Genomically encoded analog memory with precise in vivo DNA writing in living cell populations

Fahim Farzadfard and Timothy K. Lu

Science **346**, (2014);
DOI: 10.1126/science.1256272

RESEARCH ARTICLE SUMMARY

SYNTHETIC BIOLOGY

Genomically encoded analog memory with precise in vivo DNA writing in living cell populations

Fahim Farzadfard and Timothy K. Lu*

INTRODUCTION: The conversion of transient information into long-lasting responses is a common aspect of many biological processes and is crucial for the design of sophisticated synthetic circuits. Genomic DNA provides a rich medium for the storage of information in living cells. However, current cellular memory technologies are limited in their storage capacity and scalability.

RATIONALE: We converted genomic DNA into a "tape recorder" for memorizing information in living cell populations. This was achieved via SCRIBE (Synthetic Cellular Recorders Integrating Biological Events), a programmable and modular architecture for generating single-stranded DNA (ssDNA) inside of living cells in response to gene regulatory signals. When coexpressed with a recombinase, these ssDNAs address specific target loci on the basis of sequence homology and introduce precise mutations into genomic DNA, thus converting transient cellular signals into genomically encoded memory. This distributed biological memory leverages the large number of cells in bacterial cultures and encodes information into their collective genomic DNA in the form of the fraction of cells that carry specific mutations.

RESULTS: We show that SCRIBE enables the recording of arbitrary transcriptional inputs into DNA storage registers in living cells by translating regulatory signals into ssDNAs. In *Escherichia coli*, we expressed ssDNAs from engineered retrons that use a reverse transcriptase protein to produce hybrid RNA-ssDNA molecules. These intracellularly expressed ssDNAs are targeted into specific genomic loci where they are recombined and converted into permanent memory. We show that genomically stored information can be readily reprogrammed by changing the ssDNA template and controlled via both chemical and light inputs. We demonstrate that genomically encoded memory can be read with a variety of techniques, including reporter genes, functional assays, and high-throughput DNA sequencing.

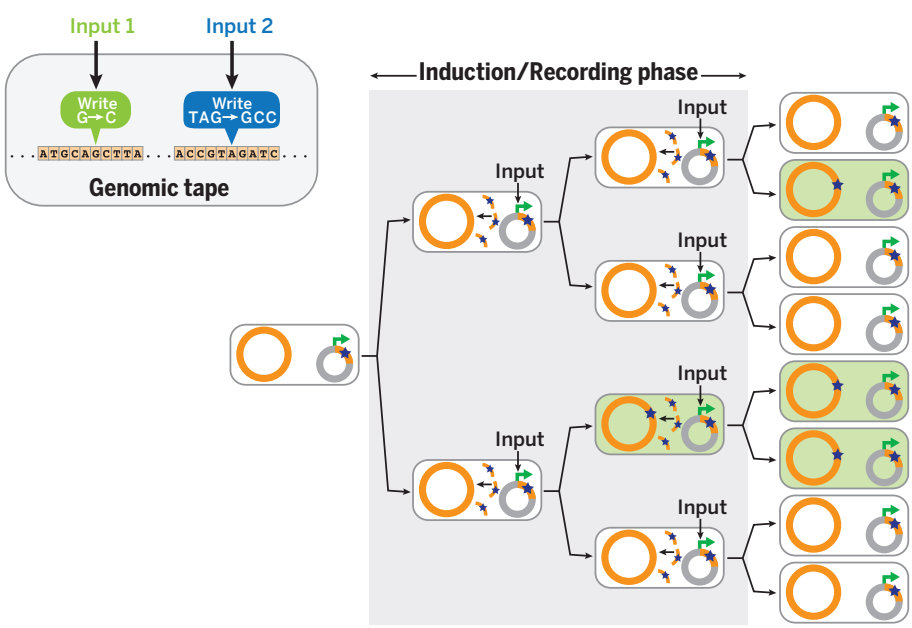
SCRIBE enables the recording of analog information such as the magnitude and time span of exposure to an input. This convenient feature is facilitated by the intermediate recombination rate of our current system (~10^{-4} recombination events per generation), which we validated via a mathematical model and computer simulations. For example, we stored the overall exposure time to chemical inducers in the DNA memory of bacterial populations for 12 days (~120 generations), independently of the induction pattern. The frequency of mutants in these populations was linearly related to the total exposure time.

Furthermore, we demonstrate that SCRIBE-induced mutations can be written and erased and can be used to record multiple inputs across the distributed genomic DNA of bacterial populations. Finally, we show that SCRIBE memory can be decomposed into independent "input," "write," and "read" operations and used to create genetic "logic-and-memory" circuits, as well as "sample-and-hold" circuits.

CONCLUSION: We describe a scalable platform that uses genomic DNA for analog, rewritable, and flexible memory distributed across living cell populations. We anticipate that SCRIBE will enable long-term cellular recorders for environmental and biomedical applications. Future optimization of recombination efficiencies achievable by SCRIBE could lead to more efficient single-cell digital memories and enhanced genome engineering technologies. Furthermore, the ability to regulate the generation of arbitrary targeted mutations with other gene-editing technologies should enable genomically encoded memory in additional organisms. ■

ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.1256272



SCRIBE enables distributed genomically encoded memory. In the presence of an input, ssDNAs (orange curved lines) are produced from a plasmid-borne cassette (gray circles) and recombined into specific genomic loci (orange circles) that are targeted on the basis of sequence homology. This results in the accumulation of precise mutations (stars in green cells) as a function of the magnitude and duration of exposure to the input.

RELATED ITEMS IN *SCIENCE*

S. Ausländer, M. Fussenegger, Dynamic genome engineering in living cells. *Science* **346**, 813–814 (2014).

The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: timlu@mit.edu
Cite this article as F. Farzadfard, T. K. Lu, *Science* **346**, 1256272 (2014). DOI: 10.1126/science.1256272

RESEARCH ARTICLE

SYNTHETIC BIOLOGY

Genomically encoded analog memory with precise in vivo DNA writing in living cell populations

Fahim Farzadfard[1,2,3] and Timothy K. Lu[1,2,3]*

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

Due to its high storage capacity, durability, ease of duplication, and high-fidelity maintenance of information, DNA has garnered much interest as an artificial storage medium (*1*, *2*). However, existing technologies for in vivo autonomous recording of information in cellular memory are limited in their storage capacity and scalability (*3*). Epigenetic memory devices such as bistable toggle switches (*4*–*7*) and positive-feedback loops (*8*) require orthogonal transcription factors and can lose their digital state due to environmental fluctuations or cell death. Recombinase-based memory devices enable the writing and storage of digital information in the DNA of living cells (*9*–*12*), where binary bits of information are stored in the orientation of large stretches of DNA. However, these devices do not efficiently exploit the full capacity of DNA for information storage: Recording a single bit of information with these devices often requires at least a few hundred base pairs of DNA, overexpression of a recombinase protein to invert the target DNA, and engineering recombinase recognition sites into target loci in advance. The scalability of this type of memory is further limited by the number of orthogonal recombinases that can be used in a single cell. Finally, epigenetic and recombinase-based memory devices described to date store digital information, and their recording capacity is exhausted within a few hours of induction. Thus, these devices have not been adapted to record analog information, such as magnitude and time course of inputs over extended periods of time (i.e., multiple days or more).

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

Single-stranded DNA expression in living cells

Previously, it was shown that synthetic oligonucleotides delivered by electroporation into cells that overexpress Beta recombinase (from bacteriophage λ) in *Escherichia coli* are specifically and efficiently recombined into homologous genomic sites (*13*–*16*). Thus, oligonucleotide-mediated recombination offers a powerful way to introduce targeted mutations in a bacterial genome (*17*, *18*). However, this technique requires the exogenous delivery of ssDNAs and cannot be used to couple arbitrary signals into genetic memory. To overcome these limitations, we developed a genome editing platform based on expressing ssDNAs inside of living cells by taking advantage of a widespread class of bacterial reverse transcriptases (RTs) called retrons (*19*, *20*).

The wild-type (WT) retron cassette encodes three components in a single transcript: a RT protein and two RNA moieties, msr and msd, which act as the primer and the template for the reverse transcriptase, respectively (Fig. 1A, left). The *msr-msd* sequence in the retron cassette is flanked by two inverted repeats. Once transcribed, the msr-msd RNA folds into a secondary structure guided by the base pairing of the inverted repeats and the msr-msd sequence. The RT recognizes this secondary structure and uses a conserved guanosine residue in the msr as a priming site to reverse transcribe the msd sequence and produce a hybrid RNA-ssDNA molecule called msDNA (i.e., multicopy single-stranded DNA) (*20*, *21*). To couple the expression of ssDNA to an external input, the WT Ec86 retron cassette from *E. coli* BL21 (*21*) was placed under the control of an isopropyl β-D-1-thiogalactopyranoside (IPTG)–inducible promoter (P_{lacO}) in *E. coli* DH5αPRO cells (*22*), which express high levels of the LacI and TetR repressors (Fig. 1A). The WT retron ssDNA [ssDNA(wt)] was readily detected in IPTG-induced cells, whereas no ssDNA was detected in noninduced cells (Fig. 1B). The identity of the detected ssDNA band was further confirmed by DNA sequencing (fig. S1). To verify that ssDNA expression depended on RT activity, point mutations [Asp^{197}→Ala^{197} (D197A) and D198A] were introduced to the active site of the RT to make a catalytically dead RT (dRT) (*23*). This modification completely abolished ssDNA production (Fig. 1B).

To engineer the msd template to express synthetic ssDNAs of interest, we initially tried to replace the whole *msd* sequence with a desired template. However, no ssDNA was detected, suggesting that some features of msd are required for ssDNA expression, as was previously noted for another retron (*24*). A variant in which the flanking regions of the msd stem remained intact (Fig. 1A, right) produced detectable amounts of ssDNA when induced by IPTG (Fig. 1B, P_{lacO}-msd($kanR$)$_{ON}$ + IPTG). The correct identity of the detected ssDNA band was further confirmed by DNA sequencing (fig. S1). Thus, the lower part of the msd stem is essential for reverse transcription, whereas the upper part of the stem and the loop are dispensable and can be replaced with desired templates to produce ssDNAs of interest in vivo.

Regulated genome editing with in vivo ssDNAs

To demonstrate that intracellularly expressed ssDNAs can be recombined into target genomic loci by concomitant expression of Beta, we developed a selectable marker reversion assay (Fig. 1C). The *kanR* gene, which encodes neomycin

[1]Synthetic Biology Group, Research Laboratory of Electronics, Department of Electrical Engineering and Computer Science and Department of Biological Engineering, Massachusetts Institute of Technology (MIT), 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [2]MIT Synthetic Biology Center, 500 Technology Square, Cambridge, MA 02139, USA. [3]MIT Microbiology Program, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.
*Corresponding author. E-mail: timlu@mit.edu

phosphotransferase II and confers resistance to kanamycin (Kan), was integrated into the *galK* locus. Two stop codons were then introduced into the genomic *kanR* to make a Kan-sensitive $kanR_{OFF}$ reporter strain (DH5αPRO $galK::kanR_{W28TAA,\ A29TAG}$). These premature stop codons could be reverted back to the WT sequence via recombination with engineered ssDNA$(kanR)_{ON}$, thus conferring kanamycin resistance (Fig. 1C). Specifically, ssDNA$(kanR)_{ON}$ contains 74 base pairs of homology to the regions of the $kanR_{OFF}$ locus flanking the premature stop codons and replaces the stop codons with the WT *kanR* gene sequence (Fig. 1C).

We cloned the Beta gene (*bet*) into a plasmid under the control of the anhydrotetracycline (aTc)–inducible P_{tetO} promoter and introduced it along with the IPTG-inducible msd$(kanR)_{ON}$ construct into the $kanR_{OFF}$ strain (Fig. 1C). Induction of cultures harboring these two plasmids with either IPTG (1 mM) or aTc (100 ng/ml) resulted in a slight increase in the frequency of Kan-resistant cells within the population (Fig. 1C). However, coexpression of both ssDNA$(kanR)_{ON}$ and Beta with IPTG and aTc resulted in a $>10^4$-fold increase in the recombinant frequency relative to the non-induced cells. This corresponded to a $>10^3$-fold increase relative to cells induced with IPTG only and a 60-fold increase relative to cells induced with aTc only. This increase in the recombinant frequency was dependent on the RT activity, as it was largely abolished with dRT. The genotypes of randomly selected Kan-resistant colonies were further confirmed by DNA sequencing to contain precise reversions of the two codons to the WT sequence (fig. S1). No Kan-resistant colonies were detected when a nonspecific ssDNA [ssDNA(wt)] was coexpressed with Beta in the $kanR_{OFF}$ reporter cells, confirming that Kan-resistant cells were not produced due to spontaneous mutations. In additional experiments, we used high-throughput sequencing (Illumina HiSeq) on the bacterial populations to analyze the genomically encoded memory (see supplementary materials and fig. S2). Comparable recombinant frequencies were obtained from both the plating assay and sequencing, confirming that genomically encoded memory can be read without the need for functional assays and reporters.

Recording input magnitudes into genomic memory

We reasoned that the rate of recombination between engineered ssDNAs and genomic DNA could be effectively modulated by changing expression levels of the engineered retron cassette and Beta. This feature would enable the recording of analog information, such as the magnitude of an input signal, in the proportion of cells in a population with a specific mutation in genomic DNA. To demonstrate this, both the ssDNA$(kanR)_{ON}$ expression cassette and *bet* were placed into a single synthetic operon [hereafter referred to as the SCRIBE$(kanR)_{ON}$ cassette] under the control of P_{lacO} (Fig. 1D). The $kanR_{OFF}$ reporter cells harboring this synthetic operon were induced with different concentrations of IPTG. The fraction of Kan-resistant recombinants increased linearly with the input inducer concentration on a log-log plot over a range of ~10^{-7} to ~10^{-5} (Fig. 1D). Statistical tests showed that at least four different concentrations of the inducer (including 0 mM IPTG) could be resolved in this experiment. Thus, the efficiency of genome writing in a population can be quantitatively tuned with external inputs.

Writing and rewriting genomic memory

We next created a complementary set of SCRIBE cassettes to write and erase (rewrite) information in the genomic *galK* locus using two different chemical inducers. Cells expressing *galK* can

