SYNTHETIC BIOLOGY

Multiplex recording of cellular events over time on CRISPR biological tape

Ravi U. Sheth,[1,2] Sung Sun Yim,[1] Felix L. Wu,[1,2] Harris H. Wang[1,3]*

Although dynamics underlie many biological processes, our ability to robustly and accurately profile time-varying biological signals and regulatory programs remains limited. Here we describe a framework for storing temporal biological information directly in the genomes of a cell population. We developed a "biological tape recorder" in which biological signals trigger intracellular DNA production that is then recorded by the CRISPR-Cas adaptation system. This approach enables stable recording over multiple days and accurate reconstruction of temporal and lineage information by sequencing CRISPR arrays. We further demonstrate a multiplexing strategy to simultaneously record the temporal availability of three metabolites (copper, trehalose, and fucose) in the environment of a cell population over time. This work enables the temporal measurement of dynamic cellular states and environmental changes and suggests new applications for chronicling biological events on a large scale.

DNA is the primary information storage medium in living organisms and can be used in synthetic cellular memory devices that convert biological signals into heritable changes in nucleotide sequences. For example, approaches using recombinases (*1*–*6*), single-stranded DNA recombineering (*7*), and CRISPR-Cas9 (*8*–*12*) have been developed to record the level of a biological signal or to track developmental lineage. However, a major outstanding challenge has been the robust recording of temporally varying biological states or signals (e.g., gene expression or metabolite fluctuations) in living cells. Such a biological recording system would have powerful applications in studying dynamic cellular processes, such as complex regulatory programs, or in engineering "sentinel" cells that track changing environmental signals over time.

The bacterial CRISPR-Cas adaptation process exemplifies a naturally occurring biological memory system. When foreign genetic elements such as plasmids and phages invade a cell, short fragments of these exogenous nucleic acids can be captured by CRISPR-Cas adaptation proteins and integrated into genomic CRISPR arrays as spacers (*13*–*15*). This spacer acquisition process occurs in a unidirectional manner; new spacers are inserted at the 5′ of CRISPR arrays (*16*, *17*) and subsequently can be used by CRISPR-Cas immunity proteins to repel invaders that exhibit matching sequence identity (*18*). The DNA-writing potential of the adaptation process was recently applied to record the sequence and ordering of chemically synthesized oligonucleotides that were serially electroporated into cell populations (*19*, *20*). However, engineering the CRISPR-Cas adaptation system to directly record biological signals and their temporal context, and not simply sequence information of exogenous DNA, has not been achieved to date.

A tape recorder converts temporal signals such as analog audio into recordable data written onto a tape substrate as it is passed at a set rate across the recorder. Inspired by this temporal data storage scheme (Fig. 1A), we set out to develop a biological realization of the system, which we call temporal recording in arrays by CRISPR expansion (TRACE). In this framework, a biological input signal is first transformed into a change in the abundance of trigger DNA within living cells. The CRISPR-Cas spacer acquisition machinery is then used to record the amount of trigger DNA into CRISPR arrays in a unidirectional manner (Fig. 1B). Through this architecture, the presence of an input signal increases the frequency of trigger spacers incorporated into arrays, which constitutes recording of the positive signal. However, in the absence of a signal, reference spacers can still be acquired into arrays at a background rate from sources other than the trigger DNA, such as the genome (*21*). These reference spacers serve as pace-denoting markers that are embedded during the recording session, akin to the physical spacing on a tape substrate that represents time intervals.

We first explored an approach to convert the presence of a biological input into an increase in the abundance of trigger DNA within a population of *Escherichia coli* cells. We used a copy number–inducible trigger plasmid (pTrig), which contained a mini-F origin for stable maintenance and the phage P1 lytic replication protein RepL placed downstream of the Lac promoter. In the presence of the test input signal, isopropyl-β-D-1-thiogalactopyranoside (IPTG), transcription from the Lac promoter increases and results in expression of RepL. The RepL protein subsequently initiates plasmid replication from an origin located within the RepL coding sequence (*22*), which in turn increases pTrig copy number (Fig. 1C). Analysis of pTrig by quantitative polymerase chain reaction (qPCR) revealed a 653 ± 5–fold increase in copy number in cells induced with IPTG for 6 hours, compared with copy number in cells with no induction (methods, Fig. 1D, and figs. S1 and S2). This demonstrates that a biological signal that elicits a transcriptional response can be coupled to the alteration of an intracellular DNA pool.

Next, we assessed whether an increase in pTrig copy number could be recorded in CRISPR arrays across a cell population. Expression of the CRISPR adaptation proteins Cas1 and Cas2 promotes unidirectional integration of ~33–base pair DNA spacers into genomic CRISPR arrays in *E. coli* (*19*, *21*, *23*). We constructed a recording plasmid (pRec) that expresses Cas1 and Cas2 upon addition of anhydrotetracycline (aTc), which results in spacer acquisition (Fig. 1E and fig. S3A). Cells with pRec or with pRec and pTrig were induced with aTc and with or without IPTG, and their CRISPR arrays were assessed by sequencing to determine the source of newly acquired spacers, either from pRec, pTrig, or the genome (methods; Fig. 1, F and G; and fig. S4). In cells with pRec, spacers were preferentially derived from the pRec plasmid, consistent with enriched spacer acquisition from plasmids in *E. coli* documented in the literature (*21*). Cells with pRec and pTrig, but without IPTG induction, resulted in similar spacer acquisitions and low pTrig spacer incorporation (0.23 ± 0.06% of spacers). However, IPTG induction of pTrig increased overall spacer acquisition (fig. S3B) and, more importantly, increased the percentage of pTrig-derived spacers (32.4 ± 0.4% of spacers). This result demonstrates that an induced increase in trigger DNA abundance can be specifically recorded in CRISPR arrays. We further explored different input IPTG concentrations and observed an increasing relationship between pTrig copy number and the resulting percentage of pTrig-derived spacers (fig. S5). Although increased pTrig spacer incorporation could be detected after 4 hours of induction, robust recording was best achieved when the signal persisted for at least 6 hours (fig. S6).

Having assessed the two main components of the system—(i) transformation of a biological signal to increase abundance of intracellular DNA and (ii) capture of the amplified pool into CRISPR arrays—we next tested whether TRACE could be used to record biological signals in the temporal domain. We performed a systematic time-course recording experiment in which cells experienced the presence or absence of IPTG across 4 sequential days, constituting 16 distinct temporal signal profiles (Fig. 2A). Sequencing the resulting CRISPR arrays confirmed an overall expansion of arrays over time (fig. S7), with 24.7 ± 5.2% of all arrays having incorporated at least one new spacer by day 4. On average, about one in 15 arrays acquired a new spacer each day. As expected, arrays with increasing numbers of spacers were detected with decreasing frequency across the population (Fig. 2B). Because longer arrays contained more temporal information, we additionally

[1]Department of Systems Biology, Columbia University, New York, NY, USA. [2]Integrated Program in Cellular, Molecular, and Biomedical Studies, Columbia University, New York, NY, USA. [3]Department of Pathology and Cell Biology, Columbia University, New York, NY, USA.
*Corresponding author. Email: hw2429@columbia.edu

implemented a size enrichment protocol (methods) that facilitated the analysis of arrays with up to five new spacers (Fig. 2B).

For TRACE to function as a useful biological tape recorder, the spacer identity (reference or trigger) and ordering within CRISPR arrays should correlate with the actual temporal signal profile. We first noted that the system can act as a simple signal counter by observing that the total percentage of pTrig spacers increased proportionally with the number of times the signal was present in the signal profile (Fig. 2C). Next, we analyzed pTrig spacer incorporation and ordering in CRISPR arrays. For example, individual arrays from a sample receiving the IPTG profile [on, on, off, off] were variable but displayed an overall enrichment of pTrig spacers at distal positions in the array (Fig. 2D and fig. S8A). To visualize these incorporation patterns across each of the 16 signal profiles, for arrays of different lengths (L1 to L5), we calculated the population average of pTrig spacers at each spacer position (Fig. 2, D and E, and fig. S8B). These patterns of pTrig frequencies exhibited a high degree of correspondence to their respective temporal signal profiles when considered in reverse (i.e., oldest to newest acquired spacers; Fig. 2F), which suggested the successful recording of temporal biological signals.

To improve the interpretation of TRACE data, we explored a method for accurate and automated inference of the input temporal signal profiles from recorded CRISPR arrays. We hypothesized that the array expansion process could be modeled to yield a useful classification scheme for matching an observed pattern of arrays to its corresponding signal profile. To test this approach, we first defined a cell population's repertoire of CRISPR arrays as a distribution of "array types." Array types constitute all possible array configurations across all array lengths with either reference or trigger spacers occupying each spacer position (Fig. 3A). We then developed a simple analytical model of the CRISPR expansion process for calculating the expected frequencies for all array types given a signal profile (methods). Only four constants are needed to parameterize the model for each array length: the rates of array expansion and pTrig incorporation per recording interval, in the presence or absence of a signal (fig. S9 and table S5). Using this model, we calculated the expected distributions of array types for all 16 temporal signal profiles and compared these distributions of array-type frequencies with those from experimentally recorded arrays. The predicted and observed array-type distributions matched closely (fig. S10). For example, for two signal profiles with an equal number of inductions but different temporal ordering, our models yielded distinctive array-type distributions that appeared to recapitulate the corresponding experimental data (Fig. 3B).

To quantitatively compare and classify the observed data with model array-type distributions, we calculated all pairwise Euclidean distances between them. An observed CRISPR array population was assigned to the most probable signal profile on the basis of the data-model pair with the shortest Euclidean distance (Fig. 3C). Using L1 arrays only, which do not contain any temporal information, only five of 16 signal profiles could be correctly classified. Using array types L2 to L4 individually resulted in much higher accuracy of assignments (13 to 14 of 16 correct). When array types L2 to L4 were used together, we could perfectly classify all 16 populations with their correct temporal signal profiles (methods and Fig. 3D). Only a few hundred arrays of a given length, corresponding to minimum populations of ~10^5 total arrays, were required to recapitulate reasonable classification accuracy (fig. S11). This demonstrates that temporal signals can be recorded and subsequently reconstructed with high accuracy from CRISPR arrays by using a simple model of the expansion process.

Beyond simply assigning spacer identity as reference or trigger, we hypothesized that spacer sequences themselves may additionally contain population lineage information, given the large pool of potential spacers. In the time-course recording experiment, cell populations were experimentally split into subpopulations each day, which resulted in a defined branching history of the 16 populations (Fig. 3E). By performing lineage reconstruction using a simple metric to assess spacer repertoire distance between populations (methods), we could reconstruct the entire experimental population lineage with nearly perfect accuracy (Fig. 3F).

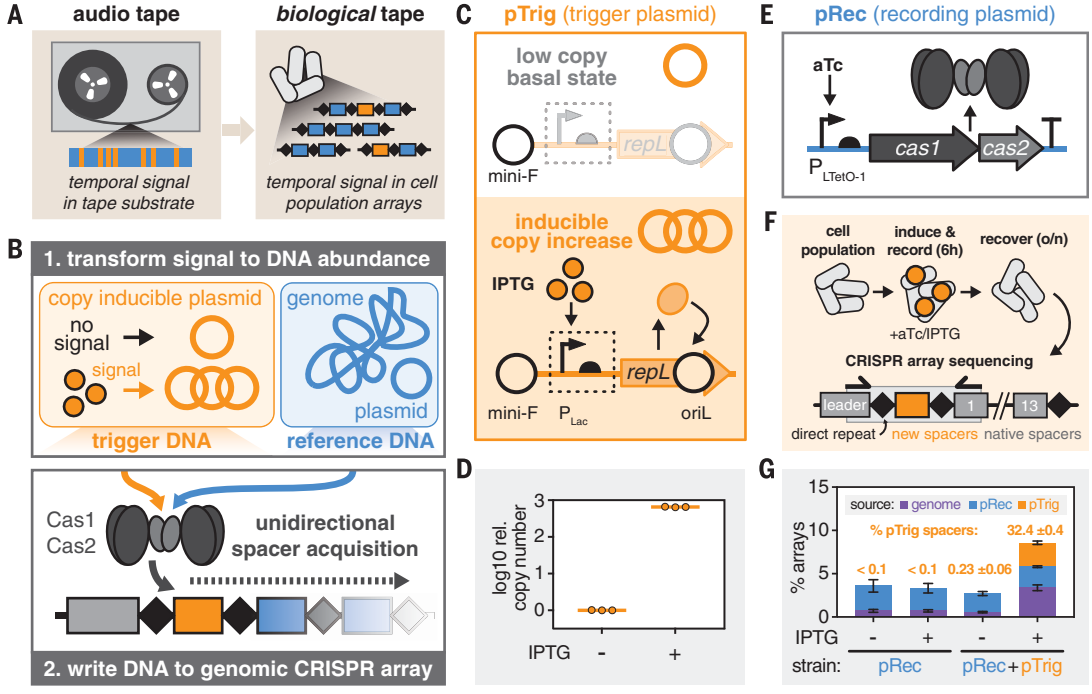

Fig. 1. Temporal recording in arrays by CRISPR expansion (TRACE). (**A**) Akin to an audio tape, temporal biological signals can be stored in DNA arrays within a cell population. (**B**) TRACE functions by first transforming an input biological signal to an altered abundance of trigger DNA (orange). This trigger DNA, alongside reference DNA (blue), is then recorded as spacers in genomic CRISPR arrays of a cell population in a unidirectional fashion, enabling capture of temporal information. (**C**) The pTrig trigger plasmid includes a mini-F origin for stable maintenance and an IPTG-inducible phage P1 replication system for copy number increase. P_{Lac}, Lac promoter. (**D**) qPCR measurement of pTrig relative copy number (log10 scale) in cells exposed to no IPTG or 1 mM IPTG for 6 hours. (**E**) The pRec recording plasmid includes an aTc-inducible *E. coli* Cas1 and Cas2 expression cassette. (**F**) Experimental induction scheme and CRISPR array sequencing approach. (o/n, overnight). (**G**) Cells with pRec or with pRec and pTrig were exposed to 100 ng/μL aTc and no or 1 mM IPTG and subjected to sequencing; resulting arrays with a single new spacer and identified source (genome, pRec, or pTrig) are plotted as a percentage of all measured CRISPR arrays. Error bars represent standard deviation of three biological replicates.

