Published in final edited form as:

Curr Opin Biomed Eng. 2017 December ; 4: 143–151. doi:10.1016/j.cobme.2017.10.003.

Scaling Computation and Memory in Living Cells

Kevin Yehl^{1,2} and Timothy Lu^{1,2,3,4,5}

¹Synthetic Biology Group, MIT Synthetic Biology Center, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

²Research Laboratory of Electronics, MIT, Cambridge, MA 02139, USA

³Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139, USA

⁴Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA 02139, USA

⁵Department of Biological Engineering, MIT, Cambridge, MA 02129, USA

Abstract

The semiconductor revolution that began in the 20th century has transformed society. Key to this revolution has been the integrated circuit, which enabled exponential scaling of computing devices using silicon-based transistors over many decades. Analogously, decreasing costs in DNA sequencing and synthesis, along with the development of robust genetic circuits, are enabling a “biocomputing revolution”. First-generation gene circuits largely relied on assembling various transcriptional regulatory elements to execute digital and analog computing functions in living cells. Basic design rules and computational tools have since been derived so that such circuits can be scaled in order to implement complex computations. In the past five years, great strides have been made in expanding the biological programming toolkit to include recombinase- and CRISPR-based gene circuits that execute complex cellular logic and memory. Recent advances have enabled increasingly dense computing and memory circuits to function in living cells while expanding the application of these circuits from bacteria to eukaryotes, including human cells, for a wide range of uses.

Introduction

Synthetic biology is a highly interdisciplinary field that focuses on engineering biological systems for a wide range of applications [1]. A major goal in synthetic biology is to perform computation and memory in living cells, thus enabling the programming of “smart” cells that have novel functions [2]. This is achieved through the design and implementation of synthetic gene circuits (Figures 1A and B). Gene circuits are composed of genetic parts that encode RNA, proteins, and gene regulatory elements. These parts direct the temporal and spatial execution of a network of chemical reactions that enable living organisms to

*Correspondence to: timlu@mit.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

dynamically sense, respond, and adapt to their environment. Being able to control and engineer such functions *de novo* would yield transformative applications in the biomedical sciences. However, limitations in the scalability and reusability of genetic parts and the lack of broadly applicable design principles have hindered progress in building high-performance synthetic circuits. The biological parts used to create these circuits often function in a context-dependent manner, thus requiring time-consuming and non-rational optimization strategies.

Now is an exciting time because advances in DNA sequencing, synthesis, and assembly technologies combined with high-throughput experimental approaches are enabling the identification and validation of robust circuit components and topologies. Such approaches are being taken to develop computational programs that can be used to design genetic circuits with much higher success rates than manual strategies [3], [4]. This will eventually lead to the democratization of biological programming technologies, thus enabling experts and non-experts alike to realize novel applications much more efficiently.

Memory is a central feature for complex computing, as it enables current behaviors to be dependent on past history as well as current events. Memory is a hallmark of natural living systems, which are able to record events through many different molecular strategies, including transcriptional pathways, epigenetic mechanisms, and protein-based memory. The ability to encode artificial memory in living cells is broadly useful. For example, the stable expression of a protein or metabolic pathway could be triggered transiently with a chemical inducer, thus reducing the cost of a bio-manufacturing process. A cancer cell's history of exposure to various environmental signals could be recorded and then read out at a later time, thus allowing scientists to determine which signals are involved in tumorigenesis. Similarly, factors controlling stem cell differentiation can be recorded and used to develop genetic circuits that enhance tissue and organ regeneration.

One of the first demonstrations of synthetic memory in living cells was the synthetic toggle switch (Figure 1C) [5]. This circuit encodes two stable states via two mutually inhibitory transcriptional repressors, thus enabling the programming of bacteria to "remember" an exposure to a chemical input by continually expressing one of the two repressors. However, this memory is transient and requires continual protein expression to function. Scaling this memory architecture is burdensome to the cell because its resources are finite, which impose potential limits on memory durability and capacity.

