Permanent genetic memory with >1-byte capacity

Lei Yang[1], Alec A K Nielsen[1], Jesus Fernandez-Rodriguez[1], Conor J McClune[2], Michael T Laub[2], Timothy K Lu[1,3] & Christopher A Voigt[1]

Genetic memory enables the recording of information in the DNA of living cells. Memory can record a transient environmental signal or cell state that is then recalled at a later time. Permanent memory is implemented using irreversible recombinases that invert the orientation of a unit of DNA, corresponding to the [0,1] state of a bit. To expand the memory capacity, we have applied bioinformatics to identify 34 phage integrases (and their cognate *attB* and *attP* recognition sites), from which we build 11 memory switches that are perfectly orthogonal to each other and the FimE and HbiF bacterial invertases. Using these switches, a memory array is constructed in *Escherichia coli* that can record 1.375 bytes of information. It is demonstrated that the recombinases can be layered and used to permanently record the transient state of a transcriptional logic gate.

Cells can 'remember' events using a variety of biochemical mechanisms embedded in their regulatory networks[1]. For engineering applications, synthetic memory enables a signal to be recorded and accessed at a later time. In a bioreactor, it can convert transient conditions (such as inducer, growth phase or glucose concentration) into the permanent induction of metabolic pathways[2,3]. Memory can also be used to record transient signals that are difficult to record *in situ*[4,5]—for example, stimuli experienced by bacteria in the gut microbiome[6,7]. Larger memory capacities enable more information to be stored, which can be used to build circuits that require storage and retrieval and to program differentiation into a large number of cell states.

Several approaches have been taken to build synthetic memory. Genetic switches incorporating feedback loops can exhibit multistability, and memory is implemented via the transition between stable steady states[8–11]. The most common implementation is a toggle switch, where repressors inhibit each other's expression and a memory state corresponds with the dominance of one repressor[9,12]. The feedback loop that maintains the memory state requires the continuous use of energy and materials for transcription and translation, analogously to volatile memory in electronic circuits.

A second approach is based on the use of recombinases that bind two recognition sites and invert the intervening DNA[3,13]. The states corresponding to the two orientations are nonvolatile and will be maintained even after cell death. A bidirectional recombinase, such as Cre or Flp, catalyzes the orientation changes in both directions, which complicates their use for memory, as their continued expression results in a distribution of states. Irreversible recombinases, such as FimE, flip the DNA in only one direction and thus implement permanent memory[14]. A memory circuit that can be both set and reset has been built with an integrase to flip DNA in one direction and the same integrase coexpressed with an excisionase to flip the DNA in the reverse direction[15].

To date, the largest *in vivo* storage capacity has been 2 bits encoded by a pair of recombinases[16,17]. The capacity has been limited by the need for the regulatory proteins underlying the memory switches to not interfere with each other. For example, a toggle switch requires two repressors; using *N* toggle switches in one cell would require 2*N* orthogonal repressors[18]. Recombinases are orthogonal if they do not bind to each other's recognition sites. To this end, we have expanded the storage capacity of a recombinase-based memory array by mining orthogonal recombinases from prophage genomes. Our focus is on irreversible large serine-type phage (LSTP) integrases, which are involved in mediating phage integration and excision into the bacterial genome between their cognate recognition sites, *attB* (bacterium) and *attP* (phage)[19] (**Fig. 1a**). By placing these sites in the opposite orientation, LSTP integrases cleave, rotate and rejoin the DNA to invert the region between sites. We applied a bioinformatics approach to discover 34 putative integrases and their *attB* and *attP* sites from prophage genomes. This set of new recombinases is used to build a memory array that is able to record 2^{11} (2,048) combinations of states (1.375 bytes of information).

RESULTS

Identification of LSTP integrases and *att* sites

The construction of a memory switch based on an integrase requires both its gene and the cognate *attB* and *attP* recognition sites. This poses two challenges. First, identifying integrases is difficult because they are closely related to other classes of DNA-modifying enzymes (such as transposases)[20,21]. Second, the *attB* and *attP* sites are difficult to find because they are small and lack an obvious sequence signature[19].

Since the discovery of phage phiC31 integrase, only a few of LSTP integrases and their cognate *attB* and *attP* sites have been identified[19].

[1]Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [3]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. Correspondence should be addressed to C.A.V. (cavoigt@gmail.com).

RECEIVED 2 JUNE; ACCEPTED 23 SEPTEMBER; PUBLISHED ONLINE 26 OCTOBER 2014; DOI:10.1038/NMETH.3147

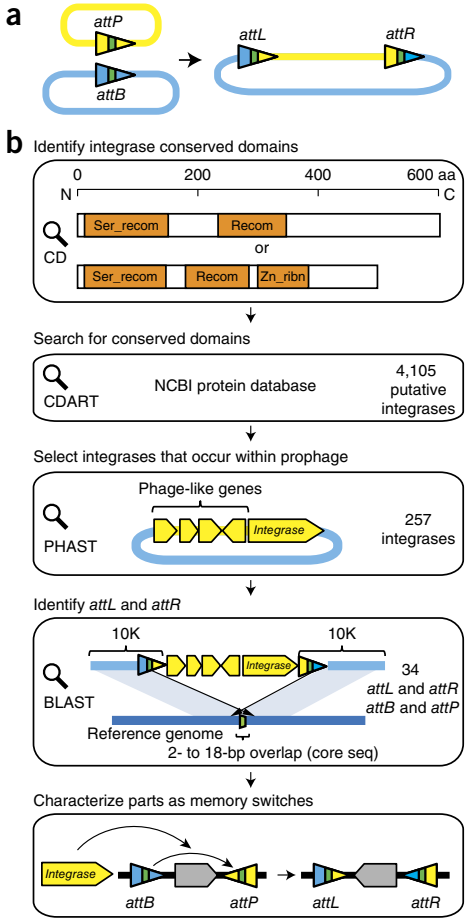

Figure 1 | Discovery of phage recombinases and their recognition sites. (**a**) LSTP integrases catalyze insertion of phage genome (yellow) into the bacterial genome (blue) between *attB* and *attP* sites, which form hybrid *attL* and *attR* sites. Multicolored arrowheads illustrate the sequence changes that occur during strand exchange, with the core sequence shown in green. (**b**) Steps from integrase discovery to the construction of a memory switch. The domain structure (orange) is shown for phiC31 (top) and Bxb1 (bottom). Blue lines indicate the bacterial genomic DNA; yellow regions correspond to prophage.