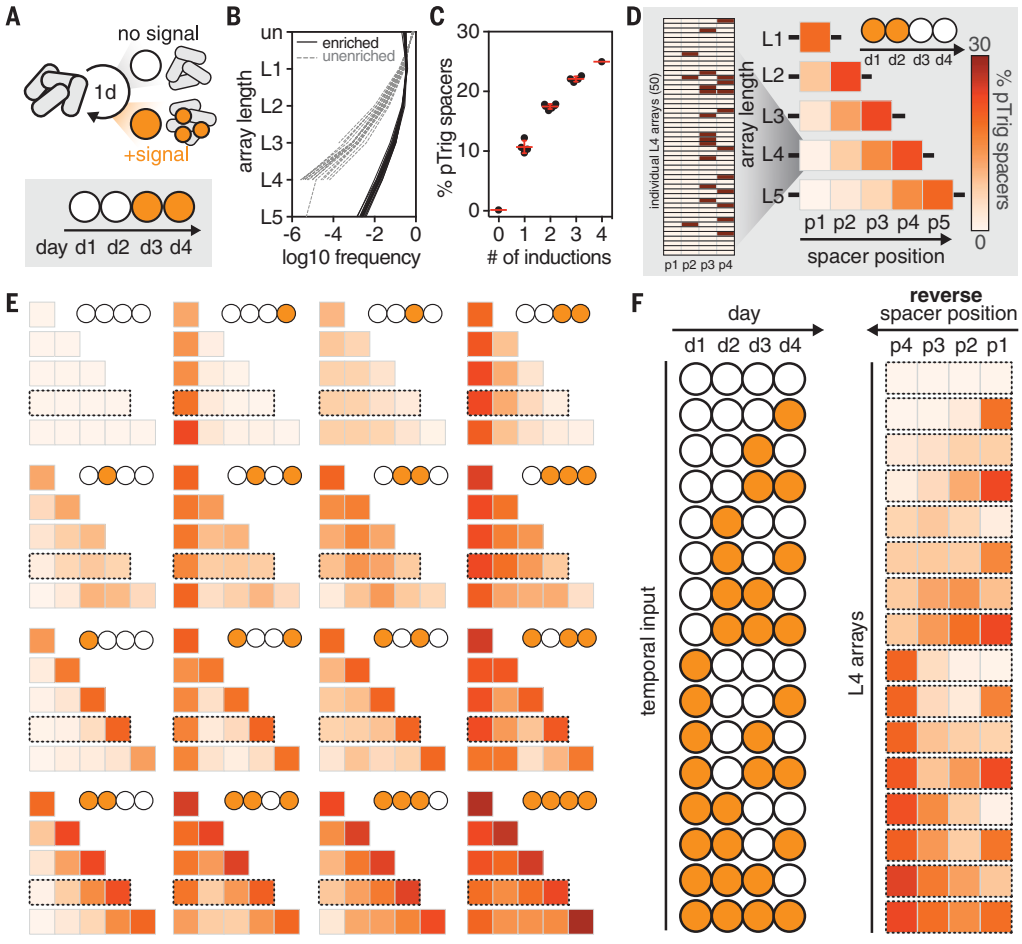

Fig. 2. Temporal recording of 4-day input profiles. (**A**) Cell populations were subjected to daily exposures over 4 sequential days (d1 to d4), constituting all 16 possible temporal signal profiles. (**B**) Resulting CRISPR arrays were sequenced with (black) and without (gray) a size-enrichment method. The frequencies (log10 scale) of unexpanded (un) and expanded arrays of different lengths (L1 to maximum detectable L5) are plotted. (**C**) Input profiles are grouped by number of pTrig inductions, and the percentage of pTrig spacers in each profile is displayed; red lines indicate means and standard deviations. (**D**) On the left, 50 L4 arrays sampled from the full data set for the input profile [on, on, off, off] are shown (shaded, pTrig spacer; unshaded, reference spacer; positions p1 to p4, 5′-to-3′ of array). Spacer incorporation can be analyzed across arrays of different lengths (L) and positions (p) as a heatmap displaying percentages of pTrig spacers detected at each location (right). (**E**) CRISPR arrays derived from recordings of all 16 temporal signal profiles. (**F**) The input signal profile (left) and corresponding L4 arrays (right, shown in reverse order to aid visual comparison) are displayed.

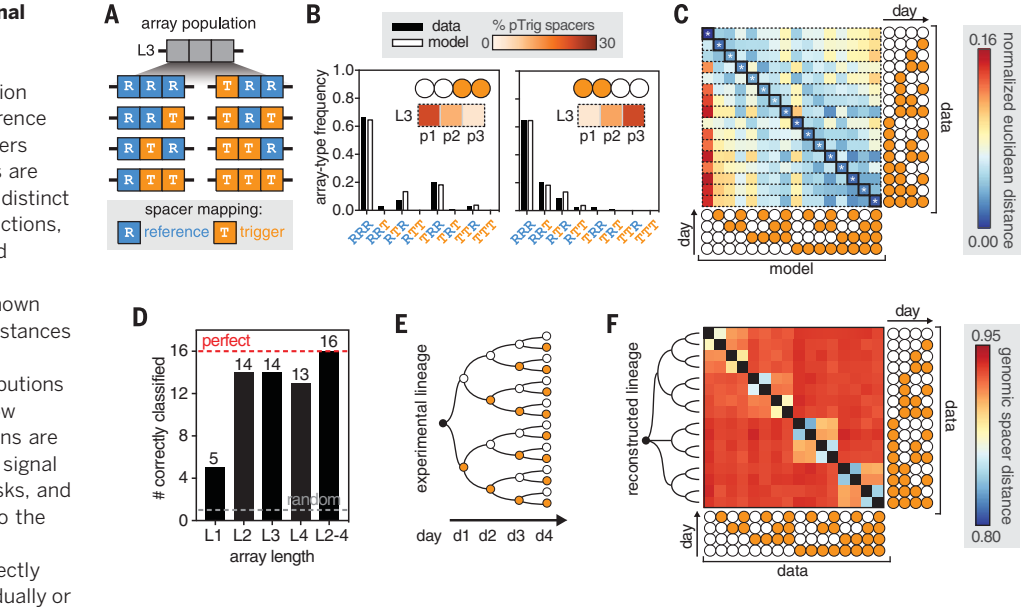

Fig. 3. Reconstructing temporal signal profiles and population lineages. (**A**) CRISPR array populations can be described as a frequency distribution consisting of all permutations of reference (R, blue) and trigger (T, orange) spacers for a given array length (L); L3 arrays are depicted. (**B**) As an example, for two distinct profiles with an equal number of inductions, observed (black) and model-predicted (white) L3 array-type frequencies are plotted; L3 positional averages are shown for reference (inset). (**C**) Euclidean distances between observed (rows) and model-predicted (columns) array-type distributions were calculated and normalized by row (L2, L3, and L4 array-type distributions are concatenated). The correct temporal signal profiles are indicated by white asterisks, and the models with minimum distance to the observed data are indicated by black outlines. (**D**) Number of profiles correctly classified using arrays L1 to L4 individually or arrays L2 to L4 together as in (C); the gray dashed line indicates the expected random classification (one of 16 correct). (**E**) A defined branching history was used in the temporal recording experiment. (**F**) The mapping locations for genomic spacers within L1 arrays were used as the sequence identity of the spacer. The Jaccard distances between all samples (1 minus the proportion of spacers shared between two samples) are displayed. Lineage reconstruction was performed using the Fitch-Margoliash method on this distance matrix and is displayed on the left; only one lineage is not fully differentiated (cells receiving induction on d1).

To further characterize the recording performance of TRACE, we assessed the stability of stored information and the potential for longer-term recordings. Propagation of recordings stored within cell populations over 8 days (~50 generations) did not appear to alter array-type distributions (fig. S12, A and B), and induction of recording showed negligible loss of previously acquired spacers (fig. S12C). These results demonstrate stable data storage. We repeated recording experiments on selected temporal signal profiles for 10 days, which showed reasonable reconstruction accuracy up to 6 days (four of seven correctly classified; fig. S13). In general, longer arrays increased the accuracy of signal profile reconstruction during longer recording sessions, which suggests that longer-read sequencing may further increase the performance of long-term recording analysis.

Last, we explored the possibility of using TRACE for multichannel temporal recording. We devised a multiplexing strategy wherein various pTrig sensor systems could be associated with uniquely barcoded CRISPR arrays within a cell population (Fig. 4A). Specifically, we chose to mutagenize the 3′ direct repeat (DR) sequence, which should not affect spacer integration (*24*), as a barcode. This allowed for multiplexing with no modification to the sequencing protocol. More importantly, this enabled more stringent calling of barcodes because the DR sequence is duplicated during each spacer incorporation event (*23*, *25*). Using MAGE [multiplex automated genome engineering (*26*)], we generated strains with new genomic DR barcodes. In distinct barcoded strains, we coupled different sensors to pTrig and screened their performance (fig. S14). Three orthogonal and robust biosensors that detected the biologically meaningful chemicals copper (heavy-metal contaminant), trehalose (dietary sugar metabolite), and fucose [associated with mammalian gut infection (*27*)] were selected for multiplex recording experiments. To assess the capacity for multichannel recording, we exposed cell populations containing a mix of all three strains to all eight combinations of the three input chemicals. The resulting CRISPR arrays were sequenced and demultiplexed using the DR barcodes. Each sensor strain displayed a robust increase in pTrig-derived spacers (>24-fold) only in the presence of their cognate input (Fig. 4B and fig. S15). Importantly, these results indicate modular compatibility of TRACE for multichannel recording with a variety of sensing systems, including engineered sensors or native promoters with endogenous transcription factor expression.

To explore multiplex temporal recording, we used the three-strain sensing system to perform a time-course exposure experiment over 3 days. Cell populations were exposed to 16 selected temporal signal profiles of 512 possible profiles, and resulting CRISPR arrays were sequenced. Sensor strains fluctuated in their final abundance but were maintained at sufficient levels to enable CRISPR array analysis (fig. S16). We parameterized models for each sensor individually as before and inferred the exposure history of each of the three inputs individually for all 16 populations by classification against model predictions (Fig. 4C). We were able to correctly classify 14, 13, and 12 of the 16 signal profiles for the copper, trehalose, and fucose sensors, respectively (Fig. 4, D and E). Classification accuracy for all three inputs simultaneously was assessed by the Hamming distance threshold to the actual temporal signal profiles; eight of 16 profiles were perfectly classified, and the rest were within a Hamming distance of 2 (Fig. 4F), implying that even incorrect predictions were close to actual signal profiles. Together, these results demonstrate accurate multichannel recording with the TRACE system.

Our work enables new applications in biological recording. TRACE could be used to record metabolite fluctuations, gene expression changes, and lineage-associated information across cell populations in difficult-to-study habitats, such as the mammalian gut, or in open settings, such as soil or marine environments. Applying inducible intracellular DNA production systems in parallel (*28*) and other CRISPR-Cas adaptation machinery (*13*, *29*) could extend our system to other bacteria (or even eukaryotes) and increase the temporal resolution of recording beyond the levels demonstrated here (6 hours, ~45 μHz). The system could be further optimized by increasing the spacer incorporation rate (*30*), increasing the sequencing length (e.g., by nanopore sequencing), and improving reconstruction algorithms. These advances could further facilitate biological recording of inputs across many signal channels, with higher temporal resolution, and in smaller populations, possibly down to single cells. TRACE and future strategies for massively parallel recording of biological states should greatly advance our ability to delineate and understand complex cellular processes across time.

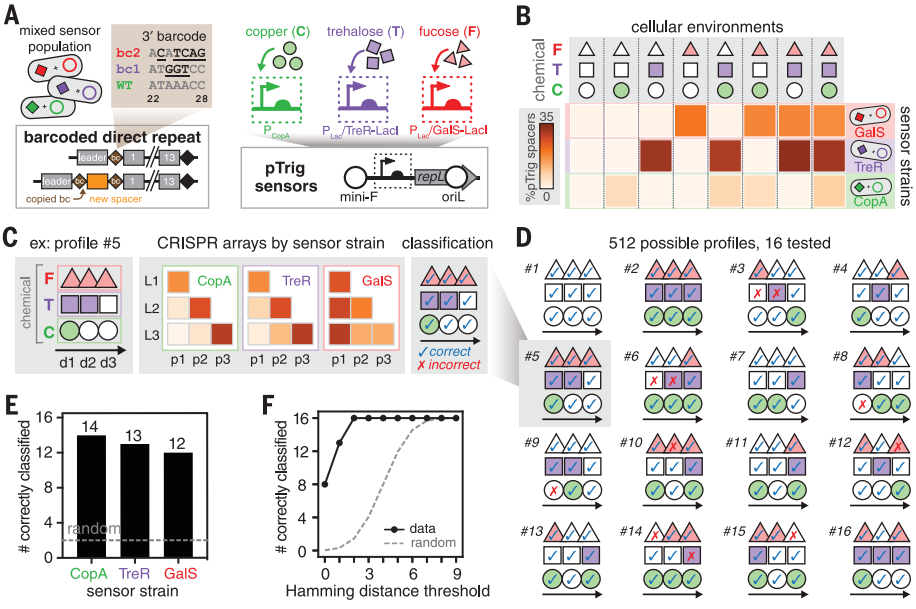

Fig. 4. Multiplex temporal recording with a barcoded sensor population. (**A**) The direct repeat (DR) of a CRISPR array can be barcoded to associate sensors with specific arrays; generated distal DR sequences with barcodes (bc) are shown. Sensors of copper, trehalose, and fucose were linked to the pTrig system and introduced into barcoded strains. The copper sensor uses a native promoter with endogenous transcription factor expression, whereas the trehalose and fucose sensors use an engineered transcription factor. (**B**) The three barcoded sensor strains were mixed and exposed to eight combinatorial inputs of the three chemicals; the resulting percentage of pTrig spacers for each barcoded sensor strain is displayed (average of three biological replicates). (**C**) The strain mixture was exposed to combinatorial inputs over 3 days. As an example, profile 5 is displayed, along with CRISPR arrays for each sensor (plotted as in Fig. 2, but the color map is rescaled for each sensor to aid visualization) and the resulting classification (correct, blue checkmark; incorrect, red X). (**D**) Of 512 (8^3) possible profiles, 16 were tested (six defined and 10 randomly generated); the resulting classification is shown as in (C) (black arrows indicate the time course, d1 to d3). (**E**) Single-channel classification accuracy. Profiles were classified for each sensor on the basis of L2 and L3 arrays; the gray dashed line indicates the expected random classification (two of 16 correct). (**F**) Multichannel classification accuracy. Predictions were considered across all three sensors, and the number classified correctly within a Hamming distance threshold is shown (black line) compared with the expected random classification (gray dashed line).

REFERENCES AND NOTES

1. A. E. Friedland *et al.*, *Science* **324**, 1199–1202 (2009).
2. J. Bonnet, P. Yin, M. E. Ortiz, P. Subsoontorn, D. Endy, *Science* **340**, 599–603 (2013).
3. L. Yang *et al.*, *Nat. Methods* **11**, 1261–1266 (2014).
4. V. Hsiao, Y. Hori, P. W. Rothemund, R. M. Murray, *Mol. Syst. Biol.* **12**, 869–14 (2016).