There are many examples of memory in natural systems that are far more sophisticated than what the simple toggle switch could achieve. The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas system allows bacteria to remember exposures to viral infections by recording viral genetic signatures in the bacterial genome, which are recalled later to fight off future infections [6]. In eukaryotes, epigenetics enable organisms to transcriptionally control gene expression by modifying the genome with various chemical groups [7]. There has been a drive to increase biological memory capacities and achieve robust recording through multiple cellular generations using synthetic gene circuits that recapitulate these biological mechanisms. A major advance in implementing such next-generation memory devices has been to record biological events into DNA as opposed to

relying on continual protein expression. DNA has the beneficial properties of stability, heritability, transferability, and high information density; moreover, DNA can be sequenced and analyzed after cell death. This review focuses on advances in synthetic biology made in the past five years using DNA for cellular computing and memory.

Recombinase-based cellular computing

One direction that is being explored for designing genetic circuits is the use of recombinases. Recombinases constitute a class of enzymes that modify DNA at specific sites known as ‘attachment’ (*att*) sites (Figure 2A) [8]. Depending on the recombinase (sometimes also referred to as an invertase or integrase) and the orientation of the *att* sites, recombinases can flip, excise, or insert DNA sequences. In some cases, depending on the recombinase or accessory protein being used, these reactions are reversible, which introduces another degree of sophistication for programming genetic circuits. By placing genetic control elements and genes of interest between *att* sites in different topologies, complex logic operations that are dependent on recombinase activity can be performed. For example, recombinases have been used to construct the first cellular counters [9]. In addition, recombinase-based circuits have been constructed to execute all possible two-input Boolean logic functions in bacteria without requiring cascades of multiple logic gates [10]. In addition, this approach was used to create genetic circuits that performed digital-to-analog conversions in which the quantitative expression of an output gene could be set to any of four levels depending on whether two independent inputs were OFF or ON. This feature could be useful for setting expression levels in bio-manufacturing settings with transient inputs. Similarly, Bonnet *et al.* constructed recombinase-based Boolean logic circuits in bacteria but with a different organization of regulatory elements [11].

Weinberg *et al.* expanded upon these studies to show that single-layered recombinase-based circuits can be built in mammalian cells (Figure 2B). Specifically, they developed a general framework for designing single-layered complex circuits in mammalian cells, which they termed BLADE (Boolean logic and arithmetic through DNA excision) [12]. Impressively, this system was readily extensible as evidenced by the high success rate in their circuit designs: 109 of the 113 circuits functioned as intended, without optimization. The general architecture of the BLADE circuit is the reason for such a high success rate. BLADE utilizes recombinase operations that act upon a single transcriptional layer that can be customized for Boolean logic operations, thus bypassing the need to tune the output signal of one gate to match the input level of another gate. In the future, recombinase-based computers could be used to create dynamic cell-based therapies that can sense and diagnose diseases [12], [13]. For example, cell-based therapies could be engineered to express therapeutic payloads in response to multiple disease biomarkers.

Recombinase-based memory devices

In addition to performing logical operations, recombinase-based circuits are excellent at manipulating DNA to store memory. As described above, toggle switches based on transcriptional repressors rely on constant gene expression to record information. In contrast, recombinases store information in DNA sequences by catalyzing DNA recombination (e.g.,

inversion or excision). Recombinase-based storage requires recombinase protein expression only to change state but not to maintain a given state, thus minimizing the use of additional cellular resources. Moreover, with recombinase-based strategies, DNA can be transferred, inherited, and recovered after cell death.

For example, Siuti et al. demonstrated that recombinase-based memory of molecular events can be stored in plasmid DNA in *Escherichia coli* and transferred to daughter cells for 90+ generations [10]. Moreover, depending on the recombinase being expressed, DNA inversion can be reversed, thus enabling rewritable memory that could be useful for complex computing. This was demonstrated by Bonnet et al., who termed this approach rewritable recombinase addressable data (RAD) [14]. In the RAD system, DNA recording/inversion is initiated by the expression of Bxb1 integrase. When Bxb1 integrase is co-expressed with Bxb1 excisionase, the reverse reaction is catalyzed, thus erasing the memory registry. However, because the same recombinase catalyzes both the forward and reverse reactions, the relative expression levels and degradation rates of recombinase and excisionase have to be tuned for the circuit to function correctly. Fernandez-Rodriguez et al. constructed memory devices that use two irreversible complementary recombinases (FimE and HbiF), in which recombination of one set of *att* sites by one recombinase results in the formation of *att* sites that can be acted upon by the complementary recombinase [15]. In this study, the authors showed that the recorder was functional through six record and reset cycles over the course of 400 hours, after which it failed. Analysis showed that the recombinase genes and regulatory elements were replaced by the transposable insertion element IS1. This work highlights the importance of investigating the evolutionary stability of synthetic gene circuits and identifying failure modes, which will aid in the design of more robust circuits.