To mine LSTP integrases from the genome database, we identified a set of conserved domains using the Conserved Domain (CD) Database[22] and then focused the search on these regions (**Fig. 1b**). The integrase from phiC31 contains two conserved domains: a 'Ser_recombinase' domain (137 aa)[20,23,24] and a 'Recombinase' domain (100 aa)[24]. The integrase from phage Bxb1 contains an additional 'Recombinase_Zinc_beta_ribbon' domain (57 aa)[24]. These three domains are present in different combinations in other known phage integrases[21]. We then used the CD Architecture Retrieval Tool (CDART) to search the NCBI Protein database to identify proteins containing at least the first two domains. This search yielded 4,105 candidate LSTP integrases.

Building memory switches requires the identification of the *attB* and *attP* recognition sites for each integrase. These sites were located using a strategy based on genome comparison. When a lytic phage integrates into a bacterial genome, the integrase recognizes *attB* and *attP*, and within these sites the DNA is cleaved and strand exchange is catalyzed[19] (**Fig. 1a**). After integration, the recombination forms new *attL* and *attR* sequences, which flank the prophage within the bacterial genome. The *attB*, *attP*, *attL* and *attR* sites share a common 2–18-bp 'core' sequence. The first step in reconstructing these sequences was to retrieve the location of the 4,105 LSTP integrases within all sequenced genomes in the NCBI database. These genomes were then scanned using PHAST, which detects clusters of phage-like proteins and provides approximate locations of prophage regions[25]. This yielded 257 integrase genes located within prophage regions.

We used genome comparisons to determine the precise prophage boundaries. The region of the bacterial genome containing the prophage and 10 kb of up- and downstream sequence was compared using BLAST against homologous genome sequences from the same genus (**Fig. 1b**). Positive hits were signified by a 2–18-bp overlap identified in the alignment of the 10-kb flanking sequences between the genome containing the prophage and the reference genome (**Fig. 1b**). The minimum *att* site length to ensure efficient recombination is 40–50 bp[21]. Therefore, we defined *attL* and *attR* as a 59–66-bp region surrounding the core sequence at the prophage boundaries. From this, we reconstructed the *attB* and *attP* sites by swapping the half-sites of *attL* and *attR* on the basis of their strand-exchange (**Fig. 1a**). We took a similar approach when the prophage occurred within a gene encoding a conserved protein (**Supplementary Note 1** and **Supplementary Fig. 1**). Using this strategy, we identified a library of 34 LSTP integrases and their *attB* and *attP* sites, including three that were gleaned from the literature[26,27] (**Supplementary Fig. 2** and **Supplementary Table 1**). Most of these integrases share <65% amino acid identity (except Int7 and Int22, Int8 and Int21) and all have distinct *attB* and/or *attP* sites, making them unlikely to cross-react.

Characterization of memory switches

We selected a subset of 13 integrases to share a maximum of 60% amino acid sequence identity (**Supplementary Fig. 2**). Their corresponding genes were codon-optimized for expression in *E. coli* and built using DNA synthesis. A two-plasmid system was constructed to test their function and rapidly screen for orthogonality (**Fig. 2a**). One plasmid contains the integrase-encoding gene under the control of the arabinose-inducible promoter, P_{Bad}. A second reporter plasmid contains the *attB* and *attP* sites flanking an inverted *gfp* reporter gene. A strong constitutive promoter (Registry of Biological Parts BBa_J23119) that transcribes in the opposite orientation from *gfp* is placed upstream of the attB site. After the integrase is expressed, the orientation of *gfp* is reversed and it is transcribed. After inversion, the *attL* site is located on the 5′ UTR and could affect *gfp* expression. To insulate against this effect, we included a spacer and the ribozyme RiboJ[28,29] . The integrase and reporter plasmids were cotransformed into *E. coli* DH10B cells and tested for function. We varied the expression of the integrase by screening 16–32 ribosome binding sites (RBSs) predicted by the RBS Calculator to widely span expression levels[30,31]. We selected RBSs that achieved the maximum GFP expression while minimizing leakage in the uninduced state (**Supplementary Note 2** and **Supplementary Table 2**). Remarkably, 11 of the 13 integrases were functional, as confirmed by fluorescence, PCR and sequencing (**Fig. 2** and **Supplementary Fig. 3**). For Int13, the *attP* site showed weak promoter activity, which was corrected by swapping its position with the *attB*