5. N. Roquet, A. P. Soleimany, A. C. Ferris, S. Aaronson, T. K. Lu, *Science* **353**, aad8559 (2016).
6. W. Pei *et al*., *Nature* **548**, 456–460 (2017).
7. F. Farzadfard, T. K. Lu, *Science* **346**, 1256272 (2014).
8. A. McKenna *et al*., *Science* **353**, aaf7907 (2016).
9. K. L. Frieda *et al*., *Nature* **541**, 107–111 (2017).
10. S. T. Schmidt, S. M. Zimmerman, J. Wang, S. K. Kim, S. R. Quake, *ACS Synth. Biol.* **6**, 936–942 (2017).
11. S. D. Perli, C. H. Cui, T. K. Lu, *Science* **353**, aag0511 (2016).
12. R. Kalhor, P. Mali, G. M. Church, *Nat. Methods* **14**, 195–200 (2017).
13. S. A. Jackson *et al*., *Science* **356**, eaal5056 (2017).
14. S. H. Sternberg, H. Richter, E. Charpentier, U. Qimron, *Mol. Cell* **61**, 797–808 (2016).
15. A. V. Wright *et al*., *Science* **357**, 1113–1118 (2017).
16. R. Barrangou *et al*., *Science* **315**, 1709–1712 (2007).
17. J. McGinn, L. A. Marraffini, *Mol. Cell* **64**, 616–623 (2016).
18. L. A. Marraffini, *Nature* **526**, 55–61 (2015).
19. S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, *Science* **353**, aaf1175 (2016).
20. S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, *Nature* **547**, 345–349 (2017).
21. A. Levy *et al*., *Nature* **520**, 505–510 (2015).
22. M. B. Łobocka *et al*., *J. Bacteriol.* **186**, 7032–7068 (2004).
23. I. Yosef, M. G. Goren, U. Qimron, *Nucleic Acids Res.* **40**, 5569–5576 (2012).
24. J. K. Nuñez, L. Bai, L. B. Harrington, T. L. Hinder, J. A. Doudna, *Mol. Cell* **62**, 824–833 (2016).
25. J. K. Nuñez, A. S. Y. Lee, A. Engelman, J. A. Doudna, *Nature* **519**, 193–198 (2015).
26. H. H. Wang *et al*., *Nature* **460**, 894–898 (2009).
27. J. M. Pickard *et al*., *Nature* **514**, 638–641 (2014).
28. J. Elbaz, P. Yin, C. A. Voigt, *Nat. Commun.* **7**, 11179 (2016).
29. S. Silas *et al*., *Science* **351**, aad4234 (2016).
30. R. Heler *et al*., *Mol. Cell* **65**, 168–175 (2017).

ACKNOWLEDGMENTS

We thank T. Blazejewski, C. Munck, and members of the Wang laboratory for advice and comments on the manuscript. Sequencing data associated with this study are available in the National Center for Biotechnology Information Sequence Read Archive under PRJNA417866, and plasmids are available through Addgene. H.H.W. acknowledges specific funding support from the Department of Defense Office of Naval Research (N00014-17-1-2353 and N00014-15-1-2704), an NIH Director's Early Independence Award (1-DP5OD009172-02), and the Sloan Foundation (FR-2015-65795) for this work. R.U.S. is supported by a Fannie and John Hertz Foundation Graduate Fellowship and an NSF Graduate Research Fellowship (DGE 16-44869). F.L.W. is supported by an NIH training grant (T32GM008224). R.U.S., S.S.Y., and H.H.W. developed the initial concept; R.U.S., S.S.Y., and F.L.W. performed experiments; R.U.S. and F.L.W. analyzed the sequencing data; R.U.S. and H.H.W. wrote the manuscript; and all authors discussed results and commented on and approved the manuscript. H.H.W. and R.U.S. are inventors on a provisional patent application filed by the Trustees of Columbia University in the City of New York regarding this work.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/358/6369/1457/suppl/DC1
Materials and Methods
Figs. S1 to S16
Tables S1 to S5
References (*31*–*44*)

16 June 2017; resubmitted 29 September 2017
Accepted 13 November 2017
Published online 23 November 2017
10.1126/science.aao0958

Multiplex recording of cellular events over time on CRISPR biological tape

Ravi U. Sheth, Sung Sun Yim, Felix L. Wu and Harris H. Wang

Science **358** (6369), 1457-1461.
DOI: 10.1126/science.aao0958originally published online November 23, 2017

A CRISPR device to record time

The CRISPR adaptation system has been used to record the sequence and ordering of exogenous oligonucleotides that are electroporated into cell populations. Sheth *et al.* engineered a system bypassing the use of exogenous DNA to directly record temporal signals. An input biological signal is transformed into the ratio of the frequency of incorporating trigger DNA to that of incorporating reference DNA into the genomes of a bacterial population. A multiplexing strategy enables simultaneous recording of three environmental signals with high temporal resolution.

Science, this issue p. 1457

www.sciencemag.org/cgi/content/full/science.aao0958/DC1

Supplementary Materials for

Multiplex recording of cellular events over time on CRISPR biological tape

Ravi U. Sheth, Sung Sun Yim, Felix L. Wu, Harris H. Wang*

*Corresponding author. Email: hw2429@columbia.edu

Published 23 November 2017 on *Science* First Release
DOI: 10.1126/science.aao0958

This PDF file includes:

Materials and Methods

Plasmid construction

All plasmids (table S1) were constructed via the Golden Gate method (*31*) with the NEB 10-beta cloning strain (NEB, C3019H) and were verified via Sanger sequencing (Eton Bioscience, Genewiz). All plasmids are deposited at Addgene. The RBS calculator (*32*) and Anderson library of promoters (http://parts.igem.org/Promoters/Catalog/Anderson) were utilized as annotated on plasmid maps.

The pTrig plasmid was generated from pSB2K3-BBa_J04450 (iGEM 2016 distribution), which itself was derived from the pSCANS vector (http://genome.bnl.gov/Vectors/pscans.php). To construct pTrig, the BBa_J04450 (RFP) sequence and Biobrick multiple cloning site were removed; the resulting plasmid contains the mini-F origin and replication machinery, P1 lytic replication element RepL placed downstream of an IPTG-inducible Lac promoter, and kanamycin resistance marker.

The pRec plasmid was generated by placing the *E. coli cas1-cas2* cassette (amplified from NEB 10-beta) downstream of the $P_{LTetO-1}$ promoter (*33*) on a ColE1 plasmid containing chloramphenicol resistance marker and constitutively expressed TetR and LacI (LacI is required to repress the Lac promoter on pTrig, see fig. S1).

For the CopA sensor, a derivative of the pTrig plasmid (pTrig-CopA) containing the *E. coli* BL21 CopA promoter (100 bp upstream sequence) with RiboJ (*34*) and B0034 RBS was constructed. This was utilized with a derivative of the pRec plasmid without LacI (pRec ΔLacI).

For the GalS and TreR sensors, derivatives of the pRec plasmid containing LacI chimeric transcription factors (pRec-TreR, pRec-GalS) were constructed by swapping the LacI ligand binding domain with either the TreR or GalS ligand binding domains and then subsequently introducing point mutations that have been characterized to improve sensor performance (TreR: V52A; GalS: Q54A, E232K) (*35*). These pRec variants were then utilized with the pTrig plasmid.

Chromosomal alteration of strains with MAGE

Given that we utilized LacI chimeric transcription factors (GalS, TreR) we generated a variant of the *E. coli* BL21 strain lacking endogenous expression of LacI to prevent interaction with the sensing systems. We utilized the MODEST tool to design a recombineering primer (MAGE_tKO_lacI, table S3) to perform a translational knockout of chromosomal LacI by introduction of three stop codons into the beginning of the *lacI* coding sequence (*36*). Briefly, the BL21 strain was transformed with pKD46 (*37*) and grown at 30 °C with 50 µg/mL Carbenicillin (Fisher BP2648). An overnight culture of this strain was back-diluted and grown for 30 min, 0.5% arabinose was added, and the culture was grown to approximately OD600 = 0.6. 1 mL of cells were then placed on ice and washed with nuclease-free water 3 times, resuspended in 2.5 µM oligonucleotide at a volume of 50 µL, and subjected to electroporation. Cells were then recovered for 1 hour at

30 °C. This process constituted one round of recombineering; after this procedure cells were plated on LB-agar with antibiotics and X-gal (200 μg/mL, Thermo FERR0404) and grown at 30 °C. Resulting clones were screened for loss of LacI expression by beta-galactosidase assay (loss of LacI expression de-represses LacZ), and a resulting clone was verified to contain the correct chromosomal alteration by Sanger sequencing. This strain was hereafter denoted BL21 LacI_tKO.

Oligo recombineering was also utilized to introduce barcodes into the genomic CRISPR array first direct repeat (DR) sequence. A recombineering primer (MAGE_BL21_DR, table S3) was designed to mutagenize the distal 7 bp of the DR sequence (inadvertently, we also targeted the first base pair of the first native genomic spacer for mutagenesis, resulting in 8 bp total targeted for mutagenesis). The BL21 LacI_tKO strain, still harboring pKD46 was subjected to five rounds of oligo recombineering as described above. The resulting cell population was then subjected to heatshock at 42 °C for 1 hour to promote loss of pKD46 and recovered overnight at 37 °C in LB without antibiotics; a cryostock of the population (15% glycerol) was saved for subsequent screening for clones with barcoded DR sequences.

Experimental conditions (induction of pRec and pTrig)

All testing was conducted in *E. coli* BL21 (NEB C2530H), a strain that contains two genomic CRISPR arrays but lacks *cas* interference machinery (*38*). For induction experiments with the LacI sensor, the *E. coli* BL21 strain was transformed with appropriate plasmids (pRec, or pRec+pTrig) via electroporation (table S2). A single colony was picked and grown to stationary phase and a cryostock (15% glycerol) was created for storage at −80 °C.

The general experimental workflow of an induction experiment was as follows:

1. A culture tube (Thomas Scientific 110158PL-TS) containing 3 mL autoclaved LB-Lennox (BD 240230) and appropriate antibiotics at indicated final concentrations (pRec: chloramphenicol 34 μg/mL [EMD Millipore Omnipur 3130, diluted in 100% ethanol], pTrig: kanamycin 50 μg/mL [Fisher BP906-5, diluted in nuclease free water]) was inoculated from the culture glycerol stock and grown overnight (>12 hours) at 37 °C in an Innova44 incubator shaker at 230 rpm.
2. The next day, this culture was diluted 1:100 into a new tube containing 3 mL LB media and appropriate antibiotics and allowed to grow in the same culture conditions for 2 hours to bring cultures into exponential phase.
3. This culture was then diluted 1:100 into a new tube containing 3 mL LB media, appropriate antibiotics, and appropriate anhydrotetracycline (aTc) and isopropyl β-D-1-thiogalactopyranoside (IPTG) inducers at indicated final concentrations (aTc: 100 ng/mL [Cayman 10009542, diluted in 100% ethanol], IPTG: 1 mM [Thermo R0392, diluted in nuclease free water]). This culture was then allowed to grow in the same culture conditions for 6 hours.

4. Finally, culture from this tube was diluted 1:100 into a new tube containing 3 mL LB media and appropriate antibiotics, and allowed to recover in the same culture conditions overnight for 16 hours.
5. At the conclusion of the experiment 500 µL of culture was transferred to a 1.5 mL tube (VWR 20170-333), the tube was spun down (15,0000 rpm, 30s) to pellet cells, media was removed, and the pellet was stored at −20 °C for subsequent analysis.

Experimental conditions (temporal recordings)

For 4 day temporal recording experiments the induction procedure as above was utilized, but after the first day, recovery cultures from the previous day were diluted. starting at step 2 of protocol. All cultures were exposed to aTc and received no IPTG or 1mM IPTG. Samples were collected from each recovery culture for analysis. As noted, the experiment was performed in a branched manner, in that a single culture from a previous day was used to inoculate two daughter cultures (one receiving IPTG inducer, one not).

For the 10 day temporal recording experiment, 8 exposure profiles were randomly generated and conducted in a similar manner over the course of 10 days (1010001010, 1001011001, 1001010101, 0111111001, 0101011010, 0100110110, 0100101010, 0001100010; 1 indicates induction and 0 indicates no induction) and samples were collected from d4 to d10. The experiment was also performed in a branching manner as above; therefore given that the starting substring of some samples were shared, some shorter time points had less than 8 samples (d4-d5:6, d6:7, d7-d10:8).

Experimental conditions (multiplexed recording)

To generate barcoded strains with the three additional sensors for the multiplexed recording experiment, 100µL of the BL21 LacI_tKO with mutagenized DR cryostock was re-inoculated into an overnight culture of LB with no antibiotics. The appropriate pRec and pTrig plasmids for the TreR and GalS sensors (table S2) were transformed into this population via electroporation. Colonies were then picked and screened for mutated DR sequence via Sanger sequencing. This yielded mutated DR sequences for TreR (ATGGTCC, underline denotes altered sequence from WT) and GalS (ACATCAG). We note that the GalS strain also contained a mutation in the first basepair of the first native genomic spacer (G to A) due to inadvertent targeting; however, this did not affect analysis given thresholds utilized in matching during sequencing analysis. The TreR background strain is referred to as BL21 LacI_tKO DR_mut_1 and the GalS background strain BL21 LacI_tKO DR_mut_2. The plasmids for the CopA sensor (table S2) were separately transformed into *E. coli* BL21. The three sensor strains were then grown separately in filter sterilized M9 media with appropriate antibiotics (1X M9 salts [BD 248510], 0.8% (wt/vol) glycerol [Fisher G33-1], 0.2% (wt/vol) casamino acids [BD 223120], 2 mM MgSO4 [Sigma-Aldrich 230391], 0.1 mM $CaCl_2$ [Sigma-Aldrich C1016]) and a cryostock (15% glycerol) was created for storage at −80 °C.

The general experimental workflow followed the temporal recording induction protocol with minor modification. All multiplexed recordings were conducted in M9 media. The three strains were grown overnight separately, optical density was measured, and the three

strains were pooled at equal densities. The initial dilution (step 2) was 1:10 rather than 1:100 given slower growth rate in M9 media compared to LB. Before recovery (step 4), cells were spun down (15,0000 rpm, 30s), media was removed and cells were resuspended in 1mL of fresh media to remove any residual inducer. Inducers for the three sensors were as follows, CopA: 100μM copper sulfate (Sigma-Aldrich 209198), TreR: 1 mM trehalose (Sigma-Aldrich T9531), GalS: 1 mM fucose (Sigma-Aldrich F8150).

qPCR assay for pTrig copy number

A qPCR plasmid copy number assay was utilized to assay pTrig copy number. Briefly, 18 µL of a qPCR master mix (10 µL 2X KAPA SYBR Fast qPCR Master Mix [KAPA KK4601], 0.6 µL 10 µM forward primer, 0.6 µL 10 µM reverse primer, 6.8 µL nuclease free water) was dispensed into a 96 well qPCR plate (Bio-Rad HSL9905) and 2 µL of template as prepared during sequencing library preparation (see protocol below) was added. Two qPCRs were performed, the first with primers targeting pTrig and the second with primers targeting the genome (derived from (*39*), see table S3 for sequences). Both primer pairs were confirmed to have >90% amplification efficiency. The PCR plates were sealed with optically transparent film (Bio-Rad MSB1001) and were placed on a qPCR system (Bio-Rad CFX96) and subjected to following cycling conditions: 95 °C 3 min, 39 cycles: 95 °C 3 s, 60 °C 20 s, 72 °C 1 s and acquisition. The Cq values were determined via the manufacturer's software, and pTrig relative enrichment was calculated with the delta delta Cq method (i.e. 2^(-1*(pTrig_Cq – 16S_Cq)), normalized to the lowest value). A melt curve was performed to ensure that only a single amplification product was present.

Design of custom CRISPR array sequencing scheme

Our custom sequencing scheme enabled highly efficient use of Illumina read lengths (up to 5 expanded spacers with a 300 cycle sequencing kit) by avoiding re-sequencing of primer sequences as required with most two-step amplification schemes. To design these primers for CRISPR BL21 sequencing (referred to as "CB"), we utilized a forward primer targeting the BL21 array I leader sequence and a reverse primer targeting the array I first native genomic spacer. The forward primer was linked to an Illumina P5 sequence and barcode sequence; we generated a series of 8 (i.e. CB501-CB508). The reverse primer was linked to an Illumina P7 sequence and barcode sequence; we generated a series of 12 (i.e. CB701-CB712). All barcode sequences were derived from Illumina Nextera indices. The combination of 8x12 primers allows for 96 samples to be uniquely barcoded via dual indexing in a single sequencing run. We generated custom read 1 (CBR1) and index 1 (CBI1) sequencing primers. All primer sequences can be found in table S4. All primers in this study were obtained from IDT with normal desalting purification.