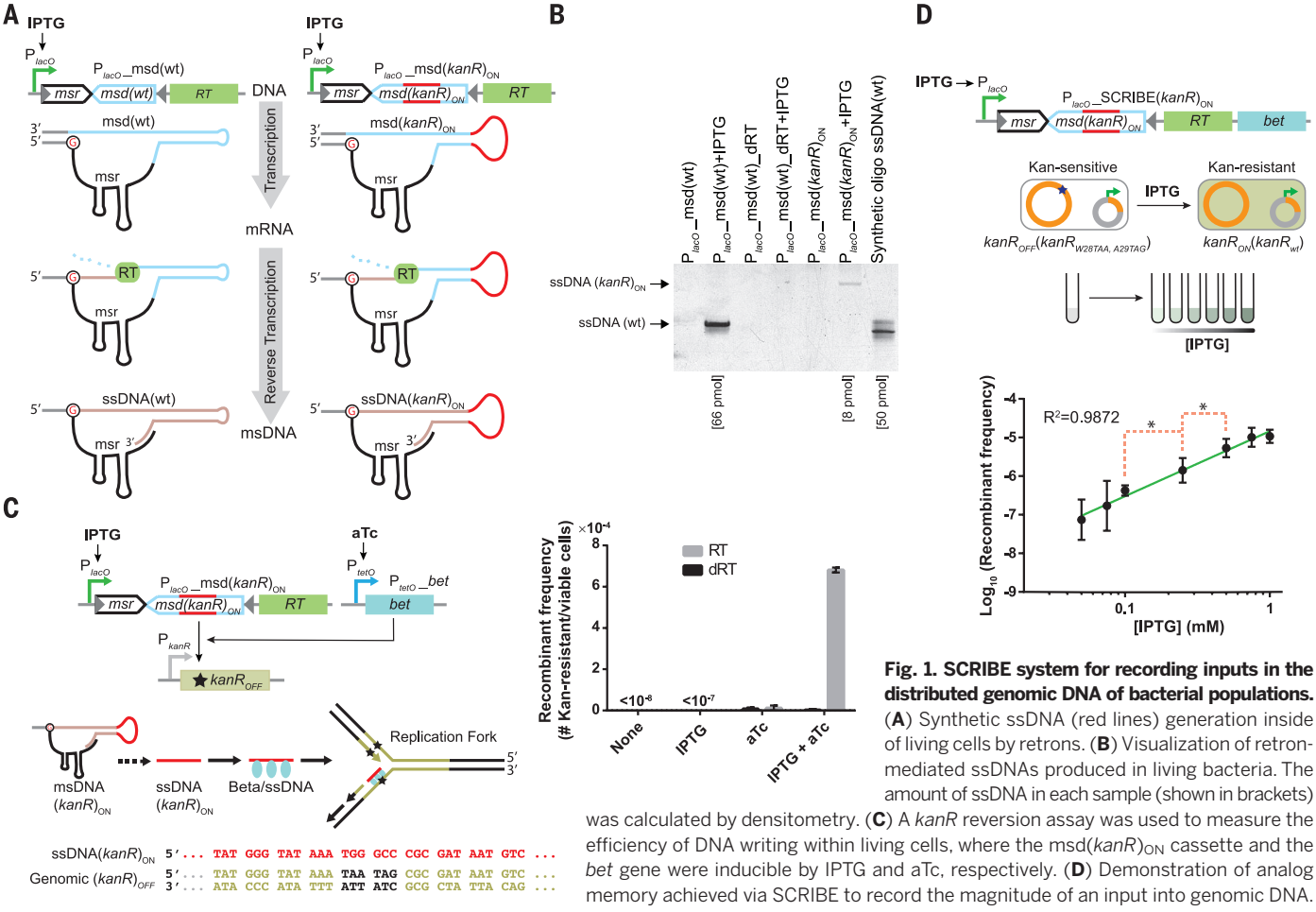
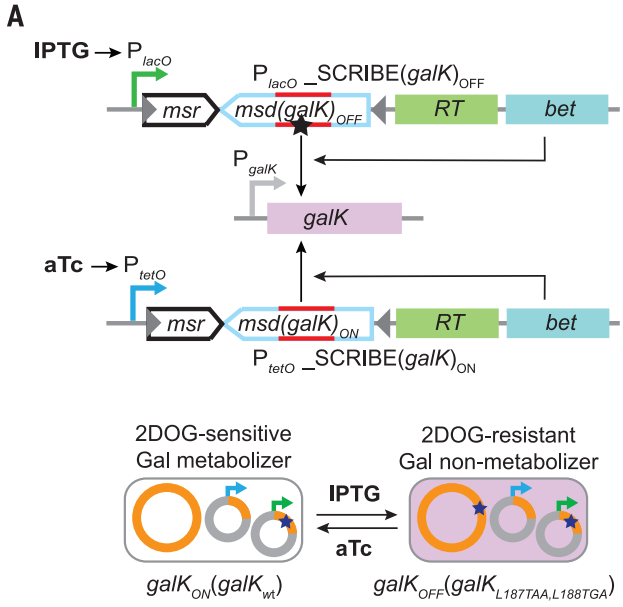


Fig. 1. SCRIBE system for recording inputs in the distributed genomic DNA of bacterial populations. (**A**) Synthetic ssDNA (red lines) generation inside of living cells by retrons. (**B**) Visualization of retron-mediated ssDNAs produced in living bacteria. The amount of ssDNA in each sample (shown in brackets) was calculated by densitometry. (**C**) A *kanR* reversion assay was used to measure the efficiency of DNA writing within living cells, where the msd$(kanR)_{ON}$ cassette and the *bet* gene were inducible by IPTG and aTc, respectively. (**D**) Demonstration of analog memory achieved via SCRIBE to record the magnitude of an input into genomic DNA. The green line is a linear regression fit. The red dashed brackets marked with asterisks connect the closest data points that are statistically significant with respect to each other ($P < 0.05$ based on one-tailed Welch's *t* test). Error bars indicate SEM for three independent biological replicates.



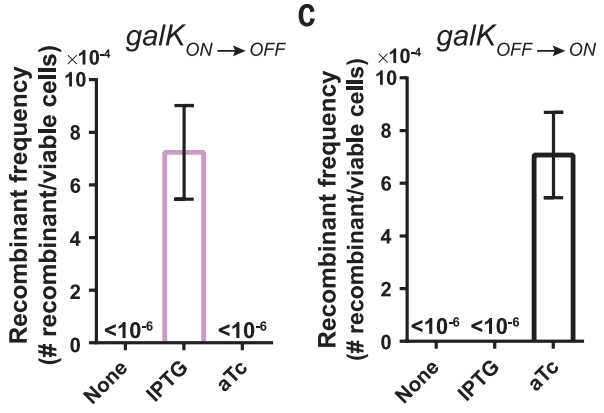


Fig. 2. SCRIBE can write multiple different DNA mutations into a common target loci (*galK*). (**A**) Schematic of the procedure (see text for details). (**B**) $galK_{ON}$ cells harboring the circuits shown in (A) were induced with either IPTG (1 mM) or aTc (100 ng/ml) for 24 hours, and the $galK_{OFF}$ frequencies in the population were determined by plating the cells on appropriate selective conditions. (**C**) $galK_{OFF}$ cells [obtained from the experiment described in (B)] were induced with IPTG (1 mM) or aTc (100 ng/ml) for 24 hours, and the $galK_{ON}$ frequencies in the population were determined by plating the cells on appropriate selective conditions. Error bars indicate SEM for three independent biological replicates.

metabolize and grow on galactose as the sole carbon source. However, these *galK*-positive ($galK_{ON}$) cells cannot metabolize 2-deoxy-galactose (2DOG) and cannot grow on plates containing glycerol (carbon source) + 2DOG. On the other hand, *galK*-negative ($galK_{OFF}$) cells cannot grow on galactose as the sole carbon source but can grow on glycerol + 2DOG plates (*25*). We transformed DH5αPRO $galK_{ON}$ cells with plasmids encoding IPTG-inducible SCRIBE$(galK)_{OFF}$ and aTc-inducible SCRIBE$(galK)_{ON}$ cassettes (Fig. 2A). Induction of SCRIBE$(galK)_{OFF}$ by IPTG resulted in the writing of two stop codons into $galK_{ON}$, leading to $galK_{OFF}$ cells that could grow on glycerol + 2DOG plates (Fig. 2B). Induction of SCRIBE$(galK)_{ON}$ in these $galK_{OFF}$ cells with aTc reversed the IPTG-induced modification, leading to $galK_{ON}$ cells that could grow on galactose plates (Fig. 2C). These results show that writing on genomic DNA with SCRIBE is reversible and that distinct information can be written and rewritten into the same locus.

Writing multiple mutations into independent loci

Scaling the capacity of previous memory devices is challenging because each additional bit of information requires additional orthogonal proteins (e.g., recombinases or transcription factors). In contrast, orthogonal SCRIBE memory devices are potentially easier to scale because they can be built by simply changing the ssDNA template (*msd*). To demonstrate this, we used SCRIBE to record multiple independent inputs into different genomic loci of bacterial population. We integrated the $kanR_{OFF}$ reporter gene into the *bioA* locus of DH5αPRO to create a $kanR_{OFF}$ $galK_{ON}$ strain. These cells were then transformed with plasmids encoding IPTG-inducible SCRIBE$(kanR)_{ON}$ and aTc-inducible SCRIBE$(galK)_{OFF}$ cassettes (Fig. 3A). Induction of these cells with IPTG or aTc resulted in the production of cells with phenotypes corresponding to $kanR_{ON}$ $galK_{ON}$ or $kanR_{OFF}$ $galK_{OFF}$ genotypes, respectively (Fig. 3, B and C). Comparable numbers of $kanR_{ON}$ $galK_{ON}$ and $kanR_{OFF}$ $galK_{OFF}$ cells (~2×10^{-4} and ~3×10^{-4} recombinant/viable cells, respectively) were produced when the cultures were induced with both aTc and IPTG (Fig. 3C, left panel). Furthermore, very few individual colonies (~3×10^{-7} recombinant/viable cells) containing both writing events ($kanR_{ON}$ $galK_{OFF}$) were obtained in the cultures that were induced with both aTc and IPTG (Fig. 3C, right panel). These data suggest that although multiplexed writing at single-cell level is rare with SCRIBE's current level of recombination efficiency, multiple independent inputs can be successfully recorded into the distributed genomic DNA of bacterial subpopulations.

Optogenetic genome editing for light-to-DNA memory

In SCRIBE, the expression of each individual ssDNA can be triggered by any endogenous or exogenous signal that can be coupled into transcriptional regulation, thus recording these inputs into long-lasting DNA storage. In addition to small-molecule chemicals, we showed that light can be used to trigger specific genome editing for genomically encoded memory. We placed the SCRIBE$(kanR)_{ON}$ cassette under the control of a previously described light-inducible promoter (P_{Dawn}) (*26*) within $kanR_{OFF}$ cells (Fig. 4A). These cultures were then grown for 4 days in the presence of light or in the dark (Fig. 4A). As Beta-mediated recombination is reportedly replication-dependent (*27*–*29*), dilutions of these cultures were made into fresh media at the end of each day to maintain active replication in the cultures. At the end of each day, samples were taken to determine the number of Kan-resistant and viable cells (Fig. 4A). Cultures grown in the dark yielded undetectable levels of Kan-resistant cells (Fig. 4A). In contrast, the frequency of Kan-resistant cells increased steadily over time in the cultures that were grown in the presence of light, indicating the successful recording of light input into long-lasting DNA memory. The analog memory faithfully stored the total time of light exposure, rather than just the digital presence or absence of light.

Recording the exposure time of inputs

The linear increase in the frequency of Kan-resistant colonies over time due to exposure to light indicates that the duration of inputs can be recorded into population-wide DNA memory using SCRIBE. To further explore population-wide genomically encoded memory whose state is a function of input exposure time, we used the $kanR_{OFF}$ strain harboring the constructs shown in Fig. 1C, where expression of ssDNA$(kanR)_{ON}$ and Beta are controlled by IPTG and aTc, respectively. These cells were subjected to four different patterns of the inputs for 12 successive days (patterns I to IV, Fig. 4B). Kan-resistant cells did not accumulate in the negative control (pattern I), which was never exposed to the inducers. The fraction of Kan-resistant cells in the three other patterns (II, III, and IV) increased linearly over their respective induction periods and remained relatively constant when the inputs were removed. These data indicate that the genomically encoded memory was stable in the absence of the inputs over the course of the experiment. The recombinant frequencies in patterns III and IV, which were induced for the same total amount of time but with different temporal patterns, reached comparable levels at the end of the experiment. These data demonstrate that the genomic memory integrates over the total induction time and is independent of the input pattern, and therefore can be used to stably record event histories over many days.

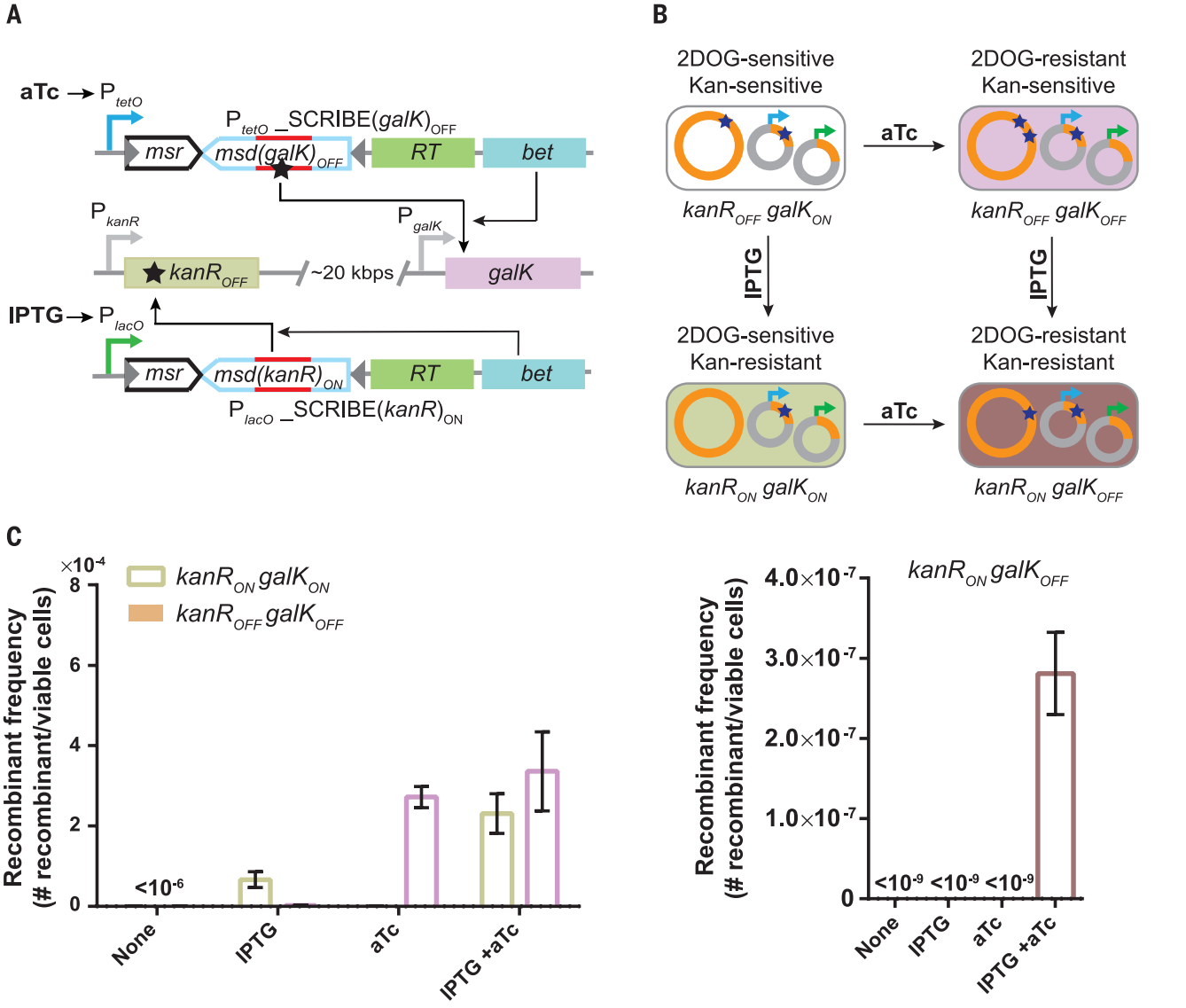


Fig. 3. Writing multiple mutations into independent target loci within population. (**A**) Constructs used to target genomic $kanR_{OFF}$ and $galK_{ON}$ loci with IPTG-inducible and aTc-inducible SCRIBE cassettes, respectively. (**B**) Induction of $kanR_{OFF}$ $galK_{ON}$ cells with IPTG or aTc generates cells with the $kanR_{ON}$ $galK_{ON}$ or $kanR_{OFF}$ $galK_{OFF}$ genotypes, respectively. Induction of $kanR_{OFF}$ $galK_{ON}$ cells with both IPTG and aTc generates cells with the $kanR_{ON}$ $galK_{OFF}$ genotype. (**C**) $kanR_{OFF}$ $galK_{ON}$ reporter cells containing the circuits in (A) were induced with different combinations of IPTG (1 mM) and aTc (100 ng/ml) for 24 hours at 30°C, and the fraction of cells with the various genotypes were determined by plating the cells on appropriate selective media. Error bars indicate SEM for three independent biological replicates.

The linear increase in the fraction of recombinants in the induced cell populations over time was consistent with a deterministic model (dashed lines in Fig. 4B, also see supplementary materials). Specifically, when triggered by inputs, SCRIBE can substantially increase the rate of recombination events at a specific target site above the WT rate [which is reportedly $<10^{-10}$ events per generation in *recA*$^-$ background (*30*)]. When recombination rates are ~10^{-4} events per generation, which is consistent with the recombination rate estimated for SCRIBE from data in Fig. 4B, a simple deterministic model and a detailed stochastic simulation both predict a linear increase in the frequency of recombinant alleles in a population over time, as long as this frequency is less than a few percent and cells in the population are equally fit over the time scale of interest (see supplementary materials and figs. S3 and S4). These models enable one to determine the ideal range of recombination rates for a given application, which depends on parameters such as the frequency of dilution, the sensitivity of the method used for reading the memory, the desired input duration to be recorded, and so forth. For example, recombination rates that are too low would be challenging to quantify and could result in loss of memory if the cultures were diluted. Moreover, higher recombination rates lead to more rapid saturation of memory capacity in which the system is unable to provide a straightforward linear relation between the input exposure time and the state of the memory (fig. S3). Thus, intermediate levels of recombination rates are desirable for population-level analog memory units that can record the time span of exposure to inputs (see supplementary materials).