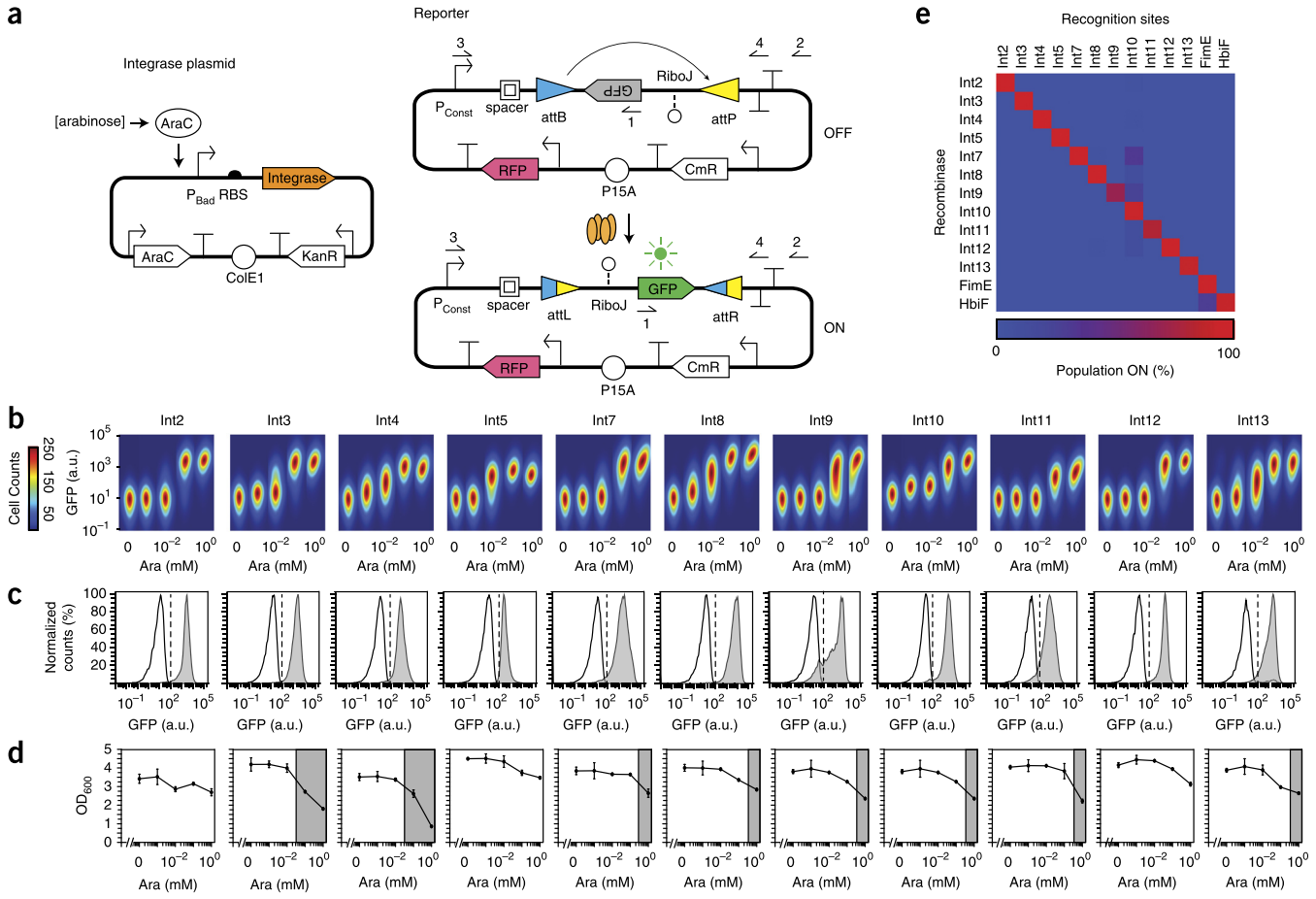

Figure 2 | Memory switch characterization. (**a**) The two-plasmid system for assaying integrases and their recognition sites. A constitutive promoter (P_{const}) controls GFP expression after DNA inversion. RFP is expressed from a constitutive promoter (Registry of Biological Parts BBa_J23101) to aid the gating of cells. The primer 1, 2, 3 and 4 locations are used to assay the inversion event by PCR and sequencing. (**b**) Induction of each functional memory switch. Five levels of arabinose induction are shown: 0, 10^{-3}, 0.01, 0.1 and 1 mM (left to right). The heat map shows the cell count; the height and width of the plot at each arabinose concentration represent fluorescence from the green and red channels, respectively. The range of red fluorescence in each plot is $10^{2.5}$–$10^{4.5}$ a.u. (log scale). (**c**) Cytometry distributions of GFP before (white) and after (gray) arabinose induction. The dashed lines show the threshold at which cells are considered to be in the ON state (10^2 for all switches). The mean ± s.d. of GFP fluorescence and fractions of GFP ON population before and after arabinose induction for three experiments done on different days are shown in **Supplementary Figure 3**. (**d**) Cell growth (optical density (OD) 600 nm) as a function of arabinose concentration. Shading indicates growth rate reductions >25% (ref. 35). Data represent the average ± s.d. for three independent experiments. (**e**) Orthogonality matrix for the recombinases and their recognition sites. Population ON, the percentage of cells above a GFP threshold of 10^2 a.u. Data represent the average of three independent replicates done on different days (averages and s.d. are provided in **Supplementary Table 5**). Ara, arabinose.

site (**Supplementary Fig. 4** and **Supplementary Note 3**). We found only two integrases (Int1 and Int6) nonfunctional and excluded them from further analysis. The sequences of the functional integrases and other genetic parts are listed in **Supplementary Table 3**.

We constructed memory switches using the 11 functional recombinases (**Fig. 2a**). Each switch responded to increasing inducer concentration (**Fig. 2b**); the majority of the integrases showed >90% switching after 8 h (**Fig. 2c** and **Supplementary Fig. 3**). There was negligible leakage for all of the switches at the uninduced state, and we observed no switching by PCR. Int8 showed the fastest switching rate, requiring <2 h to turn fully on (**Supplementary Fig. 3**). Of the remainder, six required 2–4 h and four turned on after 4–6 h. These rates include the time required to activate the arabinose-inducible promoter, the expression of the integrase, the switching rate and the expression of the GFP reporter to the steady-state level. We also measured the impact on growth (**Fig. 2d**) and found that the majority of the integrases are nontoxic except at very high levels of expression.

Memory array composed of orthogonal switches

Among the 11 functional integrases, the *attB* and *attP* sites share no nucleotide identity, and even the size and sequence of the cores differ considerably (**Supplementary Table 4**). In addition, we tested orthogonality for two bacterial invertases (FimE and HbiF) and their recognition sites[32,33]. The recombinase and reporter plasmids were cotransformed into *E. coli* DH10B in all possible combinations. We assayed each combination for activity, which is reported in terms of the percentage of the population that is expressing GFP (**Fig. 2e**, **Supplementary Fig. 5** and **Supplementary Table 5**)

