**, 1256272 (2014); published online EpubNov 14 (10.1126/science.1256272).
100. W. M. Stark, Making serine integrases work for us. *Current opinion in microbiology* **38**, 130-136 (2017); published online EpubJun 06 (10.1016/j.mib.2017.04.006).
101. E. R. Coppoolse, M. J. de Vroomen, F. van Gennip, B. J. Hersmus, M. J. van Haaren, Size does matter: cre-mediated somatic deletion efficiency depends on the distance between the target lox-sites. *Plant molecular biology* **58**, 687-698 (2005); published online EpubJul (10.1007/s11103-005-7705-7).
102. K. Nishida, T. Arazoe, N. Yachie, S. Banno, M. Kakimoto, M. Tabata, M. Mochizuki, A. Miyabe, M. Araki, K. Y. Hara, Z. Shimatani, A. Kondo, Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, (2016); published online EpubSep 16 (10.1126/science.aaf8729).
103. F. Farzadfard, S. D. Perli, T. K. Lu, Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS synthetic biology* **2**, 604-613 (2013); published online EpubOct 18 (10.1021/sb400081r).
104. L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J.

S. Weissman, L. S. Qi, CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451 (2013); published online EpubJul 18 (10.1016/j.cell.2013.06.044).
105. A. E. Briner, P. D. Donohoue, A. A. Gomaa, K. Selle, E. M. Slorach, C. H. Nye, R. E. Haurwitz, C. L. Beisel, A. P. May, R. Barrangou, Guide RNA functional modules direct Cas9 activity and orthogonality. *Molecular cell* **56**, 333-339 (2014); published online EpubOct 23 (10.1016/j.molcel.2014.09.019).
106. D. N. Nesbeth, A. Zaikin, Y. Saka, M. C. Romano, C. V. Giuraniuc, O. Kanakov, T. Laptyeva, Synthetic biology routes to bio-artificial intelligence. *Essays in biochemistry* **60**, 381-391 (2016); published online EpubNov 30 (10.1042/EBC20160014).
107. N. Gandhi, G. Ashkenasy, E. Tannenbaum, Associative learning in biochemical networks. *Journal of theoretical biology* **249**, 58-66 (2007); published online EpubNov 07 (10.1016/j.jtbi.2007.07.004).
108. D. Bray, Molecular networks: the top-down view. *Science* **301**, 1864-1865 (2003); published online EpubSep 26 (10.1126/science.1089118).
109. I. Tagkopoulos, Y. C. Liu, S. Tavazoie, Predictive behavior within microbial genetic networks. *Science* **320**, 1313-1317 (2008); published online EpubJun 06 (10.1126/science.1154456).
110. A. Chavez, J. Scheiman, S. Vora, B. W. Pruitt, M. Tuttle, P. R. I. E, S. Lin, S. Kiani, C. D. Guzman, D. J. Wiegand, D. Ter-Ovanesyan, J. L. Braff, N. Davidsohn, B. E. Housden, N. Perrimon, R. Weiss, J. Aach, J. J. Collins, G. M. Church, Highly efficient Cas9-mediated transcriptional programming. *Nature methods* **12**, 326-328 (2015); published online EpubApr (10.1038/nmeth.3312).
111. X. S. Liu, H. Wu, X. Ji, Y. Stelzer, X. Wu, S. Czauderna, J. Shu, D. Dadon, R. A. Young, R. Jaenisch, Editing DNA Methylation in the Mammalian Genome. *Cell* **167**, 233-247 e217 (2016); published online EpubSep 22 (10.1016/j.cell.2016.08.056).
112. I. B. Hilton, A. M. D'Ippolito, C. M. Vockley, P. I. Thakore, G. E. Crawford, T. E. Reddy, C. A. Gersbach, Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology* **33**, 510-517 (2015); published online EpubMay (10.1038/nbt.3199).
113. N. M. Gaudelli, A. C. Komor, H. A. Rees, M. S. Packer, A. H. Badran, D. I. Bryson, D. R. Liu, Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **advance online publication**, (2017); published online Epub10/25/online (10.1038/nature24644).
114. A. C. Komor, K. T. Zhao, M. S. Packer, N. M. Gaudelli, A. L. Waterbury, L. W. Koblan, Y. B. Kim, A. H. Badran, D. R. Liu, Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Science advances* **3**, eaao4774 (2017); published online EpubAug (10.1126/sciadv.aao4774).
115. K. L. Frieda, J. M. Linton, S. Hormoz, J. Choi, K. K. Chow, Z. S. Singer, M. W. Budde, M. B. Elowitz, L. Cai, Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107-111 (2017); published online EpubJan 05 (10.1038/nature20777).
116. R. Kalhor, P. Mali, G. M. Church, Rapidly evolving homing CRISPR barcodes. *Nature methods*, (2016); published online EpubDec 05 (10.1038/nmeth.4108).
117. A. McKenna, G. M. Findlay, J. A. Gagnon, M. S. Horwitz, A. F. Schier, J. Shendure, Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016); published online EpubJul 29 (10.1126/science.aaf7907).

118. M. L. Crowe, SeqDoC: rapid SNP and mutation detection by direct comparison of DNA sequence chromatograms. *BMC bioinformatics* **6**, 133 (2005); published online EpubMay 31 (10.1186/1471-2105-6-133).
119. J. W. Drake, A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 7160-7164 (1991); published online EpubAug 15 (
120. M. Lynch, Evolution of the mutation rate. *Trends in genetics : TIG* **26**, 345-352 (2010); published online EpubAug (10.1016/j.tig.2010.05.003).
121. M. Kirschner, J. Gerhart, Evolvability. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 8420-8427 (1998); published online EpubJul 21 (
122. H. Guo, D. Arambula, P. Ghosh, J. F. Miller, Diversity-generating Retroelements in Phage and Bacterial Genomes. *Microbiology spectrum* **2**, (2014); published online EpubDec (10.1128/microbiolspec.MDNA3-0029-2014).
123. K. W. Deitsch, S. A. Lukehart, J. R. Stringer, Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nature reviews. Microbiology* **7**, 493-503 (2009); published online EpubJul (10.1038/nrmicro2145).
124. G. H. Palmer, T. Bankhead, H. S. Seifert, Antigenic Variation in Bacterial Pathogens. *Microbiology spectrum* **4**, (2016); published online EpubFeb (10.1128/microbiolspec.VMBF-0005-2015).
125. L. Salaun, L. A. Snyder, N. J. Saunders, Adaptation by phase variation in pathogenic bacteria. *Advances in applied microbiology* **52**, 263-301 (2003).
126. P. Horvath, R. Barrangou, CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167-170 (2010); published online EpubJan 8 (10.1126/science.1179555).