The inversion of a DNA sequence by a recombinase enables an individual bit of information to be stored in that sequence. If only one set of recombinase sites is used per cognate recombinase, then the storage capacity of recombinase-based memory is limited by the number of recombinases (n) to 2^n [16], [17]. Most recombinase-based circuits have relied on only a few “workhorse” enzymes, and their memory capacity has been limited due to a lack of well-characterized recombinases. As a result, a major focus in the field has been to expand the recombinase toolkit by using strategies such as protein engineering and data mining [18]–[20]. For example, Yang et al. used bioinformatics to identify 34 large serine-type phage (LSTP) integrases, 11 of which are orthogonal. By arraying the 11 sets of orthogonal recombinase recognition sites together, they demonstrated the ability to record 1.375 bytes of information (2,048 states) (Figure 2C), thus scaling the storage capacity for recombinase-based memory devices [20].

Recombinase-based state machines

Recombinase-based computing is not inherently limited to scaling by 2^n , which is the case when only one set of recombinase sites is used per cognate recombinase. To overcome this limitation, it is possible to design multiple independent pairs of recombinase recognition sites that are acted on by the same recombinase protein [17]. When multiple independent recombinase sites per recombinase are overlapped with each other and with other orthogonal recombinase sites, the resulting DNA recombination from one enzyme can alter the

subsequent DNA recombination for another recombinase. This type of organization enables programming of complex state machine-like behavior.

State machines exist in any number of states, in which transitions between states are controlled by inputs and the current state. In such systems, the order and identities of the inputs dictate the state (Figure 2D). State machines can be used to model natural biological processes, such as disease progression and cell differentiation during development. The concept of using recombinases for constructing state machines was demonstrated by Ham et al. using the FimB and Hin invertases in *E. coli*; however, only three of the five possible states were achieved due to the reversibility and lack of robustness in recombinase activity [16]. Roquet et al. greatly expanded upon this concept by using a set of irreversible recombinases (large serine recombinases: BxbI, TP901, and A118) and developing a framework for implementing state machines by mathematically determining the relationship between the structure of circuit design and circuit scalability [17]. Specifically, they constructed a two-input, five-state machine and a three-input, 16-state machine circuit in which the bacteria adopted distinct states in response to distinct permuted substrings of inputs - that is, every possible combination and ordering of inputs (Figure 2D). Additionally, a three-input passcode switch was built that had ~97% of the cells adopting the correct state in response to the appropriate set of inputs. These recombinase-based state machines achieve $>2^n$ scaling, thus exceeding the capacity of standard recombinase-computing with only one set of recombinase recognition sites per recombinase. These biological state machines have the potential to enable the scalable programming of cellular differentiation cascades, making it possible to study how the order and identities of biological events dictate cellular functions in natural systems.

CRISPR-based memory devices

Though recombinase-based circuits have been scaled to increase their memory capacity and computing power, their ultimate capacity is limited by the DNA “real-estate” required for enzyme activity and the number of available recombinases. Recombinase recognition sites are approximately 30 bp in length and require a minimum length of DNA between sites for an inversion/recording event. Additionally, multiple orthogonal recombinases are required for large-scale memory devices, for which the expression levels have to be balanced to minimize toxicity yet maintain functionality. As a result, CRISPR-Cas nucleases are being explored as an alternative strategy to encode scalable cellular-based memory.