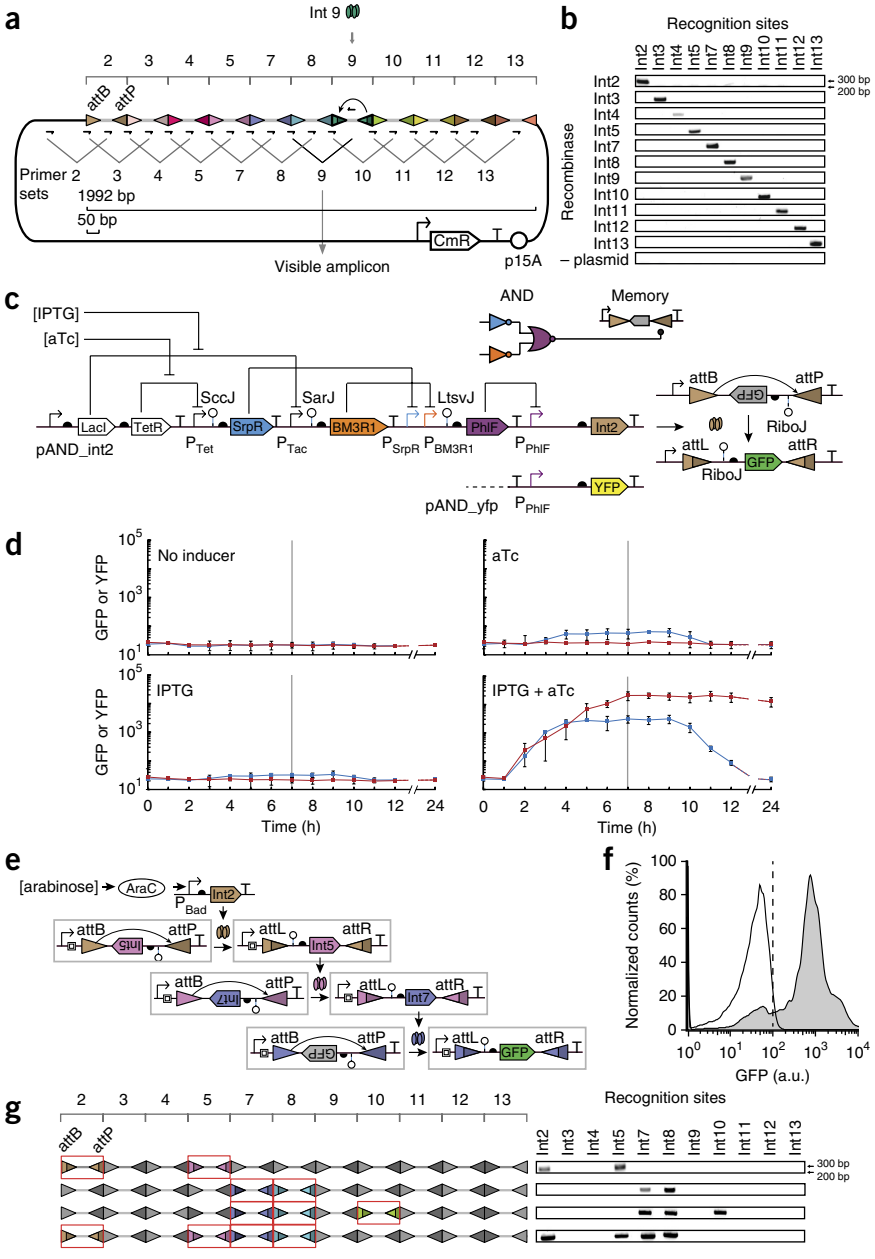

Figure 3 | Incorporation of recombinases into larger genetic circuits. (**a**) The memory array was designed as a linear concatenation of recognition sites for each integrase with a different 50-bp spacer (gray lines) between each pair of recognition sites. Primer pairs bind at the interface between recognition sites and within the spacer so that PCR amplification takes place only when the spacer is inverted. Int9 is shown as an example. (**b**) PCR products of the indicated sites; control (– plasmid) contains no integrase. (**c**) The wiring diagram and genetic system for the AND gate connected to a memory switch constructed with Int2 (top, color coding indicates repressor). The AND gate is also connected to a gene encoding YFP as a control (bottom). (**d**) Fluorescence over time with different inducer combinations. Blue lines show the fluorescence when the P_{PhlF} promoter is fused directly to *yfp*; red lines show induction of Int2. Inducers were added at time point $t = 0$ h and removed at $t = 7$ h (gray line). Dashed lines represent an extrapolation to $t = 24$ h accounting for dilution owing to cell division. Data represent mean ± s.d. of three independent experiments done on different days. (**e**) The three-layer cascade of phage integrases. Each integrase changes the orientation of the constitutive promoter (Registry of Biological Parts BBa_J23119), and the same spacer and RiboJ insulator (**Supplementary Fig. 12**) is used to insulate the integrases at each stage. (**f**) Cytometry distributions for the cascade in the absence (white) and presence (gray) of inducer. The vertical dashed line demarcates the threshold used to determine whether cells were on or off. Data represent three experiments done on different days. (**g**) Left, memory array after inversion using multiple genes encoding integrases (Int2 and Int5; Int7 and Int8; Int7, Int8 and Int10; and Int2, Int5, Int7 and Int8) organized in an operon controlled by an arabinose-inducible promoter. Colors and boxes indicate *attB* and *attP* sites that are switched. Right, the DNA bands amplified using 11 primer sets described in **a**.

and the change in fluorescence observed (**Supplementary Figs. 5** and **6** and **Supplementary Table 6**). There was no detectable cross-talk between the integrases and non-target–recognition sites except between the Int10 recognition sites and the Int7, Int8 and Int11 integrases, which showed a low level of cross-talk. These data indicate that the 13 recombinases are highly orthogonal to one another.

The orthogonality of the memory switches allows them to be used simultaneously to record different events. We constructed an array by concatenating the *attB* and *attP* sites of the 11 phage integrases to form a linear 2-kb piece of DNA (**Fig. 3a** and **Supplementary Table 7**). We included random spacers (50 bp) with 50% GC content between the *att* sites (**Supplementary Table 8**). We designed unique primers on the basis of the locations of the *attB* and *attP* sites to detect all 11 switching events possible (**Supplementary Table 9**). The final designed array is encoded in 2 kb of DNA and can record 11 bits (1.375 bytes) of information. It can be used to distinguish 2^{11} (2,048) possible combinations of events. We constructed the final array design using DNA synthesis.

The memory array plasmid (**Fig. 3a**) and each integrase plasmid (**Fig. 2a**) were cotransformed into DH10B cells. The 11 strains containing the memory array plasmid and one of the integrase plasmids were induced with 2 mM arabinose for 4 h. We assayed DNA inversion using the 11 pairs of primers designed for each recognition site. Amplification occurred only for the primer sets corresponding to the cognate pairs of integrase and *attB* and *attP* sites (**Fig. 3b**). In the absence of inducer, a negligible level of background switching occurred (**Supplementary Fig. 7**). Notably, we built only one memory array for this work (none of an intermediate size), and all of the *att* sites were functional without additional tuning or debugging[16].

1-bit memory array recording the output of a logic gate

The memory array is useful to permanently record whether a combination of environmental or cellular signals was encountered at the same time[34]. Transcriptional logic gates could perform signal integration, the output of which is recorded as memory that could be retrieved at a later stage of computation. Logic operations based solely on recombinases are irreversible (once observed, an input signal can never be forgotten) and are thus unable to resolve the order or co-occurrence of input signals.

In an example of combining digital logic with 1-bit memory, we constructed a transcriptional AND circuit by layering NOR and NOT gates (**Fig. 3c**). There are two inducible systems (based on the lac repressor LacI and the Tet repressor TetR) that serve as surrogates for environmental signals. Their corresponding P_{Tac} and P_{Tet} promoters are the two input promoters to the circuit. The AND function is composed of two NOT gates and a NOR gate using the SrpR, BM3R1 and PhlF repressors, respectively[35]. Each repressor contains a variant of the RiboJ insulator (SccJ, SarJ or LtsvJ) to reduce the effect of genetic context that arises from different promoter inputs[29]. The output promoter of the AND circuit (P_{PhlF}) is connected to the Int2 integrase, which then interacts with the array to permanently record whether both input signals (anhydrotetracycline (aTc) and isopropyl-β-D-thiogalactoside (IPTG)) are observed at the same time.