Decoupling memory operations

SCRIBE memory can be used to create more complex synthetic memory circuits. To demonstrate this, we first built a synthetic gene circuit that can record different input magnitudes into DNA memory. The memory state can then be read out later (after the initial input is removed) upon addition of a secondary signal. Specifically, we built an IPTG-inducible $lacZ_{OFF}$ ($lacZ_{A35TAA,\ S36TAG}$) reporter construct in DH5αPRO cells (Fig. 5A). Expression of this reporter is normally repressed except when IPTG ("read" signal, Fig. 5A) is added, thus enabling a convenient and switchable population-level readout of the memory based on total LacZ activity (Fig. 5B). The $lacZ_{OFF}$

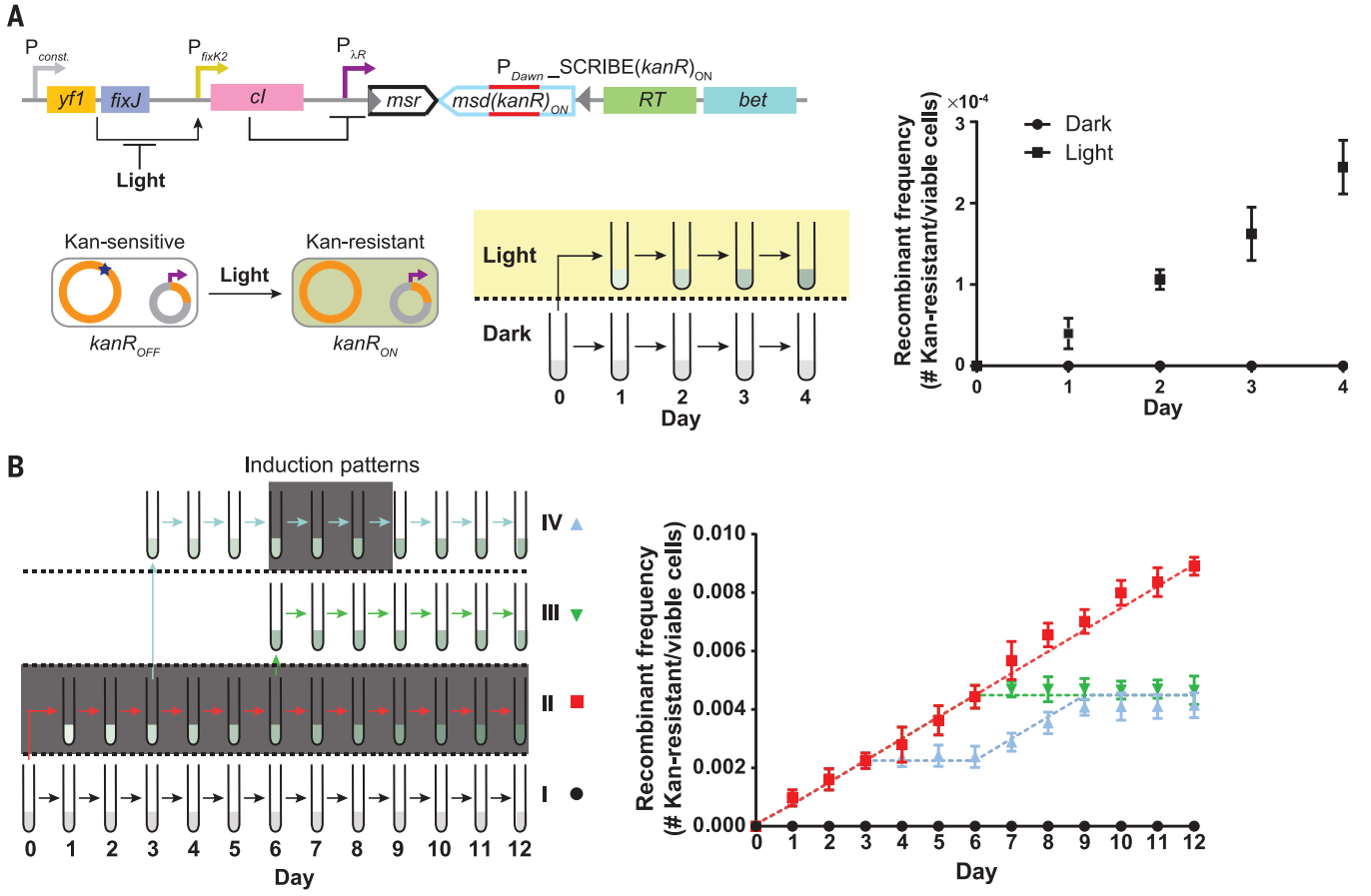


Fig. 4. Optogenetic genome editing and analog memory for long-term recording of input signal exposure times in the genomic DNA of living cell populations. (**A**) We coupled expression of SCRIBE(*kanR*)$_{ON}$ to an optogenetic system (P_{Dawn}). The *yf1/fixJ* synthetic operon was expressed from a constitutive promoter: In dark conditions, YF1 interacts with and phosphorylates FixJ. Phosphorylated FixJ activates the P_{fixK2} promoter, which drives λ repressor (*cI*) expression, which subsequently represses the SCRIBE(*kanR*)$_{ON}$ cassette. Light inhibits the interaction between YF1 and FixJ, leading to the generation of ssDNA(*kanR*)$_{ON}$ and Beta expression and, thus, the conversion of $kanR_{OFF}$ to $kanR_{ON}$. Cells harboring this circuit were grown overnight at 37°C in the dark, diluted 1:1000, and then incubated for 24 hours at 30°C in the dark (no shading) or in the presence of light (yellow shading). Subsequently, cells were diluted by 1:1000 and grown for another 24 hours at 30°C in the dark or in the presence of light. The dilution-regrowth cycle was performed for four consecutive days. The *kanR* allele frequencies in the populations were determined by sampling the cultures after each 24-hour period. (**B**) SCRIBE analog memory records the total time exposure to a given input, regardless of the underlying induction pattern. Cells harboring the circuit shown in Fig. 1C were grown in four different patterns (I to IV) over a 12-day period, where induction by IPTG (1 mM) and aTc (100 ng/mL) is represented by dark gray shading. At the end of each 24-hour incubation period, cells were diluted by 1:1000 into fresh media. The frequency of Kan-resistant cells in the cultures was determined at the end of each day. Dashed lines represent the recombinant allele frequencies predicted by the model (see supplementary materials). Error bars indicate SEM for three independent biological replicates.

reporter cells were transformed with a plasmid encoding an aTc-inducible SCRIBE(*lacZ*)$_{ON}$ cassette (Fig. 5A). Overnight cultures were diluted and induced with various amounts of aTc to write the genomic memory (Fig. 5B). These cells were grown up to saturation and then diluted into fresh media in the presence or absence of IPTG to read the genomic memory (Fig. 5B). In the absence of IPTG, the total LacZ activity remained low, regardless of the aTc concentration. In the presence of IPTG, cultures that had been exposed to higher aTc concentrations had greater total LacZ activity. These results show that population-level reading of genomically encoded memory can be decoupled from writing and controlled externally. Furthermore, this circuit enables the magnitude of the inducer (aTc) to be stably recorded in the distributed genomic memory of a cellular population. Independent control over the read memory operation as shown in this experiment could help to minimize fitness costs associated with the expression of reporter genes until needed.

We have shown that (i) both ssDNA expression and Beta are required for writing into genomic memory (Fig. 1C), (ii) multiple ssDNAs can be used to independently address different memory units (Fig. 3), and (iii) genomic memory is stably recorded into DNA and can be used to modify functional genes whose expression can be controlled by external inducers (Figs. 1 to 4). Thus, SCRIBE memory units can be conceptually decomposed into separate "input," "write," and "read" operations to facilitate greater control and the integration of logic with memory. The separation of these signals could enable master control over the writing of multiple independent inputs into genomic memory. To achieve this, we placed the msd(*lacZ*)$_{ON}$ cassette under the control of an acyl homoserine lactone (AHL)–inducible promoter (P_{luxR}) (*31*) and cotransformed this plasmid with an aTc-inducible Beta-expressing plasmid into the $lacZ_{OFF}$ reporter strain (Fig. 5D). Using this design, information on the input (ssDNA expression via addition of AHL) can be written into DNA memory only in the presence of the write signal (Beta expression via addition of aTc). The information recorded in the memory register (i.e., the state of *lacZ* across the population) can be retrieved by adding the read signal (IPTG).

To demonstrate this, overnight $lacZ_{OFF}$ cultures harboring the circuit shown in Fig. 5D were diluted and then grown to saturation in the presence of all four possible combinations of AHL and aTc (Fig. 5E). The saturated cultures were then diluted into fresh media in the absence or presence of IPTG. As shown in Fig. 5F, only cultures that had been exposed to both the input and write signals simultaneously showed substantial LacZ activity, and only when they were

induced with the read signal. These results indicate that short stretches of DNA of living organisms can be used as addressable read/write memory registers to record transcriptional inputs. Furthermore, SCRIBE memory can be combined with logic, such as the AND function between the input and write signals shown here. The logic in Fig. 5D enables this circuit to act as a "sample-and-hold" system in which information about an input can be recorded in the presence of another signal and read out at will. Additional inputs in the form of orthogonal ssDNAs under the control

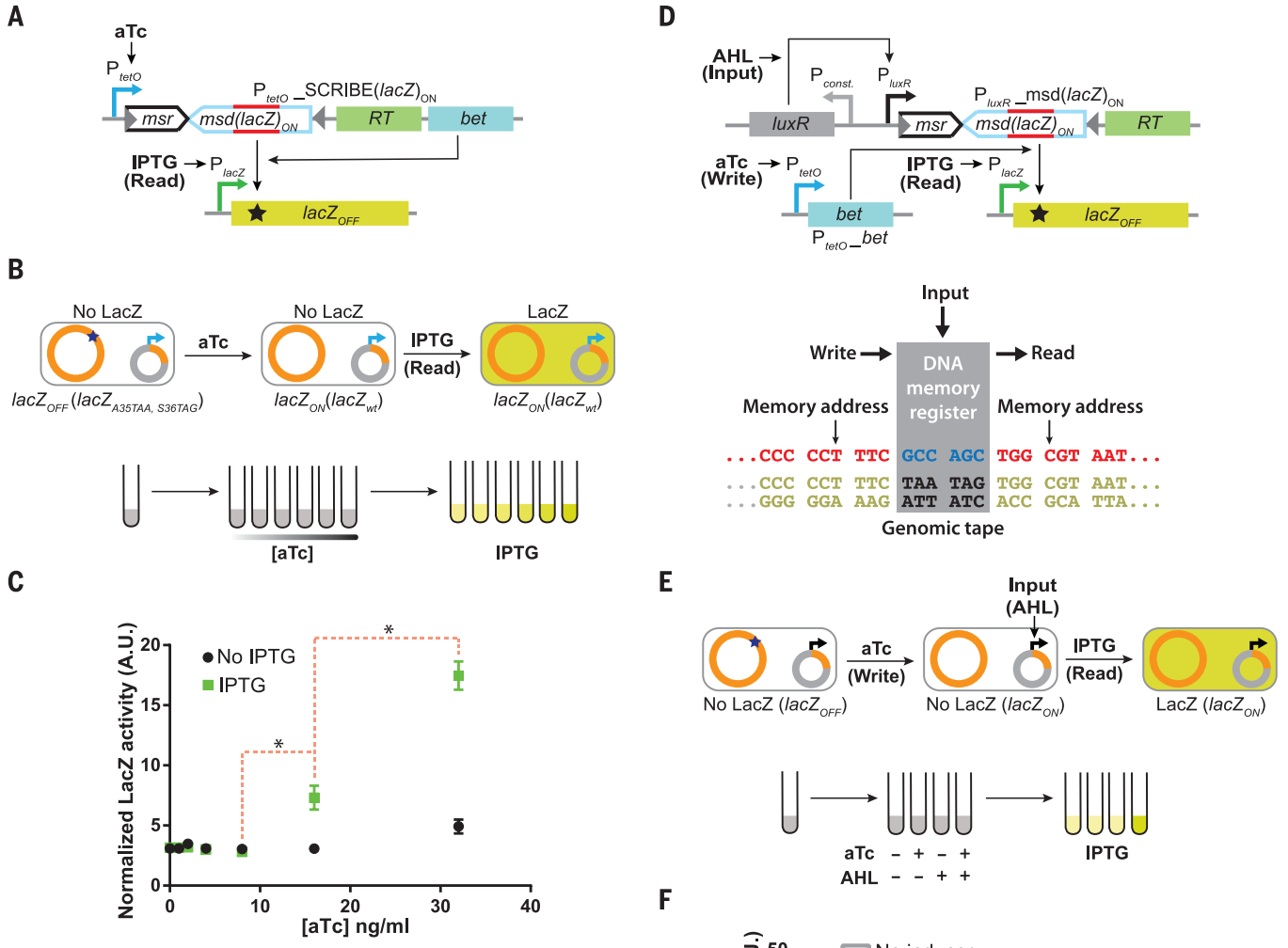


Fig. 5. SCRIBE memory operations can be decoupled into independent "input," "write," and "read" operations, thus facilitating greater control over addressable memory registers in genomic tape recorders and the creation of sample-and-hold circuits. (**A**) We built a circuit where information about the first inducer (aTc) is recorded in the population, which can then be read later upon addition of a second inducer (IPTG) that triggers a read operation. We created an IPTG-inducible $lacZ_{OFF}$ locus in the DH5αPRO background, which contains the full-length *lacZ* gene with two premature stop codons inside the open-reading frame. Expression of ssDNA$(lacZ)_{ON}$ from the aTc-inducible SCRIBE$(lacZ)_{ON}$ cassette results in the reversion of the stop codons inside $lacZ_{OFF}$ to yield the $lacZ_{ON}$ genotype. (**B**) Cells harboring the circuit shown in (A) were grown in the presence of different levels of aTc for 24 hours at 30°C to enable recording into genomic DNA. Subsequently, cell populations were diluted into fresh media without or with IPTG (1 mM) and incubated at 37°C for 8 hours. (**C**) Total LacZ activity in these cultures was measured using a fluorogenic *lacZ* substrate (FDG) assay. The red dashed brackets marked with asterisks connect the closest data points of IPTG-induced samples that are statistically significant ($P < 0.05$ based on one-tailed Welch's *t* test). A.U., arbitrary units. (**D**) We extended the circuit in (A) to create a sample-and-hold circuit where input, write, and read operations are independently controlled. This feature enables the creation of addressable read/write memory registers in the genomic DNA tape. Induction of cells with the input signal (AHL) produces ssDNA$(lacZ)_{ON}$, which targets the genomic $lacZ_{OFF}$ locus for reversion to the WT sequence. In the presence of the write signal (aTc), which expresses Beta, ssDNA$(lacZ)_{ON}$ is recombined into the $lacZ_{OFF}$ locus and produces the $lacZ_{ON}$ genotype. Thus, the write signal enables the input signal to be sampled and held in memory. The total LacZ activity in the cell populations is retrieved by adding the read signal (IPTG). (**E**) Cells harboring the circuit shown in (D) were induced with different combinations of aTc (100 ng/ml) and AHL (50 ng/ml) for 24 hours, after which the cultures were diluted in fresh media with or without IPTG (1 mM). These cultures were then incubated at 37°C for 8 hours and assayed for total LacZ activity with the FDG assay. (**F**) Cell populations that received both the input and write signals followed by the read signal exhibited enhanced levels of total LacZ activity. Error bars indicate SEM for three independent biological replicates.

of other inducible promoters (e.g., Fig. 3), could be written into genomic memory only when the write signal (Beta expression) is present. Thus, SCRIBE memory units can be readily reprogrammed, integrated with logic circuits, and decomposed into independent input, write, and read operations. We anticipate that more complex logic circuits could be combined with SCRIBE-based memory to create analog memory and computation systems capable of storing the results of multi-input calculations (*32*, *33*).

Discussion

We described a scalable platform that uses genomic DNA for analog, rewritable, and flexible memory distributed across living cell populations. One current limitation is the number of orthogonal inducible promoters that can be used as inputs, but this could be addressed by the development of additional inducible transcriptional regulatory devices (*34*). Additionally, ssDNA expression can be coupled to endogenous promoters to sense and record native cellular events. Although we primarily targeted mutations into functional genes to facilitate convenient functional and reporter assays, natural or synthetic noncoding DNA segments could also be used to record memory within genomic DNA. The recorded memory could then be read by high-throughput sequencing (fig. S2). A potential benefit of using synthetic DNA segments as memory registers is the ability to introduce mutations for memory storage that are neutral in terms of fitness costs.