127. R. Sorek, C. M. Lawrence, B. Wiedenheft, CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual review of biochemistry* **82**, 237-266 (2013)10.1146/annurev-biochem-072911-172315).
128. S. H. Sternberg, H. Richter, E. Charpentier, U. Qimron, Adaptation in CRISPR-Cas Systems. *Molecular cell* **61**, 797-808 (2016); published online EpubMar 17 (10.1016/j.molcel.2016.01.030).
129. A. Mayer, T. Mora, O. Rivoire, A. M. Walczak, Diversity of immune strategies explained by adaptation to pathogen statistics. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 8630-8635 (2016); published online EpubAug 2 (10.1073/pnas.1600663113).
130. J. M. Di Noia, M. S. Neuberger, Molecular mechanisms of antibody somatic hypermutation. *Annual review of biochemistry* **76**, 1-22 (2007)10.1146/annurev.biochem.76.061705.090740).
131. J. K. Rogers, N. D. Taylor, G. M. Church, Biosensor-based engineering of biosynthetic pathways. *Current opinion in biotechnology* **42**, 84-91 (2016); published online EpubMar 18 (10.1016/j.copbio.2016.03.005).
132. D. J. Jin, C. A. Gross, Mapping and sequencing of mutations in the Escherichia coli rpoB gene that lead to rifampicin resistance. *Journal of molecular biology* **202**, 45-58 (1988); published online EpubJul 5 (
133. Y. A. Ovchinnikov, G. S. Monastyrskaya, S. O. Guriev, N. F. Kalinina, E. D. Sverdlov, A. I. Gragerov, I. A. Bass, I. F. Kiver, E. P. Moiseyeva, V. N. Igumnov, S. Z. Mindlin, V. G. Nikiforov, R. B. Khesin, RNA polymerase rifampicin resistance mutations in Escherichia coli: sequence changes and dominance. *Molecular & general genetics : MGG* **190**, 344-348 (1983).
134. J. Hrebenda, H. Heleszko, K. Brzostek, J. Bielecki, Mutation affecting resistance of Escherichia coli K12 to nalidixic acid. *Journal of general microbiology* **131**, 2285-2292 (1985); published online EpubSep (10.1099/00221287-131-9-2285).

135. S. K. Petersen-Mahrt, R. S. Harris, M. S. Neuberger, AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418**, 99-103 (2002); published online EpubJul 4 (10.1038/nature00862).
136. B. M. Hall, C. X. Ma, P. Liang, K. K. Singh, Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* **25**, 1564-1565 (2009); published online EpubJun 15 (10.1093/bioinformatics/btp253).
137. S. Sarkar, W. T. Ma, G. H. Sandri, On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica* **85**, 173-179 (1992).
138. A. S. Bhagwat, W. Hao, J. P. Townes, H. Lee, H. Tang, P. L. Foster, Strand-biased cytosine deamination at the replication fork causes cytosine to thymine mutations in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 2176-2181 (2016); published online EpubFeb 23 (10.1073/pnas.1522325113).
139. B. Medhekar, J. F. Miller, Diversity-generating retroelements. *Current opinion in microbiology* **10**, 388-395 (2007); published online EpubAug (10.1016/j.mib.2007.06.004).
140. K. Nishikura, Functions and regulation of RNA editing by ADAR deaminases. *Annual review of biochemistry* **79**, 321-349 (2010)10.1146/annurev-biochem-060208-105251).
141. V. A. Risso, J. A. Gavira, D. F. Mejia-Carmona, E. A. Gaucher, J. M. Sanchez-Ruiz, Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian beta-lactamases. *Journal of the American Chemical Society* **135**, 2899-2902 (2013); published online EpubFeb 27 (10.1021/ja311630a).
142. J. W. Thornton, Resurrecting ancient genes: experimental analysis of extinct molecules. *Nature reviews. Genetics* **5**, 366-375 (2004); published online EpubMay (10.1038/nrg1324).
143. T. M. Jermann, J. G. Opitz, J. Stackhouse, S. A. Benner, Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature* **374**, 57-59 (1995); published online EpubMar 2 (10.1038/374057a0).
144. D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111-114 (2006); published online EpubApr 7 (10.1126/science.1123539).
145. C. Pal, B. Papp, G. Posfai, The dawn of evolutionary genome engineering. *Nature reviews. Genetics* **15**, 504-512 (2014); published online EpubJul (10.1038/nrg3746).
146. B. G. Hall, H. Acar, A. Nandipati, M. Barlow, Growth rates made easy. *Molecular biology and evolution* **31**, 232-238 (2014); published online EpubJan (10.1093/molbev/mst187).
147. W. Chan, N. Costantino, R. Li, S. C. Lee, Q. Su, D. Melvin, D. L. Court, P. Liu, A recombineering based approach for high-throughput conditional knockout targeting vector construction. *Nucleic acids research* **35**, e64 (2007)10.1093/nar/gkm163).
148. K. M. Lelli, M. Slattery, R. S. Mann, Disentangling the many layers of eukaryotic transcriptional regulation. *Annu Rev Genet* **46**, 43-68 (2012)10.1146/annurev-genet-110711-155437).
149. K. Ellwood, W. Huang, R. Johnson, M. Carey, Multiple layers of cooperativity regulate enhanceosome-responsive RNA polymerase II transcription complex assembly. *Mol Cell Biol* **19**, 2613-2623 (1999); published online EpubApr (
150. L. Chen, Combinatorial gene regulation by eukaryotic transcription factors. *Curr Opin Struct Biol* **9**, 48-55 (1999); published online EpubFeb (S0959-440X(99)80007-4 [pii]).

151. W. Bacchus, M. Lang, M. D. El-Baba, W. Weber, J. Stelling, M. Fussenegger, Synthetic two-way communication between mammalian cells. *Nat Biotechnol* **30**, 991-996 (2012); published online EpubOct (10.1038/nbt.2351 nbt.2351 [pii]).
152. C. J. Bashor, A. A. Horwitz, S. G. Peisajovich, W. A. Lim, Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annu Rev Biophys* **39**, 515-537 (2010)10.1146/annurev.biophys.050708.133652).
153. N. Nandagopal, M. B. Elowitz, Synthetic biology: integrated gene circuits. *Science* **333**, 1244-1248 (2011); published online EpubSep 2 (10.1126/science.1207084 333/6047/1244 [pii]).
154. C. C. Guet, M. B. Elowitz, W. Hsing, S. Leibler, Combinatorial synthesis of genetic networks. *Science* **296**, 1466-1470 (2002); published online EpubMay 24 (10.1126/science.1067407 296/5572/1466 [pii]).