The dynamics of the AND gate in the presence of different combinations of inputs is shown in **Figure 3d** and **Supplementary Figure 8**. We measured the activation of P_{PhlF} without memory separately (**Fig. 3d**, blue lines). The circuit turns on only in the presence of both inducers , and after 7 h it reaches 84-fold induction. At this time point, the cells are diluted into media lacking inducer that returns the inputs to the [0,0] state. This reversal takes ~2 h to begin because the induced repressors must be diluted over time via cell division. The circuit returns to the off state with a timescale limited by degradation and dilution and reverts completely within 24 h. When the output of the AND gate is connected to the memory array, the circuit still relaxes to the off state (P_{PhlF}), but the transient on state is recorded permanently (**Fig. 3d**, red line). There was a ~1-h delay in triggering the memory switch, and once the inversion occurs it produced stable expression for >24 h. The switch also functions as a filter, where aTc alone causes a small increase in the output, and this leakiness did not cross the threshold required for the memory switch. The strong constitutive promoter in the switch amplified the output of the circuit by increasing the dynamic range to more than 1,000-fold (**Fig. 3d**).

Circuits composed of multiple recombinases

We constructed circuits to demonstrate that multiple orthogonal recombinases can be used in a single cell without interference. We arranged three of the new integrases (Int2, Int5 and Int7) to form a cascade (**Fig. 3e**). When we induced the cells with arabinose for 12 h, almost the entire population (92%) progressed to the final layer of the cascade (**Fig. 3f**). We constructed a shorter two-integrase cascade based on Int5 and Int7, and 89% of the cells were induced after 8 h (**Supplementary Fig. 9**). The average fluorescence of the induced population remained similar for the two- and three-integrase cascades (611 ± 43 and 908 ± 10 mean ± s.d. arbitrary units (a.u.), respectively. This was expected, as the same constitutive promoter dictates the expression level at each layer. This is in contrast to transcription factor–based cascades, where the signal properties change at each layer owing to ultrasensitivity[36], mismatches in the transfer function[35] and the response properties of the final output promoter[37,38]. We constructed operons containing two, three or four genes encoding integrases (Int2 and Int5; Int7 and Int8; Int7, Int8 and Int10; and Int2, Int5, Int7 and Int8) (**Fig. 3g**). The integrases were induced with 2 mM arabinose, and DNA inversion was assayed with 11 pairs of primers (**Fig. 3a**). We detected amplification only for the primer sets of the corresponding integrases. This demonstrates that the memory array is able to write multiple bits of information according to the expression of specific integrases. Use of multiple integrases did not result in slower growth (**Supplementary Fig. 10**).

DISCUSSION

This work expands the programmable memory capacity in a living cell to beyond 1 byte of information. This allows engineering bacteria to permanently record multiple environmental and cellular stimuli that can be recalled at a later stage of the computation or interrogation of the exposure of the bacteria to particular conditions. The design of the memory array is simple and robust, requiring only the stringing together of recognition sites into a linear DNA sequence. The combination of the recognition sites to build the array worked in the first attempt. The state of these memory devices can be read through reporter genes or nucleic acid–based assays even if the chassis organism is dead, which is beneficial for real-world applications. A memory array can be connected to an environmental sensor to elucidate how bacteria respond to difficult-to-assay environments, such as niches within the human body or within biofilms and microbial communities[39].

The recombinases demonstrate high orthogonality, with essentially no measureable cross-talk with off-target recognition sites. This differs from other DNA-binding proteins, where small operators and sequence degeneracy leads to cross-talk such that many variants have to be screened to obtain a small orthogonal set[35,40,41]. All 11 functional integrases were also highly orthogonal, and none had to be eliminated owing to cross-talk. This allowed the recognition sites for all functional integrases to be used together to build an 11-bit array. This approach can be scaled to higher capacities, as there are >4,000 integrases in the sequence databases, and our bioinformatics approach yielded 34 predicted *att* sites that are diverse and likely to be orthogonal. A higher memory capacity enables new classes of computing that can be performed in cells. Memory allows intermediate calculations to be stored so that the same computing units can be used repetitively, rather than having specialized computing-memory circuits. Almost all modern computers incorporate architectures analogous to that shown in **Figure 3c**, in which combinatorial logic circuits store the output of their computations in memory, which can then be accessed by later computing steps. This feature enables the creation of complex sequential logic systems and state machines.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

C.A.V., T.K.L. and L.Y. are supported by the US Defense Advanced Research Projects Agency (DARPA CLIO N66001-12-C-4016). C.A.V. and A.A.K.N. are supported by DARPA CLIO N66001-12-C-4018. C.A.V., M.T.L., A.A.K.N. and J.F.-R. are supported by the Office of Naval Research Multidisciplinary University Research Initiative (N00014-13-1-0074; Boston University MURI award 4500000552). C.A.V. is also supported by US National Institutes of Health (GM095765), the US National Institute of General Medical Sciences (P50 GM098792) and the US National Science Foundation Synthetic Biology Engineering Research Center (SynBERC EEC0540879). A.A.K.N. receives government support FA9550-11-C-0028 and is supported by the National Defense Science and Engineering Graduate Fellowship 32 CFR 168a from the US Department of Defense Air Force Office of Scientific Research.