The CRISPR-Cas system is a prokaryotic immune system that confers protection against foreign genetic elements, such as viruses and plasmids (Figure 3A) [6]. Each palindromic repeat of DNA is followed by a spacer sequence that encodes a guide RNA (gRNA) that directs the Cas protein to degrade a complementary DNA or RNA sequence that contains a protospacer adjacent motif (PAM). This PAM sequence allows Cas nucleases to differentiate the invading genetic threat from the CRISPR spacer locus. Unlike the restriction modification system, in which each enzyme recognizes a particular nucleic-acid sequence, the CRISPR-Cas system is highly modular in that nuclease activity can be directed to specific foreign genetic elements based on sequence-specific pairing with gRNAs provided in *trans*. Therefore, rather than having to express multiple restriction enzymes, the CRISPR-

Cas system can selectively target multiple sequences at once through the expression of a single nuclease along with various gRNAs. The CRISPR-Cas9 system introduces double-strand breaks into DNA that can ultimately result in the deletion, editing, or addition of genetic information when coupled with intracellular repair machinery.

CRISPR-based memory devices have the potential to store vast amounts of information in DNA with high scalability given that they can be readily reprogrammed at the nucleic-acid level. Kalhor et al. demonstrated this novel concept by using genome editing to track cell lineages in the zebrafish (*Danio rerio*), a technique they termed GESTALT (Genome Editing of Synthetic Target Arrays for Lineage Tracing) (Figure 3B) [21]. The authors arrayed 9 to 12 CRISPR target sites together, which functioned as the memory registry, downstream of a green fluorescent protein (GFP) transcript and co-expressed Cas9 with gRNAs complementary to those regions. The ribonucleoprotein complex randomly introduced double-strand breaks amongst the target sites, which were then repaired through the native non-homologous end joining (NHEJ) pathway to produce insertion-deletion mutations (indels) that varied in position and length. Over successive cellular divisions, these mutations irreversibly accumulated in a combinatorial fashion, generating compact DNA barcodes that were informative of cell lineage and could be read by DNA sequencing. The rate of indel formation and mutational patterns were qualitatively shown to be tunable based on the delivery method and concentration of the editing reagents. Using this technique, Kalhor et al. showed that the majority of cells in each organ of the zebrafish had derived from a small number of progenitor cells [21]. However, in this memory system, if multiple double-strand breaks occur prior to NHEJ repair, then inter-target deletions can occur (dropouts), resulting in loss of memory. Additionally, this memory device could only record for approximately 4 hours. Junker et al. expanded upon these studies to show that distributing multiple CRISPR target sites throughout the genome of zebrafish can prove to be an effective strategy at preventing memory loss for cell lineage tracking [22]. They termed their technique “scartrace,” referring to the genetic barcodes as genetic scars. In their studies, they targeted a single site within multiple GFP transgenes so that a loss in fluorescence would also be indicative of gene editing. The authors showed that not all barcode sequences are created equal and that some sequences are more likely to be generated than others. However, their CRISPR-based recorder was limited to recording for 10 hours.

Because multiple types of indel mutations and resulting sequences can occur within a CRISPR-targeted memory registry, each registry can encode for over a bit of information. Therefore, Schmidt et al. used *Caenorhabditis elegans*, whose cell lineage tree is known, as a model reference organism to develop an analytical mathematical model to quantify the storage capacity of CRISPR-based registries and to better predict the introduction of mutations [23]. They showed that cell lineage trees for the development of *C. elegans* can be precisely reproduced by using CRISPR recording with 86% efficiency (Figure 3B). Cell-lineage tracing using this system is imperfect due to sequence excision events between registry sites that result in loss of information. However, these CRISPR-based approaches have vastly increased the scalability and resolution for cell lineage tracking and cellular recording. Additionally, models and simulations described by Schmidt et al. showed that each CRISPR target site encodes 3.8 bits of information out of an upper limit of 6.6 bits

[23]. They attributed the inability to reach 6.6 bits storage due to non-uniform indel formation that is biased to the 3' end of the double-strand DNA cleavage site.