155. S. Bandyopadhyay, M. Mehta, D. Kuo, M. K. Sung, R. Chuang, E. J. Jaehnig, B. Bodenmiller, K. Licon, W. Copeland, M. Shales, D. Fiedler, J. Dutkowski, A. Guenole, H. van Attikum, K. M. Shokat, R. D. Kolodner, W. K. Huh, R. Aebersold, M. C. Keogh, N. J. Krogan, T. Ideker, Rewiring of genetic networks in response to DNA damage. *Science* **330**, 1385-1389 (2010); published online EpubDec 3 (10.1126/science.1195618 330/6009/1385 [pii]).
156. E. J. Kwon, A. Laderoute, K. Chatfield-Reed, L. Vachon, J. Karagiannis, G. Chua, Deciphering the transcriptional-regulatory network of flocculation in Schizosaccharomyces pombe. *PLoS Genet* **8**, e1003104 (2012)10.1371/journal.pgen.1003104 PGENETICS-D-12-01358 [pii]).
157. J. J. Tabor, H. M. Salis, Z. B. Simpson, A. A. Chevalier, A. Levskaya, E. M. Marcotte, C. A. Voigt, A. D. Ellington, A synthetic genetic edge detection program. *Cell* **137**, 1272-1281 (2009); published online EpubJun 26 (10.1016/j.cell.2009.04.048 S0092-8674(09)00509-1 [pii]).
158. C. Lou, X. Liu, M. Ni, Y. Huang, Q. Huang, L. Huang, L. Jiang, D. Lu, M. Wang, C. Liu, D. Chen, C. Chen, X. Chen, L. Yang, H. Ma, J. Chen, Q. Ouyang, Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch. *Mol Syst Biol* **6**, 350 (2010)10.1038/msb.2010.2 msb20102 [pii]).
159. M. Leisner, L. Bleris, J. Lohmueller, Z. Xie, Y. Benenson, Rationally designed logic integration of regulatory signals in mammalian cells. *Nat Nanotechnol* **5**, 666-670 (2010); published online EpubSep (10.1038/nnano.2010.135 nnano.2010.135 [pii]).
160. L. Nissim, R. H. Bar-Ziv, A tunable dual-promoter integrator for targeting of cancer cells. *Mol Syst Biol* **6**, 444 (2010); published online EpubDec 21 (10.1038/msb.2010.99 msb201099 [pii]).
161. M. Gossen, H. Bujard, Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* **89**, 5547-5551 (1992); published online EpubJun 15 (
162. G. Belli, E. Gari, L. Piedrafita, M. Aldea, E. Herrero, An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. *Nucleic Acids Res* **26**, 942-947 (1998); published online EpubFeb 15 (gkb206 [pii]).

163. C. A. Cronin, W. Gluba, H. Scrable, The lac operator-repressor system is functional in the mouse. *Genes Dev* **15**, 1506-1517 (2001); published online EpubJun 15 (10.1101/gad.892001).
164. S. Urlinger, U. Baron, M. Thellmann, M. T. Hasan, H. Bujard, W. Hillen, Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. *Proc Natl Acad Sci U S A* **97**, 7963-7968 (2000); published online EpubJul 5 (10.1073/pnas.130192197
130192197 [pii]).
165. A. S. Khalil, T. K. Lu, C. J. Bashor, C. L. Ramirez, N. C. Pyenson, J. K. Joung, J. J. Collins, A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647-658 (2012); published online EpubAug 3 (10.1016/j.cell.2012.05.045
S0092-8674(12)00780-5 [pii]).
166. Y. Li, R. Moore, M. Guinn, L. Bleris, Transcription activator-like effector hybrids for conditional control and rewiring of chromosomal transgene expression. *Sci Rep* **2**, 897 (2012)10.1038/srep00897).
167. K. H. Bae, Y. D. Kwon, H. C. Shin, M. S. Hwang, E. H. Ryu, K. S. Park, H. Y. Yang, D. K. Lee, Y. Lee, J. Park, H. S. Kwon, H. W. Kim, B. I. Yeh, H. W. Lee, S. H. Sohn, J. Yoon, W. Seol, J. S. Kim, Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol* **21**, 275-280 (2003); published online EpubMar (10.1038/nbt796
nbt796 [pii]).
168. R. Morbitzer, P. Romer, J. Boch, T. Lahaye, Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc Natl Acad Sci U S A* **107**, 21617-21622 (2010); published online EpubDec 14 (10.1073/pnas.1013133107
1013133107 [pii]).
169. T. Ellis, X. Wang, J. J. Collins, Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotechnol* **27**, 465-471 (2009); published online EpubMay (10.1038/nbt.1536
nbt.1536 [pii]).
170. J. J. Lohmueller, T. Z. Armel, P. A. Silver, A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. *Nucleic Acids Res* **40**, 5180-5187 (2012); published online EpubJun (10.1093/nar/gks142
gks142 [pii]).
171. P. Perez-Pinera, D. G. Ousterout, J. M. Brunger, A. M. Farin, K. A. Glass, F. Guilak, G. E. Crawford, A. J. Hartemink, C. A. Gersbach, Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat Methods* **10**, 239-242 (2013); published online EpubMar (10.1038/nmeth.2361
nmeth.2361 [pii]).
172. T. Cermak, E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove, D. F. Voytas, Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**, e82 (2011); published online EpubJul (10.1093/nar/gkr218
gkr218 [pii]).
173. M. L. Maeder, S. J. Linder, D. Reyon, J. F. Angstman, Y. Fu, J. D. Sander, J. K. Joung, Robust, synergistic regulation of human gene expression using TALE activators. *Nat Methods* **10**, 243-245 (2013); published online EpubMar (10.1038/nmeth.2366
nmeth.2366 [pii]).

174. N. E. Sanjana, L. Cong, Y. Zhou, M. M. Cunniff, G. Feng, F. Zhang, A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* **7**, 171-192 (2012); published online EpubJan (10.1038/nprot.2011.431 nprot.2011.431 [pii]).
175. D. Reyon, S. Q. Tsai, C. Khayter, J. A. Foden, J. D. Sander, J. K. Joung, FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* **30**, 460-465 (2012); published online EpubMay (10.1038/nbt.2170 nbt.2170 [pii]).
176. B. R. Glick, Metabolic load and heterologous gene expression. *Biotechnol Adv* **13**, 247-261 (1995)073497509500004A [pii]).
177. B. Wiedenheft, S. H. Sternberg, J. A. Doudna, RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331-338 (2012); published online EpubFeb 16 (10.1038/nature10886 nature10886 [pii]).
178. D. Bhaya, M. Davison, R. Barrangou, CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* **45**, 273-297 (2011)10.1146/annurev-genet-110410-132430).
179. W. Y. Hwang, Y. Fu, D. Reyon, M. L. Maeder, S. Q. Tsai, J. D. Sander, R. T. Peterson, J. R. Yeh, J. K. Joung, Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31**, 227-229 (2013); published online EpubMar (10.1038/nbt.2501).