AUTHOR CONTRIBUTIONS

C.A.V. and L.Y. conceived of the study and designed the experiments. L.Y., A.A.K.N., C.J.M. and J.F.-R. performed the experiments and analyzed the data. C.A.V., L.Y., A.A.K.N. and C.J.M. wrote the manuscript. C.A.V., T.K.L. and M.T.L. managed the project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Burrill, D.R. & Silver, P.A. Making cellular memories. *Cell* **140**, 13–18 (2010).
2. Yamanishi, M. & Matsuyama, T. A modified Cre-*lox* genetic switch to dynamically control metabolic flow in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* **1**, 172–180 (2012).
3. Ham, T.S., Lee, S.K., Keasling, J.D. & Arkin, A.P. A tightly regulated inducible expression system utilizing the fim inversion recombination switch. *Biotechnol. Bioeng.* **94**, 1–4 (2006).
4. Kawashima, T. *et al.* Functional labeling of neurons and their projections using the synthetic activity-dependent promoter E-SARE. *Nat. Methods* **10**, 889–895 (2013).
5. Zariwala, H.A. *et al.* A Cre-dependent GCaMP3 reporter mouse for neuronal imaging *in vivo*. *J. Neurosci.* **32**, 3131–3141 (2012).
6. Kotula, J.W. *et al.* Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc. Natl. Acad. Sci. USA* **111**, 4838–4843 (2014).
7. Archer, E.J., Robinson, A.B. & Suel, G.M. Engineered *E. coli* that detect and respond to gut inflammation through nitric oxide sensing. *ACS Synth. Biol.* **1**, 451–457 (2012).
8. Ingolia, N.T. & Murray, A.W. Positive-feedback loops as a flexible biological module. *Curr. Biol.* **17**, 668–677 (2007).
9. Gardner, T.S., Cantor, C.R. & Collins, J.J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
10. Ajo-Franklin, C.M. *et al.* Rational design of memory in eukaryotic cells. *Genes Dev.* **21**, 2271–2276 (2007).
11. Burrill, D.R., Inniss, M.C., Boyle, P.M. & Silver, P.A. Synthetic memory circuits for tracking human cell fate. *Genes Dev.* **26**, 1486–1497 (2012).
12. Greber, D., El-Baba, M.D. & Fussenegger, M. Intronically encoded siRNAs improve dynamic range of mammalian gene regulation systems and toggle switch. *Nucleic Acids Res.* **36**, e101 (2008).
13. Ham, T.S., Lee, S.K., Keasling, J.D. & Arkin, A.P. Design and construction of a double inversion recombination switch for heritable sequential genetic memory. *PLoS ONE* **3**, e2815 (2008).
14. Moon, T.S. *et al.* Construction of a genetic multiplexer to toggle between chemosensory pathways in *Escherichia coli*. *J. Mol. Biol.* **406**, 215–227 (2011).
15. Bonnet, J., Subsoontorn, P. & Endy, D. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. USA* **109**, 8884–8889 (2012).
16. Bonnet, J., Yin, P., Ortiz, M.E., Subsoontorn, P. & Endy, D. Amplifying genetic logic gates. *Science* **340**, 599–603 (2013).
17. Friedland, A.E. *et al.* Synthetic gene networks that count. *Science* **324**, 1199–1202 (2009).
18. Nielsen, A.A., Segall-Shapiro, T.H. & Voigt, C.A. Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression. *Curr. Opin. Chem. Biol.* **17**, 878–892 (2013).
19. Brown, W.R., Lee, N.C., Xu, Z. & Smith, M.C. Serine recombinases as tools for genome engineering. *Methods* **53**, 372–379 (2011).
20. Smith, M.C. & Thorpe, H.M. Diversity in the serine recombinases. *Mol. Microbiol.* **44**, 299–307 (2002).
21. Smith, M.C., Brown, W.R., McEwan, A.R. & Rowley, P.A. Site-specific recombination by phiC31 integrase and other large serine recombinases. *Biochem. Soc. Trans.* **38**, 388–394 (2010).
22. Marchler-Bauer, A. & Bryant, S.H. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* **32**, W327–W331 (2004).
23. Li, W. *et al.* Structure of a synaptic γδ resolvase tetramer covalently linked to two cleaved DNAs. *Science* **309**, 1210–1215 (2005).
24. Rutherford, K., Yuan, P., Perry, K., Sharp, R. & Van Duyne, G.D. Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res.* **41**, 8341–8356 (2013).
25. Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J. & Wishart, D.S. PHAST: a fast phage search tool. *Nucleic Acids Res.* **39**, W347–W352 (2011).
26. Canchaya, C. *et al.* Genome analysis of an inducible prophage and prophage remnants integrated in the *Streptococcus pyogenes* strain SF370. *Virology* **302**, 245–258 (2002).
27. Brenciani, A. *et al.* Phim46.1, the main *Streptococcus pyogenes* element carrying *mef*(A) and *tet*(O) genes. *Antimicrob. Agents Chemother.* **54**, 221–229 (2010).
28. Deuschle, U., Kammerer, W., Gentz, R. & Bujard, H. Promoters of *Escherichia coli*: a hierarchy of *in vivo* strength indicates alternate structures. *EMBO J.* **5**, 2987–2994 (1986).
29. Lou, C., Stanton, B., Chen, Y.J., Munsky, B. & Voigt, C.A. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol.* **30**, 1137–1142 (2012).
30. Salis, H.M., Mirsky, E.A. & Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
31. Farasat, I. *et al.* Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Mol. Syst. Biol.* **10**, 731 (2014).
32. Klemm, P. Two regulatory *fim* genes, *fimB* and *fimE*, control the phase variation of type 1 fimbriae in *Escherichia coli*. *EMBO J.* **5**, 1389–1393 (1986).
33. Xie, Y., Yao, Y., Kolisnychenko, V., Teng, C.H. & Kim, K.S. HbiF regulates type 1 fimbriation independently of FimB and FimE. *Infect. Immun.* **74**, 4039–4047 (2006).
34. Clarke, E.J. & Voigt, C.A. Characterization of combinatorial patterns generated by multiple two-component sensors in *E. coli* that respond to many stimuli. *Biotechnol. Bioeng.* **108**, 666–675 (2011).
35. Stanton, B.C. *et al.* Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.* **10**, 99–105 (2014).
36. Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl. Acad. Sci. USA* **102**, 3581–3586 (2005).
37. Moon, T.S., Lou, C., Tamsir, A., Stanton, B.C. & Voigt, C.A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249–253 (2012).
38. Tamsir, A., Tabor, J.J. & Voigt, C.A. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* **469**, 212–215 (2011).
39. Costello, E.K., Stagaman, K., Dethlefsen, L., Bohannan, B.J. & Relman, D.A. The application of ecological theory toward an understanding of the human microbiome. *Science* **336**, 1255–1262 (2012).
40. Rhodius, V.A. *et al.* Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. *Mol. Syst. Biol.* **9**, 702 (2013).
41. Khalil, A.S. *et al.* A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647–658 (2012).

ONLINE METHODS

Strains and media. *E. coli* DH10B (F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L *nup*G λ–)[42] was used for genetic manipulation and characterization. *E. coli* DH10B has the *fimE/fimB* and *fim* structural genes deleted[42]. Cells were grown in LB Miller broth (Difco, 90003-350) for functional assays and SOB (Teknova, S0210) (2% Bacto-tryptone, 0.5% Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl) for cloning. Chloramphenicol (34 μg/ml) (Alfa Aesar, AAB20841-14), kanamycin (50 μg/ml) (GoldBio, K-120-10) or spectinomycin sulfate (50 μg/mL) (MP Biomedicals LLC, 158993) was supplemented where appropriate. Arabinose (Sigma-Aldrich, MO, A3256), IPTG (GoldBio, I2481C25) and aTc (Sigma-Aldrich, 37919) were used as inducers. For arabinose induction system, 0.5% (w/v) glucose was used to reduce the leakiness of the uninduced state. Three fluorescence proteins, GFPmut3 (ref. 43), mRFP1 (ref. 44) and YFP[43], were used as reporters.