SCRIBE enables conditional increases in the recombination rate at specific loci beyond background levels. The maximum observed recombination rate of the current SCRIBE platform (~10^{-4} recombination events per generation) is suitable for long-term recording of analog memory distributed across the collective genomes of cellular populations (fig. S3). However, it is not high enough to allow recording of digital information and efficient genome editing at the single-cell level. In principle, population-level analog memory could be achieved by other types of DNA memory switches, such as site-specific recombinases, if they were tuned to achieve intermediate recombination rates. Further investigation is required to determine the exact mechanisms involved in processing retron-based ssDNAs for recombination into genomic DNA and the effects of different growth conditions on SCRIBE memory. Because Beta-mediated recombination is replication-dependent (*27*–*29*) and ssDNA is believed to be recombined into the genome during passage of the replication fork (*27*), we speculate that only actively dividing cells are likely to participate in the described population-level memory. Future optimization of SCRIBE [e.g., by modulating the mismatch repair system (*14*) and cellular exonucleases (*35*)] could lead to more efficient single-cell digital memories. This could enable other useful applications, including recording extracellular and intracellular events at the single-cell level for biological studies, dynamic engineering of cellular phenotypes, experimental evolution and population dynamics studies, single-cell computation and memory, the construction of complex cellular state machines and biological Turing machines, and enhanced genome engineering techniques.

Additionally, because retrons have been found in a diverse range of microorganisms (*20*), in vivo ssDNA expression could potentially be extended to hard-to-transform organisms in which SCRIBE plasmids could be introduced by conjugation or transduction. Because retrons have also been shown to be functional in eukaryotes (*24*, *36*, *37*), they could potentially be used with other genome editing tools for memory. Moreover, by using error-prone RNA polymerases (*38*) and reverse transcriptases (*39*, *40*), we anticipate that mutagenized ssDNA libraries could be generated inside cells for in vivo continuous evolution (*41*) and cellular barcoding applications. Finally, in vivo ssDNA generation could potentially be used to create DNA nanosystems (*42*–*48*) and ssDNA-protein hybrid nanomachines in living cells (*49*) or could be optimized and scaled-up to create an economical source of ssDNAs for DNA nanotechnology (*50*). In summary, we envision that in vivo ssDNA production and SCRIBE platforms will open up a broad range of new capabilities for engineering biology.

Materials and methods

Strains and plasmids

Conventional cloning methods were used to construct the plasmids. Lists of strains and plasmids used in this study and the construction procedures are provided in tables S1 and S2, respectively. The sequences for the synthetic parts and primers are provided in tables S3 and S4.

Cells and antibiotics

Chemically competent *E. coli* DH5α was used for cloning. Unless otherwise noted, antibiotics were used at the following concentrations: carbenicillin (50 μg/ml), kanamycin (20 μg/ml), chloramphenicol (30 μg/ml), and spectinomycin (100 μg/ml). In the experiment shown in Fig. 2, kanamycin (15 μg/ml) and chloramphenicol (15 μg/ml) were used.

Detection of single-stranded DNA

Overnight cultures harboring IPTG-inducible plasmids encoding msd(wt), msd(wt) with deactivated RT [msd(wt)_dRT], or msd($kanR$)$_{ON}$ were grown overnight with or without IPTG (1 mM). Total RNA samples were prepared from non-induced or induced cultures using TRIzol reagent (Invitrogen) according to the manufacturer's protocol. Total RNA (10 μg) from each sample was treated with ribonuclease A (37°C, 2 hours) to remove RNA species and the msr moiety. The samples were then resolved on 10% tris-borate EDTA–urea denaturing gel and visualized with SYBR-Gold. A polyacrylamide gel electrophoresis–purified synthetic oligo (FF_oligo347, 50 pmol) with the same sequence as ssDNA(wt) was used as a molecular size marker. The band intensities were measured by Fiji software (*51*). The intensities were normalized to the intensity of the marker oligo, and normalized intensities were used to calculate the amount of ssDNA in each sample.

Induction of cells and plating assays

For each experiment, three transformants were separately inoculated in Luria broth (LB) media plus appropriate antibiotics and grown overnight [37°C, 700 revolutions per minute (RPM)] to obtain seed cultures. Unless otherwise noted, inductions were performed by diluting the seed cultures (1:1000) in 2 ml of prewarmed LB plus appropriate antibiotics with or without inducers followed by 24 hours incubation (30°C, 700 RPM). Aliquots of the samples were then serially diluted, and appropriate dilutions were plated on selective media to determine the number of recombinants and viable cells in each culture. For each sample, the recombinant frequency was reported as the mean of the ratio of recombinants to viable cells for three independent replicates.

In all experiments, the number of viable cells was determined by plating aliquots of cultures on LB-plus-spectinomycin plates. LB-plus-kanamycin plates were used to determine the number of recombinants in the *kanR* reversion assay. For the *galK* reversion assay (Fig. 2), the numbers of $galK_{ON}$ recombinants were determined by plating the cells on MOPS EZ rich–defined media (Teknova) plus galactose (0.2%). The numbers of $galK_{OFF}$ recombinants were determined by plating the cells on MOPS EZ rich–defined media plus glycerol (0.2%) plus 2-DOG (2%). For the experiment shown in Fig. 3, the numbers of $kanR_{ON}$ $galK_{ON}$ and $kanR_{OFF}$ $galK_{OFF}$ cells were determined by using LB-plus-kanamycin plates and MOPS EZ rich–defined media plus glycerol (0.2%), 2-DOG (2%), and D-biotin (0.01%), respectively. The numbers of $kanR_{ON}$ $galK_{OFF}$ cells were determined by plating the cells on MOPS EZ–rich defined media plus glycerol (0.2%), 2-DOG (2%), kanamycin, and D-biotin (0.01%).

For the light-inducible SCRIBE experiment (Fig. 4A), induction was performed with white light (using the built-in fluorescent lamp in a VWR 1585 shaker incubator). The "dark" condition was achieved by wrapping aluminum foil around the tubes. Growth of and sampling from these cultures were performed as described earlier.

LacZ assay

Overnight seed cultures were diluted (1:1000) in prewarmed LB plus appropriate antibiotics and inducers [with different concentrations of aTc or without aTc (Fig. 5, A to C) and with all the four possible combinations of aTc (100 ng/ml) and AHL (50 ng/ml) (Fig. 5, D to F)] and incubated for 24 hours (30°C, 700 RPM). These cultures were then diluted (1:50) in prewarmed LB plus appropriate antibiotics with or without IPTG (1 mM) and incubated for 8 hours (37°C, 700 RPM). To measure LacZ activity, 60 μl of each culture was mixed with 60 μl of B-PER II reagent (Pierce Biotechnology) and fluorescein di-β-D-galactopyranoside (FDG) (0.05 mg/ml final concentration). The fluorescence signal (absorption/emission: 485/515) was monitored in a plate reader with continuous shaking for 2 hours. The LacZ activity was calculated by

normalizing the rate of FDG hydrolysis (obtained from fluorescence signal) to the initial optical density. For each sample, LacZ activity was reported as the mean of three independent biological replicates.

REFERENCES AND NOTES

1. N. Goldman *et al*., Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **494**, 77–80 (2013). doi: 10.1038/nature11875; pmid: 23354052
2. G. M. Church, Y. Gao, S. Kosuri, Next-generation digital information storage in DNA. *Science* **337**, 1628 (2012). doi: 10.1126/science.1226355; pmid: 22903519
3. M. C. Inniss, P. A. Silver, Building synthetic memory. *Curr. Biol.* **23**, R812–R816 (2013). doi: 10.1016/j.cub.2013.06.047; pmid: 24028965
4. T. S. Gardner, C. R. Cantor, J. J. Collins, Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000). doi: 10.1038/35002131; pmid: 10659857
5. J. W. Kotula *et al*., Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4838–4843 (2014). doi: 10.1073/pnas.1321321111; pmid: 24639514
6. D. R. Burrill, M. C. Inniss, P. M. Boyle, P. A. Silver, Synthetic memory circuits for tracking human cell fate. *Genes Dev.* **26**, 1486–1497 (2012). doi: 10.1101/gad.189035.112; pmid: 22751502
7. B. P. Kramer *et al*., An engineered epigenetic transgene switch in mammalian cells. *Nat. Biotechnol.* **22**, 867–870 (2004). doi: 10.1038/nbt980; pmid: 15184906
8. C. M. Ajo-Franklin *et al*., Rational design of memory in eukaryotic cells. *Genes Dev.* **21**, 2271–2276 (2007). doi: 10.1101/gad.1586107; pmid: 17875664
9. P. Siuti, J. Yazbek, T. K. Lu, Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31**, 448–452 (2013). doi: 10.1038/nbt.2510; pmid: 23396014
10. J. Bonnet, P. Subsoontorn, D. Endy, Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8884–8889 (2012). doi: 10.1073/pnas.1202344109; pmid: 22615351
11. 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; pmid: 23539178
12. A. E. Friedland *et al*., Synthetic gene networks that count. *Science* **324**, 1199–1202 (2009). doi: 10.1126/science.1172005; pmid: 19478183
13. H. M. Ellis, D. Yu, T. DiTizio, D. L. Court, High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6742–6746 (2001). doi: 10.1073/pnas.121164898; pmid: 11381128
14. N. Costantino, D. L. Court, Enhanced levels of λ Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15748–15753 (2003). doi: 10.1073/pnas.2434959100; pmid: 14673109
15. D. Yu *et al*., An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5978–5983 (2000). doi: 10.1073/pnas.100127597; pmid: 10811905
16. J. A. Sawitzke *et al*., Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *J. Mol. Biol.* **407**, 45–59 (2011). doi: 10.1016/j.jmb.2011.01.030; pmid: 21256136
17. B. Swingle *et al*., Oligonucleotide recombination in Gram-negative bacteria. *Mol. Microbiol.* **75**, 138–148 (2010). doi: 10.1111/j.1365-2958.2009.06976.x; pmid: 19943907
18. S. Datta, N. Costantino, X. Zhou, D. L. Court, Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1626–1631 (2008). doi: 10.1073/pnas.0709089105; pmid: 18230724
19. T. Yee, T. Furuichi, S. Inouye, M. Inouye, Multicopy single-stranded DNA isolated from a gram-negative bacterium, *Myxococcus xanthus*. *Cell* **38**, 203–209 (1984). doi: 10.1016/0092-8674(84)90541-5; pmid: 6088065
20. B. C. Lampson, M. Inouye, S. Inouye, Retrons, msDNA, and the bacterial genome. *Cytogenet. Genome Res.* **110**, 491–499 (2005). doi: 10.1159/000084982; pmid: 16093702
21. D. Lim, W. K. Maas, Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. *Cell* **56**, 891–904 (1989). doi: 10.1016/0092-8674(89)90693-4; pmid: 2466573
22. R. Lutz, H. Bujard, Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997). doi: 10.1093/nar/25.6.1203; pmid: 9092630
23. P. L. Sharma, V. Nurpeisov, R. F. Schinazi, Retrovirus reverse transcriptases containing a modified YXDD motif. *Antivir. Chem. Chemother.* **16**, 169–182 (2005).pmid: 16004080
24. J. R. Mao, M. Shimada, S. Inouye, M. Inouye, Gene regulation by antisense DNA produced in vivo. *J. Biol. Chem.* **270**, 19684–19687 (1995). doi: 10.1074/jbc.270.34.19684; pmid: 7544343
25. S. Warming, N. Costantino, D. L. Court, N. A. Jenkins, N. G. Copeland, Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* **33**, e36 (2005). doi: 10.1093/nar/gni035; pmid: 15731329
26. R. Ohlendorf, R. R. Vidavski, A. Eldar, K. Moffat, A. Möglich, From dusk till dawn: One-plasmid systems for light-regulated gene expression. *J. Mol. Biol.* **416**, 534–542 (2012). doi: 10.1016/j.jmb.2012.01.001; pmid: 22245580
27. M. S. Huen *et al*., The involvement of replication in single stranded oligonucleotide-mediated gene repair. *Nucleic Acids Res.* **34**, 6183–6194 (2006). doi: 10.1093/nar/gkl852; pmid: 17088285
28. A. R. Poteete, Involvement of DNA replication in phage lambda Red-mediated homologous recombination. *Mol. Microbiol.* **68**, 66–74 (2008). doi: 10.1111/j.1365-2958.2008.06133.x; pmid: 18333884
29. A. R. Poteete, Involvement of *Escherichia coli* DNA replication proteins in phage Lambda Red-mediated homologous recombination. *PLOS ONE* **8**, e67440 (2013). doi: 10.1371/journal.pone.0067440; pmid: 23840702
30. B. E. Dutra, V. A. Sutera Jr., S. T. Lovett, RecA-independent recombination is efficient but limited by exonucleases. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 216–221 (2007). doi: 10.1073/pnas.0608293104; pmid: 17182742
31. S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, R. Weiss, A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–1134 (2005). doi: 10.1038/nature03461; pmid: 15858574
32. S. Ausländer, D. Ausländer, M. Müller, M. Wieland, M. Fussenegger, Programmable single-cell mammalian biocomputers. *Nature* **487**, 123–127 (2012).pmid: 22722847
33. R. Daniel, J. R. Rubens, R. Sarpeshkar, T. K. Lu, Synthetic analog computation in living cells. *Nature* **497**, 619–623 (2013). doi: 10.1038/nature12148; pmid: 23676681
34. E. J. Olson, L. A. Hartsough, B. P. Landry, R. Shroff, J. J. Tabor, Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nat. Methods* **11**, 449–455 (2014). doi: 10.1038/nmeth.2884; pmid: 24608181
35. J. A. Mosberg, C. J. Gregg, M. J. Lajoie, H. H. Wang, G. M. Church, Improving lambda red genome engineering in Escherichia coli via rational removal of endogenous nucleases. *PLOS ONE* **7**, e44638 (2012). doi: 10.1371/journal.pone.0044638; pmid: 22957093
36. S. Miyata, A. Ohshima, S. Inouye, M. Inouye, In vivo production of a stable single-stranded cDNA in *Saccharomyces cerevisiae* by means of a bacterial retron. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5735–5739 (1992). doi: 10.1073/pnas.89.13.5735; pmid: 1378616
37. O. Mirochnitchenko, S. Inouye, M. Inouye, Production of single-stranded DNA in mammalian cells by means of a bacterial retron. *J. Biol. Chem.* **269**, 2380–2383 (1994).pmid: 7507924
38. S. Brakmann, S. Grzeszik, An error-prone T7 RNA polymerase mutant generated by directed evolution. *ChemBioChem* **2**, 212–219 (2001). doi: 10.1002/1439-7633(20010302)2:3<212::AID-CBIC212>3.0.CO;2-R; pmid: 11828447
39. J. D. Roberts, K. Bebenek, T. A. Kunkel, The accuracy of reverse transcriptase from HIV-1. *Science* **242**, 1171–1173 (1988). doi: 10.1126/science.2460925; pmid: 2460925
40. K. Bebenek, J. Abbotts, S. H. Wilson, T. A. Kunkel, Error-prone polymerization by HIV-1 reverse transcriptase. Contribution of template-primer misalignment, miscoding, and termination probability to mutational hot spots. *J. Biol. Chem.* **268**, 10324–10334 (1993).pmid: 7683675
41. K. M. Esvelt, J. C. Carlson, D. R. Liu, A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011). doi: 10.1038/nature09929; pmid: 21478873
42. Y. Amir *et al*., Universal computing by DNA origami robots in a living animal. *Nat. Nanotechnol.* **9**, 353–357 (2014). doi: 10.1038/nnano.2014.58; pmid: 24705510
43. L. Qian, E. Winfree, J. Bruck, Neural network computation with DNA strand displacement cascades. *Nature* **475**, 368–372 (2011). doi: 10.1038/nature10262; pmid: 21776082
44. G. Seelig, D. Soloveichik, D. Y. Zhang, E. Winfree, Enzyme-free nucleic acid logic circuits. *Science* **314**, 1585–1588 (2006). doi: 10.1126/science.1132493; pmid: 17158324
45. P. W. Rothemund, Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006). doi: 10.1038/nature04586; pmid: 16541064
46. S. M. Douglas *et al*., Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009). doi: 10.1038/nature08016; pmid: 19458720
47. S. M. Douglas, I. Bachelet, G. M. Church, A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335**, 831–834 (2012). doi: 10.1126/science.1214081; pmid: 22344439
48. S. M. Chirieleison, P. B. Allen, Z. B. Simpson, A. D. Ellington, X. Chen, Pattern transformation with DNA circuits. *Nat. Chem.* **5**, 1000–1005 (2013). doi: 10.1038/nchem.1764; pmid: 24256862
49. C. A. Brosey *et al*., A new structural framework for integrating replication protein A into DNA processing machinery. *Nucleic Acids Res.* **41**, 2313–2327 (2013). doi: 10.1093/nar/gks1332; pmid: 23303776
50. S. Kosuri, G. M. Church, Large-scale de novo DNA synthesis: Technologies and applications. *Nat. Methods* **11**, 499–507 (2014). doi: 10.1038/nmeth.2918; pmid: 24781323
51. J. Schindelin *et al*., Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012). doi: 10.1038/nmeth.2019; pmid: 22743772

ACKNOWLEDGMENTS

This work was supported by the NIH New Innovator Award (1DP2OD008435), NIH National Centers for Systems Biology (1P50GM098792), the U.S. Office of Naval Research (N000141310424), and the Defense Advanced Research Projects Agency. Sequencing data have been deposited in GenBank with accession numbers KM923743 to KM923754.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6211/1256272/suppl/DC1
Supplementary Text
Figs. S1 to S6
Tables S1 to S4
References (*52*–*61*)

19 May 2014; accepted 16 October 2014
10.1126/science.1256272

Supplementary Materials for

Genomically encoded analog memory with precise in vivo DNA writing in living cell populations

Fahim Farzadfard and Timothy K. Lu*

*Corresponding author. E-mail: timlu@mit.edu

Published 14 November 2014, Science **346**, 1256272 (2014)
DOI: 10.1126/science.1256272

This PDF file includes:

Supplementary Text

High-Throughput Sequencing of Genomically Encoded Memory

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

The obtained reads were processed and demultiplexed by the MIT BMC-BCC Pipeline. These reads then were trimmed to remove the added 10 bp randomized sequence. To filter out any reads that could have been produced by non-specific binding of primers during PCR, we discarded reads that lacked the expected "CGCGNNNNNATTT" motif, where "NNNNN" corresponds to the 5 base-pair *kanR* memory register. Furthermore, any reads that contained ambiguous bases within this 5 base-pair memory register were discarded. The frequencies of the obtained variants (either GGCCC ($kanR_{ON}$) or CTATT ($kanR_{OFF}$), which constitute the two states of the *kanR* memory register (Fig. 1C)), were then calculated for each sample.