180. J. E. Dicarlo, J. E. Norville, P. Mali, X. Rios, J. Aach, G. M. Church, Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res* **41**, 4336-4343 (2013); published online EpubApr 1 (10.1093/nar/gkt135 gkt135 [pii]).
181. L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, F. Zhang, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013); published online EpubFeb 15 (10.1126/science.1231143).
182. P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, G. M. Church, RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013); published online EpubFeb 15 (10.1126/science.1232033).
183. S. W. Cho, S. Kim, J. M. Kim, J. S. Kim, Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**, 230-232 (2013); published online EpubMar (10.1038/nbt.2507).
184. D. Bikard, W. Jiang, P. Samai, A. Hochschild, F. Zhang, L. A. Marraffini, Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res*, (2013); published online EpubJun 12 (gkt520 [pii] 10.1093/nar/gkt520).
185. R. R. Beerli, D. J. Segal, B. Dreier, C. F. Barbas, 3rd, Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc Natl Acad Sci U S A* **95**, 14628-14633 (1998); published online EpubDec 8 (
186. W. Z. Li, F. Sherman, Two types of TATA elements for the CYC1 gene of the yeast Saccharomyces cerevisiae. *Mol Cell Biol* **11**, 666-676 (1991); published online EpubFeb (
187. L. Guarente, R. R. Yocum, P. Gifford, A GAL10-CYC1 hybrid yeast promoter identifies the GAL4 regulatory region as an upstream site. *Proc Natl Acad Sci U S A* **79**, 7410-7414 (1982); published online EpubDec (

188. D. H. Burke, D. G. Nickens, Expressing RNA aptamers inside cells to reveal proteome and ribonome function. *Brief Funct Genomic Proteomic* **1**, 169-188 (2002); published online EpubJul (
189. L. Kuras, K. Struhl, Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. *Nature* **399**, 609-613 (1999); published online EpubJun 10 (10.1038/21239).
190. X. Y. Li, A. Virbasius, X. Zhu, M. R. Green, Enhancement of TBP binding by activators and general transcription factors. *Nature* **399**, 605-609 (1999); published online EpubJun 10 (10.1038/21232).
191. Q. Huang, C. Gong, J. Li, Z. Zhuo, Y. Chen, J. Wang, Z. C. Hua, Distance and helical phase dependence of synergistic transcription activation in cis-regulatory module. *PLoS One* **7**, e31198 (2012)10.1371/journal.pone.0031198 PONE-D-11-13312 [pii]).
192. Y. D. Kwak, H. Koike, K. Sugaya, RNA interference with small hairpin RNAs transcribed from a human U6 promoter-driven DNA vector. *J Pharmacol Sci* **93**, 214-217 (2003); published online EpubOct (
193. M. Sawadogo, R. G. Roeder, Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell* **43**, 165-175 (1985); published online EpubNov (0092-8674(85)90021-2 [pii]).
194. C. N. Adra, P. H. Boer, M. W. McBurney, Cloning and expression of the mouse pgk-1 gene and the nucleotide sequence of its promoter. *Gene* **60**, 65-74 (1987)0378-1119(87)90214-9 [pii]).
195. A. Dorn, J. Bollekens, A. Staub, C. Benoist, D. Mathis, A multiplicity of CCAAT box-binding proteins. *Cell* **50**, 863-872 (1987); published online EpubSep 11 (0092-8674(87)90513-7 [pii]).
196. M. C. Blake, R. C. Jambou, A. G. Swick, J. W. Kahn, J. C. Azizkhan, Transcriptional initiation is controlled by upstream GC-box interactions in a TATAA-less promoter. *Mol Cell Biol* **10**, 6632-6641 (1990); published online EpubDec (
197. R. R. Yocum, S. Hanley, R. West, Jr., M. Ptashne, Use of lacZ fusions to delimit regulatory elements of the inducible divergent GAL1-GAL10 promoter in Saccharomyces cerevisiae. *Mol Cell Biol* **4**, 1985-1998 (1984); published online EpubOct (
198. S. Labbe, D. J. Thiele, Copper ion inducible and repressible promoter systems in yeast. *Methods Enzymol* **306**, 145-153 (1999)S0076-6879(99)06010-3 [pii]).
199. A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4**, e309 (2006); published online EpubOct (06-PLBI-RA-0625R2 [pii] 10.1371/journal.pbio.0040309).
200. G. Bak, S. W. Hwang, Y. Ko, J. Lee, Y. Kim, K. Kim, S. K. Hong, Y. Lee, On-off controllable RNA hybrid expression vector for yeast three-hybrid system. *BMB Rep* **43**, 110-114 (2010); published online EpubFeb (
201. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012); published online EpubAug 17 (10.1126/science.1225829).
202. Y. Fu, J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, J. K. Joung, J. D. Sander, High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*, (2013); published online EpubJun 23 (10.1038/nbt.2623 nbt.2623 [pii]).
203. Luke A. Gilbert, Matthew H. Larson, L. Morsut, Z. Liu, Gloria A. Brar, Sandra E. Torres, N. Stern-Ginossar, O. Brandman, Evan H. Whitehead, Jennifer A. Doudna, Wendell A. Lim, Jonathan S. Weissman, Lei S. Qi,

CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*, (2013).
204. M. L. Maeder, S. J. Linder, V. M. Cascio, Y. Fu, Q. H. Ho, J. K. Joung, CRISPR RNA-guided activation of endogenous human genes. *Nat Methods*, (2013); published online EpubJul 25 (10.1038/nmeth.2598 nmeth.2598 [pii]).
205. P. Perez-Pinera, D. D. Kocak, C. M. Vockley, A. F. Adler, A. M. Kabadi, L. R. Polstein, P. I. Thakore, K. A. Glass, D. G. Ousterout, K. W. Leong, F. Guilak, G. E. Crawford, T. E. Reddy, C. A. Gersbach, RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods*, (2013); published online EpubJul 25 (10.1038/nmeth.2600 nmeth.2600 [pii]).
206. A. A. Cheng, T. K. Lu, Synthetic biology: an emerging engineering discipline. *Annu Rev Biomed Eng* **14**, 155-178 (2012)10.1146/annurev-bioeng-071811-150118).
207. M. Costanzo, B. VanderSluis, E. N. Koch, A. Baryshnikova, C. Pons, G. Tan, W. Wang, M. Usaj, J. Hanchard, S. D. Lee, V. Pelechano, E. B. Styles, M. Billmann, J. van Leeuwen, N. van Dyk, Z. Y. Lin, E. Kuzmin, J. Nelson, J. S. Piotrowski, T. Srikumar, S. Bahr, Y. Chen, R. Deshpande, C. F. Kurat, S. C. Li, Z. Li, M. M. Usaj, H. Okada, N. Pascoe, B. J. San Luis, S. Sharifpoor, E. Shuteriqi, S. W. Simpkins, J. Snider, H. G. Suresh, Y. Tan, H. Zhu, N. Malod-Dognin, V. Janjic, N. Przulj, O. G. Troyanskaya, I. Stagljar, T. Xia, Y. Ohya, A. C. Gingras, B. Raught, M. Boutros, L. M. Steinmetz, C. L. Moore, A. P. Rosebrock, A. A. Caudy, C. L. Myers, B. Andrews, C. Boone, A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, (2016); published online EpubSep 23 (10.1126/science.aaf1420).