Bioinformatics for integrase discovery. To identify LSTP integrases from the protein database, we analyzed the conserved domains of several known LSTP integrases using the Conserved Domains search tool[22] (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) with default parameters ($E < 0.01$). The NCBI protein database was searched for proteins that contain two or three of the identified domains using CDART with default parameters (http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi).

For each LSTP integrase candidate, the genome was retrieved and scanned using PHAST[25] (http://phast.wishartlab.com/) with default parameters to detect clusters of phage-like genes adjacent to the integrase-encoding genes of interest. Integrase genes that are known or located in plasmids and phages were manually removed from the list. All the databases were updated until 22 October 2012. To identify the *attL* and *attR* sites (prophage boundary), we searched the prophage genome together with 10-kb up- and downstream sequences against all the homologous genomes belonging to the same genus using Megablast provided by NCBI (default parameters). The BLAST results were manually scanned for patterns described in **Figure 1b** and **Supplementary Figure 1**. Finally, a library of 34 integrases and their cognate *attL* and *attR* and *attB* and *attP* sites were identified. The phylogenetic analysis of the 34 integrase protein sequences was performed using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) (default parameters). The tree view was constructed using the 'letsmakeatree' function in Matlab (MathWorks Inc.).

Codon optimization and DNA synthesis. The recombinase library was codon optimized by GeneArt for *E. coli* K12 MG1655, synthesized and assembled into the parent vectors using the one-step isothermal DNA assembly method[45]. The memory array was designed by concatenating the *attB* and *attP* sites of integrases 2, 3, 4, 5, 7, 8 ,9, 10 and 11 in numerical order. Between each *attB* and *attP* site pair, a random 50-bp spacer was inserted. The spacer was designed to have a GC content of 50% using the Sequence Manipulation Suite[46]. The 1,992-bp design was synthesized by GeneArt and then assembled into the parent plasmid by golden-gate method (**Fig. 3a**)[47].

Flow cytometry analysis. Fluorescence was measured using an LSRII flow cytometer (BD Biosciences) or MACSQuant VYB (Miltenyi Biotec) with a 488-nm laser for GFP and YFP measurement. For each sample at least 10^4 events were recorded using a flow rate of 0.5 μl/s. FlowJo v10 (TreeStar Inc.) was used to analyze the data. All events were gated by forward scatter and side scatter. RFP fluorescence (10^2–10^5 a.u.) was also used for gating cells containing the reporter plasmids. Events corresponding to negative GFP fluorescence were excluded. The background fluorescence of *E. coli* DH10B cells without plasmids was subtracted before calculating the fold change.

Characterization of memory switches. *E. coli* DH10B cells containing only the reporter plasmid were made chemically competent using Z-competent reagents (Zymo Research, T3001) and transformed with plasmids containing different integrases. The transformants were selected on LB agar (1%) supplemented with kanamycin, chloramphenicol and 0.5% glucose. Three colonies were picked for biological replication. The overnight cultures were prepared in LB supplemented with kanamycin, chloramphenicol in the presence of 0.5% glucose. For functional assays, all cultures were grown at 37 °C in V-bottom 96-well plates (Nunc, 249952) covered with air-permeable membranes AeraSeal (E&K Scientific) in an ELMI Digital Thermos Microplates shaker (1,000 r.p.m.) (Elmi Ltd.). This transformation method and culture conditions were used for data shown in **Figure 2** and **Supplementary Figures 3–5**.

To characterize the function of the integrases (**Fig. 2b–d** and **Supplementary Fig. 3**), the transformants with the cognate pair of integrase-reporter plasmids were used to test the function of each integrase. The overnight cultures were washed twice with LB and diluted 200:1 in 200 μl LB containing 50 μg/ml kanamycin and 34 μg/ml chloramphenicol in the presence of various inducers. For the induction of the integrases (**Fig. 2b**), cells were induced with 0 (with 0.5% glucose, control), 0.001, 0.01, 0.1 or 1 mM arabinose for 8 h. To characterize the function and efficiency of the integrases (**Fig. 2c** and **Supplementary Fig. 3a–c**), cells were induced with 0 (with 0.5% glucose) or 1 mM arabinose for 8 h. To characterize the induction of integrases over time (Supplementary **Fig. 3d**), cells were induced with 1 mM arabinose for 15 h. For flow cytometer analysis, a 2–20 μl aliquot of each culture was added to 198 μl PBS containing 2 mg/ml kanamycin and stored at 4 °C for 16 h.

To test the toxicity of different single integrases (**Fig. 2d**) cells were induced with 0 (with 0.5% glucose), 0.001, 0.01, 0.1 or 1mM arabinose for 8 h. 150 μl of cultures were transferred into a flat-bottom 96-well plate (Nunc, 165305) to measure OD (600 nm) using a Synergy H1 Hybrid Microplate Reader (BioTek).

To test of the orthogonality of the integrases (**Fig. 2e** and **Supplementary Fig. 6**), cells containing all combinations of integrase plasmids and reporter plasmids were induced with 0 mM (with 0.5% glucose) or 1 mM arabinose for 6 h and 2 μl of cultures were prepared as for **Figure 2** and used for flow cytometer analysis.

For testing the impact of expressing multiple integrases on cell growth (**Supplementary Fig. 10**), overnight cultures were adjusted to the same OD and diluted 1:200 into LB (supplemented with 50 μg/ml kanamycin and 34 μg/ml chloramphenicol) containing various concentrations of arabinose (0.5% glucose was supplemented at 0 mM arabinose) in 96-well plates sealed with Breathe-Easy sealing membrane (Sigma-Aldrich) for 12 h at

37 °C. A Synergy H1 Hybrid Microplate Reader (BioTek) was used for cultivation and OD measurement.

PCR and sequencing verification of DNA inversion. For colony PCR, 25 µl aliquots of the cultures were heated at 95 °C for 10 min and 2 µl of supernatants were used for PCR analysis. A set of primers was designed to confirm the DNA inversion catalyzed by the novel integrases by PCR analysis. Primer 1 (CAATACCTTTT AACTCGATTCTATTAACAAG) was located in the middle of the *gfp* coding sequence and primer 2 (CAGTGCCAACATAGTAAG CCAGTAT) was located downstream of the DNA inversion region (**Fig. 2a**). A PCR product was generated only when the cognate integrase was induced and the DNA fragment between *attB* and *attP* was flipped. The PCR products were also sequenced using primer 1 to confirm that they contained the *attR* or *attL* sequences, as predicted. Primer 3 (TTGACAGCTAGCTCAGT CCTAGGTATAATGC) and primer 4 (GGGGTTTTTTTTTGG GTATGGGCCCTAG) (**Fig. 2a**) were also used to verify the OFF state by PCR and sequencing. The sequences of OFF state are the same as designed and an example of Int2 reporter (from the P_{const} to *attB*) is listed in **Supplementary Table 3**.