CRISPR array sequencing library preparation protocol

To perform sequencing of CRISPR arrays from populations of cells, we developed a library preparation and sequencing pipeline consisting of three steps: (1) gDNA preparation, (2) PCR amplification and (3) sample pooling, purification, and quality control.

To purify gDNA from cell pellets obtained at the end of an experiment, we developed a modified protocol utilizing the prepGEM Bacteria kit (ZyGEM PBA0500; VWR 95044-082). Cell pellets were removed from storage at −20 °C in 1.5 mL tubes and resuspended in 100 µL of TE (10 mM Tris-HCl pH 8.0 [Fisher BP1758], 1 mM EDTA pH 8.0 [Sigma-Aldrich 03690] in nuclease free water [Ambion AM9937]). 10 µL of the resulting suspension was pipetted into a 96-well skirted PCR plate (Eppendorf 951020401). 20 µL of a prepGEM master mix (0.30 µL prepGEM enzyme; 0.30 µL lysozyme enzyme, 3.0 µL 10X Green Buffer, 16.4 µL nuclease free water) was then added to each well with a multichannel pipette, and the plate was heat sealed (Vitl V901004 and Vitl V902001). The plate was then spun down for 30 seconds on a plate microfuge (Axygen C1000-AXY) and incubated on a PCR thermocycler (Bio-Rad S1000) with the following program: 37 °C 15 min, 75 °C 15 min, 95 °C 15 min, 4 °C infinite. 70 µL of TW (10 mM Tris-HCl pH 8.0 in nuclease free water) was then added to each well with a multichannel pipette.

To prepare uniquely barcoded amplicons for each sample, we performed PCR amplification using the CB50X and CB7XX sequencing primers (table S4). First, a master primer-plate was prepared by arraying the CB50X primers across rows of a 96-well PCR plate and CB7XX primers down columns of the same 96-well PCR plate at a final concentration of 10 µM for each primer in 50 µL. Thus, each well contained a unique combination of CB50X and CB7XX primers. A PCR reaction was then set up for each sample by pipetting 2 µL of the mix from the master primer-plate, 5 µL of gDNA from prepared genomic DNA plate and 13 µL of a PCR master-mix (10 µL NEB Next Q5 Hot Start HiFi PCR Master Mix [NEB M0543L], 2.96 µL nuclease free water, 0.04 µL SYBR Green I 100X [1:100 dilution in nuclease free water of 10,000X SYBR Green I concentrate, ThermoFisher S7567]) into a new 96-well PCR plate. Alongside each set of samples, a no template control (NTC) was performed as a quality control measure utilizing nuclease-free water rather than gDNA as template. The plate was sealed with optically transparent film (Bio-Rad MSB1001), spun down for 30 seconds on a plate microfuge, placed on a qPCR system (Bio-Rad CFX96), and the following PCR program was performed: 98 °C 30 s, 29 cycles: 98 °C 10 s, 65 °C 75 s, 65 °C 5 min, 4 °C infinite. Amplification was observed and stopped while samples remained in exponential amplification (typically 12-15 cycles).

To perform pooling and quality control of the resulting sample amplicons, representative samples and the NTC were assessed on a 2% E-Gel (ThermoFisher G402002 and G6465) for presence of the expected product (164 bp unexpanded CRISPR array product, and expanded products; each new spacer expansion results in addition of ~61 bp) and no observable product in the NTC. Next, a SYBR Green I plate assay was performed to quantify the relative concentration of amplicon present in each PCR product. Concentrated 10,000X SYBR Green I stock was diluted to a final concentration of 1X in TE, and 198 µL was pipetted with a multichannel pipette into wells of a black optically transparent 96 well plate (ThermoFisher 165305). 2 µL of PCR product was added to each well, and the plate was allowed to incubate in a dark location for 10 minutes. Fluorescence of each well (excitation: 485 nm, emission: 535 nm) was measured on a microplate reader (Tecan Infinite F200), and fluorescence values for individual samples were background subtracted with the fluorescence value of the NTC to control for presence of primers in each PCR. Using this background subtracted fluorescence value, samples were pooled using a Biomek

4000 robot such that equal arbitrary fluorescence units of each sample were present in the final pool.

To remove primers from the pooled product in a manner that did not affect abundance of different amplicon products, the pool was then subjected to gel electrophoresis (2% agarose gel, 100 V) and gel extracted (Promega A9282) from size ranges ~150 bp to ~1 kb, and eluted in 30 µL TW in an LoBind tube (Eppendorf 022431021). The amount of DNA present in purified pool was quantified (Qubit dsDNA HS Assay Kit, ThermoFisher Q32854 with Qubit 3.0 Fluorometer, ThermoFisher Q33216) with at least two replicates performed with different pipettes and the average fragment size was quantified on an Agilent Bioanalyzer 2100 with Bioanalyzer High Sensitivity DNA kit (Agilent 5067-4626). The molar concentration of the pool was determined with use of Qubit fluorometric quantification and Bioanalyzer size determination.

Size-enrichment of CRISPR array libraries

For selected libraries, a size-enrichment protocol was performed to enrich for expanded arrays and deplete unexpanded arrays. We utilized SPRI bead-based size selection with AMPureXP beads (Beckman Coulter A63881); altering the ratio of AMPureXP added to a particular sample can allow for size selection of a particular library. Rather than performing gel extraction as in the normal library preparation protocol, pooled PCR products were subject to two AmpureXP cleanups with 0.75X ratio of AmpureXP beads to volume of PCR product. These cleanups were performed as per the manufacturer's recommendations with minor modifications: 80% ethanol rather than 70% ethanol, elution into 33 µL TW and removal of 30 µL (to reduce carryover of beads).

We found that the resulting libraries displayed enrichment of larger DNA products which did not appear to be CRISPR arrays and were presumably plasmid or degraded genomic DNA carrying through from the template. This did not alter quality of the resulting library, but to better assess concentration of the library, a qPCR quantification (NEB E7630L) was utilized in addition to fluorometric quantification.

Sequencing CRISPR array libraries

Sequencing was performed on the Illumina MiSeq platform (reagent kits: V3 150 cycle, V2 300 cycle, Micro V2 300 cycle depending on the experiment). All runs included at least a 20% PhiX spike-in (PhiX Sequencing Control V3) which was necessary for run completion given relatively low sequence diversity and variable amplicon size. For V3 kits, samples were loaded at 15 pM final concentration, while for V2 kits samples were loaded at 10-12 pM final concentration following the manufacturer's instructions with the following modifications. First, to spike in custom sequencing primers, 6 µL of a 100 µM stock of the CBR1 primer (table S4) was spiked into well 12 of the reagent cartridge utilizing an extended length tip (Rainin RT-L200XF). Similarly, 6 µL of a 100 µM stock of the CBI1 primer (table S4) was spiked into well 13 of the reagent cartridge. This spike-in procedure was necessary (rather than utilizing custom primer wells) to allow for the PhiX control to be sequenced with primers already present in the standard primer wells (*40*). Second, we note that significant amounts of sample may be retained in the sample

loading line from run to run (*Illumina Technical Support Note: Reducing Run-to-Run Carryover on the MiSeq Using Dilute Sodium Hypochlorite Solution*), which may result in contamination of samples indexed with similar barcodes. Therefore, after every run we performed an optional template line wash, and where possible utilized unique barcodes for adjacent runs.

CRISPR spacer extraction and mapping from sequencing data

Raw sequencing reads were analyzed with a custom Python analysis pipeline. Code utilized for sequencing analysis can be found at https://github.com/ravisheth/trace. Briefly, the pipeline comprised the following steps: (1) raw reads were subjected to spacer extraction, (2) extracted spacers were then mapped against genome and plasmid references to determine their origin, (3) uniquely mapping spacers were determined from mapping results.

To extract spacers (*spacer_extraction.py*), we began with raw reads (given the low error rates of the Illumina platform, and highly structured nature of sequences, we found filtering of raw sequences to be unnecessary). For each read, we checked the beginning 12 bp of the read to ensure that this matched the expected DR sequence. If this criterion was passed, the DR sequence was stripped from the 5′ of read and the remaining sequence was passed into a spacer extraction loop. First, the 5′ of the remaining read sequence was compared to the native genomic first spacer sequence (i.e. end of potential newly acquired spacers); if a match was found we considered the read terminated and recorded any spacers extracted, or that the array was unexpanded if no spacers were extracted. If the sequence did not match, we attempted to find a DR sequence given different possible spacer lengths, in this case 32-34 bp (*19*). If a DR sequence was identified, the spacer was extracted, the spacer and DR sequence were stripped from the 5′ of the read, and the extraction loop was repeated for the remaining sequence. For sequencing runs with 150-159 bp read length, we utilized the full DR sequence during matching, which enabled extraction of up to two new spacers. However, for sequencing runs with 309 bp read length (i.e. maximum possible with 300 cycle reagent kit), only 15 bp of the 5′ of the DR sequence was utilized for matching given read length constraints (using full length DR sequences would only allow for extraction of 4 new spacers). For all multiplexed temporal recordings, the full length DR sequence was utilized to enable differentiation of DR sequences. This extraction routine allowed for high efficiency read extraction (for example, on average >97% of all reads could be extracted without error for each sample).

To map spacers against reference (*blast_search.sh*), the extracted spacers were searched against reference databases of the genome (NCBI GenBank CP001509.3) and plasmids (as appropriate given the sample) using NCBI BLAST 2.6.0 (*41*). Extracted spacer files generated by the extraction pipeline were passed to the *blastn* command, using the flag -evalue 0.0001 to threshold spurious mapping results (*21*) .

Finally, the resulting BLAST output files were analyzed and spacers mapping to only one reference were determined (*unique_spacers.py*). This was necessary given that the plasmids may share sequence homology with the reference genome. The resulting uniquely mapping spacers were saved to an output file for further analysis. For analysis of array

types frequencies, only arrays with all spacers uniquely mapping to one reference were analyzed.

Model of CRISPR array expansion and reconstruction of temporal input profiles

We utilized a simple model of CRISPR expansion. We consider a population of CRISPR arrays that undergoes an expansion process during each round of induction. The parameters governing the expansion process are dependent on the identity of the round (if pTrig is activated or not). Specifically:

- Each array can undergo expansion with probability p_{exp}. The acquired spacer can be:
    - A trigger spacer with probability p_T
    - A reference spacer with probability $p_R = 1 - p_T$
- The probability of an array not undergoing expansion is $1 - p_{exp}$

Therefore, for each state (0: no pTrig activation; 1: pTrig activation), two parameters govern the expansion process (p_{exp}, p_T) for a total of four parameters ($p_{exp,0}$, $p_{T,0}$, $p_{exp,1}$, $p_{T,1}$) governing the entire model. To determine these parameters, we utilized control experiments as well as the "1111" and "0000" samples; all model parameters can be found in table S5. To calculate $p_{exp,0}$ and $p_{exp,1}$, the average proportion of singly expanded arrays after a single round of induction (with and without pTrig activation) was determined from control experiments. To calculate $p_{T,0}$, we used the average pTrig incorporation rate across all array lengths and positions (L1 to L5, p1 to p5) from the "0000" sample. To calculate $p_{T,1}$, we similarly utilized pTrig incorporation frequencies from the "1111" sample. However, the pTrig incorporation rate appeared to decrease with array length; this is likely due to the fact that CRISPR expansion precedes full pTrig activation in our experimental scheme, resulting in highly expanded arrays containing a lower proportion of pTrig spacers (fig. S9). To account for these differences, we parameterized an "apparent pTrig incorporation rate" for different array lengths based on the "1111" sample by calculating the average pTrig incorporation at each array length. When simulating expected array-type frequencies for different array lengths, the corresponding $p_{T,1}$ for that array length was utilized (i.e. $p_{T,1}^{L1}$ to $p_{T,1}^{L5}$).

We then calculated predicted array-type frequencies given a particular temporal input profile and parameterized model. Specifically, all possible array-types were enumerated for a given array-length. We calculated the probability of generating each array-type by enumerating all possible incorporation patterns leading to the array-type (i.e. an array of length 2 during a 3 day temporal input pattern could result from expansion on days {1,2}, {2,3}, or {1,3}) and then analytically calculated the sum of the probabilities of each incorporation pattern. This value was treated as the "global" array-type probability. After all array-type probabilities were calculated, the "global" probabilities for all array-types of a particular length were normalized to unity, resulting in the final predicted array-type frequency vector.

As an example of the model, for a single day of induction (state =1), the probability of an array containing an expanded spacer derived from pTrig (i.e. L1 array, T) is simply $p_{exp,1} * p_{T,1}^{L1}$. For one day of induction followed by one day of no induction (state = 10) the probability of an array containing two expanded spacers derived from pTrig (i.e. L2 array, TT) is simply $(p_{exp,1} * p_{T,1}^{L2}) * (p_{exp,0} * p_{T,0})$. For three days of induction (state = 111) the probability of an array containing two expanded spacers, one derived from the genome and the next derived from pTrig (i.e. L2 array, RT) is the sum of all incorporation patterns leading to RT arrays (incorporation on days {1,2}, {2,3}, {1,3}) or:

$$[p_{exp,1} * (1 - p_{T,1}^{L2})] * (p_{exp,1} * p_{T,1}^{L2}) * (1 - p_{exp,1}) +$$
$$(1 - p_{exp,1}) * [p_{exp,1} * (1 - p_{T,1}^{L2})] * (p_{exp,1} * p_{T,1}^{L2}) +$$
$$[p_{exp,1} * (1 - p_{T,1}^{L2})] * (1 - p_{exp,1}) * (p_{exp,1} * p_{T,1}^{L2})$$

Array type frequencies can be calculated for any input profile and array-type in a similar manner.

We used the array-type frequencies calculated from the model to classify the observed data. The Euclidean distance between observed array-type frequencies and predicted array-type frequencies was calculated, and the model with minimum distance to the observed data was selected as the predicted temporal input. This procedure can be repeated for different array lengths. To consider multiple array lengths simultaneously, aggregate array-type vectors were constructed by concatenating array-type vectors of different array lengths of interest (both observed and model) and the same procedure was used to calculate distance and predict temporal inputs.

Population lineage reconstruction using CRISPR array information

To perform lineage reconstruction, we identified genomic spacers within L1 arrays for the 16 4-day temporal recording samples (pooled from enriched and unenriched samples). Genomic spacers were utilized as they contain the highest sequence diversity, and L1 arrays were utilized given that they were observed with the highest frequencies in populations. These spacers were randomly subsampled for each sample to the minimum number of spacers detected (14,715). The location that each spacer mapped to on the reference genome was utilized as the identity of the spacer; the Jaccard distance between two samples (i.e. 1 – proportion of unique spacers in a sample shared with another sample) was calculated for all samples in a pairwise fashion. This 16x16 distance matrix was then utilized for lineage reconstruction using the Fitch-Margoliash method (*42*). Specifically, an tool implementing the PHYLIP program was utilized with default settings (http://www.trex.uqam.ca/index.php?action=phylip&app=fitch).

Multiplexed recording analysis and reconstruction

For all multiplexed temporal recordings, the full length DR sequence was utilized to enable differentiation of DR sequences. Given the strict criteria for DR matching utilized (no more than Hamming distance 2), this allowed for extraction of individual sensors from the CRISPR array populations.