Self-targeting CRISPR barcodes

With the conventional CRISPR-Cas system, once a target site is edited and contains an indel, it can no longer record information because the gRNA used to target that site is no longer complementary to the newly mutated sequence. This limits the recording capacity to only a few hours and a single cleavage event. To overcome this problem, Perli et al. engineered the PAM sequence just downstream of the specificity-determining sequence (SDS) of the gRNA, such that the DNA locus encoding the gRNA could itself be targeted by the gRNA expressed from that locus, thus allowing it to serve as both a recorder and a memory registry [24]. They termed these engineered gRNA molecules as self-targeted guide RNAs (stgRNAs) (Figure 3C). This approach compacts the memory unit because the gRNA locus and the memory registry are combined into a single unit, enabling longer-term continuous recording at the stgRNA site. Longer recording is enabled because each mutated stgRNA-encoding locus then expresses a stgRNA that exactly matches the locus, thus allowing for repeated cleavage at the same locus. The authors showed that by increasing the length of the self-targeted region, the CRISPR recorders could undergo multiple rounds of mutagenesis for over 16 days, vastly improving the recording capacity. Moreover, the authors were able to record the duration and/or intensity of biological events by coupling stgRNA and/or Cas9 expression to desired biological events, thus expanding the scope of applications beyond cell lineage tracking. For example, the authors built tumor necrosis factor- α (TNF α)-inducible cellular memory devices, delivered these into human cells, and used them to record lipopolysaccharide (LPS)-induced acute inflammation events over time in mice (Figure 3B). This technique was termed mSCRIBE for Mammalian Synthetic Cellular Recorders Integrating Biological Events.

Kalhor et al. expanded upon this work by quantifying the improvement in coding capacity for self-targeting gRNAs compared to traditional gRNAs, which was shown to be an 8-fold improvement. This improvement could be implemented because these barcodes can be repeatedly targeted [25]. Additionally, Kalhor et al. showed that this technique can be coupled with fluorescence *in situ* sequencing (FISSEQ) to preserve spatial information while achieving barcode readout. To scale this approach will require improvements in high-throughput imaging capabilities and sample automation.

Summary and Perspective

Cellular memory is critical for encoding complex computing systems in living cells, as it enables cellular behaviors to be dictated not only by current events, but also by past events. In recent years, many new cellular computing and memory strategies that manipulate DNA in living cells, rather than relying solely on transcriptional activities, have been described. These advancements have dramatically improved the information storage capacity and the resolution at which we can store biological information. As DNA sequencing and synthesis technologies decrease in cost, these advancements should continue to scale. However, strategies based solely on DNA recording and computing have limitations. For example,

memory retrieval can be cumbersome if it relies on next-generation sequencing for readouts. Additionally, recording and computing are slow processes, limiting the sense-and-respond capabilities of engineered cells. Also, the unpredictable nature of mutations that accrue in living organisms can lead to memory corruption. New memory architectures as well as read and write strategies are needed to address these challenges.

Recent advances in memory circuits include recombinase-based recorders that have much denser storage capacities than traditional toggle switch-based memory devices. In addition, CRISPR-Cas-based recorders now enable vast numbers of barcodes to be generated over time in order to record cell lineage and important biological events. Based on the studies by Kalhor et al. that quantify the information that a single CRISPR site can encode, CRISPR-based recorders can produce $\sim(2^5)^n$ unique barcodes, where n is equal to the number of arrayed target sites [25]. Arraying ten sites together yields more than enough barcodes to label every cell in the human body ($\sim 3.7 \times 10^{13}$) [26]. However, a fundamental limitation in the current implementation of this technology is the loss of information that results from the excision of CRISPR sites. These dropouts are a byproduct of the NHEJ repair mechanism, which is error-prone. Additionally, it is difficult to control how each Cas9 cleavage event translates into a defined mutation. Though CRISPR-based recorders have increased the efficiency of information being encoded into DNA, future efforts should continue to push this efficiency further. For example, future ultra-dense recorders should be enabled by CRISPR-Cas base editors that produce predictable mutations without requiring double-strand breaks or NHEJ [27].

In addition to increasing recording capacity for cellular memory, progress towards complex computing has drastically improved. Using recombinase-based circuits, state machines can be programmed to efficiently achieve up to 16 unique states using three inputs. However, there still is a need to push the limit for increasing the number of states that cellular state machines can encode and to apply these state machines to solve complex biological questions. Envisioned applications include determining how the order of gene regulatory or signaling events dictates biological functions in areas such as development and cancer progression, to just name a few.