208. M. Boutros, J. Ahringer, The art and design of genetic screens: RNA interference. *Nature reviews. Genetics* **9**, 554-566 (2008); published online EpubJul (10.1038/nrg2364).
209. A. E. Carpenter, D. M. Sabatini, Systematic genome-wide screens of gene function. *Nature reviews. Genetics* **5**, 11-22 (2004); published online EpubJan (10.1038/nrg1248).
210. S. L. Forsburg, The art and design of genetic screens: yeast. *Nature reviews. Genetics* **2**, 659-668 (2001); published online EpubSep (10.1038/35088500).
211. P. Blancafort, E. I. Chen, B. Gonzalez, S. Bergquist, A. Zijlstra, D. Guthy, A. Brachat, R. H. Brakenhoff, J. P. Quigley, D. Erdmann, C. F. Barbas, 3rd, Genetic reprogramming of tumor cells by zinc finger transcription factors. *Proc Natl Acad Sci U S A* **102**, 11716-11721 (2005); published online EpubAug 16 (0501162102 [pii] 10.1073/pnas.0501162102).
212. K. S. Park, D. K. Lee, H. Lee, Y. Lee, Y. S. Jang, Y. H. Kim, H. Y. Yang, S. I. Lee, W. Seol, J. S. Kim, Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. *Nat Biotechnol* **21**, 1208-1214 (2003); published online EpubOct (10.1038/nbt868 nbt868 [pii]).
213. D. Carroll, Genome engineering with targetable nucleases. *Annual review of biochemistry* **83**, 409-439 (2014)10.1146/annurev-biochem-060713-035418).
214. F. Zhang, L. Cong, S. Lodato, S. Kosuri, G. M. Church, P. Arlotta, Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* **29**, 149-153 (2011); published online EpubFeb (10.1038/nbt.1775).
215. F. Farzadfard, S. D. Perli, T. K. Lu, Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas. *ACS Synth Biol*, (2013); published online EpubSep 11 (10.1021/sb400081r).

216. P. Mali, J. Aach, P. B. Stranges, K. M. Esvelt, M. Moosburner, S. Kosuri, L. Yang, G. M. Church, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**, 833-838 (2013); published online EpubSep (10.1038/nbt.2675).
217. H. Nishimasu, F. A. Ran, P. D. Hsu, S. Konermann, S. I. Shehata, N. Dohmae, R. Ishitani, F. Zhang, O. Nureki, Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949 (2014); published online EpubFeb 27 (10.1016/j.cell.2014.02.001).
218. J. G. Zalatan, M. E. Lee, R. Almeida, L. A. Gilbert, E. H. Whitehead, M. La Russa, J. C. Tsai, J. S. Weissman, J. E. Dueber, L. S. Qi, W. A. Lim, Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339-350 (2015); published online EpubJan 15 (10.1016/j.cell.2014.11.052).
219. L. A. Gilbert, M. A. Horlbeck, B. Adamson, J. E. Villalta, Y. Chen, E. H. Whitehead, C. Guimaraes, B. Panning, H. L. Ploegh, M. C. Bassik, L. S. Qi, M. Kampmann, J. S. Weissman, Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647-661 (2014); published online EpubOct 23 (10.1016/j.cell.2014.09.029).
220. M. A. Horlbeck, L. A. Gilbert, J. E. Villalta, B. Adamson, R. A. Pak, Y. Chen, A. P. Fields, C. Y. Park, J. E. Corn, M. Kampmann, J. S. Weissman, Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* **5**, (2016); published online EpubSep 23 (10.7554/eLife.19760).
221. S. Konermann, M. D. Brigham, A. E. Trevino, J. Joung, O. O. Abudayyeh, C. Barcena, P. D. Hsu, N. Habib, J. S. Gootenberg, H. Nishimasu, O. Nureki, F. Zhang, Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588 (2015); published online EpubJan 29 (10.1038/nature14136).
222. A. S. Wong, G. C. Choi, A. A. Cheng, O. Purcell, T. K. Lu, Massively parallel high-order combinatorial genetics in human cells. *Nat Biotechnol* **33**, 952-961 (2015); published online EpubSep (10.1038/nbt.3326).
223. A. S. Wong, G. C. Choi, C. H. Cui, G. Pregernig, P. Milani, M. Adam, S. D. Perli, S. W. Kazer, A. Gaillard, M. Hermann, A. K. Shalek, E. Fraenkel, T. K. Lu, Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proc Natl Acad Sci U S A* **113**, 2544-2549 (2016); published online EpubMar 1 (10.1073/pnas.1517883113).
224. O. Parnas, M. Jovanovic, T. M. Eisenhaure, R. H. Herbst, A. Dixit, C. J. Ye, D. Przybylski, R. J. Platt, I. Tirosh, N. E. Sanjana, O. Shalem, R. Satija, R. Raychowdhury, P. Mertins, S. A. Carr, F. Zhang, N. Hacohen, A. Regev, A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* **162**, 675-686 (2015); published online EpubJul 30 (10.1016/j.cell.2015.06.059).
225. O. Shalem, N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. S. Mikkelsen, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench, F. Zhang, Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87 (2014); published online EpubJan 3 (10.1126/science.1247005 science.1247005 [pii]).
226. R. Cencic, H. Miura, A. Malina, F. Robert, S. Ethier, T. M. Schmeing, J. Dostie, J. Pelletier, Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. *PLoS One* **9**, e109213 (2014)10.1371/journal.pone.0109213).
227. R. L. Frock, J. Hu, R. M. Meyers, Y. J. Ho, E. Kii, F. W. Alt, Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol* **33**, 179-186 (2015); published online EpubFeb (10.1038/nbt.3101).

228. H. O'Geen, I. M. Henry, M. S. Bhakta, J. F. Meckler, D. J. Segal, A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic acids research* **43**, 3389-3404 (2015); published online EpubMar 31 (10.1093/nar/gkv137).
229. X. Wu, D. A. Scott, A. J. Kriz, A. C. Chiu, P. D. Hsu, D. B. Dadon, A. W. Cheng, A. E. Trevino, S. Konermann, S. Chen, R. Jaenisch, F. Zhang, P. A. Sharp, Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* **32**, 670-676 (2014); published online EpubJul (10.1038/nbt.2889).