Models were parameterized for each of the three sensors independently. Expansion rates in the absence and presence of signal ($p_{exp,0}$ and $p_{exp,1}$) were calculated as the average proportion of singly expanded arrays after 1 day for no input and input of all three chemicals (C,T,F) and the same value was utilized for all three sensors. pTrig incorporation rates in the absence of input ($p_{T,0}$) were calculated for each sensor from profile #1 (i.e. no input throughout the recording) as the average of pTrig spacers at all positions within L1 to L3 arrays. pTrig incorporation rates in the presence of input ($p_{T,1}^{L2}$, $p_{T,1}^{L3}$) were calculated for each sensor in a similar manner from profile #2 for L2 and L3 arrays separately. For the CopA sensor, we observed that pTrig spacer incorporation was higher when other inducers (T, F) were both present compared to other conditions. Therefore, the pTrig incorporation rate in the presence of input was calculate from profile #6, where the copper was present for three days but other inducers varied. All parameters utilized can be found in table S5.

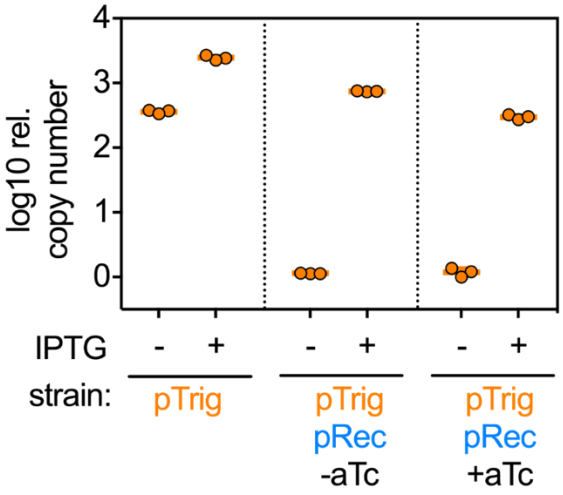

Fig. S1: pTrig copy number induction. To assess pTrig copy number increase in the context of recording, we measured pTrig copy number by qPCR. Cells with only pTrig displayed high copy number in the absence of inducer and low fold increase in copy number, since only genomic expression of LacI was present to repress the Lac promoter upstream of RepL on pTrig. The addition of pRec (which expresses LacI) resulted in repression of copy number in the absence of inducer and high fold increase in copy number after induction. Addition of aTc slightly decreased pTrig copy number during induction, indicating that Cas1 and Cas2 expression may reduce the apparent copy number of pTrig, for example by degradation.

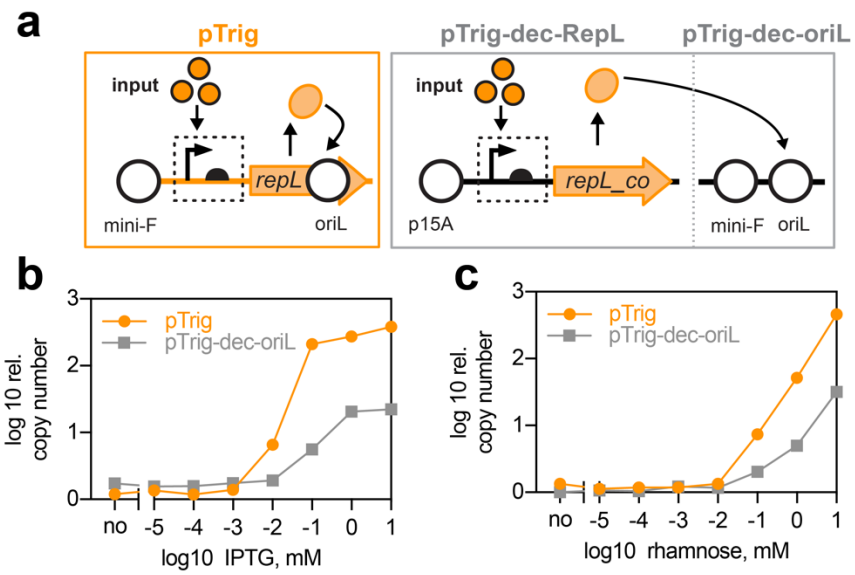

Fig. S2: Decoupling pTrig copy number induction. (a) To better understand the mechanism of pTrig copy number induction, RepL expression was decoupled from the amplifying effects of pTrig copy number increase. RepL was codon optimized for *E. coli* (to remove the origin of replication located within the RepL coding sequence) and placed on a p15A plasmid (pTrig-dec-RepL). The pTrig plasmid was then modified to remove the upstream promoter and first 100 bp of the RepL coding sequence, and a terminator (L3S1P52) was placed immediately upstream (to retain the RepL oriL origin of replication but eliminate expression; pTrig-dec-oriL). **(b)** The Lac promoter along with RiboJ and B0034 RBS was placed upstream of the RepL (pTrig-dec-RepL-Lac), and the decoupled system (pTrig-dec-RepL-Lac + pTrig-dec-oriL + pRec) was exposed to aTc and varying concentrations of IPTG for 6 hours alongside the pTrig system utilized in the main text. Plasmid copy number of pTrig or pTrig-dec-oriL was then measured by qPCR. The decoupled system displayed reduced range in copy induction and lower sensitivity to input compared to pTrig, suggesting that a positive feedback loop may mediate induction of the pTrig system. **(c)** To assess generality of the result, the experiment was repeated with a second inducible promoter. A rhamnose inducible promoter was swapped into the pTrig system (pTrig-Rha, 150 bp upstream sequence of *E. coli* RhaB, see also fig. S14). The same promoter with the addition of RiboJ and B0034 RBS was swapped into the RepL expression plasmid (pTrig-dec-RepL-Rha), and the decoupled system (pTrig-dec-RepL-Lac + pTrig-dec-oriL + pRec ΔLacI) was compared to the pTrig system (pTrig-Rha + pRec ΔLacI) as in (a) with aTc and varying concentrations of rhamnose inducer for 6 hours. A similar reduced copy number induction range and lower input sensitivity in the decoupled system compared to the pTrig system was also observed.

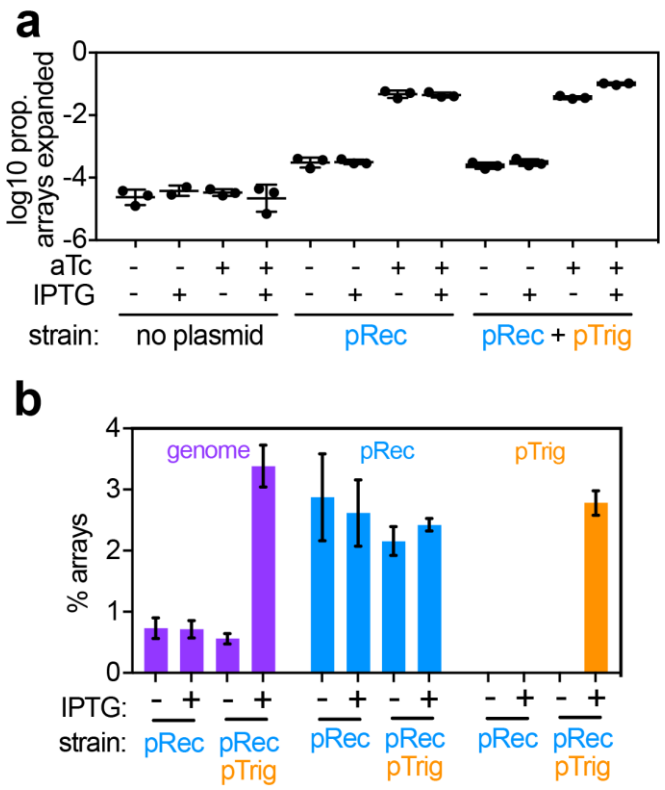

Fig. S3: CRISPR spacer acquisition. (a) CRISPR expansion, calculated as the log10 proportion of arrays detected as expanded, was assessed over the course of a single recording round. As a control, a strain harboring no plasmids was first tested. We detected a very low amount of expansion, presumably from index swapping between samples that can occur at background levels on the Illumina sequencing platform. For one of the no plasmid samples receiving only IPTG inducer, no expanded spacers were detected, therefore this replicate not plotted. Addition of pRec increased CRISPR expansion above background levels, likely due to leaky expression of Cas1 and Cas2; addition of aTc inducer greatly increased CRISPR expansion. The addition of pTrig did not affect CRISPR expansion without copy number induction by IPTG, but overall expansion increased when IPTG was added. **(b)** An alternative visualization of Fig. 1g. With the pTrig plasmid in the presence of IPTG, pTrig spacer acquisition greatly increases (other pTrig bars are not present as they are too small to be visualized on this Y-axis scale). pTrig induction did not appear to affect pRec spacer acquisition, but increased genomic spacer acquisition, indicating that pTrig copy number increase may interact with genomic replication or spacer acquisition processes. Error bars represent standard deviation of three biological replicates.

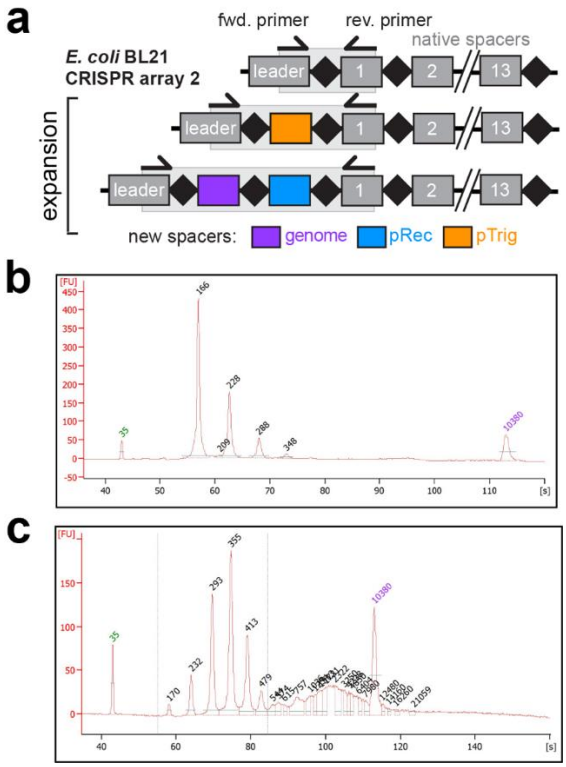

Fig. S4: CRISPR array sequencing. (a) Schematic of custom CRISPR amplicon sequencing approach. **(b)** Example library size distribution by Bioanalyzer HS DNA assay; the smallest product (~166 bp) corresponds to unexpanded arrays, and products of larger sizes, ~228, ~288, ~348 bp, correspond to expanded arrays of 1, 2, 3 spacers respectively. **(c)** Size-enriched library size distribution determined in the same manner as (b); expanded arrays are enriched. Contaminating high molecular weight DNA was observed (presumably plasmid or genomic DNA from the PCR template) but did not affect sequencing.

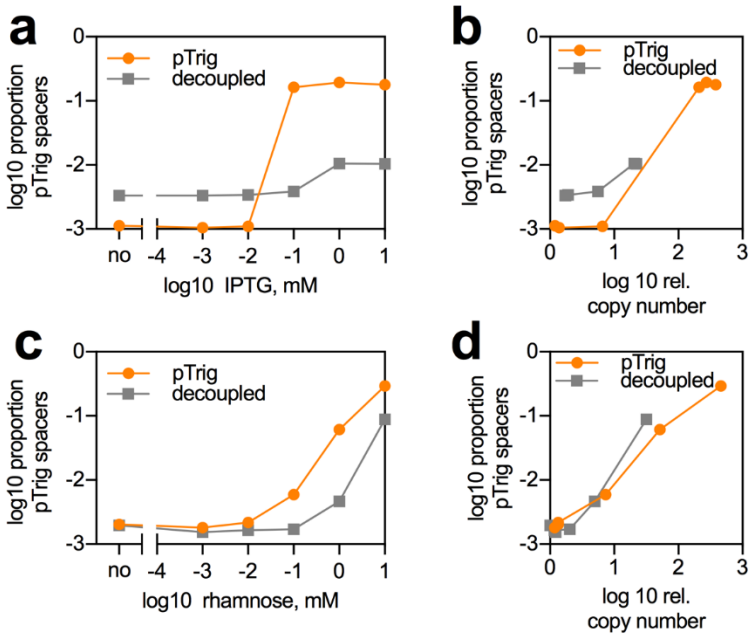

Fig. S5: Relationship between pTrig copy number and pTrig spacer incorporation. For the same experiment shown in fig. S2, cells were recovered and subjected to CRISPR array sequencing. **(a)** For the Lac pTrig and decoupled system, the resulting proportion of pTrig spacers is displayed in log 10 scale. **(b)** Spacer incorporation was directly compared to measured pTrig or pTrig-dec-oriL copy number (shown in fig. S2), and displayed an increasing relationship as expected. **(c,d)** The same data is shown for the Rha pTrig and decoupled system.

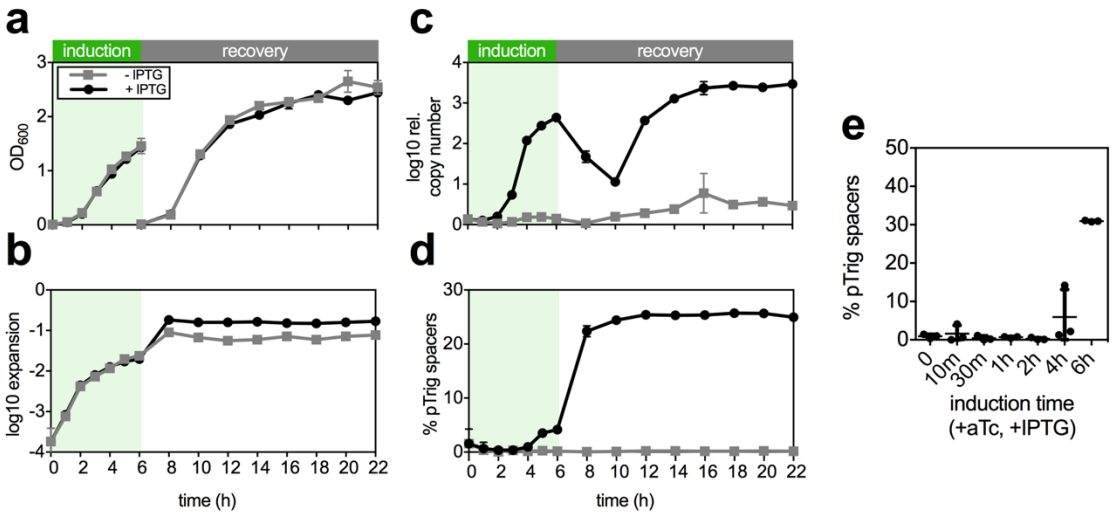

Fig. S6: CRISPR expansion and pTrig incorporation over a single induction round. We tracked culture growth, pTrig copy number and spacer acquisition over the course of induction and recovery to assess response dynamics of the system. Cells received aTc induction and were exposed to no IPTG or IPTG inducer; all points display the mean and standard deviation of three biological replicates. **(a)** Induction of pTrig did not appear to affect cell growth as measured by optical density compared to basal maintenance of the system. **(b)** Array expansion was observed after 1 hour of induction, and a large increase in spacer acquisition was observed during recovery. **(c)** pTrig copy number as measured by qPCR displayed an increase beginning 3 hours after induction. In addition, copy number increased during the recovery period only when cells had been previously induced, likely due to residual IPTG inducer in recovery media. Further dilution on the subsequent day prevents this re-activation from interfering with multi-day recordings. **(d)** The percentage of pTrig spacers appeared to increase after 4 hours, consistent with pTrig copy number dynamics. **(e)** The duration of induction (with aTc and IPTG) was varied between 0 to 6 hours and the recovery time was adjusted such that all samples were collected at the same time. Robust recording required the full 6 hours of induction.