As shown in Fig. S2A, the frequency of reads mapping to $kanR_{ON}$ in the induced samples expressing ssDNA($kanR$)$_{ON}$ was comparable to the frequency of Kan-resistant colonies obtained from the plating assay in the *KanR* reversion assay (Fig. 1C). Very few reads mapping to ssDNA($kanR$)$_{ON}$ were observed in the non-induced samples. Interestingly, a few reads mapping to ssDNA($kanR$)$_{ON}$ were observed in induced samples expressing ssDNA($kanR$)$_{OFF}$. To better understand the source of these reads we analyzed the variants observed in the 5 bp *kanR* memory register. These variants and their corresponding frequencies are shown for one representative sample for P_{lacO}_msd($kanR$)$_{OFF}$ + P_{tetO}_*bet* + IPTG + aTc Rep#1 in Fig. S2B. In all the samples, less than 25 variants out of the total 4^5=1024 possible variants were observed. Reads mapping exactly to $kanR_{OFF}$ constituted the majority of reads, as expected. Reads with single or two bp mutations relative to $kanR_{OFF}$ were observed in all the samples, with frequencies ranging from 10^{-7}-10^{-3}. These reads were likely produced by the relatively high mutation rate of high-throughput sequencing (*52*) or during library preparation steps. We did not observe any reads with more than 2 bps of mismatch to both $kanR_{ON}$ and $kanR_{OFF}$. In the negative control sample of Fig. S2B (in which ssDNA($KanR$)$_{OFF}$ was expressed and no $kanR_{ON}$ sequence was present), the absence of reads with 3 or 4 mismatches to $kanR_{OFF}$ suggests that the

observed $kanR_{ON}$ reads were likely an artifact of multiplexed sequencing, such as barcode misassignment or recombination during the sequencing protocol.

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

Modeling and Simulation

Deterministic Model

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Overall, our model predicts a linear increase (with a slope = $\frac{r}{2}$) in the recombinant frequency as long as the cells in the population are equally fit and as long as $\frac{r}{2}tf_0$ is relatively small. However, in reality, mutations can occur within populations over time, which can potentially affect the fitness of individual cells. In the absence of recombination in asexual populations, two beneficial mutations that arise independently cannot be combined into a single, superior genotype (*56, 57*). Hence, these carriers could compete with each other, a phenomenon known as clonal interference that is important in shaping the evolutionary trajectory of large asexual populations with high mutation rates over prolonged growth. Under these circumstances, the model assumption that all the cells in the population are equally fit does not hold and deviation from the model is expected. However, since the natural rate of beneficial mutations is low (~10^{-9} per bp per generation for *E. coli (57)*), the probability of mutations with significant fitness effects and clonal interference is relatively low, at least over the timescales of our experiments. Similarly, a linear increase in mutant frequencies during exponential growth of a bacterial culture was previously predicted (*58, 59*).

Stochastic Simulations

To further validate the model, we performed stochastic simulations of a growing bacterial population with three different recombination rates (r=10^{-9}, 0.00015, or 0.005 events/generation) for 250 generations (Fig. S3). Growth was simulated for 25 serial iterations, with 10 generations in each iteration. The simulation started with a clonal population of bacteria (10^6 cells). During each generation, each cell could stochastically

produce a recombinant allele with a likelihood equal to the recombination rate. The wild-type and recombinant cells were assumed to be equally fit. We also assumed that all the cells in the population followed the same growth profile (no clonal interference). After 10 generations, a sample of ~10^6 cells was taken from the population to start a new culture in order to simulate the serial batch culture procedure.

As shown in Fig. S3A, the model predicts a linear increase in the frequency of recombinants with a very low mutation rate ($r = 10^{-9}$). However, the simulation results were not consistent with the deterministic model; instead, the simulation showed stochastic fluctuations in the recombinant frequency since samples taken after 10 generations may not contain representative numbers of recombinants due to the low recombination rate. This condition is representative of the recombinant frequencies observed in the absence of SCRIBE. Major recombination pathways in *E. coli* are *recA*-dependent and knocking out RecA activity can severely affect the recombination rate (*30, 54*). In a recombination-deficient background (*recA*$^-$), such as DH5α, recombination is a very rare, stochastic event (<10^{-10} events/generation (*30, 54*)). These data are consistent with Fig. 4B, where no significant increase in recombinant frequencies was observed in the absence of SCRIBE activation (induction pattern I).

In contrast, at a higher targeted recombination rate ($r = 0.00015$ events/generation), a linear increase in the frequency of recombinants is predicted by both the model and simulation (Fig. S3B). This rate is representative of cells containing a specific locus targeted by SCRIBE memory. SCRIBE enables control over the recombination rate at a specific locus by external inputs, thus increasing the recombination rate by multiple orders of magnitude over the background rate. For example, using data shown in Fig. 4B for cells induced with both aTc and IPTG (induction pattern II), $r = 0.00015$ events/generation was calculated based on the linear regression of the recombination frequency versus generation (Fig. S4). This recombination rate ensures that samples taken from an induced culture contain a representative number of recombinant cells. Thus, successive sampling and regrowth of cells results in the gradual accumulation of recombinants in the population over time in the presence of the inputs (Fig. S3B and Fig. 4B).

Finally, as the recombination rate increases ($r = 0.005$ events/generation, Fig. S3C), the model and simulation predict a linear increase in the recombination frequency at initial times. However, they both start to deviate from the linear approximation as the frequency of recombinants increases (above ~ 5%) since the cultures are increasingly depleted of the wild-type alleles. These models demonstrate that the upper and lower limits of recombination rates that can be used for analog memory depend on the time-span that is desired for recording. In our current configuration, recombination rates lower than 10^{-7} resulted in stochastic fluctuations in the recombination frequency while recombination rates higher than 10^{-2} quickly led to saturated memory that deviated from the linear regime within less than 10 generations (Fig. S5). Intermediate recombination rates (e.g., 10^{-5}-10^{-4} events/generation) enabled the recording of input exposures with a simple linear relationship for hundreds of generation without saturation.

Fig. S5 shows stochastic simulations of populations with 9 different recombination rates, with 10 independent runs for each recombination rate. At very low recombination rates, the number of recombinants in simulated populations increases in a noisy fashion.

However, this increase becomes less noisy as the recombination rate increases. Thus, recombinant frequencies measured after exposure to low recombination rates (e.g., low inducer levels) are likely to have higher relative standard deviations than recombinant frequencies measured after exposure to higher recombination rates (e.g., higher inducer levels). This trend was observed in the data shown in Fig. 1D and is also shown in the simulation results of Fig. S6A. In other words, measurements at low recombination rates are inherently noisier and have lower signal-to-noise ratios than measurements at higher recombination rates (Fig. S5 and S6).

Finally, when starting from a clonal population (or when the initial number of recombinants in the population is negligible), for limited number of generations, the model and simulation both predict a linear increase in the recombinant frequency as a function of the recombination rate (Equation 2, Fig. S6A). However, the recombination rate is not generally a linear function with respect to the concentration of an input inducer, but rather depends on the input-output transfer function of the inducible system. The transfer functions of many commonly used inducible promoters have outputs that undergo a sharp transition within a narrow range of inputs (60). Using promoters that are can be titrated over a wider range of concentrations (*33*) in combination with more sensitive assays (e.g., high-throughput sequencing) could help to achieve a wider input dynamic range with SCRIBE.

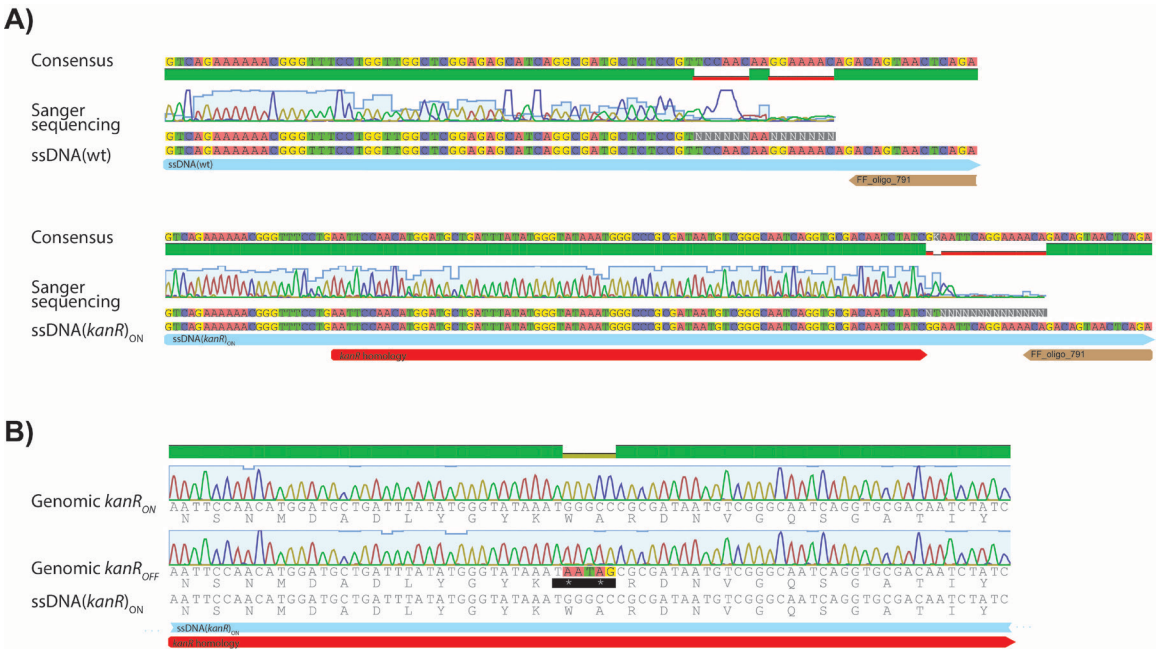


Fig. S1 | Sanger sequencing results for *in vivo* expressed ssDNAs and for recombined target loci.

A) ssDNA bands shown in Fig. 1B were purified and sequenced. The ssDNA($kanR$)$_{ON}$ construct contained the expected engineered DNA sequence (red arrow). **B)** Sanger sequencing for the genomic *kanR* locus. The *kanR* locus in $kanR_{OFF}$ cells as well as the *kanR* locus in Kan-resistant ($kanR_{ON}$) cells obtained from induction of ssDNA($kanR$)$_{ON}$ in $kanR_{OFF}$ reporter cells (Fig. 1C) were PCR amplified and sequenced.

A)

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

B)

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

Fig. S2 | Using high-throughput Illumina HiSeq sequencing to read out the genomically encoded memory at the *kanR* memory register.

(A) Frequency of reads that perfectly match to $kanR_{ON}$ or $kanR_{OFF}$ after writing with SCRIBE. Note that the sequences attributed to $kanR_{ON}$ and $kanR_{OFF}$ here are reverse complemented with respect to the sequences in Fig. 1C. (**B)** Sequencing variants and their corresponding frequencies observed in the 5 bp *kanR* memory register in one representative sample from cells induced to express ssDNA($kanR$)$_{OFF}$ within a genomic $kanR_{OFF}$ background (P_{lacO}_msd($kanR$)$_{OFF}$ + P_{tetO}_*bet* + IPTG + aTc Rep#1).

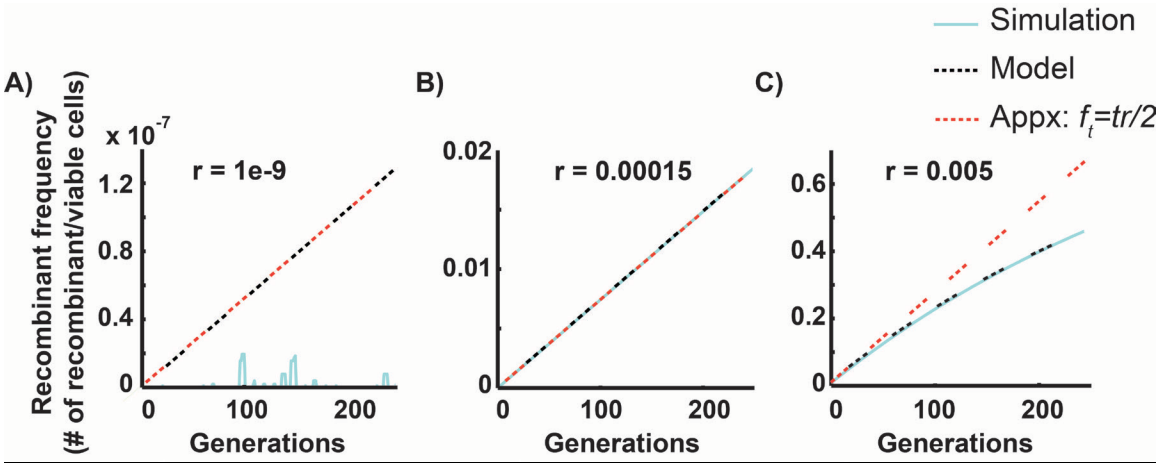


Fig. S3 | Deterministic model, stochastic simulations, and mathematical approximations describing the long-term recording of information into genomically encoded memory with the SCRIBE system at three different recombination rates. **(A)** $r = 10^{-9}$ **(B)** $r = 0.00015$, and **(C)** $r = 0.005$. The parameters in the approximation, f_t, r and t, correspond to the recombinant frequency, recombination rate, and number of generations, respectively.

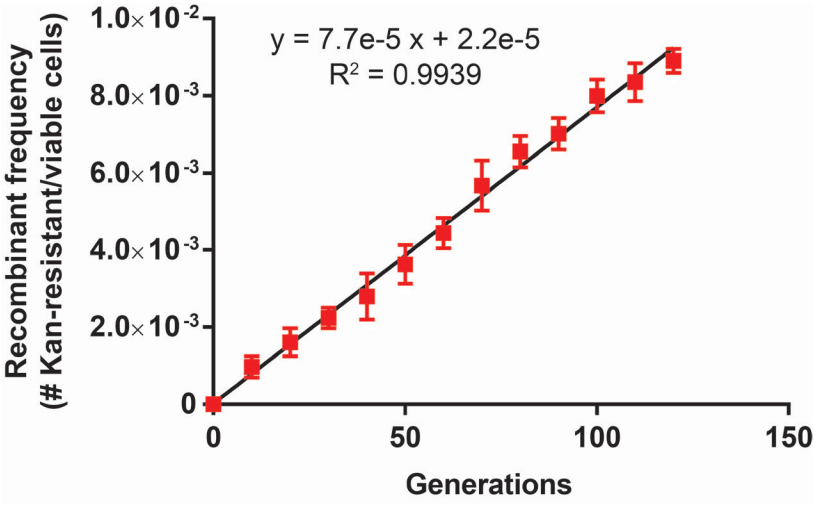


Fig. S4 | Estimating the recombination rate of the SCRIBE system.
The recombination rate for the SCRIBE circuit (shown in Fig. 1C) when the system is induced with both IPTG (1 mM) and aTc (100 ng/ml) was estimated by calculating the slope of the regression line for the data shown in Fig. 4B (induction pattern II) and multiplying that slope by a factor of two as described in the deterministic model ($r = 2\frac{df}{dt} = 2(7.7 * 10^{-5}) = 1.54 * 10^{-4}$). In the experiment shown in Fig. 4B, the cultures were diluted 1:1000 at the beginning of each day and grown to saturation by the end of the day. Thus, the unit of the x-axis in Fig. 4B corresponds to $\log_2(1000) \approx 10$ generations per day.