A future challenge for the field will be to test the evolutionary stability of these cellular computers and memory devices in real-world contexts. We envision that a particularly exciting application will involve combining these devices with other components, such as sensors and effectors, to create living theranostics that can dynamically sense, diagnose, and treat disease.

Acknowledgments

We thank Karen Pepper for reading over the document and providing input. The Lu lab acknowledges financial support by the National Science Foundation (MCB-1350625, #1521925), the National Institutes of Health (5-P50-GM098792-05), the Office of Naval Research (N00014-13-1-0424), the Army Research Office (W911NF-11-1-0281), the Broad Institute, the Koch Institute, the Defense Threat Reduction Agency (HDTRA1-14-1-0007), the Rainin Foundation (2016-3066), the Ellison Foundation (AG-NS-0948-12), the Wertheimer Fund, the Institute for Soldier Nanotechnologies (W9111NF-13-D-0001), J-WAFS, Novartis (14081955), the Desphande Center at MIT, the Singapore-MIT Alliance for Research and Technology, and the Center for Microbiome Informatics and Technology (15127713).

Abbreviations

ADC	analog-to-digital converter
att	attachment sites
BLADE	Boolean logic and arithmetic through DNA excision
Cas	CRISPR-associated nuclease
CRISPR	clustered regularly interspaced short palindromic repeats
DNA	deoxyribonucleic acid
FISSEQ	fluorescence <i>in situ</i> sequencing
GESTALT	<u>G</u> enome <u>E</u> editing of <u>S</u> ynthetic <u>T</u> arget <u>A</u> rrays for <u>L</u> ineage <u>T</u> racing
GFP	green fluorescent protein
gRNA	guide RNA
hgRNA	homing guide RNA
indels	insertion-deletion mutations
LPS	lipopolysaccharide
LSTP	large serine-type phage
mSCRIBE	mammalian synthetic cellular recorders integrating biological events
NHEJ	non-homologous end joining
PAM	protospacer adjacent motif
RAD	rewritable recombinase addressable data
RNA	ribonucleic acid
RSM	recombinase-based state machine
RT	reverse transcriptase
SCRIBE	synthetic cellular recorders integrating biological events
ssDNA	single stranded deoxyribonucleic acid
stgRNA	self-targeting guide RNA
TALEN	transcription activator-like effector nuclease
TNFα	tumor necrosis factor- α
ZNF	zinc finger nuclease