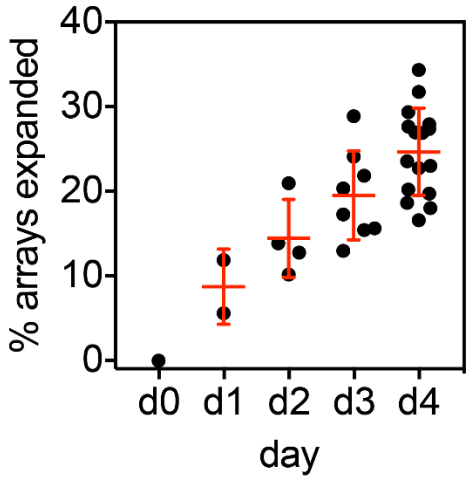

Fig. S7: CRISPR array expansion over multiple days. Samples from intermediate states (d1, d2, d3) were sequenced in addition to d4. The percent of CRISPR arrays detected as expanded in each sample is plotted; increasing array expansion was observed over the course of the experiment.

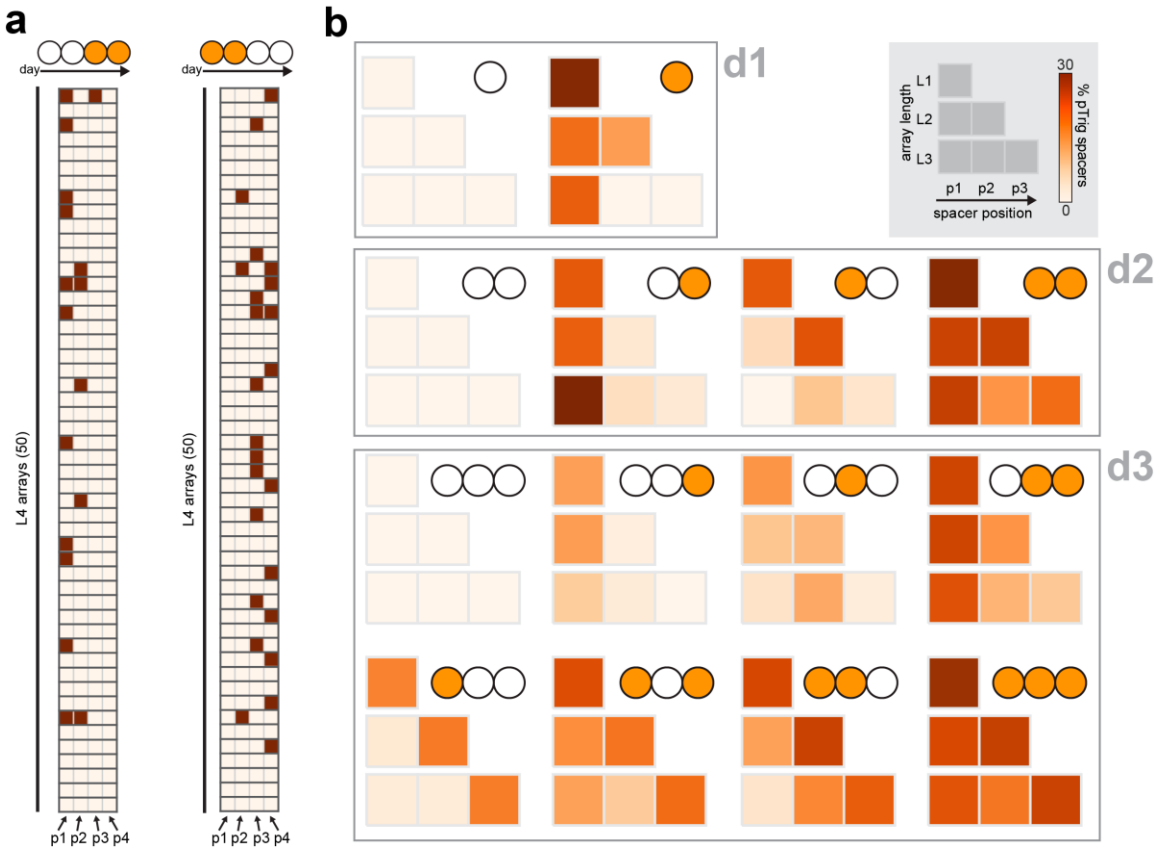

Fig. S8: pTrig spacer incorporation. (a) To visualize the information encoded within individual arrays, 50 L4 arrays sampled from two representative temporal input profiles are shown as a heatmap (as in Fig. 2d) where rows are individual arrays and columns are positions in the array (shaded: pTrig spacer, unshaded: reference spacer). The individual array information can be then visualized as positional averages as shown in the main text (Fig. 2e). **(b)** Samples from d1, d2 and d3 were additionally sequenced. The resulting %pTrig spacers detected for different array lengths (L1 to L3) at different positions (p1 to p3) is plotted as in Fig. 2e. L4 and L5 arrays are omitted as a low number were detected (intermediate samples were sequenced without the enrichment protocol).

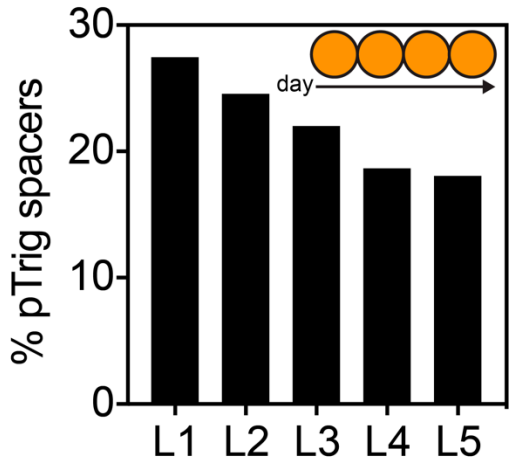

Fig. S9: CRISPR array length-dependence of pTrig incorporation for model parameterization. pTrig incorporation appeared to differ across different array lengths; for example the average percentage of pTrig spacers at each position of each array length for the sample receiving inducer for four days is shown. This is presumably due to the delayed activation of pTrig compared to array expansion; the first incorporations during a recording round are less likely to contain pTrig spacers as pTrig copy number has not yet increased. Therefore, highly expanded arrays may display slightly lower levels of pTrig incorporation. Given this trend, we individually parameterized models for each array-length by empirically using the average percentage of pTrig spacers found in each array length (i.e. using the values above, see table S5).

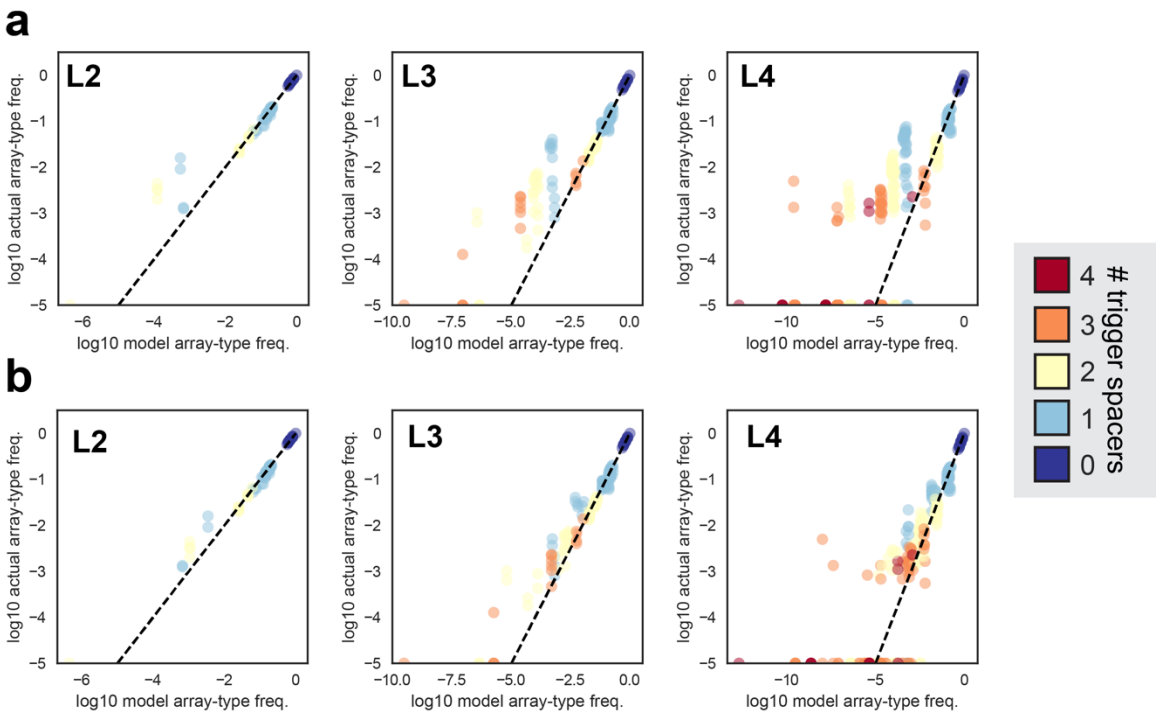

Fig. S10: Observed and predicted array-type frequencies for all four day input profiles. (a) The observed array-type frequencies from experimental data and modeled array-type frequencies are displayed for all 16 input profiles as an aggregate scatter plot, with both axes in log scale. The shading of each point indicates the number of trigger spacers the specific array-type contains. Array-types not observed in a particular sample are plotted on the X axis (i.e. log frequency = -5). Results for different array lengths are shown: L2 arrays, L3 arrays, L4 arrays. A close correspondence between the observed and predicted array-type frequencies is apparent, although a subset of low frequency array-types occur more often than predicted. **(b)** We hypothesized that the model assumption of only up to one expansion per day contributed to the discrepancy between data and model. We altered the model to allow a second expansion for singly expanded arrays at with the same probability as the first expansion but scaled by a constant value. This scaling factor (0.02402) was calculated as the proportion of doubly expanded arrays observed to singly expanded arrays observed from the same control experiment utilized to parameterize expansion rates. The same plots in (a) are shown for this two expansion model, and visually display better model recapitulation of low frequency array-types, suggesting better modeling of the CRISPR expansion process. For the sake of simplicity the single expansion model is still utilized for classification, but we note that more nuanced models of CRISPR expansion could allow for improved reconstruction performance.

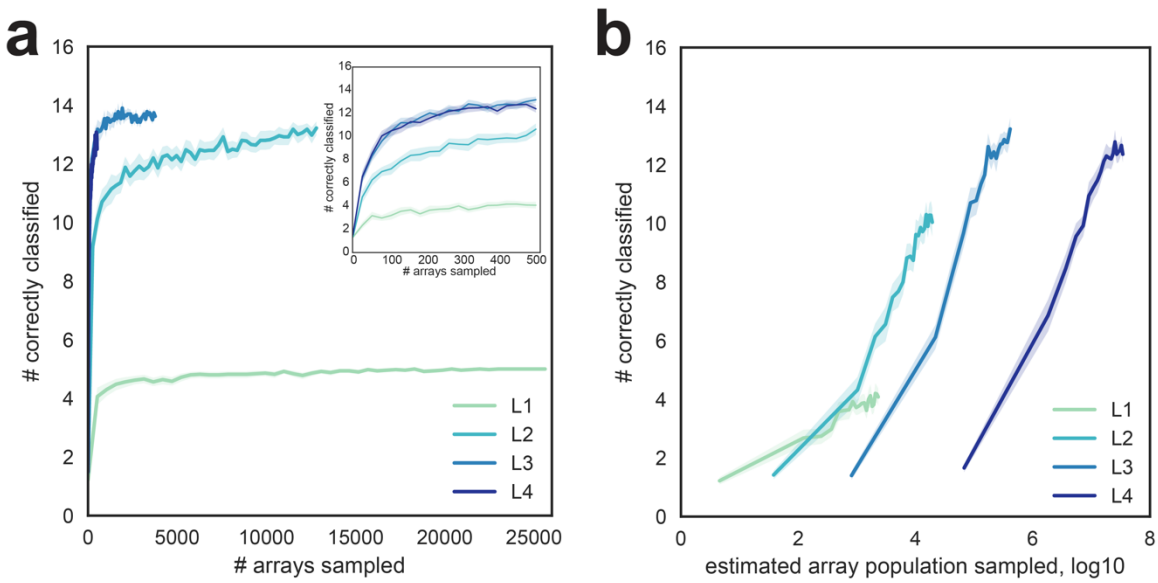

Fig. S11: Number of arrays required for classification of temporal input profiles. (a) For each of the 16 temporal profiles, arrays were subsampled to the minimum number detected for each array length (array lengths L1 to L4; decreasing arrays detected for increasing array lengths) and classification was performed. Lines display the mean and error bands display the 95% confidence interval of 50 iterations of subsampling and classification. In the inset, arrays are subsampled to 508 arrays (the minimum number arrays detected in any sample of any array length). These results demonstrate that only a few hundred arrays of a given length are required for reasonable classification performance. **(b)** The same subsampling and classification accuracy analysis was replotted with an estimate of the total array population required (log10 scale), rather than the number of arrays of a given length as in (a). Specifically, the x-axis was rescaled utilizing the average proportion of arrays of a given length (L1 to L4) observed across the 16 temporal profile samples sequenced without size enrichment (Fig. 2b, dotted lines). These results demonstrate that a population of ~10^5 arrays (using L3 arrays for classification) can recapitulate reasonable accuracy (~75% or 12/16 correctly classified).

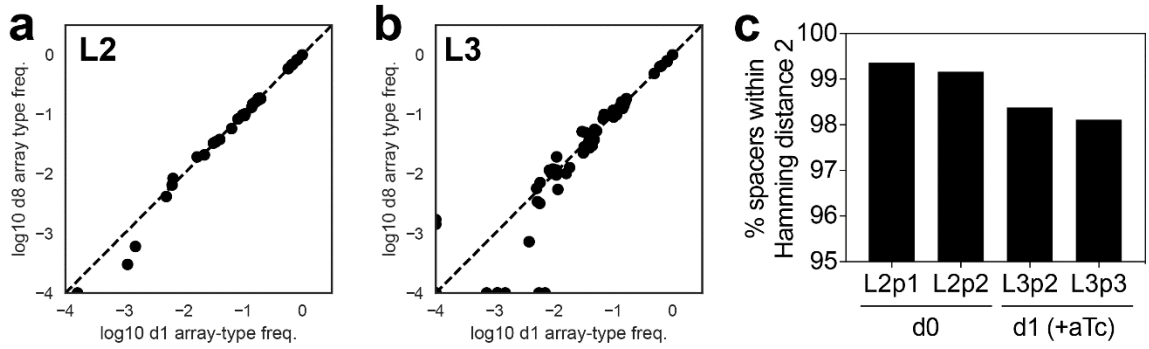

Fig. S12: Stability of TRACE recordings. Cell populations were subjected to 3 day temporal recording, resulting in 8 temporal profiles. The 8 populations were subsequently diluted 1:100 every 24 hours into 3mL fresh LB media with antibiotics for a total of 8 days (~6.6 generations per day, ~50 total generations). Array-type frequencies at d1 and d8 were then compared for all of the 8 profiles in aggregate for **(a)** array length 2 (L2) **(b)** and L3 arrays. Array types not detected in a sample are plotted on the axis (i.e. log frequency = -4). Array type frequencies appeared stable over the course of the experiment, although some low frequency array types exhibited variability likely due to population fluctuations. **(c)** A strain containing an array with two expanded spacers was clonally isolated, and induced with aTc with the same induction protocol utilized for recording. The strain before induction (d0) and after induction (d1) was sequenced. The percentage of extracted L2 spacer sequences within Hamming distance 2 of the actual expected sequence at d0 is displayed and was >99% at each position; other spacers likely represent sequencing errors or background levels of spacer loss. After induction, L3 arrays (i.e. arrays receiving a new spacer) were analyzed; the distal p2 and p3 positions largely contained the expected spacers with a small but measureable loss (~1%) compared to the background rate before induction. In sum, these experiments demonstrate stability of array type frequencies and thus recorded information, and a low rate of loss of previously recorded spacers.