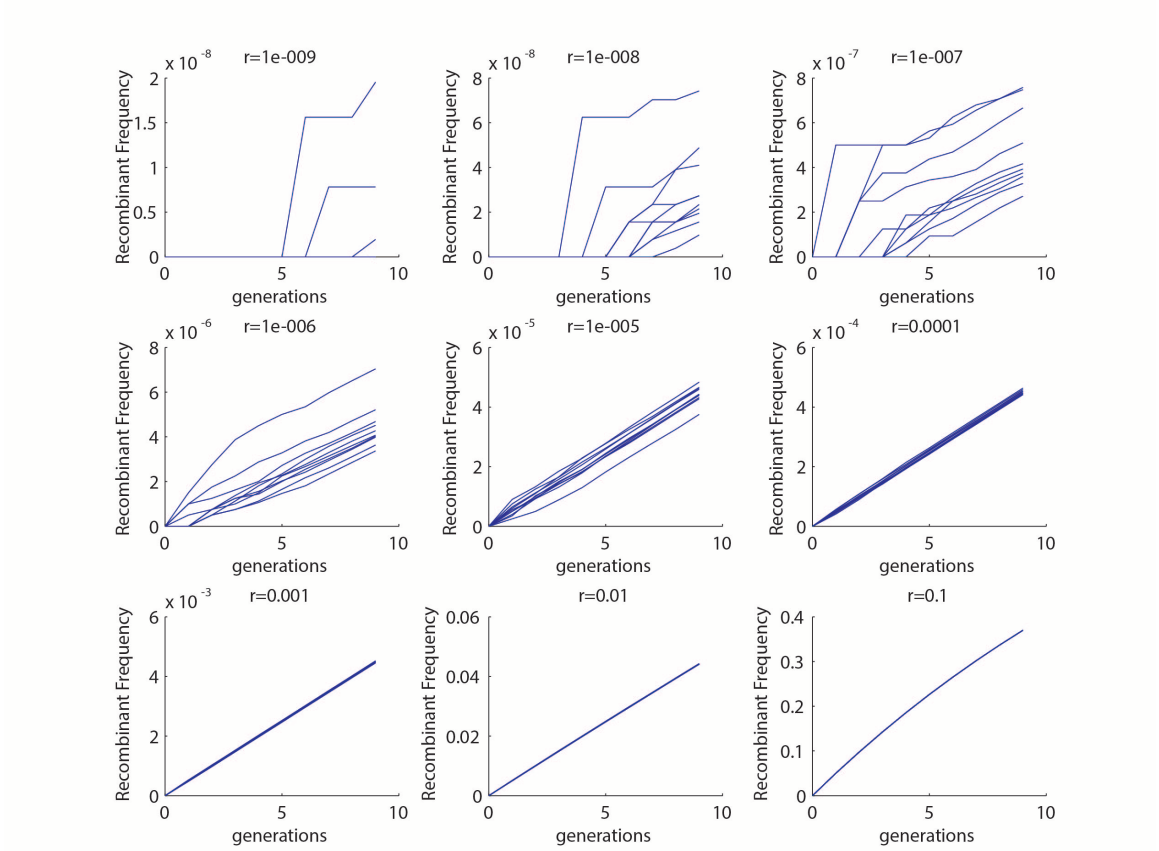


Fig. S5 | Stochastic simulations of analog memory with different recombination rates.

Recombinant frequencies within populations with 9 different recombination rates are shown, with 10 independent runs for each recombination rate. As the recombination rate increases, the increase in the frequency of recombinants in the population becomes less variable.

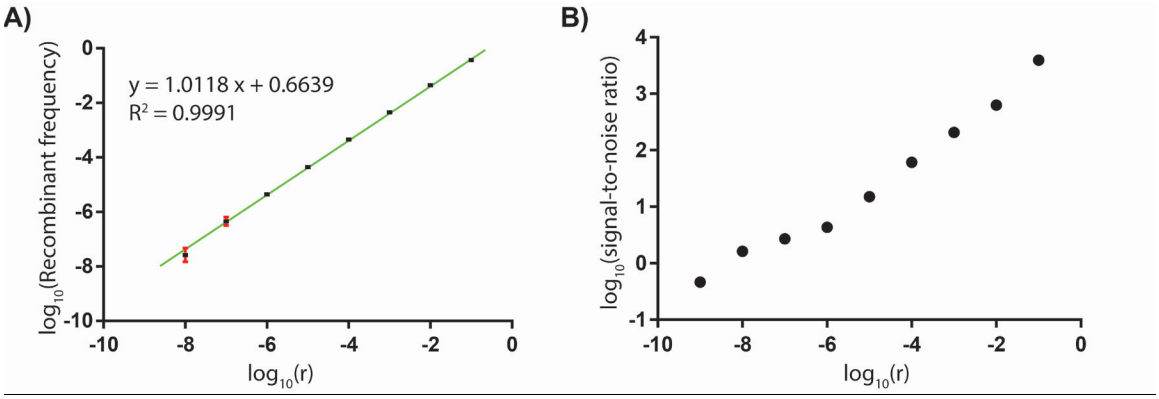


Fig. S6 | The signal-to-noise ratio of the population-level analog memory increases as the recombination rate increases.

(A) The mean of recombinant frequencies *(f)* obtained from the simulation results shown in Fig. S5 (after 9 generations) were plotted on a log-log scale. The error bars (red bars) indicate the standard deviation of the recombinant frequency for 10 independent replicate simulations (n = 10). As the recombination rate increases, the relative standard deviation of the measurements is reduced. **(B)** The signal-to-noise ratio ($\frac{mean(f)}{std(f)}$) of the recombinant frequency increases as the recombination rate increases. Due to the inherent noisiness of the system at low recombination rates, the first three data points are highly variable and new iterations may result in different values.

Table S1 | List of the reporter strains used in this study

Name	Strain Code	Construction method	Genotype	Used In
$kanR_{OFF}$ reporter strain	FFF144	The *kanR* cassette was PCR amplified from the pBT3-SUC (Dualsystems Biotech) plasmid using FF_oligo183 and FF_oligo184 primers followed by a second round of PCR with FF_oligo185 and FF_oligo186 to add additional sequences with homology to the sequences flanking the *galK* locus. The fragment then was integrated into the *galK* locus of a DH5α strain (with an integrated PRO cassette (*22*)) by recombineering. Two premature stop codons then were introduced into this *kanR* cassette using oligo-mediated recombineering with FF_oligo187 to make the $kanR_{OFF}$ strain.	DH5αPRO *galK::kanR*$_{W28TAA, A29TAG}$	Fig. 1 Fig. 4
$kanR_{OFF}$ $galK_{ON}$ reporter strain	FFF774	The $kanR_{OFF}$ cassette was PCR amplified from FFF144 and integrated into the *bioA* locus of DH5α. The cells were then transformed with the PRO plasmid (pZS4Int-LacI/TetR (*22*)).	DH5α *bioA::kanR*$_{W28TAA, A29TAG}$ + PRO plasmid	Fig. 3
galK reporter strain	FFF762	DH5α cells transformed with the PRO plasmid (*22*).	DH5α + PRO plasmid	Fig. 2

$lacZ_{OFF}$ reporter strain	FFF798	The *lacZ* α-fragment was introduced into the DH5α *lacZ* locus by recombineering using a PCR fragment amplified from *E. coli* MG1655 (using FF_oligo1069 and FF_oligo1070). Two premature stop codons were then introduced into the *lacZ* ORF using oligo-mediated recombineering with FF_oligo220 to make the $lacZ_{OFF}$ strain. This strain was then transformed with the PRO plasmid (*22*).	DH5α $lacZ_{A35TAA,\ S36TAG}$ + PRO plasmid	Fig. 5

Table S2 | List of the plasmids used in this study

Name	Plasmid Code	Construction method	Used in
P_{lacO}_msd(wt)	pFF753	The wild-type retron Ec86 cassette was PCR-amplified from *E. coli* BL21 and cloned downstream of the P_{lacO} promoter (PacI and BamHI sites) in the pZE32 (*22*) plasmid.	Fig. 1B
P_{lacO}_msd(wt)_dRT	pFF758	This plasmid was produced by QuikChange site-directed mutagenesis (using FF_oligo912 and FF_oligo913) to mutate the YADD active site of the RT to YAAA (D197A and D198A mutations) in the P_{lacO}_msd(wt) plasmid.	Fig. 1B
P_{lacO}_msd(*kanR*)$_{ON}$	pFF530	This plasmid was produced by introducing a 79-bp fragment with homology to the *kanR* ORF (template for ssDNA(*kanR*)$_{ON}$) and flanked by EcoRI sites into the P_{lacO}_msd(wt) plasmid using QuikChange site-directed mutagenesis.	Fig. 1B Fig. 1C Fig. 4B
P_{lacO}_msd(*kanR*)$_{ON}$_dRT	pFF749	This plasmid was produced by QuikChange site-directed mutagenesis (using FF_oligo912 and FF_oligo913) primers to mutate the YADD active site of the RT to YAAA (D197A and D198A mutations) in the P_{lacO}_msd(*kanR*)$_{ON}$ plasmid.	Fig. 1C
P_{tetO}_*bet*	pFF145	This plasmid was constructed by cloning the *bet* ORF from pKD46 (*61*) plasmid downstream of the P_{tetO} promoter (KpnI and BamHI sites) in the pZA11 (*22*) plasmid.	Fig. 1C Fig. 4B Fig. 5D-F
P_{lacO}_SCRIBE(*kanR*)$_{ON}$	pFF745	This plasmid was constructed by cloning *bet* and its natural ribosome binding site (RBS) downstream of the RT in the P_{lacO}_msd(*kanR*)$_{ON}$ plasmid (BamH1, MluI sites). 18 bp upstream of the *bet* start codon in the pKD46 (*61*) plasmid was used as the *bet* RBS.	Fig. 1D Fig. 3
P_{Dawn}_SCRIBE(*kanR*)$_{ON}$	pFF706	This plasmid was constructed by replacing the	Fig. 4A

(light inducible)		P_{lacO} promoter in SCRIBE(*kanR*)$_{ON}$ with a PCR fragment containing the light-regulated cassettes (*yf1/fixJ* operon and *cI* and their corresponding promoters as shown in Fig. 4A) from pDawn plasmid (Addgene # 43796 (*26*)).	
P_{lacO}_SCRIBE(*galK*)$_{OFF}$	pFF714	This plasmid was constructed by replacing the 79-bp *kanR* homology in P_{lacO}_SCRIBE(*kanR*)$_{ON}$ with a 78-bp fragment containing two stop codons flanked by 72 bp homology to *galK* using QuikChange site-directed mutagenesis.	Fig. 2
P_{tetO}_SCRIBE(*galK*)$_{OFF}$	pFF761	This plasmid was constructed by cloning the SCRIBE(*galK*)$_{OFF}$ cassette into the pZA11 plasmid (*22*) downstream of the P_{tetO} promoter.	Fig. 3
P_{tetO}_SCRIBE(*galK*)$_{ON}$	pFF746	This plasmid constructed by cloning SCRIBE(*galK*)$_{OFF}$ in the pZA21 backbone (*22*) (downstream of P_{tetO}) followed by a QuikChange *in vitro* mutagenesis step to revert the two stop codons in the msd(*galK*)$_{OFF}$ back to the wild-type sequence.	Fig. 2
P_{tetO}_SCRIBE(*lacZ*)$_{ON}$	pFF838	This plasmid was made by cloning a 78-bp fragment from the *lacZ* ORF into EcoRI sites of the SCRIBE cassette in P_{tetO}_SCRIBE(*galK*)$_{ON}$, replacing the *galK* homology with *lacZ* homology. The obtained SCRIBE cassette then was cloned into the pZA31 (*22*) backbone.	Fig. 5A-C
P_{luxR}_msd(*lacZ*)$_{ON}$	pFF828	This plasmid was made by replacing the P_{lacO} in the P_{lacO}_msd(*kanR*)$_{ON}$ plasmid with an AHL-inducible promoter (*luxR* cassette and P_{luxR} promoter (*31*)) followed by the replacement of the ssDNA(*kanR*)$_{ON}$ template with a 78-bp fragment from the *lacZ* ORF.	Fig. 5D-F

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

Part name	**Type**	**Sequence**	**Ref**
P_{lacO}	Promoter	AATTGTGAGCGGATAACAATTGACATTGTGAGC GGATAACAAGATACTGAGCACATCAGCAGGAC GCACTGACC	(*22*)
P_{tetO}	Promoter	TCCCTATCAGTGATAGAGATTGACATCCCTATC AGTGATAGAGATACTGAGCACATCAGCAGGAC GCACTGACC	(*22*)
P_{luxR}	Promoter	ACCTGTAGGATCGTACAGGTTTACGCAAGAAAA TGGTTTGTTATAGTCGAATA	(*31*)
$P_{\lambda R}$	Promoter	TAACACCGTGCGTGTTGACTATTTTACCTCTGGC GGTGATAATGGTTGC	(*26*)
P_{fixK2}	Promoter	ACGCCCGTGATCCTGATCACCGGCTATCCGGAC GAAAACATCTCGACCCGGGCCGCCGAGGCCGG CGTAAAAGACGTGGTTTTGAAGCCGCTTCTCGA CGAAAACCTGCTCAAGCGTATCCGCCGCGCCAT CCAGGACCGGCCTCGGGCATGACCTACGGGGTT CTACGTAAGGCACCCCCCTTAAGATATCGCTCG AAATTTTCGAACCTCCCGATACCGCGTACCAAT GCGTCATCACAACGGAG	(*26*)
msr	Primer for the RT	ATGCGCACCCTTAGCGAGAGGTTTATCATTAAG GTCAACCTCTGGATGTTGTTTCGGCATCCTGCAT TGAATCTGAGTTACT	This work
msd(*wt*)	Template for the RT	GTCAGAAAAAACGGGTTTCCTGGTTGGCTCGGA GAGCATCAGGCGATGCTCTCCGTTCCAACAAGG AAAACAGACAGTAACTCAGA	This work
msd(*kanR*)$_{ON}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCCAACAT GGATGCTGATTTATATGGGTATAAATGGGCCCG CGATAATGTCGGGCAATCAGGTGCGACAATCTA TCGGAATTCAGGAAAACAGACAGTAACTCAGA	This work
msd(*galK*)$_{OFF}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCCAGCTA ATTTCCGCGCTCGGCAAGAAAGATCATGCCTAA TGAATCGATTGCCGCTCACTGGGGACCAAAGCA GTTTCCGAATTCAGGAAAACAGACAGTAACTCA	This work