230. T. Wang, J. J. Wei, D. M. Sabatini, E. S. Lander, Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80-84 (2014); published online EpubJan 3 (10.1126/science.1246981 science.1246981 [pii]).
231. E. Yeger-Lotem, L. Riva, L. J. Su, A. D. Gitler, A. G. Cashikar, O. D. King, P. K. Auluck, M. L. Geddie, J. S. Valastyan, D. R. Karger, S. Lindquist, E. Fraenkel, Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. *Nature genetics* **41**, 316-323 (2009); published online EpubMar (10.1038/ng.337).
232. V. Khurana, J. Peng, C. Y. Chung, P. K. Auluck, S. Fanning, D. F. Tardiff, T. Bartels, M. Koeva, S. W. Eichhorn, H. Benyamini, Y. Lou, A. Nutter-Upham, V. Baru, Y. Freyzon, N. Tuncbag, M. Costanzo, B. J. San Luis, D. C. Schondorf, M. I. Barrasa, S. Ehsani, N. Sanjana, Q. Zhong, T. Gasser, D. P. Bartel, M. Vidal, M. Deleidi, C. Boone, E. Fraenkel, B. Berger, S. Lindquist, Genome-Scale Networks Link Neurodegenerative Disease Genes to alpha-Synuclein through Specific Molecular Pathways. *Cell systems* **4**, 157-170.e114 (2017); published online EpubFeb 22 (10.1016/j.cels.2016.12.011).
233. M. Goedert, M. G. Spillantini, K. Del Tredici, H. Braak, 100 years of Lewy pathology. *Nature reviews. Neurology* **9**, 13-24 (2013); published online EpubJan (10.1038/nrneurol.2012.242).
234. M. G. Spillantini, M. L. Schmidt, V. M. Lee, J. Q. Trojanowski, R. Jakes, M. Goedert, Alpha-synuclein in Lewy bodies. *Nature* **388**, 839-840 (1997); published online EpubAug 28 (10.1038/42166).
235. H. A. Lashuel, C. R. Overk, A. Oueslati, E. Masliah, The many faces of alpha-synuclein: from structure and toxicity to therapeutic target. *Nature reviews. Neuroscience* **14**, 38-48 (2013); published online EpubJan (10.1038/nrn3406).
236. E. Maries, B. Dass, T. J. Collier, J. H. Kordower, K. Steece-Collier, The role of alpha-synuclein in Parkinson's disease: insights from animal models. *Nature reviews. Neuroscience* **4**, 727-738 (2003); published online EpubSep (10.1038/nrn1199).
237. Y. C. Wong, D. Krainc, alpha-synuclein toxicity in neurodegeneration: mechanism and therapeutic strategies. *Nature medicine* **23**, 1-13 (2017); published online EpubFeb 07 (10.1038/nm.4269).
238. V. Khurana, S. Lindquist, Modelling neurodegeneration in Saccharomyces cerevisiae: why cook with baker's yeast? *Nature reviews. Neuroscience* **11**, 436-449 (2010); published online EpubJun (10.1038/nrn2809).
239. D. F. Tardiff, V. Khurana, C. Y. Chung, S. Lindquist, From yeast to patient neurons and back again: powerful new discovery platform. *Movement disorders : official journal of the Movement Disorder Society* **29**, 1231-1240 (2014); published online EpubSep (10.1002/mds.25989).
240. A. D. Gitler, A. Chesi, M. L. Geddie, K. E. Strathearn, S. Hamamichi, K. J. Hill, K. A. Caldwell, G. A. Caldwell, A. A. Cooper, J. C. Rochet, S. Lindquist, Alpha-synuclein is part of a diverse and highly conserved interaction network that includes PARK9 and manganese toxicity. *Nature genetics* **41**, 308-315 (2009); published online EpubMar (10.1038/ng.300).

241. S. Tenreiro, R. Rosado-Ramos, E. Gerhardt, F. Favretto, F. Magalhaes, B. Popova, S. Becker, M. Zweckstetter, G. H. Braus, T. F. Outeiro, Yeast reveals similar molecular mechanisms underlying alpha- and beta-synuclein toxicity. *Hum Mol Genet* **25**, 275-290 (2016); published online EpubJan 15 (10.1093/hmg/ddv470).
242. A. A. Cooper, A. D. Gitler, A. Cashikar, C. M. Haynes, K. J. Hill, B. Bhullar, K. Liu, K. Xu, K. E. Strathearn, F. Liu, S. Cao, K. A. Caldwell, G. A. Caldwell, G. Marsischky, R. D. Kolodner, J. Labaer, J. C. Rochet, N. M. Bonini, S. Lindquist, Alpha-synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's models. *Science* **313**, 324-328 (2006); published online EpubJul 21 (10.1126/science.1129462).
243. C. Y. Chung, V. Khurana, P. K. Auluck, D. F. Tardiff, J. R. Mazzulli, F. Soldner, V. Baru, Y. Lou, Y. Freyzon, S. Cho, A. E. Mungenast, J. Muffat, M. Mitalipova, M. D. Pluth, N. T. Jui, B. Schule, S. J. Lippard, L. H. Tsai, D. Krainc, S. L. Buchwald, R. Jaenisch, S. Lindquist, Identification and rescue of alpha-synuclein toxicity in Parkinson patient-derived neurons. *Science* **342**, 983-987 (2013); published online EpubNov 22 (10.1126/science.1245296 science.1245296 [pii]).
244. D. F. Tardiff, N. T. Jui, V. Khurana, M. A. Tambe, M. L. Thompson, C. Y. Chung, H. B. Kamadurai, H. T. Kim, A. K. Lancaster, K. A. Caldwell, G. A. Caldwell, J. C. Rochet, S. L. Buchwald, S. Lindquist, Yeast reveal a "druggable" Rsp5/Nedd4 network that ameliorates alpha-synuclein toxicity in neurons. *Science* **342**, 979-983 (2013); published online EpubNov 22 (10.1126/science.1245321 science.1245321 [pii]).
245. T. F. Outeiro, S. Lindquist, Yeast cells provide insight into alpha-synuclein biology and pathobiology. *Science* **302**, 1772-1775 (2003); published online EpubDec 5 (10.1126/science.1090439).
246. V. Bonifati, P. Rizzu, M. J. van Baren, O. Schaap, G. J. Breedveld, E. Krieger, M. C. Dekker, F. Squitieri, P. Ibanez, M. Joosse, J. W. van Dongen, N. Vanacore, J. C. van Swieten, A. Brice, G. Meco, C. M. van Duijn, B. A. Oostra, P. Heutink, Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. *Science* **299**, 256-259 (2003); published online EpubJan 10 (10.1126/science.1077209).