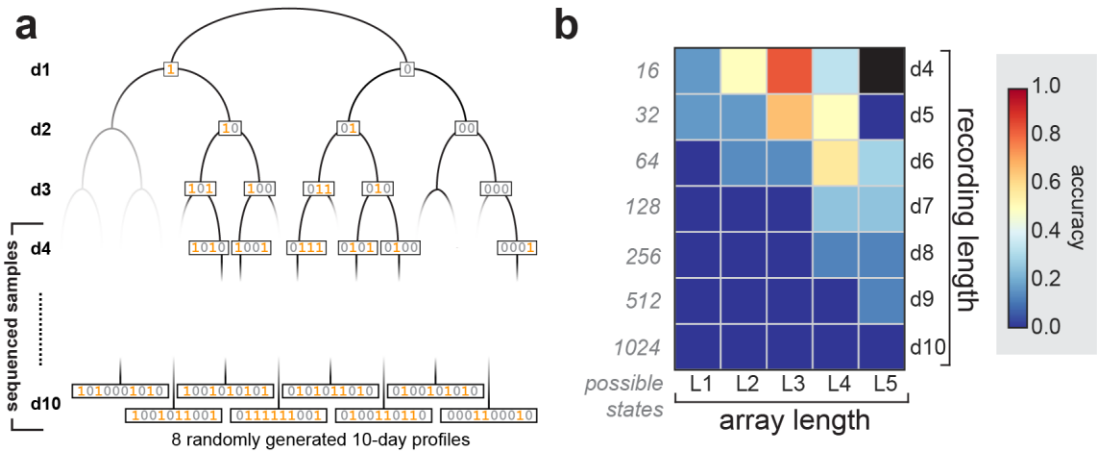

Fig. S13: 10-day temporal recording. A 10 day recording (~150 generations, ~15 generations per day) was performed to assess the limits of long term recording. **(a)** We randomly selected 8 of the 1024 (2^10) possible 10-day temporal input profiles (bottom boxes) and experimentally exposed 8 corresponding lineages to these input profiles in a similar manner to the 4 day experiment, utilizing a branching lineage method. Samples were collected at each time point from d4 to d10 for sequencing; given that some of the early time point substrings were shared between samples, not all early days contained 8 distinct samples (minimum 6 samples each day). Here, input exposures are displayed as a binary string (1 indicates induction and 0 indicates no induction) for clarity. **(b)** The data was then classified against models of all potential profiles and for array lengths L1-L5. We could obtain reasonable reconstruction accuracies up to d6 (L4 arrays: 4/7 tested correct, 1/64 expected by random guessing). In addition, arrays with more spacers appeared to enable better classification of input profiles of longer duration.

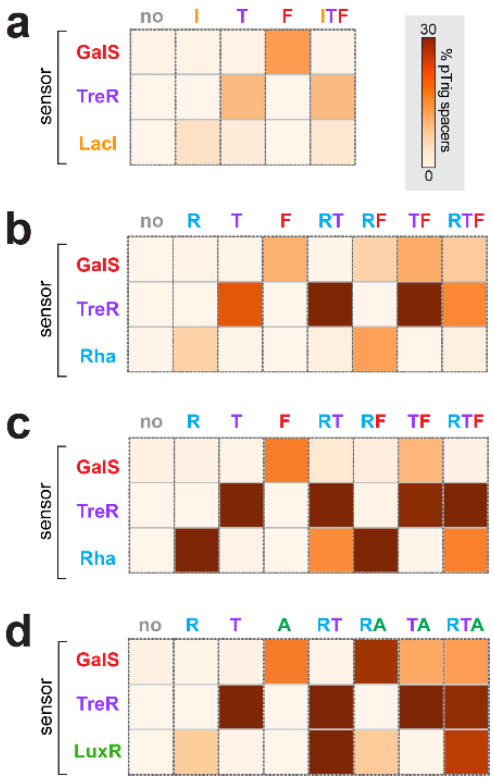

Fig. S14: Screening for orthogonal TRACE sensor systems. In preliminary experiments, we screened sensors and exposure conditions to identify three sensors displaying orthogonal function. In total, we demonstrated functionality of multi-channel sensing with six distinct sensing systems. **(a)** We first utilized the GalS and TreR sensor strains alongside the LacI sensor system. We found that each strain responded to its cognate inducer, however the GalS sensor displayed inactivation in the presence of IPTG and trehalose. A previous study has reported inactivation of the GalS sensor in response to IPTG, consistent with this result (*43*). **(b)** We constructed a rhamnose sensor, consisting of the 150 bp upstream sequence of *E. coli* RhaB swapped in place of the Lac promoter pTrig (pTrig-Rha); no transcription factor overexpression was utilized for this sensor system. This sensor was tested alongside the GalS and TreR sensor strains containing barcoded DR sequences. Populations of cells were exposed to combinatorial inputs (as in the main text; 1 mM rhamnose [R] was used as inducer for the Rha sensor). Again, we observed cognate response of each sensor to its ligand; however, we observed inactivation of the Rha sensor in the presence of trehalose. **(c)** We repeated the experiment in (b), but utilized 10 mM rhamnose rather than 1 mM rhamnose in an attempt to avoid trehalose inactivation of the Rha sensor strain. However, with these inducer conditions, we observed inactivation of the GalS sensor in the presence of rhamnose. These results highlight the complex interplay of endogenous sensing systems in *E. coli*, likely reflecting host sugar utilization hierarchies. **(d)** Finally, we constructed a 3OC6-HSL (i.e. AHL) sensor by swapping the D49 promoter (*44*) in place of the Lac promoter (pTrig-D49) and expressing the LuxR transcription factor on a variant of the pRec plasmid (pRec-LuxR). Populations of cells were exposed to combinatorial inputs (as in the main text; 100nM 3OC6-HSL [A] was used as inducer for the LuxR sensor). Each sensor displayed a response only to its cognate input.

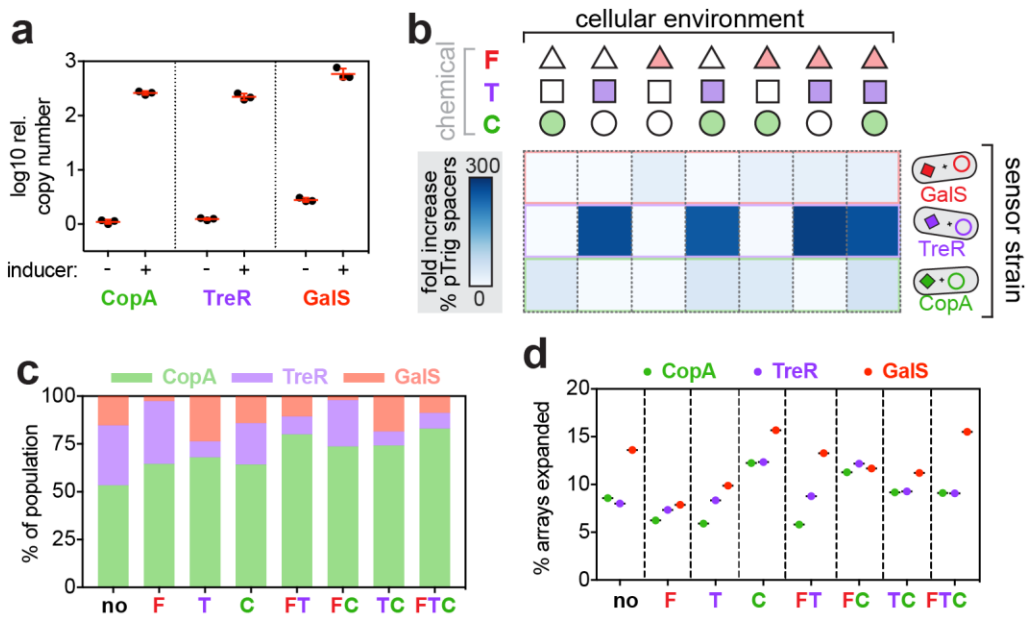

Fig. S15: Multi-channel recording with the TRACE system. (a) pTrig copy number characterization by qPCR for each of the three sensing systems individually exposed to their cognate input for 6 hours. **(b)** Fold increase (linear scale) in the percentage of pTrig spacers after recording for 7 input conditions compared to no inducer; all systems display >24 fold increase (each value displays the average of three biological replicates). The TreR sensor displays a higher fold increase compared to the two other systems. **(c)** The frequency of each of the three barcoded CRISPR arrays after the 8 inducer input exposures. All sensors are detected in each of conditions although with differing frequencies, suggesting that subtle fitness differences between sensor strains and during pTrig activation may result in altered population abundances. **(d)** The percentage of expanded arrays detected for each of the three sensors; the two barcoded arrays (TreR and GalS sensors) display similar expansion to the wild type array (CopA). This demonstrates that the barcoding does not impede the CRISPR expansion process for the two barcode sequences tested.

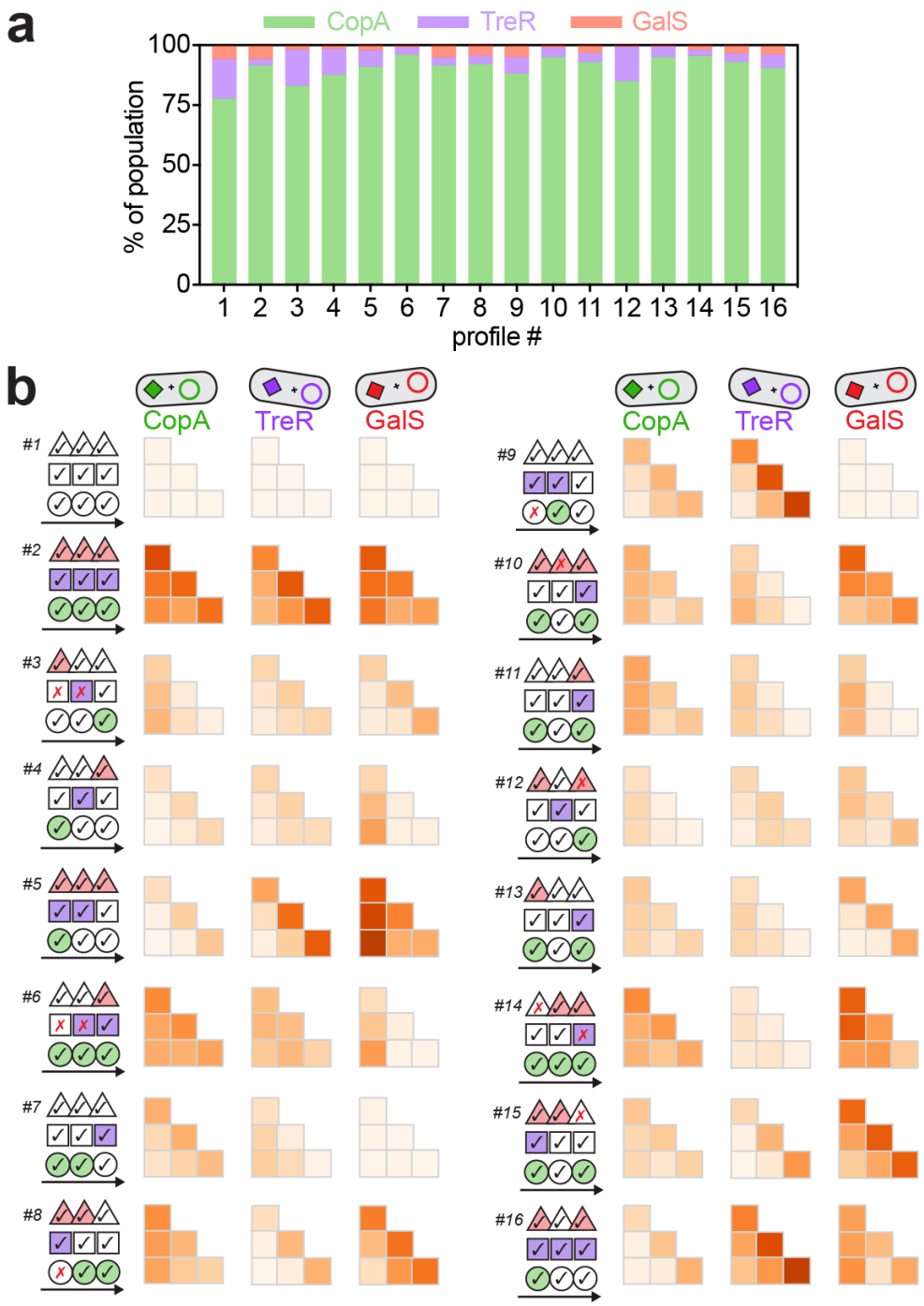

Fig. S16: Population frequencies and pTrig spacer incorporation for the multiplex temporal recording experiment. (a) The final frequency of each of the three barcoded CRISPR arrays at d3 is displayed for all 16 temporal profiles tested; frequencies vary per profile and sensor but all three are detected in each sample at a frequency of at least ~0.4%. **(b)** Average pTrig spacer incorporation for different array lengths (L1 to L3) and positions (p1 to p3) plotted as in Fig. 2e. To aid visualization, the color map for the CopA sensor ranges from 0 to 8%, while the color map for the TreR and GalS sensors ranges from 0 to 30%.

Table S1. Plasmids used in this study.

ref name	plasmid	resistance marker	origin	description
pRS001	pRec	CmR	ColE1	PTet-cas12, tetR, lacI
pRS002	pRec-GalS	CmR	ColE1	PTet-cas12, tetR, lacI-galS
pRS003	pRec-TreR	CmR	ColE1	PTet-cas12, tetR, lacI-treR
pRS004	pRec ΔLacI	CmR	ColE1	PTet-cas12, tetR
pRS005	pTrig	KanR	mini-F	PLac-repL
pRS006	pTrig-CopA	KanR	mini-F	PCopA-RiboJ-B0034-repL

ref name	plasmid	map
pRS001	pRec	https://benchling.com/s/seq-A9McFCX7BXXXI9vSBrRe
pRS002	pRec-GalS	https://benchling.com/s/seq-I8zistPSzTIXMneP5V4h
pRS003	pRec-TreR	https://benchling.com/s/seq-30jc7WJzBGX8fNKZp7Pz
pRS004	pRec ΔLacI	https://benchling.com/s/seq-cv8by55ejdFb4xCZD4d1
pRS005	pTrig	https://benchling.com/s/seq-ISWVXtHWPPuY5zCBNceM
pRS006	pTrig-CopA	https://benchling.com/s/seq-JBz03HXz2h1sDNJ4ob9P

Table S2. Strains used in this study.

strain	plasmid1	plasmid2
BL21		
BL21	pRec	
BL21	pRec	pTrig
BL21	pRec ΔLacI	pTrig-CopA
BL21 LacI_tKO DR_mut_1	pRec-TreR	pTrig
BL21 LacI_tKO DR_mut_2	pRec-GalS	pTrig

Table S3. Primers used in this study.

