Cite as: R. U. Sheth *et al*., *Science* 10.1126/science.aao0958 (2017).

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]*

[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**

While 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 to store temporal biological information directly into the genomes of a cell population. A "biological tape recorder" is developed 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, 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 utilized 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 is the robust recording of temporally varying biological states or signals (e.g., gene expression, metabolite fluctuations) in living cells. Such a biological recording system would have powerful applications in studying dynamic cellular processes including complex regulatory programs, or 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 can be subsequently used by CRISPR-Cas immunity proteins to repel future invaders that exhibit matching sequence identity (*18*). The DNA writing potential of the adaptation process has been recently explored 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 to 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 a trigger DNA pool within living cells. The CRISPR-Cas spacer acquisition machinery is then employed 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 a trigger DNA pool within a population of *Escherichia coli* cells. A copy number inducible trigger plasmid (pTrig) was utilized, 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 PCR (qPCR) revealed a 653 ± 5 fold increase in copy number in cells induced with IPTG for 6 hours compared to 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 alteration of an intracellular DNA pool.

Next, we assessed whether an increase in pTrig copy number could be recorded into CRISPR arrays across a cell population. Expression of the CRISPR adaptation proteins Cas1 and Cas2 promotes unidirectional integration of ~33 bp 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 pRec+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 (Fig. 1, F and G; fig. S4; and methods). In cells with pRec, spacers were preferentially derived from the pRec plasmid, consistent with enriched spacer acquisitions from plasmids in *E. coli* documented in the literature (*21*). Cells with pRec+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 into 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). While 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 (transformation of a biological signal to increase abundance of an intracellular DNA pool, and 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 four sequential days (d1-d4) constituting 16 unique temporal signal profiles (Fig. 2A). Sequencing the resulting CRISPR arrays confirmed an overall expansion in array sizes over time (fig. S7) with 24.7 ± 5.2% of all arrays having incorporated at least one new spacer by d4. On average, ~1 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). Since longer arrays contained more temporal information, we additionally implemented a size enrichment protocol (methods) that facilitated the analysis of arrays with up to 5 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 (p1 to p5, Fig. 2, D and E, and fig. S8B). Strikingly, 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 (R) or trigger (T) spacers occupying each spacer position (Fig. 3A). A simple analytical model of the CRISPR expansion process was then developed 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, given 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.

Interestingly, the predicted and observed array-type distributions matched closely (fig. S10). For example, for two signal profiles of equal number of inductions but differing temporal ordering, our models yielded distinctive array-type distributions that appeared to recapitulate their 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 based on the data-model pair with the shortest Euclidean distance (Fig. 3C). Using L1 arrays only, which do not contain any temporal information, only 5 of 16 signal profiles could be correctly classified. In contrast, using L2 to L4 array-types individually resulted in much higher accuracy of assignments (13-14 of 16 correct). When L2-L4 array-types were simultaneously 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 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 sub-populations each day, which resulted in a defined branching history of the 16 populations (Fig. 3E). Interestingly, 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 near perfect accuracy (Fig. 3F).

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), while induction of recording showed negligible loss of previously acquired spacers (fig. S12C). Thus, 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 (4 of 7 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.

Finally, we explored the possibility of 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 since the DR sequence is duplicated during each spacer incorporation event (*23*, *25*). Using MAGE (*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 biologically meaningful chemicals, copper (heavy-metal contaminant), trehalose (dietary sugar metabolite) and fucose (associated with mammalian gut infection (*27*)), were eventually selected for multiplex recording experiments. To assess the capacity for multi-channel recording, we exposed cell populations containing a mix of all three strains to all 8 combinations of the 3 input chemicals. The resulting CRISPR arrays were sequenced and demultiplexed using the DR barcodes. Each sensor strain displayed robust increase of 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 multi-channel 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 three 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 of all three inputs simultaneously was assessed by the Hamming distance threshold to the actual temporal signal profiles; 8 of 16 profiles were perfectly classified and the rest were within Hamming distance 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 utilized 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. The system could employ inducible intracellular DNA production systems in parallel (*28*) and other CRISPR-Cas adaptation machinery (*13*, *29*), which may be needed for extension to other bacteria (or even eukaryotes) and to 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.