247. R. M. Canet-Aviles, M. A. Wilson, D. W. Miller, R. Ahmad, C. McLendon, S. Bandyopadhyay, M. J. Baptista, D. Ringe, G. A. Petsko, M. R. Cookson, The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. *Proc Natl Acad Sci U S A* **101**, 9103-9108 (2004); published online EpubJun 15 (10.1073/pnas.0402959101).
248. L. Zondler, L. Miller-Fleming, M. Repici, S. Goncalves, S. Tenreiro, R. Rosado-Ramos, C. Betzer, K. R. Straatman, P. H. Jensen, F. Giorgini, T. F. Outeiro, DJ-1 interactions with alpha-synuclein attenuate aggregation and cellular toxicity in models of Parkinson's disease. *Cell Death Dis* **5**, e1350 (2014)10.1038/cddis.2014.307).
249. C. J. Tsai, K. Aslam, H. M. Drendel, J. M. Asiago, K. M. Goode, L. N. Paul, J. C. Rochet, T. R. Hazbun, Hsp31 Is a Stress Response Chaperone That Intervenes in the Protein Misfolding Process. *J Biol Chem* **290**, 24816-24834 (2015); published online EpubOct 9 (10.1074/jbc.M115.678367).
250. S. Escusa, D. Laporte, A. Massoni, H. Boucherie, A. Dautant, B. Daignan-Fornier, Skp1-Cullin-F-box-dependent degradation of Aah1p requires its interaction with the F-box protein Saf1p. *J Biol Chem* **282**, 20097-20103 (2007); published online EpubJul 13 (10.1074/jbc.M702425200).
251. K. G. Mark, M. Simonetta, A. Maiolica, C. A. Seller, D. P. Toczyski, Ubiquitin ligase trapping identifies an SCF(Saf1) pathway targeting

unprocessed vacuolar/lysosomal proteins. *Mol Cell* **53**, 148-161 (2014); published online EpubJan 9 (10.1016/j.molcel.2013.12.003).
252. S. Hadano, R. Kunita, A. Otomo, K. Suzuki-Utsunomiya, J. E. Ikeda, Molecular and cellular function of ALS2/alsin: implication of membrane dynamics in neuronal development and degeneration. *Neurochemistry international* **51**, 74-84 (2007); published online EpubJul-Sep (10.1016/j.neuint.2007.04.010).
253. J. Chandran, J. Ding, H. Cai, Alsin and the molecular pathways of amyotrophic lateral sclerosis. *Molecular neurobiology* **36**, 224-231 (2007); published online EpubDec (10.1007/s12035-007-0034-x).
254. H. Takatsu, K. Yoshino, K. Toda, K. Nakayama, GGA proteins associate with Golgi membranes through interaction between their GGAH domains and ADP-ribosylation factors. *The Biochemical journal* **365**, 369-378 (2002); published online EpubJul 15 (10.1042/bj20020428).
255. O. Zhdankina, N. L. Strand, J. M. Redmond, A. L. Boman, Yeast GGA proteins interact with GTP-bound Arf and facilitate transport through the Golgi. *Yeast (Chichester, England)* **18**, 1-18 (2001); published online EpubJan 15 (10.1002/1097-0061(200101)18:1¡1::aid-yea644¿3.0.co;2-5).
256. M. Kosicek, P. Wunderlich, J. Walter, S. Hecimovic, GGA1 overexpression attenuates amyloidogenic processing of the amyloid precursor protein in Niemann-Pick type C cells. *Biochemical and biophysical research communications* **450**, 160-165 (2014); published online EpubJul 18 (10.1016/j.bbrc.2014.05.083).
257. B. von Einem, A. Wahler, T. Schips, A. Serrano-Pozo, C. Proepper, T. M. Boeckers, A. Rueck, T. Wirth, B. T. Hyman, K. M. Danzer, D. R. Thal, C. A. von Arnim, The Golgi-Localized gamma-Ear-Containing ARF-Binding (GGA) Proteins Alter Amyloid-beta Precursor Protein (APP) Processing through Interaction of Their GAE Domain with the Beta-Site APP Cleaving Enzyme 1 (BACE1). *PLoS One* **10**, e0129047 (2015)10.1371/journal.pone.0129047).
258. M. J. Vos, J. Hageman, S. Carra, H. H. Kampinga, Structural and functional diversities between members of the human HSPB, HSPH, HSPA, and DNAJ chaperone families. *Biochemistry* **47**, 7001-7011 (2008); published online EpubJul 8 (10.1021/bi800639z).
259. J. Gillis, S. Schipper-Krom, K. Juenemann, A. Gruber, S. Coolen, R. van den Nieuwendijk, H. van Veen, H. Overkleeft, J. Goedhart, H. H. Kampinga, E. A. Reits, The DNAJB6 and DNAJB8 protein chaperones prevent intracellular aggregation of polyglutamine peptides. *J Biol Chem* **288**, 17225-17237 (2013); published online EpubJun 14 (10.1074/jbc.M112.421685).
260. K. Vekrellis, M. Xilouri, E. Emmanouilidou, L. Stefanis, Inducible over-expression of wild type alpha-synuclein in human neuronal cells leads to caspase-dependent non-apoptotic death. *Journal of neurochemistry* **109**, 1348-1362 (2009); published online EpubJun (10.1111/j.1471-4159.2009.06054.x).
261. G. P. Dietz, K. V. Stockhausen, B. Dietz, B. H. Falkenburger, P. Valbuena, F. Opazo, P. Lingor, K. Meuer, J. H. Weishaupt, J. B. Schulz, M. Bahr, Membrane-permeable Bcl-xL prevents MPTP-induced dopaminergic neuronal loss in the substantia nigra. *Journal of neurochemistry* **104**, 757-765 (2008); published online EpubFeb (10.1111/j.1471-4159.2007.05028.x).
262. C. Henchcliffe, M. F. Beal, Mitochondrial biology and oxidative stress in Parkinson disease pathogenesis. *Nature clinical practice. Neurology* **4**, 600-609 (2008); published online EpubNov (10.1038/ncpneuro0924).
263. W. Neupert, J. M. Herrmann, Translocation of proteins into mitochondria. *Annual review of biochemistry* **76**, 723-749 (2007)10.1146/annurev.biochem.76.052705.163409).
264. A. A. Borisy, P. J. Elliott, N. W. Hurst, M. S. Lee, J. Lehar, E. R. Price, G. Serbedzija, G. R. Zimmermann, M. A. Foley, B. R. Stockwell, C. T. Keith,

Systematic discovery of multicomponent therapeutics. *Proc Natl Acad Sci U S A* **100**, 7977-7982 (2003); published online EpubJun 24 (10.1073/pnas.1337088100).
265. B. K. Slinker, The statistics of synergism. *Journal of molecular and cellular cardiology* **30**, 723-731 (1998); published online EpubApr (10.1006/jmcc.1998.0655).