		GA	
msd(*galK*)$_{ON}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCCAGCTA ATTTCCGCGCTCGGCAAGAAAGATCATGCCCTC TTGATCGATTGCCGCTCACTGGGGACCAAAGCA GTTTCCGAATTCAGGAAAACAGACAGTAACTCA GA	This work
msd(*lacZ*)$_{ON}$	Template for the RT	GTCAGAAAAAACGGGTTTCCTGAATTCACCCAA CTTAATCGCCTTGCAGCACATCCCCCTTTCGCCA GCTGGCGTAATAGCGAAGAGGCCCGCACCGAT CGCCCTGAATTCAGGAAAACAGACAGTAACTCA GA	This work
RT Ec86	Reverse Transcriptase	ATGAAATCCGCTGAATATTTGAACACTTTTAGA TTGAGAAATCTCGGCCTACCTGTCATGAACAAT TTGCATGACATGTCTAAGGCGACTCGCATATCT GTTGAAACACTTCGGTTGTTAATCTATACAGCT GATTTTCGCTATAGGATCTACACTGTAGAAAAG AAAGGCCCAGAGAAGAGAATGAGAACCATTTA CCAACCTTCTCGAGAACTTAAAGCCTTACAAGG ATGGGTTCTACGTAACATTTTAGATAAACTGTC GTCATCTCCTTTTTCTATTGGATTTGAAAAGCAC CAATCTATTTTGAATAATGCTACCCCGCATATTG GGGCAAACTTTATACTGAATATTGATTTGGAGG ATTTTTTCCCAAGTTTAACTGCTAACAAAGTTTT TGGAGTGTTCCATTCTCTTGGTTATAATCGACTA ATATCTTCAGTTTTGACAAAAATATGTTGTTATA AAAATCTGCTACCACAAGGTGCTCCATCATCAC CTAAATTAGCTAATCTAATATGTTCTAAACTTG ATTATCGTATTCAGGGTTATGCAGGTAGTCGGG GCTTGATATATACGAGATATGCCGATGACCTCA CCTTATCTGCACAGTCTATGAAAAAGGTTGTTA AAGCACGTGATTTTTTATTTTCTATAATCCCAAG TGAAGGATTGGTTATTAACTCAAAAAAAACTTG TATTAGTGGGCCTCGTAGTCAGAGGAAAGTTAC AGGTTTAGTTATTTCACAAGAGAAAGTTGGGAT AGGTAGAGAAAAATATAAAGAAATTAGAGCAA AGATACATCATATATTTTGCGGTAAGTCTTCTG AGATAGAACACGTTAGGGGATGGTTGTCATTTA TTTTAAGTGTGGATTCAAAAAGCCATAGGAGAT TAATAACTTATATTAGCAAATTAGAAAAAAAAT ATGGAAAGAACCCTTTAAATAAAGCGAAGACC TAA	This work
Beta	ssDNA-specific recombinase	ATGAGTACTGCACTCGCAACGCTGGCTGGGAAG CTGGCTGAACGTGTCGGCATGGATTCTGTCGAC CCACAGGAACTGATCACCACTCTTCGCCAGACG GCATTTAAAGGTGATGCCAGCGATGCGCAGTTC	(*61*)

	protein	ATCGCATTACTGATCGTTGCCAACCAGTACGGC CTTAATCCGTGGACGAAAGAAATTTACGCCTTT CCTGATAAGCAGAATGGCATCGTTCCGGTGGTG GGCGTTGATGGCTGGTCCCGCATCATCAATGAA AACCAGCAGTTTGATGGCATGGACTTTGAGCAG GACAATGAATCCTGTACATGCCGGATTTACCGC AAGGACCGTAATCATCCGATCTGCGTTACCGAA TGGATGGATGAATGCCGCCGCGAACCATTCAAA ACTCGCGAAGGCAGAGAAATCACGGGGCCGTG GCAGTCGCATCCCAAACGGATGTTACGTCATAA AGCCATGATTCAGTGTGCCCGTCTGGCCTTCGG ATTTGCTGGTATCTATGACAAGGATGAAGCCGA GCGCATTGTCGAAAATACTGCATACACTGCAGA ACGTCAGCCGGAACGCGACATCACTCCGGTTAA CGATGAAACCATGCAGGAGATTAACACTCTGCT GATCGCCCTGGATAAAACATGGGATGACGACTT ATTGCCGCTCTGTTCCCAGATATTTCGCCGCGAC ATTCGTGCATCGTCAGAACTGACACAGGCCGAA GCAGTAAAAGCTCTTGGATTCCTGAAACAGAAA GCCGCAGAGCAGAAGGTGGCAGCATGA	
cI	λ repressor	ATGAGCACAAAAAAGAAACCATTAACACAAGA GCAGCTTGAGGACGCACGTCGCCTTAAAGCAAT TTATGAAAAAAAGAAAAATGAACTTGGCTTATC CCAGGAATCTGTCGCAGACAAGATGGGGATGG GGCAGTCAGGCGTTGGTGCTTTATTTAATGGCA TCAATGCATTAAATGCTTATAACGCCGCATTGC TTGCAAAAATTCTCAAAGTTAGCGTTGAAGAAT TTAGCCCTTCAATCGCCAGAGAAATCTACGAGA TGTATGAAGCGGTTAGTATGCAGCCGTCACTTA GAAGTGAGTATGAGTACCCTGTTTTTTCTCATGT TCAGGCAGGGATGTTCTCACCTGAGCTTAGAAC CTTTACCAAAGGTGATGCGGAGAGATGGGTAA GCACAACCAAAAAAGCCAGTGATTCTGCATTCT GGCTTGAGGTTGAAGGTAATTCCATGACCGCAC CAACAGGCTCCAAGCCGAGCTTTCCTGACGGAA TGTTAATTCTCGTTGACCCTGAGCAGGCTGTTG AGCCAGGTGATTTCTGCATAGCCAGACTTGGGG GTGATGAGTTTACCTTCAAGAAACTGATCAGGG ATAGCGGTCAGGTGTTTTTACAACCACTAAACC CACAGTACCCAATGATCCCATGCAATGAGAGTT GTTCCGTTGTGGGGAAAGTTATCGCTAGTCAGT GGCCTGAAGAGACGTTTGGCGCTGCAAACGAC GAAAACTACGCTTTAGTAGCTTAA	(*26*)
kanR$_{OFF}$	Reporter gene (premature stop codons are	ATGAGCCATATTCAACGGGAAACGTCTTGCTCG AGGCCGCGATTAAATTCCAACATGGATGCTGAT TTATATGGGTATAAATAATAGCGCGATAATGTC GGGCAATCAGGTGCGACAATCTATCGATTGTAT GGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAA	This work

	highlighted)	CATGGCAAAGGTAGCGTTGCCAATGATGTTACA GATGAGATGGTCAGACTAAACTGGCTGACGGA ATTTATGCCTCTTCCGACCATCAAGCATTTTATC CGTACTCCTGATGATGCATGGTTACTCACCACT GCGATCCCCGGGAAAACAGCATTCCAGGTATTA GAAGAATATCCTGATTCAGGTGAAAATATTGTT GATGCGCTGGCAGTGTTCCTGCGCCGGTTGCAT TCGATTCCTGTTTGTAATTGTCCTTTTAACAGCG ATCGCGTATTTCGTCTCGCTCAGGCGCAATCAC GAATGAATAACGGTTTGGTTGATGCGAGTGATT TTGATGACGAGCGTAATGGCTGGCCTGTTGAAC AAGTCTGGAAAGAAATGCATAAACTTTTGCCAT TCTCACCGGATTCAGTCGTCACTCATGGTGATTT CTCACTTGATAACCTTATTTTTGACGAGGGGAA ATTAATAGGTTGTATTGATGTTGGACGAGTCGG AATCGCAGACCGATACCAGGATCTTGCCATCCT ATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTA CAGAAACGGCTTTTTCAAAAATATGGTATTGAT AATCCTGATATGAATAAATTGCAGTTTCATTTG ATGCTCGATGAGTTTTTCTAA	
galK$_{OFF}$	Reporter gene (premature stop codons are highlighted)	ATGAGTCTGAAAGAAAAAACACAATCTCTGTTT GCCAACGCATTTGGCTACCCTGCCACTCACACC ATTCAGGCGCCTGGCCGCGTGAATTTGATTGGT GAACACACCGACTACAACGACGGTTTCGTTCTG CCCTGCGCGATTGATTATCAAACCGTGATCAGT TGTGCACCACGCGATGACCGTAAAGTTCGCGTG ATGGCAGCCGATTATGAAAATCAGCTCGACGAG TTTTCCCTCGATGCGCCCATTGTCGCACATGAA AACTATCAATGGGCTAACTACGTTCGTGGCGTG GTGAAACATCTGCAACTGCGTAACAACAGCTTC GGCGGCGTGGACATGGTGATCAGCGGCAATGT GCCGCAGGGTGCCGGGTTAAGTTCTTCCGCTTC ACTGGAAGTCGCGGTCGGAACCGTATTGCAGCA GCTTTATCATCTGCCGCTGGACGGCGCACAAAT CGCGCTTAACGGTCAGGAAGCAGAAAACCAGT TTGTAGGCTGTAACTGCGGGATCATGGATCAGC TAATTTCCGCGCTCGGCAAGAAAGATCATGCCT AATGAATCGATTGCCGCTCACTGGGGACCAAAG CAGTTTCCATGCCCAAAGGTGTGGCTGTCGTCA TCATCAACAGTAACTTCAAACGTACCCTGGTTG GCAGCGAATACAACACCCGTCGTGAACAGTGC GAAACCGGTGCGCGTTTCTTCCAGCAGCCAGCC CTGCGTGATGTCACCATTGAAGAGTTCAACGCT GTTGCGCATGAACTGGACCCGATCGTGGCAAAA CGCGTGCGTCATATACTGACTGAAAACGCCCGC ACCGTTGAAGCTGCCAGCGCGCTGGAGCAAGG CGACCTGAAACGTATGGGCGAGTTGATGGCGG AGTCTCATGCCTCTATGCGCGATGATTTCGAAA TCACCGTGCCGCAAATTGACACTCTGGTAGAAA	

		TCGTCAAAGCTGTGATTGGCGACAAAGGTGGCG TACGCATGACCGGCGGCGGATTTGGCGGCTGTA TCGTCGCGCTGATCCCGGAAGAGCTGGTGCCTG CCGTACAGCAAGCTGTCGCTGAACAATATGAAG CAAAAACAGGTATTAAAGAGACTTTTTACGTTT GTAAACCATCACAAGGAGCAGGACAGTGCTGA	
lacZ$_{OFF}$	Reporter gene (premature stop codons are highlighted)	ATGACCATGATTACGGATTCACTGGCCGTCGTT TTACAACGTCGTGACTGGGAAAACCCTGGCGTT ACCCAACTTAATCGCCTTGCAGCACATCCCCCT TTCTAATAGTGGCGTAATAGCGAAGAGGCCCGC ACCGATCGCCCTTCCCAACAGTTGCGCAGCCTG AATGGCGAATGGCGCTTTGCCTGGTTTCCGGCA CCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTG CGATCTTCCTGAGGCCGATACTGTCGTCGTCCC CTCAAACTGGCAGATGCACGGTTACGATGCGCC CATCTACACCAACGTGACCTATCCCATTACGGT CAATCCGCCGTTTGTTCCCACGGAGAATCCGAC GGGTTGTTACTCGCTCACATTTAATGTTGATGA AAGCTGGCTACAGGAAGGCCAGACGCGAATTA TTTTTGATGGCGTTAACTCGGCGTTTCATCTGTG GTGCAACGGGCGCTGGGTCGGTTACGGCCAGG ACAGTCGTTTGCCGTCTGAATTTGACCTGAGCG CATTTTTACGCGCCGGAGAAAACCGCCTCGCGG TGATGGTGCTGCGCTGGAGTGACGGCAGTTATC TGGAAGATCAGGATATGTGGCGGATGAGCGGC ATTTTCCGTGACGTCTCGTTGCTGCATAAACCG ACTACACAAATCAGCGATTTCCATGTTGCCACT CGCTTTAATGATGATTTCAGCCGCGCTGTACTG GAGGCTGAAGTTCAGATGTGCGGCGAGTTGCGT GACTACCTACGGGTAACAGTTTCTTTATGGCAG GGTGAAACGCAGGTCGCCAGCGGCACCGCGCC TTTCGGCGGTGAAATTATCGATGAGCGTGGTGG TTATGCCGATCGCGTCACACTACGTCTGAACGT CGAAAACCCGAAACTGTGGAGCGCCGAAATCC CGAATCTCTATCGTGCGGTGGTTGAACTGCACA CCGCCGACGGCACGCTGATTGAAGCAGAAGCCT GCGATGTCGGTTTCCGCGAGGTGCGGATTGAAA ATGGTCTGCTGCTGCTGAACGGCAAGCCGTTGC TGATTCGAGGCGTTAACCGTCACGAGCATCATC CTCTGCATGGTCAGGTCATGGATGAGCAGACGA TGGTGCAGGATATCCTGCTGATGAAGCAGAACA ACTTTAACGCCGTGCGCTGTTCGCATTATCCGA ACCATCCGCTGTGGTACACGCTGTGCGACCGCT ACGGCCTGTATGTGGTGGATGAAGCCAATATTG AAACCCACGGCATGGTGCCAATGAATCGTCTGA CCGATGATCCGCGCTGGCTACCGGCGATGAGCG AACGCGTAACGCGAATGGTGCAGCGCGATCGT AATCACCCGAGTGTGATCATCTGGTCGCTGGGG AATGAATCAGGCCACGGCGCTAATCACGACGC	This work

		GCTGTATCGCTGGATCAAATCTGTCGATCCTTCC CGCCCGGTGCAGTATGAAGGCGGCGGAGCCGA CACCACGGCCACCGATATTATTTGCCCGATGTA CGCGCGCGTGGATGAAGACCAGCCCTTCCCGGC TGTGCCGAAATGGTCCATCAAAAAATGGCTTTC GCTACCTGGAGAGACGCGCCCGCTGATCCTTTG CGAATACGCCCACGCGATGGGTAACAGTCTTGG CGGTTTCGCTAAATACTGGCAGGCGTTTCGTCA GTATCCCCGTTTACAGGGCGGCTTCGTCTGGGA CTGGGTGGATCAGTCGCTGATTAAATATGATGA AAACGGCAACCCGTGGTCGGCTTACGGCGGTGA TTTTGGCGATACGCCGAACGATCGCCAGTTCTG TATGAACGGTCTGGTCTTTGCCGACCGCACGCC GCATCCAGCGCTGACGGAAGCAAAACACCAGC AGCAGTTTTTCCAGTTCCGTTTATCCGGGCAAA CCATCGAAGTGACCAGCGAATACCTGTTCCGTC ATAGCGATAACGAGCTCCTGCACTGGATGGTGG CGCTGGATGGTAAGCCGCTGGCAAGCGGTGAA GTGCCTCTGGATGTCGCTCCACAAGGTAAACAG TTGATTGAACTGCCTGAACTACCGCAGCCGGAG AGCGCCGGGCAACTCTGGCTCACAGTACGCGTA GTGCAACCGAACGCGACCGCATGGTCAGAAGC CGGGCACATCAGCGCCTGGCAGCAGTGGCGTCT GGCGGAAAACCTCAGTGTGACGCTCCCCGCCGC GTCCCACGCCATCCCGCATCTGACCACCAGCGA AATGGATTTTTGCATCGAGCTGGGTAATAAGCG TTGGCAATTTAACCGCCAGTCAGGCTTTCTTTCA CAGATGTGGATTGGCGATAAAAAACAACTGCTG ACGCCGCTGCGCGATCAGTTCACCCGTGCACCG CTGGATAACGACATTGGCGTAAGTGAAGCGACC CGCATTGACCCTAACGCCTGGGTCGAACGCTGG AAGGCGGCGGGCCATTACCAGGCCGAAGCAGC GTTGTTGCAGTGCACGGCAGATACACTTGCTGA TGCGGTGCTGATTACGACCGCTCACGCGTGGCA GCATCAGGGGAAAACCTTATTTATCAGCCGGAA AACCTACCGGATTGATGGTAGTGGTCAAATGGC GATTACCGTTGATGTTGAAGTGGCGAGCGATAC ACCGCATCCGGCGCGGATTGGCCTGAACTGCCA GCTGGCGCAGGTAGCAGAGCGGGTAAACTGGC TCGGATTAGGGCCGCAAGAAAACTATCCCGACC GCCTTACTGCCGCCTGTTTTGACCGCTGGGATCT GCCATTGTCAGACATGTATACCCCGTACGTCTT CCCGAGCGAAAACGGTCTGCGCTGCGGGACGC GCGAATTGAATTATGGCCCACACCAGTGGCGCG GCGACTTCCAGTTCAACATCAGCCGCTACAGTC AACAGCAACTGATGGAAACCAGCCATCGCCATC TGCTGCACGCGGAAGAAGGCACATGGCTGAAT ATCGACGGTTTCCATATGGGGATTGGTGGCGAC GACTCCTGGAGCCCGTCAGTATCGGCGGAATTC CAGCTGAGCGCCGGTCGCTACCATTACCAGTTG	