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

ACKNOWLEDGMENTS

We thank T. Blazejewski, C. Munck and members of the Wang lab for advice and comments on the manuscript. Sequencing data associated with this study is available at NCBI SRA under PRJNA417866, and plasmids are available through Addgene. H.H.W. acknowledges specific funding support from the DoD ONR (N00014-17-1-2353, N00014-15-1-2704), 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. All authors discussed results and commented on and approved the manuscript. H.H.W. and R.U.S. have filed a provisional patent based on the work described in this paper.

SUPPLEMENTARY MATERIALS

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

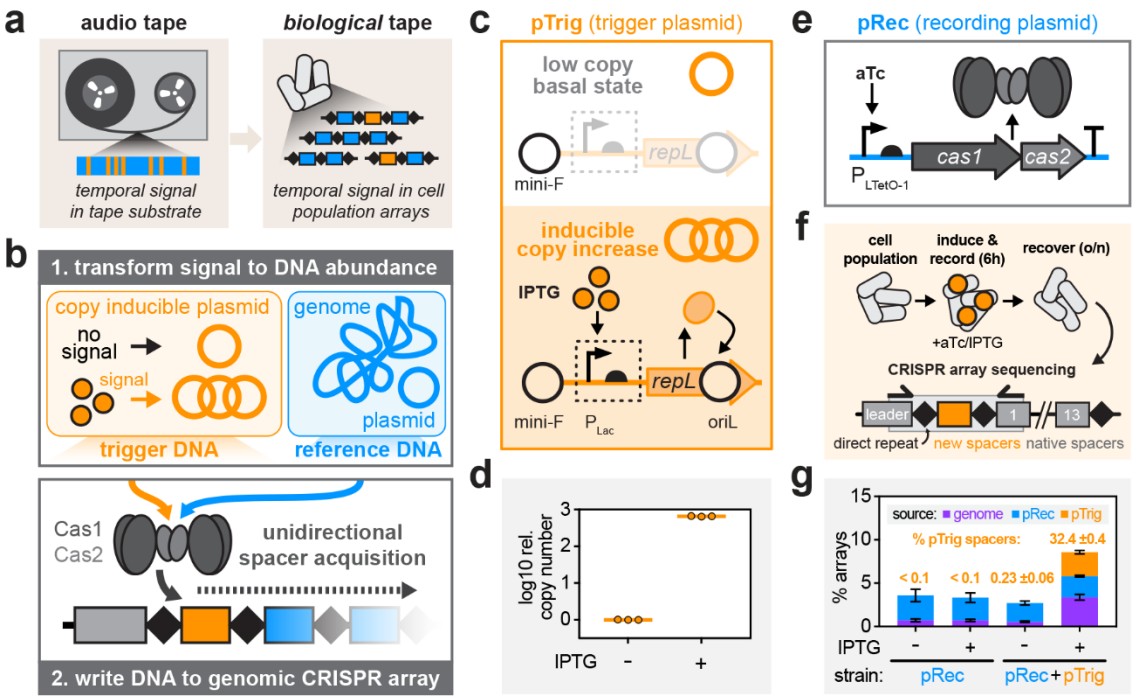

Fig. 1. Temporal recording in arrays by CRISPR expansion (TRACE). (**A**) Akin to an audio tape, temporal biological signals can be stored in DNA arrays within a cell population. (**B**) TRACE functions by first transforming an input biological signal to an altered abundance of a trigger DNA pool (orange). This trigger DNA pool, alongside reference DNA (blue) is then recorded as spacers into genomic CRISPR arrays of a cell population in a unidirectional fashion, enabling recording 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. (**D**) qPCR measurement of pTrig relative copy number (log10 scale) in cells exposed to no IPTG or 1mM 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. (**G**) Cells with pRec or pRec+pTrig were exposed to 100 ng/μL aTc and no or 1mM 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.