References and recommended reading

- of special interest
 - of outstanding interest
1. Keasling JD. Synthetic Biology for Synthetic Chemistry. *ACS Chem. Biol.* 2008; 3:64–76. [PubMed: 18205292]
 2. Brophy JAN, Voigt CA. Principles of genetic circuit design. *Nat. Methods.* 2014; 11:508–520. [PubMed: 24781324]
 3. Rodrigo G, Jaramillo A. AutoBioCAD: Full Biodesign Automation of Genetic Circuits. *ACS Synth. Biol.* 2013; 2:230–236. [PubMed: 23654253]
 4. Nielsen AAK, et al. Genetic circuit design automation. *Science.* 2016; 352:aac7341. [PubMed: 27034378]
 5. Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in *Escherichia coli*. *Nature.* 2000; 403:339–342. [PubMed: 10659857]
 6. Barrangou R, et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science.* 2007; 315:1709–1712. [PubMed: 17379808]
 7. Gibney ER, Nolan CM. Epigenetics and gene expression. *Heredity.* 2010; 105:4–13. [PubMed: 20461105]
 8. Grindley NDF, Whiteson KL, Rice PA. Mechanisms of Site-Specific Recombination. *Annu. Rev. Biochem.* 2006; 75:567–605. [PubMed: 16756503]
 9. Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ. Synthetic Gene Networks That Count. *Science.* 2009; 324:1199–1202. [PubMed: 19478183]
 10. Siuti P, Yazbek J, Lu TK. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* 2013; 31:448–452. [PubMed: 23396014]
 11. Bonnet J, Yin P, Ortiz ME, Subsoontorn P, Endy D. Amplifying Genetic Logic Gates. *Science.* 2013; 340:599–603. [PubMed: 23539178]
 - 12•• Weinberg BH, et al. Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat. Biotechnol.* 2017; 35:453–462. This paper describes a robust, general, and scalable system to build complex recombinase-based genetic circuits that function efficiently in mammalian cells. The authors showed that these circuits can perform logic and arithmetic and can be combined with CRISPR-Cas9 to regulate gene expression. [PubMed: 28346402]
 13. Müller M, et al. Designed cell consortia as fragrance-programmable analog-to-digital converters. *Nat. Chem. Biol.* 2017; 13:309–316. [PubMed: 28092361]
 14. Bonnet J, Subsoontorn P, Endy D. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci.* 2012; 109:8884–8889. [PubMed: 22615351]
 - 15• Fernandez-Rodriguez J, Yang L, Gorochofski T, Gordon D, Voigt C. Memory and Combinatorial Logic Based on DNA Inversions: Dynamics and Evolutionary Stability. *ACS Synthetic Biology.* 2015; 4:1361–1372. The authors built rewritable recombinase-based memory devices that performed AND and NOT logic. The authors measured the evolutionary stability of these circuits and showed that there were multiple causes for failure, thus emphasizing the need to determine failure modes and design counter-measures to build next-generation evolutionarily robust circuits. [PubMed: 26548807]
 16. Ham TS, Lee SK, Keasling JD, Arkin AP. Design and Construction of a Double Inversion Recombination Switch for Heritable Sequential Genetic Memory. *PLOS ONE.* 2008; 3:e2815. [PubMed: 18665232]
 - 17•• Roquet N, Soleimany AP, Ferris AC, Aaronson S, Lu TK. Synthetic recombinase-based state machines in living cells. *Science.* 2016; 353:aad8559. This paper describes a framework for implementing recombinase-based state machines that can perform temporal-based logic in living cells, in which the order and input type determine the cellular state. The authors built a two-input, five-state machine and a three-input, 16-state machine gene circuit. In addition, the authors showed it is possible to design multiple independent pairs of recombinase recognition sites that

are acted on by the same recombinase protein, thus scaling the recording capacity to 2^{kn} , which is greater than the 2^n scaling that is possible with standard recombinase-computing, in which only one set of recombinase recognition sites is used per recombinase (k is the number of independent pairs of recombinase sites per recombinase and n is equal to the number of orthogonal recombinases). [PubMed: 27463678]

18. Chaikind B, Bessen JL, Thompson DB, Hu JH, Liu DR. A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. *Nucleic Acids Res.* 2016; 44:9758–9770. [PubMed: 27515511]
19. Karpinski J, et al. Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity. *Nat. Biotechnol.* 2016; 34:401–409. [PubMed: 26900663]
- 20•. Yang L, Nielsen A, Fernandez-Rodriguez J, McClune C, Laub M, Lu T, Voigt C. Permanent genetic memory with >1-byte capacity. *Nat. Methods.* 2014; 11:1261–1266. The authors applied bioinformatics to identify 34 new phage integrases and their cognate *attB* and *attP* recognitions sites, 11 of which were orthogonal. In addition, they constructed a memory device by arraying together these 11 recognition sites to store 1.375 bytes of information in *E. coli*. Consequently, the recombinase toolkit has been greatly expanded. [PubMed: 25344638]
- 21••. McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J. Whole organism lineage tracing by combinatorial and cumulative genome editing. *Science.* 2016:aaf7907. This paper demonstrated the first CRISPR-based memory device, which was applied towards cell lineage tracing in zebrafish (*Danio rerio*). Specifically, the authors showed that CRISPR could be combined with the NHEJ repair pathway to produce thousands of lineage-informative barcodes in living organisms. [PubMed: 27229144]
22. Junker JP, et al. Massively parallel whole-organism lineage tracing using CRISPR/Cas9 induced genetic scars. *bioRxiv.* 2016:56499.
23. Schmidt ST, Zimmerman SM, Wang J, Kim SK, Quake SR. Cell lineage tracing using nuclease barcoding. *ArXiv160600786 Q-Bio.* 2016
- 24••. Perli SD, Cui CH, Lu TK. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science.* 2016; 353:aag0511. The authors engineered a PAM sequence into the DNA locus that encodes gRNA. This enabled the production of stgRNA that could repeatedly target its own locus to produce evolvable barcodes that can record molecular information for extended periods of time (>16 days), thus expanding the recording capacity for CRISPR-based memory devices. The authors coupled stgRNA and Cas expression to TNF α -responsive units to record LPS-induced acute inflammation in mice over time using engineered human cells. [PubMed: 27540006]
25. Kalhor R, Mali P, Church GM. Rapidly evolving homing CRISPR barcodes. *Nat. Methods.* 2017; 14:195–200. [PubMed: 27918539]
26. Bianconi E, et al. An estimation of the number of cells in the human body. *Ann. Hum. Biol.* 2013; 40:463–471. [PubMed: 23829164]
27. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature.* 2016; 533:420–424. [PubMed: 27096365]