		GTCTGGTGTCAAAAATAA	
SCRIBE(*kanR*)$_{ON}$	The synthetic operon for writing into the *kanR* locus. The msd(*kanR*)$_{ON}$ region is highlighted. The region flanked by EcoRI sites (red) can be replaced with a template for ssDNA of interest.	ATGCGCACCCTTAGCGAGAGGTTTATCATTAAG GTCAACCTCTGGATGTTGTTTCGGCATCCTGCAT TGAATCTGAGTTACTGTCTGTTTTCCTGAATTCC GATAGATTGTCGCACCTGATTGCCCGACATTAT CGCGGGCCCATTTATACCCATATAAATCAGCAT CCATGTTGGAATTCAGGAAACCCGTTTTTTCTG ACGTAAGGGTGCGCAACTTTCATGAAATCCGCT GAATATTTGAACACTTTTAGATTGAGAAATCTC GGCCTACCTGTCATGAACAATTTGCATGACATG TCTAAGGCGACTCGCATATCTGTTGAAACACTT CGGTTGTTAATCTATACAGCTGATTTTCGCTATA GGATCTACACTGTAGAAAAGAAAGGCCCAGAG AAGAGAATGAGAACCATTTACCAACCTTCTCGA GAACTTAAAGCCTTACAAGGATGGGTTCTACGT AACATTTTAGATAAACTGTCGTCATCTCCTTTTT CTATTGGATTTGAAAAGCACCAATCTATTTTGA ATAATGCTACCCCGCATATTGGGGCAAACTTTA TACTGAATATTGATTTGGAGGATTTTTTCCCAAG TTTAACTGCTAACAAAGTTTTTGGAGTGTTCCAT TCTCTTGGTTATAATCGACTAATATCTTCAGTTT TGACAAAAATATGTTGTTATAAAAATCTGCTAC CACAAGGTGCTCCATCATCACCTAAATTAGCTA ATCTAATATGTTCTAAACTTGATTATCGTATTCA GGGTTATGCAGGTAGTCGGGGCTTGATATATAC GAGATATGCCGATGACCTCACCTTATCTGCACA GTCTATGAAAAAGGTTGTTAAAGCACGTGATTT TTTATTTTCTATAATCCCAAGTGAAGGATTGGTT ATTAACTCAAAAAAAACTTGTATTAGTGGGCCT CGTAGTCAGAGGAAAGTTACAGGTTTAGTTATT TCACAAGAGAAAGTTGGGATAGGTAGAGAAAA ATATAAAGAAATTAGAGCAAAGATACATCATAT ATTTTGCGGTAAGTCTTCTGAGATAGAACACGT TAGGGGATGGTTGTCATTTATTTTAAGTGTGGA TTCAAAAAGCCATAGGAGATTAATAACTTATAT TAGCAAATTAGAAAAAAAATATGGAAAGAACC CTTTAAATAAAGCGAAGACCTAAGGATCCGGTT GATATTGATTCAGAGGTATAAAACGAATGAGTA CTGCACTCGCAACGCTGGCTGGGAAGCTGGCTG AACGTGTCGGCATGGATTCTGTCGACCCACAGG AACTGATCACCACTCTTCGCCAGACGGCATTTA AAGGTGATGCCAGCGATGCGCAGTTCATCGCAT TACTGATCGTTGCCAACCAGTACGGCCTTAATC CGTGGACGAAAGAAATTTACGCCTTTCCTGATA AGCAGAATGGCATCGTTCCGGTGGTGGGCGTTG ATGGCTGGTCCCGCATCATCAATGAAAACCAGC AGTTTGATGGCATGGACTTTGAGCAGGACAATG AATCCTGTACATGCCGGATTTACCGCAAGGACC GTAATCATCCGATCTGCGTTACCGAATGGATGG	This work

		ATGAATGCCGCCGCGAACCATTCAAAACTCGCG AAGGCAGAGAAATCACGGGGCCGTGGCAGTCG CATCCCAAACGGATGTTACGTCATAAAGCCATG ATTCAGTGTGCCCGTCTGGCCTTCGGATTTGCTG GTATCTATGACAAGGATGAAGCCGAGCGCATTG TCGAAAATACTGCATACACTGCAGAACGTCAGC CGGAACGCGACATCACTCCGGTTAACGATGAAA CCATGCAGGAGATTAACACTCTGCTGATCGCCC TGGATAAAACATGGGATGACGACTTATTGCCGC TCTGTTCCCAGATATTTCGCCGCGACATTCGTGC ATCGTCAGAACTGACACAGGCCGAAGCAGTAA AAGCTCTTGGATTCCTGAAACAGAAAGCCGCAG AGCAGAAGGTGGCAGCATGA	

Table S4 | List of the synthetic oligos used in this study

Name	Sequence
FF_oligo183	GCGATATCCATTTTCGCGAATCCGGAGTGTAAGAAGAGCTCCTGACTCCCCGTCGTGTAG
FF_oligo184	GACCGCAGAACAGGCAGCAGAGCGTTTGCGCGCAGTCAGCGATATCCATTTTCGCGAATC
FF_oligo185	CGGCTGACCATCGGGTGCCAGTGCGGGAGTTTCGTGACGTCGTTAAGCCAGCCCCGACAC
FF_oligo186	ACTACCATCCCTGCGTTGTTACGCAAAGTTAACAGTCGGTACGGCTGACCATCGGGTGCC
FF_oligo187 (* shows phosphorothioate bond)	C*G*CGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATAATAGCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCG*A*T
FF_oligo220	CAACTTAATCGCCTTGCAGCACATCCCCCTTTCTAATAGTGGCGTAATAGCGAAGAGGCCCGCACCGATCGC
FF_oligo912	GATATATACGAGATATGCCGCTGCTCTCACCTTATCTGCAC
FF_oligo913	GTGCAGATAAGGTGAGAGCAGCGGCATATCTCGTATATATC
FF_oligo1069	AATACGCAAACCGCCTCTCC
FF_oligo1070	CGGCGGATTGACCGTAATGG
FF_oligo1291	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNGCCCGACATTATCGCG
FF_oligo1292	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTCAACATGGATGCTGATTTATATGGGT
FF_oligo347 (PAGE purified, used as ssDNA size marker in Fig. 1B)	GTCAGAAAAAACGGGTTTCCTGGTTGGCTCGGAGAGCATCAGGCGATGCTCTCCGTTCCAACAAGGAAAACAGACAGTAACTCAGA

References and Notes

1. N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipos, E. Birney, Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **494**, 77–80 (2013). Medline doi:10.1038/nature11875

2. G. M. Church, Y. Gao, S. Kosuri, Next-generation digital information storage in DNA. *Science* **337**, 1628 (2012). Medline doi:10.1126/science.1226355

3. M. C. Inniss, P. A. Silver, Building synthetic memory. *Curr. Biol.* **23**, R812–R816 (2013). Medline doi:10.1016/j.cub.2013.06.047

4. T. S. Gardner, C. R. Cantor, J. J. Collins, Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000). Medline doi:10.1038/35002131

5. J. W. Kotula, S. J. Kerns, L. A. Shaket, L. Siraj, J. J. Collins, J. C. Way, P. A. Silver, Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4838–4843 (2014). Medline doi:10.1073/pnas.1321321111

6. D. R. Burrill, M. C. Inniss, P. M. Boyle, P. A. Silver, Synthetic memory circuits for tracking human cell fate. *Genes Dev.* **26**, 1486–1497 (2012). Medline doi:10.1101/gad.189035.112

7. B. P. Kramer, A. U. Viretta, M. Daoud-El Baba, D. Aubel, W. Weber, M. Fussenegger, An engineered epigenetic transgene switch in mammalian cells. *Nat. Biotechnol.* **22**, 867–870 (2004). Medline doi:10.1038/nbt980

8. C. M. Ajo-Franklin, D. A. Drubin, J. A. Eskin, E. P. Gee, D. Landgraf, I. Phillips, P. A. Silver, Rational design of memory in eukaryotic cells. *Genes Dev.* **21**, 2271–2276 (2007). Medline doi:10.1101/gad.1586107

9. P. Siuti, J. Yazbek, T. K. Lu, Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31**, 448–452 (2013). Medline doi:10.1038/nbt.2510

10. J. Bonnet, P. Subsoontorn, D. Endy, Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8884–8889 (2012). Medline doi:10.1073/pnas.1202344109

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

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

13. H. M. Ellis, D. Yu, T. DiTizio, D. L. Court, High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6742–6746 (2001). Medline doi:10.1073/pnas.121164898

14. N. Costantino, D. L. Court, Enhanced levels of λ Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15748–15753 (2003). Medline doi:10.1073/pnas.2434959100

15. D. Yu, H. M. Ellis, E. C. Lee, N. A. Jenkins, N. G. Copeland, D. L. Court, An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5978–5983 (2000). Medline doi:10.1073/pnas.100127597

16. J. A. Sawitzke, N. Costantino, X. T. Li, L. C. Thomason, M. Bubunenko, C. Court, D. L. Court, Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *J. Mol. Biol.* **407**, 45–59 (2011). Medline doi:10.1016/j.jmb.2011.01.030

17. B. Swingle, E. Markel, N. Costantino, M. G. Bubunenko, S. Cartinhour, D. L. Court, Oligonucleotide recombination in Gram-negative bacteria. *Mol. Microbiol.* **75**, 138–148 (2010). Medline doi:10.1111/j.1365-2958.2009.06976.x

18. S. Datta, N. Costantino, X. Zhou, D. L. Court, Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1626–1631 (2008). Medline doi:10.1073/pnas.0709089105

19. T. Yee, T. Furuichi, S. Inouye, M. Inouye, Multicopy single-stranded DNA isolated from a gram-negative bacterium, *Myxococcus xanthus*. *Cell* **38**, 203–209 (1984). Medline doi:10.1016/0092-8674(84)90541-5

20. B. C. Lampson, M. Inouye, S. Inouye, Retrons, msDNA, and the bacterial genome. *Cytogenet. Genome Res.* **110**, 491–499 (2005). Medline doi:10.1159/000084982

21. D. Lim, W. K. Maas, Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. *Cell* **56**, 891–904 (1989). Medline doi:10.1016/0092-8674(89)90693-4

22. R. Lutz, H. Bujard, Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997). Medline doi:10.1093/nar/25.6.1203

23. P. L. Sharma, V. Nurpeisov, R. F. Schinazi, Retrovirus reverse transcriptases containing a modified YXDD motif. *Antivir. Chem. Chemother.* **16**, 169–182 (2005). Medline

24. J. R. Mao, M. Shimada, S. Inouye, M. Inouye, Gene regulation by antisense DNA produced in vivo. *J. Biol. Chem.* **270**, 19684–19687 (1995). Medline doi:10.1074/jbc.270.34.19684

25. S. Warming, N. Costantino, D. L. Court, N. A. Jenkins, N. G. Copeland, Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* **33**, e36 (2005). Medline doi:10.1093/nar/gni035

26. R. Ohlendorf, R. R. Vidavski, A. Eldar, K. Moffat, A. Möglich, From dusk till dawn: One-plasmid systems for light-regulated gene expression. *J. Mol. Biol.* **416**, 534–542 (2012). Medline doi:10.1016/j.jmb.2012.01.001

27. M. S. Huen, X. T. Li, L. Y. Lu, R. M. Watt, D. P. Liu, J. D. Huang, The involvement of replication in single stranded oligonucleotide-mediated gene repair. *Nucleic Acids Res.* **34**, 6183–6194 (2006). Medline doi:10.1093/nar/gkl852

28. A. R. Poteete, Involvement of DNA replication in phage lambda Red-mediated homologous recombination. *Mol. Microbiol.* **68**, 66–74 (2008). Medline doi:10.1111/j.1365-2958.2008.06133.x

29. A. R. Poteete, Involvement of *Escherichia coli* DNA replication proteins in phage Lambda Red-mediated homologous recombination. *PLOS ONE* **8**, e67440 (2013). Medline doi:10.1371/journal.pone.0067440

30. B. E. Dutra, V. A. Sutera Jr., S. T. Lovett, RecA-independent recombination is efficient but limited by exonucleases. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 216–221 (2007). Medline doi:10.1073/pnas.0608293104

31. S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, R. Weiss, A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–1134 (2005). Medline doi:10.1038/nature03461

32. S. Ausländer, D. Ausländer, M. Müller, M. Wieland, M. Fussenegger, Programmable single-cell mammalian biocomputers. *Nature* **487**, 123–127 (2012). Medline

33. R. Daniel, J. R. Rubens, R. Sarpeshkar, T. K. Lu, Synthetic analog computation in living cells. *Nature* **497**, 619–623 (2013). Medline doi:10.1038/nature12148

34. E. J. Olson, L. A. Hartsough, B. P. Landry, R. Shroff, J. J. Tabor, Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nat. Methods* **11**, 449–455 (2014). Medline doi:10.1038/nmeth.2884

35. J. A. Mosberg, C. J. Gregg, M. J. Lajoie, H. H. Wang, G. M. Church, Improving lambda red genome engineering in Escherichia coli via rational removal of endogenous nucleases. *PLOS ONE* **7**, e44638 (2012). Medline doi:10.1371/journal.pone.0044638

36. S. Miyata, A. Ohshima, S. Inouye, M. Inouye, In vivo production of a stable single-stranded cDNA in *Saccharomyces cerevisiae* by means of a bacterial retron. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5735–5739 (1992). Medline doi:10.1073/pnas.89.13.5735

37. O. Mirochnitchenko, S. Inouye, M. Inouye, Production of single-stranded DNA in mammalian cells by means of a bacterial retron. *J. Biol. Chem.* **269**, 2380–2383 (1994). Medline

38. S. Brakmann, S. Grzeszik, An error-prone T7 RNA polymerase mutant generated by directed evolution. *ChemBioChem* **2**, 212–219 (2001). Medline doi:10.1002/1439-7633(20010302)2:3<212::AID-CBIC212>3.0.CO;2-R

39. J. D. Roberts, K. Bebenek, T. A. Kunkel, The accuracy of reverse transcriptase from HIV-1. *Science* **242**, 1171–1173 (1988). Medline doi:10.1126/science.2460925

40. K. Bebenek, J. Abbotts, S. H. Wilson, T. A. Kunkel, Error-prone polymerization by HIV-1 reverse transcriptase. Contribution of template-primer misalignment, miscoding, and termination probability to mutational hot spots. *J. Biol. Chem.* **268**, 10324–10334 (1993). Medline

41. K. M. Esvelt, J. C. Carlson, D. R. Liu, A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011). Medline doi:10.1038/nature09929

42. Y. Amir, E. Ben-Ishay, D. Levner, S. Ittah, A. Abu-Horowitz, I. Bachelet, Universal computing by DNA origami robots in a living animal. *Nat. Nanotechnol.* **9**, 353–357 (2014). Medline doi:10.1038/nnano.2014.58

43. L. Qian, E. Winfree, J. Bruck, Neural network computation with DNA strand displacement cascades. *Nature* **475**, 368–372 (2011). Medline doi:10.1038/nature10262

44. G. Seelig, D. Soloveichik, D. Y. Zhang, E. Winfree, Enzyme-free nucleic acid logic circuits. *Science* **314**, 1585–1588 (2006). Medline doi:10.1126/science.1132493

45. P. W. Rothemund, Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006). Medline doi:10.1038/nature04586

46. S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, W. M. Shih, Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009). Medline doi:10.1038/nature08016

47. S. M. Douglas, I. Bachelet, G. M. Church, A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335**, 831–834 (2012). Medline doi:10.1126/science.1214081

48. S. M. Chirieleison, P. B. Allen, Z. B. Simpson, A. D. Ellington, X. Chen, Pattern transformation with DNA circuits. *Nat. Chem.* **5**, 1000–1005 (2013). Medline doi:10.1038/nchem.1764

49. C. A. Brosey, C. Yan, S. E. Tsutakawa, W. T. Heller, R. P. Rambo, J. A. Tainer, I. Ivanov, W. J. Chazin, A new structural framework for integrating replication protein A into DNA processing machinery. *Nucleic Acids Res.* **41**, 2313–2327 (2013). Medline doi:10.1093/nar/gks1332

50. S. Kosuri, G. M. Church, Large-scale de novo DNA synthesis: Technologies and applications. *Nat. Methods* **11**, 499–507 (2014). Medline doi:10.1038/nmeth.2918

51. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012). Medline doi:10.1038/nmeth.2019

52. A. E. Minoche, J. C. Dohm, H. Himmelbauer, Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol.* **12**, R112 (2011). Medline doi:10.1186/gb-2011-12-11-r112

53. D. I. Lou, J. A. Hussmann, R. M. McBee, A. Acevedo, R. Andino, W. H. Press, S. L. Sawyer, High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19872–19877 (2013). Medline doi:10.1073/pnas.1319590110

54. S. T. Lovett, R. L. Hurley, V. A. Sutera Jr., R. H. Aubuchon, M. A. Lebedeva, Crossing over between regions of limited homology in *Escherichia coli*. RecA-dependent and RecA-independent pathways. *Genetics* **160**, 851–859 (2002). Medline

55. K. C. Murphy, M. G. Marinus, RecA-independent single-stranded DNA oligonucleotide-mediated mutagenesis. *F1000 Biol. Rep.* **2**, 56 (2010). Medline

56. C. A. Fogle, J. L. Nagle, M. M. Desai, Clonal interference, multiple mutations and adaptation in large asexual populations. *Genetics* **180**, 2163–2173 (2008). Medline doi:10.1534/genetics.108.090019

57. M. Imhof, C. Schlotterer, Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1113–1117 (2001). Medline doi:10.1073/pnas.98.3.1113

58. P. L. Foster, Methods for determining spontaneous mutation rates. *Methods Enzymol.* **409**, 195–213 (2006). Medline doi:10.1016/S0076-6879(05)09012-9

59. S. E. Luria, The frequency distribution of spontaneous bacteriophage mutants as evidence for the exponential rate of phage reproduction. *Cold Spring Harb. Symp. Quant. Biol.* **16**, 463–470 (1951). Medline doi:10.1101/SQB.1951.016.01.033

60. A. Tamsir, J. J. Tabor, C. A. Voigt, Robust multicellular computing using genetically encoded NOR gates and chemical ‘wires’. *Nature* **469**, 212–215 (2011). Medline doi:10.1038/nature09565

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