To analyze the memory array, *E. coli* DH10B cells were made chemically competent and cotransformed with the memory device plasmid and each of the functional controller plasmids. Colonies were picked into LB and 0.4% glucose and grown overnight. 10 µl of each culture was then added to 100 µl LB supplemented with 0.4% glucose (uninduced) or 100 µl LB supplemented with 2 mM arabinose (induced). After 4 h, 2 µl of each culture was added to a 25 µl GoTaq PCR mix (Promega) for each pair of analytical primers. All PCR reactions were run at 72 °C for 25 cycles before electrophoresis in a 1% agarose gel. Inversion was indicated by the presence of a ~300-bp band. Each of the induced cultures where flipping was observed was further confirmed by MiniPrep (Qiagen) and sequencing. Moreover, the sequence of memory array after induction of the three-integrase operon (Int7, Int8 and Int10, encoded in plasmid pCis_7+10+8) and four-integrase operon (Int2, Int5, Int7 and Int10, encoded in plasmid pCis_2+7+8+5) were confirmed by sequencing the plasmids using primers JFR57 (CATTTTAGCTTCCTTAGCTCCTG), CM209 (CATTAGAGGTCGTATCCTATCGCGATAATTCC) and CM213 (GCATGAGGCTGCCTGAGATCCTCTA). The sequence of memory array after induction of four-integrase operon (pCis_2+7+8+5) is shown in **Supplementary Table 7**.

Recording of the digital AND circuit. *E. coli* DH10B cells were cotransformed so that they contained either pAND-yfp and a control plasmid (pSpec) or pAND_Int2 and pAND_reporter, which contains the *attB* and *attP* sites of integrase Int2 (**Fig. 3c** and **Supplementary Fig. 11**). The cultures were grown overnight in LB media with 50 µg/ml kanamycin and 50 µg/ml spectinomycin at 37 °C and 1,000 r.p.m. using 2-ml 96 deep-well plates (USA Scientific, 1896-2000) in a Multitron Pro shaker-incubator (*In vitro* Technologies). After overnight growth, cells were diluted 1:500 into 500 µl of LB with antibiotic in a new 96-deep-well plate and grown for 3 h in the shaker-incubator. Next, a 10-µl aliquot of culture was suspended in 190 µl PBS with 2 mg/ml kanamycin and stored for cytometry analysis (time point 0 h). The remaining culture was divided into four 300-µl cultures by diluting cells 1:3 into medium with antibiotic. Each of the four cultures contained a different combination of inducers. A time point was taken every hour for 7 h by storing 10 µl of cells into PBS with kanamycin. During this time course, every hour, the cultures were diluted 1:3 into fresh LB with corresponding inducers and antibiotics. After 7 h, the cultures were spun down at 4,000 r.c.f. and resuspended in LB medium without inducers; this wash step was repeated a second time. Samples continued to be taken every hour for another 5 h by storing 10 µl of cells in PBS with kanamycin, and then diluting cultures 1:3 into fresh LB with antibiotics and without inducers. Subsequently, the cells were grown overnight for 12 h without dilution. Finally, 1 µl of cells was stored in 199 µl PBS with 2 mg/ml kanamycin for flow cytometry analysis.

Analysis of integrase cascades. *E. coli* DH10B cells were transformed with plasmids encoding the two-integrase cascade (pInt5 and pCasc_5+7_gfp) or three-integrase cascade (pCas_2+5 and pCas_5+7_gfp) (**Supplementary Fig. 12**). Three colonies were picked for biological replication. The overnight cultures were prepared in LB supplemented with 50 µg/ml kanamycin and 34 µg/ml chloramphenicol. The overnight cultures were diluted 1:200 into 200 µl LB media supplemented with kanamycin and chloramphenicol and cultivated with 0 mM arabinose or with 1 mM arabinose. The cells were grown aerobically at 37 °C, 1,000 r.p.m. for 12 h in V-bottom 96-well plates (Nunc, 249952) covered with air-permeable membranes AeraSeal (E&K scientific) in an ELMI Digital Thermos Microplates shaker (Elmi Ltd). For FACS analysis 2 µl of each culture was added to 198 µl PBS (stored at 4 °C) containing 2mg/ml kanamycin for flow cytometer analysis, and RFP fluorescence was used to assist gating cells.

Characterization of the memory array controlled by multiple integrases. Different combinations of integrase-encoding genes were cloned in a polycistron under the control of the P_{Bad} promoter, leaving intact the original RBS (listed in **Supplementary Table 2**). Plasmids containing these combinations were transformed into cells containing an array of 11 integrase sites (**Fig. 3a**) and cultured onto LB agar plates containing kanamycin and chloramphenicol and supplemented with 0.5% glucose. After 16 h, single colonies were inoculated into 500 µl LB with kanamycin and chloramphenicol plus 0.5% glucose and grown overnight. These overnight cultures were then diluted 1:20 into 5 ml of fresh LB containing kanamycin and chloramphenicol and supplemented with either 0.5% glucose (uninduced conditions) or 1 mM arabinose (induced conditions). Cells containing a two-integrase plasmid (pCis_2+5 or pCis_7+8) were grown under these conditions for 4 h; cells with a three-integrase plasmid (pCis_7+10+8) were grown in the same conditions overnight; and cells containing the four-integrase plasmid (pCis_2+7+8+5) were grown overnight under induced and uninduced conditions and diluted the next day 1:20 into fresh LB plus antibiotics to be subjected to a second cycle of induction. For all integrase combinations, 4 µl of the final culture was used in a PCR reaction to query the state of their cognate site and all the other integrase sites using primers listed in **Supplementary Table 9**. The plasmid maps are shown in **Supplementary Figure 13**. The plasmids used in this study are available on request from Addgene (https://www.addgene.org/browse/article/9134/).

doi:10.1038/nmeth.3147

42. Durfee, T. *et al.* The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *J. Bacteriol.* **190**, 2597–2606 (2008).
43. Cormack, B.P., Valdivia, R.H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).
44. Campbell, R.E. *et al.* A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. USA* **99**, 7877–7882 (2002).
45. Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
46. Stothard, P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**, 1104 (2000).
47. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS ONE* **4**, e5553 (2009).