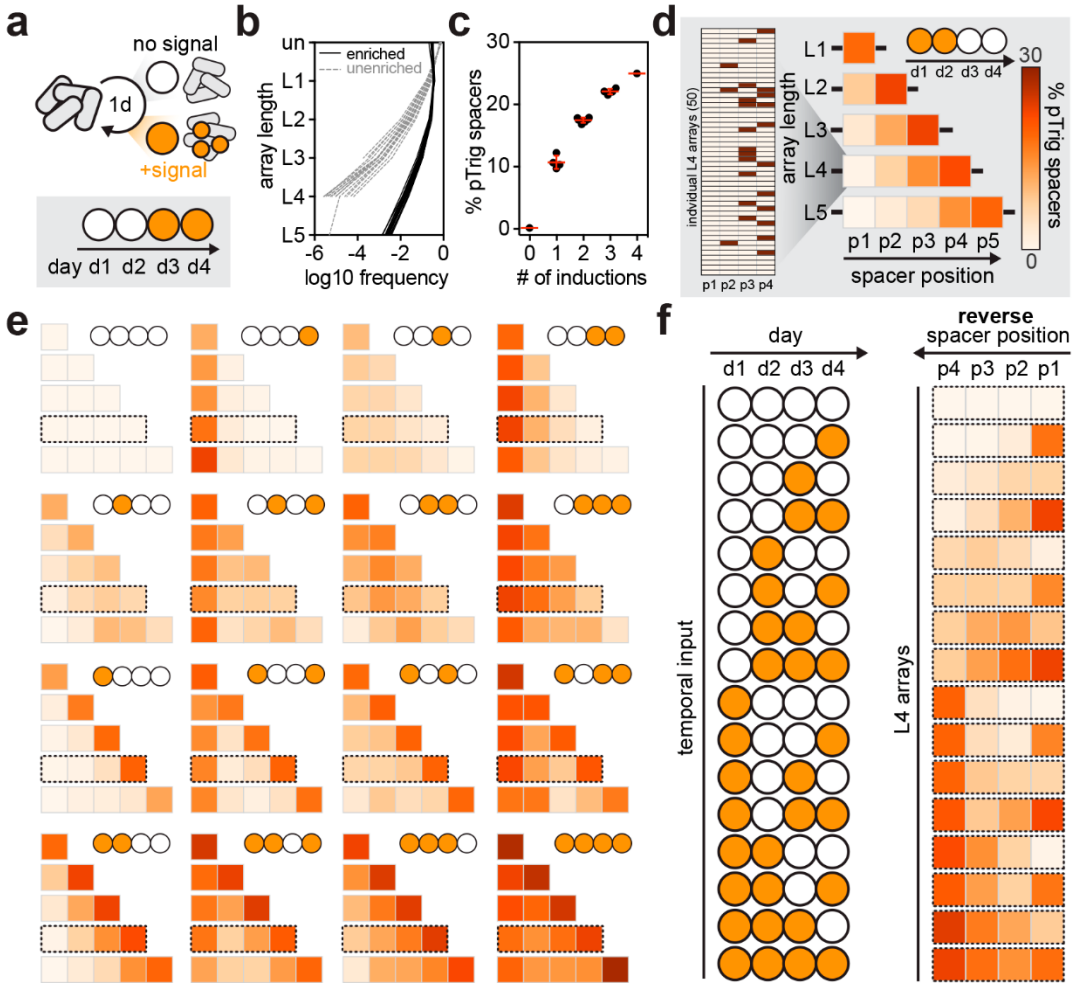

Fig. 2. Temporal recording of four day input profiles. (**A**) Cell populations were subjected to daily exposures over four sequential days, constituting all 16 possible temporal signal profiles. (**B**) Resulting CRISPR arrays were sequenced with (black line) and without (grey dashed line) a size-enrichment method and the frequency (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 line indicates mean and standard deviation. (**D**) 50 L4 arrays sampled from the full dataset 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 percentage of pTrig spacers detected at each location. (**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 improve visual comparison) are displayed.