Highlights

- Recombinase-based circuits execute all 16 2-input, 1-output Boolean logic functions
- BLADE is a general framework for building single-layer circuits in mammalian cells
- Recombinase-based machines can achieve scalable programming of cell states
- CRISPR memory devices trace cell lineages and record biological inputs in vivo

Figure 1. Synthetic gene circuits

Schematic illustrations showing designs for synthetic gene circuits that encode various logic functions and a memory device: **A)** a NOR gate, a genetic circuit that produces an output signal (GFP) only when neither input (chemical inducers) is present [2]; **B)** an AND gate, a genetic circuit that requires both inputs to produce an output signal [2]; **C)** a toggle switch, a genetic memory device that encodes two stable states via two mutually inhibitory transcriptional repressors. In the presence of neither input, 50% of the cells express GFP and 50% are repressed. Gene expression is turned OFF upon exposure to chemical inducer 1 and is turned ON upon exposure to chemical inducer 2. The gene is continually expressed, even in the absence of inducer 1, and is repressed only when cells are exposed to inducer 2 [5].

Figure 2. Recombinase-based gene circuits

A) A schematic illustrating the various DNA modifications that recombinase Bxb1 can perform. When the recognition sites are aligned (parallel), excision occurs; when they are anti-aligned (antiparallel), inversion occurs. **B)** Summary of the field programmable, read only memory (FPRM) genetic circuit designed using the BLADE architecture. The Boolean logic is a 6-input-one-output circuit that receives two data inputs, A and B, and is controlled by four select inputs (S_{1-4}) to produce a GFP output. All 16 Boolean logic functions can be carried out based on the combination of inputs [10]. **C)** Heat map illustrating the selectivity of the 11 newly mined integrases, in addition to FimE and HbiF, at targeting their cognate sites compared to other sites, thus enabling the encoding of 1.375 bytes of information [20]. **D)** A schematic showing an example of a state machine and the general mechanism of operation for recombinase-based state machines [17].

Figure 3. CRISPR-based memory devices

A) A schematic summarizing the general mechanism of CRISPR-Cas immunity. Upon viral infection (1), viral genetic signatures are incorporated into the CRISPR locus (2). After acquisition, gRNAs are produced and processed (3), and assembled into ribonucleoprotein complexes (4) that prevent future infections by selectively degrading viral genetic elements (5). **B)** A schematic illustrating how CRISPR barcoding can be used for cell lineage tracking. Multiple CRISPR sites are arrayed together, which function as a memory registry. When co-expressed with Cas9 and complementary gRNAs to those regions, double-strand breaks are randomly introduced; these are repaired through the native non-homologous end joining (NHEJ) pathway to produce indels that vary in position and length. Over successive cellular divisions, these mutations accumulate, generating compact DNA barcodes that are informative of cell lineage [21]. **C)** A schematic illustrating the general mechanism for the self-targeting CRISPR-Cas recorder system. The stgRNAs target the locus at which the stgRNAs are transcribed. Upon cleavage, error-prone repair occurs, resulting in a new locus that can produce another round of stgRNAs that are complementary to the mutated locus. This process can be reiterated for multiple cycles, providing a molecular record of cellular events for long periods of time [24].