Primer	sequence (5′-3′)
MAGE_tKO_LacI	G*G*A*A*GAGAGTCAATTCAGGGTGGTGAATGTGAAACCAGTA TAGTGATAAGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTC
MAGE_BL21_DR	G*G*G*GAACACCCGTAAGTGGTTTGAGCGATGATATTTGTGCT NNNNNNNNCCCCGCTGGCGCGGGGAACACTCTAAACATAACCTATTATT
genome_fwd	GCGAGCGATCCAGAAGATCT
genome_rev	GGGTAAAGGATGCCACAGACA
pTrig_fwd	CGCTCTATGATCCAGTCGATTT
pTrig_rev	TCCGTATGCCATGCGTTTAT

For the MAGE_tKO_LacI primer, underlined bases indicate mismatch with genomic LacI sequence. For the MAGE_BL21_DR primer, underlined bases indicate mismatch with genomic sequence designed to barcode individual arrays (note the last N base erroneously targets the first base pair of the first genomic spacer in the array). * indicates that the base immediately preceding symbol is phosphorothioated.

Table S4. CRISPR array sequencing primers.

primer	sequence (5′-3′)
CB501	AATGATACGGCGACCACCGAGATCTACAC TAGATCGC ctggcttaaaaaatcattaattaataataggttatgtttaga
CB502	AATGATACGGCGACCACCGAGATCTACAC CTCTCTAT ctggcttaaaaaatcattaattaataataggttatgtttaga
CB503	AATGATACGGCGACCACCGAGATCTACAC TATCCTCT ctggcttaaaaaatcattaattaataataggttatgtttaga
CB504	AATGATACGGCGACCACCGAGATCTACAC AGAGTAGA ctggcttaaaaaatcattaattaataataggttatgtttaga
CB505	AATGATACGGCGACCACCGAGATCTACAC GTAAGGAG ctggcttaaaaaatcattaattaataataggttatgtttaga
CB506	AATGATACGGCGACCACCGAGATCTACAC ACTGCATA ctggcttaaaaaatcattaattaataataggttatgtttaga
CB507	AATGATACGGCGACCACCGAGATCTACAC AAGGAGTA ctggcttaaaaaatcattaattaataataggttatgtttaga
CB508	AATGATACGGCGACCACCGAGATCTACAC CTAAGCCT ctggcttaaaaaatcattaattaataataggttatgtttaga
CB701	CAAGCAGAAGACGGCATACGAGAT TCGCCTTA ggtttgagcgatgatatttgtgct
CB702	CAAGCAGAAGACGGCATACGAGAT CTAGTACG ggtttgagcgatgatatttgtgct
CB703	CAAGCAGAAGACGGCATACGAGAT TTCTGCCT ggtttgagcgatgatatttgtgct
CB704	CAAGCAGAAGACGGCATACGAGAT GCTCAGGA ggtttgagcgatgatatttgtgct
CB705	CAAGCAGAAGACGGCATACGAGAT AGGAGTCC ggtttgagcgatgatatttgtgct
CB706	CAAGCAGAAGACGGCATACGAGAT CATGCCTA ggtttgagcgatgatatttgtgct
CB707	CAAGCAGAAGACGGCATACGAGAT GTAGAGAG ggtttgagcgatgatatttgtgct
CB708	CAAGCAGAAGACGGCATACGAGAT CCTCTCTG ggtttgagcgatgatatttgtgct
CB709	CAAGCAGAAGACGGCATACGAGAT AGCGTAGC ggtttgagcgatgatatttgtgct
CB710	CAAGCAGAAGACGGCATACGAGAT CAGCCTCG ggtttgagcgatgatatttgtgct
CB711	CAAGCAGAAGACGGCATACGAGAT TGCCTCTT ggtttgagcgatgatatttgtgct
CB712	CAAGCAGAAGACGGCATACGAGAT TCCTCTAC ggtttgagcgatgatatttgtgct
CBR1	CTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTCCCCGCGCCAG
CBI1	CGGGGATAAACCGAGCACAAATATCATCGCTCAAACC

For all samples, underlined bases indicate barcode sequence (derived from Illumina Nextera barcodes).

Table S5: Parameters utilized in CRISPR expansion models

sensor	state	parameter	value	calculated from?
LacI	1	p_T, L1	0.27490	"1111" sample, pTrig proportion in L1 arrays
LacI	1	p_T, L2	0.24570	"1111" sample, average of pTrig proportion in L2 arrays
LacI	1	p_T, L3	0.22020	"1111" sample, average of pTrig proportion in L3 arrays
LacI	1	p_T, L4	0.18650	"1111" sample, average of pTrig proportion in L4 arrays
LacI	1	p_T, L5	0.18090	"1111" sample, average of pTrig proportion in L5 arrays
LacI	0	p_T	0.00070	"0000" sample, average of pTrig proportion at all positions (L1-L5)
LacI	1	p_exp	0.09880	average proportion singly expanded after single round (control experiment)
LacI	0	p_exp	0.03560	average proportion singly expanded after single round (control experiment)
CopA	1	p_T, L2	0.03542	profile #6, average of pTrig proportion in CopA sensor in L2 arrays
CopA	1	p_T, L3	0.03092	profile #6, average of pTrig proportion in CopA sensor in L3 arrays
CopA	0	P_T	0.00107	profile #1, average of pTrig proportion in CopA sensor at all array positions L1-L3
TreR	1	p_T, L2	0.16064	profile #2, average of pTrig proportion in TreR sensor in L2 arrays
TreR	1	p_T, L3	0.14790	profile #2, average of pTrig proportion in TreR sensor in L3 arrays
TreR	0	p_T	0.00065	profile #1, average of pTrig proportion in TreR sensor at all array positions L1-L3
GalS	1	p_T, L2	0.17542	profile #2, average of pTrig proportion in GalS sensor in L2 arrays
GalS	1	p_T, L3	0.13966	profile #2, average of pTrig proportion in GalS sensor in L3 arrays
GalS	0	p_T	0.00306	profile #1, average of pTrig proportion in GalS sensor at all array positions L1-L3
CopA/TreR/GalS	1	p_exp	0.10155	average proportion singly expanded after one day (across all sensors) with no inducer
CopA/TreR/GalS	0	p_exp	0.09829	average proportion singly expanded after one day (across all sensors) with all three inducers

References and Notes

1. 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). doi:10.1126/science.1172005 Medline

2. J. Bonnet, P. Yin, M. E. Ortiz, P. Subsoontorn, D. Endy, Amplifying genetic logic gates. *Science* **340**, 599–603 (2013). doi:10.1126/science.1232758 Medline

3. L. Yang, A. A. K. Nielsen, J. Fernandez-Rodriguez, C. J. McClune, M. T. Laub, T. K. Lu, C. A. Voigt, Permanent genetic memory with >1-byte capacity. *Nat. Methods* **11**, 1261–1266 (2014). doi:10.1038/nmeth.3147 Medline

4. V. Hsiao, Y. Hori, P. W. Rothemund, R. M. Murray, A population-based temporal logic gate for timing and recording chemical events. *Mol. Syst. Biol.* **12**, 869–14 (2016). doi:10.15252/msb.20156663 Medline

5. 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). doi:10.1126/science.aad8559 Medline

6. W. Pei, T. B. Feyerabend, J. Rössler, X. Wang, D. Postrach, K. Busch, I. Rode, K. Klapproth, N. Dietlein, C. Quedenau, W. Chen, S. Sauer, S. Wolf, T. Höfer, H.-R. Rodewald, Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* **548**, 456–460 (2017). doi:10.1038/nature23653 Medline

7. F. Farzadfard, T. K. Lu, Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* **346**, 1256272 (2014). doi:10.1126/science.1256272 Medline

8. 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). doi:10.1126/science.aaf7907 Medline

9. 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). doi:10.1038/nature20777 Medline

10. S. T. Schmidt, S. M. Zimmerman, J. Wang, S. K. Kim, S. R. Quake, Quantitative analysis of synthetic cell lineage tracing using nuclease barcoding. *ACS Synth. Biol.* **6**, 936–942 (2017). doi:10.1021/acssynbio.6b00309 Medline

11. S. D. Perli, C. H. Cui, T. K. Lu, Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* **353**, aag0511 (2016). doi:10.1126/science.aag0511 Medline

12. R. Kalhor, P. Mali, G. M. Church, Rapidly evolving homing CRISPR barcodes. *Nat. Methods* **14**, 195–200 (2017). doi:10.1038/nmeth.4108 Medline

13. S. A. Jackson, R. E. McKenzie, R. D. Fagerlund, S. N. Kieper, P. C. Fineran, S. J. J. Brouns, CRISPR-Cas: Adapting to change. *Science* **356**, eaal5056 (2017). doi:10.1126/science.aal5056 Medline

14. S. H. Sternberg, H. Richter, E. Charpentier, U. Qimron, Adaptation in CRISPR-Cas Systems. *Mol. Cell* **61**, 797–808 (2016). doi:10.1016/j.molcel.2016.01.030 Medline

15. A. V. Wright, J.-J. Liu, G. J. Knott, K. W. Doxzen, E. Nogales, J. A. Doudna, Structures of the CRISPR genome integration complex. *Science* **357**, 1113–1118 (2017). doi:10.1126/science.aao0679 Medline

16. R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, P. Horvath, CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007). doi:10.1126/science.1138140 Medline

17. J. McGinn, L. A. Marraffini, CRISPR-Cas systems optimize their immune response by specifying the site of spacer integration. *Mol. Cell* **64**, 616–623 (2016). doi:10.1016/j.molcel.2016.08.038 Medline

18. L. A. Marraffini, CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015). doi:10.1038/nature15386 Medline

19. S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016). doi:10.1126/science.aaf1175 Medline

20. 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). doi:10.1038/nature23017 Medline

21. A. Levy, M. G. Goren, I. Yosef, O. Auster, M. Manor, G. Amitai, R. Edgar, U. Qimron, R. Sorek, CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**, 505–510 (2015). doi:10.1038/nature14302 Medline

22. M. B. Łobocka, D. J. Rose, G. Plunkett 3rd, M. Rusin, A. Samojedny, H. Lehnherr, M. B. Yarmolinsky, F. R. Blattner, Genome of bacteriophage P1. *J. Bacteriol.* **186**, 7032–7068 (2004). doi:10.1128/JB.186.21.7032-7068.2004 Medline

23. I. Yosef, M. G. Goren, U. Qimron, Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* **40**, 5569–5576 (2012). doi:10.1093/nar/gks216 Medline

24. J. K. Nuñez, L. Bai, L. B. Harrington, T. L. Hinder, J. A. Doudna, CRISPR immunological memory requires a host factor for specificity. *Mol. Cell* **62**, 824–833 (2016). doi:10.1016/j.molcel.2016.04.027 Medline

25. J. K. Nuñez, A. S. Y. Lee, A. Engelman, J. A. Doudna, Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**, 193–198 (2015). doi:10.1038/nature14237 Medline

26. 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). doi:10.1038/nature08187 Medline

27. J. M. Pickard, C. F. Maurice, M. A. Kinnebrew, M. C. Abt, D. Schenten, T. V. Golovkina, S. R. Bogatyrev, R. F. Ismagilov, E. G. Pamer, P. J. Turnbaugh, A. V. Chervonsky, Rapid fucosylation of intestinal epithelium sustains host-commensal symbiosis in sickness. *Nature* **514**, 638–641 (2014). doi:10.1038/nature13823 Medline

28. J. Elbaz, P. Yin, C. A. Voigt, Genetic encoding of DNA nanostructures and their self-assembly in living bacteria. *Nat. Commun.* **7**, 11179 (2016). doi:10.1038/ncomms11179 Medline

29. S. Silas, G. Mohr, D. J. Sidote, L. M. Markham, A. Sanchez-Amat, D. Bhaya, A. M. Lambowitz, A. Z. Fire, Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* **351**, aad4234 (2016). doi:10.1126/science.aad4234 Medline

30. R. Heler, A. V. Wright, M. Vucelja, D. Bikard, J. A. Doudna, L. A. Marraffini, Mutations in Cas9 enhance the rate of acquisition of viral spacer sequences during the CRISPR-Cas immune response. *Mol. Cell* **65**, 168–175 (2017). doi:10.1016/j.molcel.2016.11.031 Medline

31. C. Engler, R. Kandzia, S. Marillonnet, A one pot, one step, precision cloning method with high throughput capability. *PLOS ONE* **3**, e3647 (2008). doi:10.1371/journal.pone.0003647 Medline

32. H. M. Salis, E. A. Mirsky, C. A. Voigt, Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009). doi:10.1038/nbt.1568 Medline

33. 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 Res.* **25**, 1203–1210 (1997). doi:10.1093/nar/25.6.1203 Medline

34. C. Lou, B. Stanton, Y.-J. Chen, B. Munsky, C. A. Voigt, Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol.* **30**, 1137–1142 (2012). doi:10.1038/nbt.2401 Medline

35. S. Meinhardt, M. W. Manley Jr., N. A. Becker, J. A. Hessman, L. J. Maher 3rd, L. Swint-Kruse, Novel insights from hybrid LacI/GalR proteins: Family-wide functional attributes and biologically significant variation in transcription repression. *Nucleic Acids Res.* **40**, 11139–11154 (2012). doi:10.1093/nar/gks806 Medline

36. M. T. Bonde, M. S. Klausen, M. V. Anderson, A. I. N. Wallin, H. H. Wang, M. O. A. Sommer, MODEST: A web-based design tool for oligonucleotide-mediated genome engineering and recombineering. *Nucleic Acids Res.* **42**, W408–W415 (2014). doi:10.1093/nar/gku428 Medline

37. 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). doi:10.1073/pnas.120163297 Medline

38. S. J. Brouns, M. M. Jore, M. Lundgren, E. R. Westra, R. J. H. Slijkhuis, A. P. L. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin, J. van der Oost, Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008). doi:10.1126/science.1159689 Medline

39. M. Škulj, V. Okršlar, S. Jalen, S. Jevsevar, P. Slanc, B. Strukelj, V. Menart, Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes. *Microb. Cell Fact.* **7**, 6 (2008). doi:10.1186/1475-2859-7-6 Medline

40. J. J. Kozich, S. L. Westcott, N. T. Baxter, S. K. Highlander, P. D. Schloss, Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013). doi:10.1128/AEM.01043-13 Medline

41. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009). doi:10.1186/1471-2105-10-421 Medline

42. W. M. Fitch, E. Margoliash, Construction of phylogenetic trees. *Science* **155**, 279–284 (1967). doi:10.1126/science.155.3760.279 Medline

43. D. L. Shis, F. Hussain, S. Meinhardt, L. Swint-Kruse, M. R. Bennett, Modular, multi-input transcriptional logic gating with orthogonal LacI/GalR family chimeras. *ACS Synth. Biol.* **3**, 645–651 (2014). doi:10.1021/sb500262f Medline

44. R. S. Cox 3rd, M. G. Surette, M. B. Elowitz, Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.* **3**, 145 (2007). doi:10.1038/msb4100187 Medline