266. W. R. Greco, G. Bravo, J. C. Parsons, The search for synergy: a critical review from a response surface perspective. *Pharmacological reviews* **47**, 331-385 (1995); published online EpubJun (
267. J. Moffat, D. A. Grueneberg, X. Yang, S. Y. Kim, A. M. Kloepfer, G. Hinkle, B. Piqani, T. M. Eisenhaure, B. Luo, J. K. Grenier, A. E. Carpenter, S. Y. Foo, S. A. Stewart, B. R. Stockwell, N. Hacohen, W. C. Hahn, E. S. Lander, D. M. Sabatini, D. E. Root, A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283-1298 (2006); published online EpubMar 24 (S0092-8674(06)00238-8 [pii] 10.1016/j.cell.2006.01.040).
268. R. D. Paulsen, D. V. Soni, R. Wollman, A. T. Hahn, M. C. Yee, A. Guan, J. A. Hesley, S. C. Miller, E. F. Cromwell, D. E. Solow-Cordero, T. Meyer, K. A. Cimprich, A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. *Mol Cell* **35**, 228-239 (2009); published online EpubJul 31 (10.1016/j.molcel.2009.06.021 S1097-2765(09)00459-6 [pii]).
269. C. N. Santos, G. Stephanopoulos, Combinatorial engineering of microbes for optimizing cellular phenotype. *Curr Opin Chem Biol* **12**, 168-176 (2008); published online EpubApr (10.1016/j.cbpa.2008.01.017 S1367-5931(08)00007-0 [pii]).
270. D. E. Root, N. Hacohen, W. C. Hahn, E. S. Lander, D. M. Sabatini, Genome-scale loss-of-function screening with a lentiviral RNAi library. *Nat Methods* **3**, 715-719 (2006); published online EpubSep (nmeth924 [pii] 10.1038/nmeth924).
271. A. W. Whitehurst, B. O. Bodemann, J. Cardenas, D. Ferguson, L. Girard, M. Peyton, J. D. Minna, C. Michnoff, W. Hao, M. G. Roth, X. J. Xie, M. A. White, Synthetic lethal screen identification of chemosensitizer loci in cancer cells. *Nature* **446**, 815-819 (2007); published online EpubApr 12 (nature05697 [pii] 10.1038/nature05697).
272. K. Demir, M. Boutros, Cell perturbation screens for target identification by RNAi. *Methods Mol Biol* **910**, 1-13 (2012)10.1007/978-1-61779-965-5˙1).
273. R. Durigon, Q. Wang, E. Ceh Pavia, C. M. Grant, H. Lu, Cytosolic thioredoxin system facilitates the import of mitochondrial small Tim proteins. *EMBO reports* **13**, 916-922 (2012); published online EpubOct (10.1038/embor.2012.116).
274. M. T. Lin, M. F. Beal, Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature* **443**, 787-795 (2006); published online EpubOct 19 (10.1038/nature05292).
275. M. E. Tanenbaum, L. A. Gilbert, L. S. Qi, J. S. Weissman, R. D. Vale, A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635-646 (2014); published online EpubOct 23 (10.1016/j.cell.2014.09.039).
276. N. A. Kearns, H. Pham, B. Tabak, R. M. Genga, N. J. Silverstein, M. Garber, R. Maehr, Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods* **12**, 401-403 (2015); published online EpubMay (10.1038/nmeth.3325).
277. P. I. Thakore, A. M. D'Ippolito, L. Song, A. Safi, N. K. Shivakumar, A. M. Kabadi, T. E. Reddy, G. E. Crawford, C. A. Gersbach, Highly specific

epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* **12**, 1143-1149 (2015); published online EpubDec (10.1038/nmeth.3630).
278. K. Han, E. E. Jeng, G. T. Hess, D. W. Morgens, A. Li, M. C. Bassik, Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat Biotechnol* **35**, 463-474 (2017); published online EpubMay (10.1038/nbt.3834).
279. J. P. Shen, D. Zhao, R. Sasik, J. Luebeck, A. Birmingham, A. Bojorquez-Gomez, K. Licon, K. Klepper, D. Pekin, A. N. Beckett, K. S. Sanchez, A. Thomas, C. C. Kuo, D. Du, A. Roguev, N. E. Lewis, A. N. Chang, J. F. Kreisberg, N. Krogan, L. Qi, T. Ideker, P. Mali, Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. *Nat Methods*, (2017); published online EpubMar 20 (10.1038/nmeth.4225).
280. C. Kuscu, S. Arslan, R. Singh, J. Thorpe, M. Adli, Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* **32**, 677-683 (2014); published online EpubJul (10.1038/nbt.2916).
281. K. L. MacQuarrie, A. P. Fong, R. H. Morse, S. J. Tapscott, Genome-wide transcription factor binding: beyond direct target regulation. *Trends in genetics : TIG* **27**, 141-148 (2011); published online EpubApr (10.1016/j.tig.2011.01.001).
282. C. Gouarne, J. Tracz, M. G. Paoli, V. Deluca, M. Seimandi, G. Tardif, M. Xilouri, L. Stefanis, T. Bordet, R. M. Pruss, Protective role of olesoxime against wild-type alpha-synuclein-induced toxicity in human neuronally differentiated SHSY-5Y cells. *British journal of pharmacology* **172**, 235-245 (2015); published online EpubJan (10.1111/bph.12939).
283. C. Lois, E. J. Hong, S. Pease, E. J. Brown, D. Baltimore, Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* **295**, 868-872 (2002); published online EpubFeb 1 (10.1126/science.1067081).
284. Y. C. Chen, J. Kenworthy, C. Gabrielse, C. Hanni, P. Zegerman, M. Weinreich, DNA replication checkpoint signaling depends on a Rad53-Dbf4 N-terminal interaction in Saccharomyces cerevisiae. *Genetics* **194**, 389-401 (2013); published online EpubJun (10.1534/genetics.113.149740).
285. Y. Benjamini, and Yosef Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *ournal of the royal statistical society. Series B (Methodological)*, 289-300 (1995).
286. M. D. Robinson, J. Grigull, N. Mohammad, T. R. Hughes, FunSpec: a web-based cluster interpreter for yeast. *BMC bioinformatics* **3**, 35 (2002); published online EpubNov 13 (
287. Y. C. Chen, M. Weinreich, Dbf4 regulates the Cdc5 Polo-like kinase through a distinct non-canonical binding interaction. *J Biol Chem* **285**, 41244-41254 (2010); published online EpubDec 31 (10.1074/jbc.M110.155242).
288. A. Xiao, Z. Cheng, L. Kong, Z. Zhu, S. Lin, G. Gao, B. Zhang, CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics*, (2014); published online EpubJan 21 (10.1093/bioinformatics/btt764).