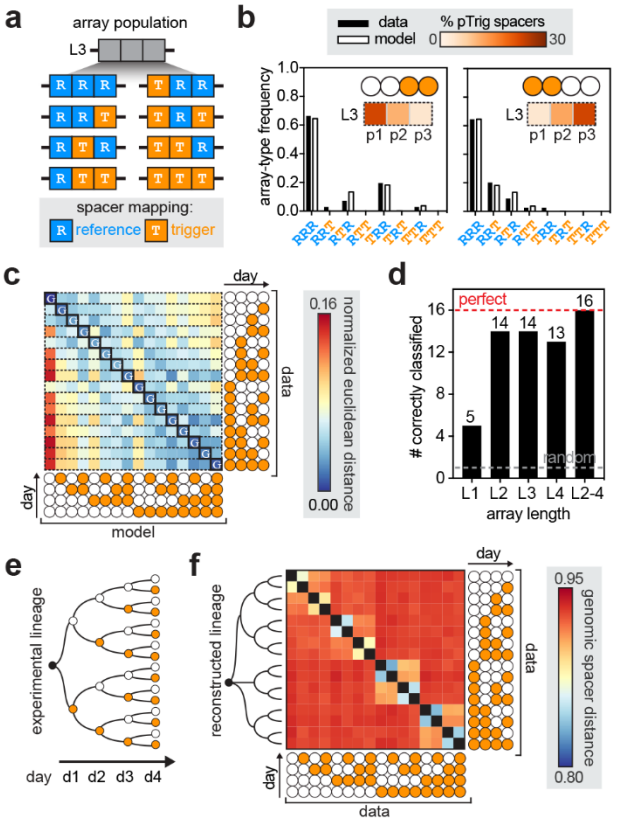

Fig. 3. Reconstructing temporal signal profiles and population lineages. (**A**) CRISPR array populations can be described as a frequency distribution constituting of all permutations of reference (R, blue) and trigger (T, orange) spacers for a given array length (L); L3 arrays are depicted. (**B**) As an example, for two distinct profiles of 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 distance between observed data (rows) and predicted model (columns) array type distributions (L2, L3 and L4 array distributions concatenated) was calculated and normalized by row. The correct temporal signal profile is indicated by a white asterisk, and the model with minimum distance to the data is indicated by black outline on the diagonal. (**D**) Number of profiles correctly classified utilizing L1 to L4 arrays individually or L2-L4 arrays together as in (C); grey dashed line indicates expected random classification (1/16). (**E**) A defined branching history was utilized when performing the temporal recording experiment. (**F**) The mapping locations for genomic spacers within L1 arrays was utilized as sequence identity of the spacer and the Jaccard distance between all samples (1 – proportion of spacers shared between two samples) is 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).

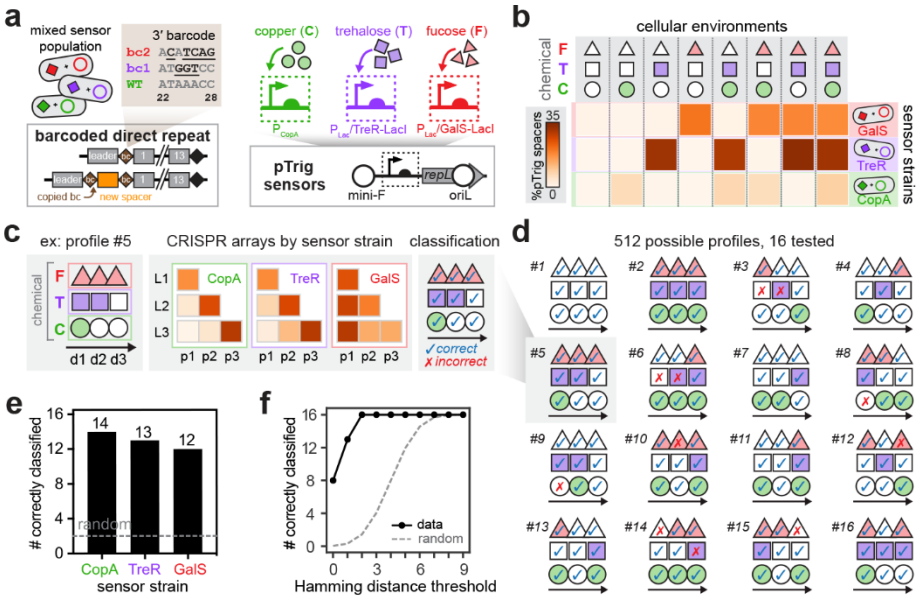

Fig. 4. Multiplex temporal recording with a barcoded sensor population. (**A**) The direct repeat (DR) of a CRISPR array can be barcoded to associate sensors with specific arrays; generated distal DR barcode sequences are shown. Sensors of copper, trehalose and fucose were linked to the pTrig system and introduced into barcoded strains. The copper sensor utilizes a native promoter with endogenous transcription factor expression, while the trehalose and fucose sensors utilize an engineered transcription factor. (**B**) The three barcoded sensor strains were mixed and exposed to 8 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 three 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 resulting classification (correct: blue checkmark or incorrect: red X). (**D**) 16 profiles were tested (6 defined, 10 randomly generated) of 512 (8^3) possible profiles; the resulting classification is shown as in (c). (**E**) Single channel classification accuracy: profiles were classified for each sensor using L2 and L3 arrays; grey dashed line indicates expected random classification (2/16). (**F**) Multi-channel classification accuracy: predictions were considered across all three sensors, and the number classified correctly within a Hamming distance threshold is shown (black line) compared to the expected random classification (grey dashed line).

Multiplex recording of cellular events over time on CRISPR biological tape

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

published online November 23, 2017