Published in final edited form as:

Science. 2018 August 31; 361(6405): 870–875. doi:10.1126/science.aat9249.

Emerging applications for DNA writers and molecular recorders

Fahim Farzadfard, Timothy K. Lu

Synthetic Biology Group, Research Laboratory of Electronics, Department of Electrical Engineering and Computer Science and Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.

Abstract

Natural life is encoded by evolvable, DNA-based memory. Recent advances in dynamic genome-engineering technologies, which we collectively refer to as *in vivo* DNA writing, have opened new avenues for investigating and engineering biology. This Review surveys these technological advances, outlines their prospects and emerging applications, and discusses the features and current limitations of these technologies for building various genetic circuits for processing and recording information in living cells.

Genomic DNA is an ideal medium for artificial biological information storage because of its ubiquitous presence, durability, and compatibility with biological functions, especially as the throughput of DNA sequencing has substantially increased along with drops in cost (1). With the advent of genome-editing technologies, we can now dynamically change genetic information and harness the vast capacity of genomic DNA for information processing and storage in living cells. These dynamic *in vivo* DNA-writing technologies have opened new avenues for investigating and engineering biology, ranging from building molecular recorders and living biosensors for the longitudinal study of signaling dynamics in biological processes (2–6) to rationally designing genetic memory elements and computation operations in living cells (6–10) to tracing cellular lineages during development and differentiation (11–13). Here, we first review the applications, prospects, and potential uses of these technologies in various biological and biomedical settings. We then outline current *in vivo* DNA-writing technologies, summarize the memory architectures and features that each of these technologies offers, and discuss their current limitations.

DNA writers

DNA writers are genetically encoded devices that enable targeted, dynamic, and recurring modifications of DNA in living cells (2–7,14). These modifications can take the form of targeted insertions, deletions, inversions, or base substitution mutations and can serve as distinct DNA memory states (Fig. 1A). On the basis of the mutational outcomes, these devices can be broadly categorized into two classes: precise and pseudorandom writers. Precise DNA writers generate predetermined mutational outcomes, resulting in well-defined transitions between memory states in cell populations. Pseudorandom DNA writers generate

ffard@mit.edu (F.F.); timlu@mit.edu (T.K.L.).

Competing interests: F.F. and T.K.L. have filed patent applications on some works related to this Review.

targeted but stochastic mutational outcomes, resulting in unpredictable mutation signatures in cell populations. These two classes of DNA writers offer different levels of encoding capacity and control over memory states and operations, making them suitable for different sets of applications (summarized in Table 1).

DNA-writing applications

Molecular recording

Many molecular events that occur in biological systems are transient and thus difficult to monitor and study within their native context. DNA writing can be used to create molecular recorders that capture these transient signals and stably encode them into the DNA of cell populations or individual cells in vivo and in situ (Fig. 1B, left). The accumulated mutations can then be retrieved by DNA sequencing or functional assays to infer information about the original signals, even after the original signals are gone. This principle converts living cells into recording devices that memorize the history of their own signaling dynamics into permanent DNA records, which in turn can provide longitudinal insights into biological processes in their native contexts, as opposed to snapshots in time obtained by current approaches.

Several strategies, including conditional transcriptional or posttranscriptional activation of DNA writer components, can be used to couple the activity of a given DNA writer to signals of interest (Fig. 1B, bottom left). For example, by using signal-responsive promoters, information regarding the presence, duration, intensity, order, and timing of biological cues (such as metabolites and cytokines) or environmental cues (such as light, pollutants, exposure to phages, or changes in temperature) can be recorded in DNA (2–6). Naturally occurring signal-responsive promoters could be linked to DNA writer activity and used as a proxy to record and study the dynamics of the corresponding signaling pathways. If desired, rational design or directed evolution could be used to decouple natural promoters from unwanted overlapping pathways (e.g., by removing binding sites of corresponding transcription factors) or to engineer synthetic promoters with altered response dynamics (15–17). Alternatively, conditional activation of DNA writers in response to a desired signal can be achieved posttranscriptionally: for example, by implementing signal-dependent changes in conformation or interactions between DNA writer components.

Basic research—By offering an unprecedented ability to capture transient spatiotemporal molecular events in their native contexts, molecular-recording technologies could have broad utility across various disciplines (Fig. 1B, middle). For example, developmental biologists could use these DNA recorders to study the dynamics of differentiation cues and developmental pathways. Cancer biologists could use these recorders to study tumor development and to gain deeper insight into the cellular and environmental cues in tumor microenvironments that are involved in cancer heterogeneity. Immunologists could use these recorders to study signaling in immune cell maturation, memory formation, and immune responses. Microbiologists could use these recorders to study signaling dynamics and molecular interactions within bacterial communities and biofilms.

Various biological signals, ranging from small molecules to immunological cues to light, have been successfully recorded in both prokaryotic and eukaryotic cells (2–6). However, those recordings have been applied mainly to in vitro settings and relied on population-averaged read-outs (see Box 1). Future work is needed to improve these technologies for single-cell recording or to demonstrate the transformative use of molecular recorders in live animals, where the longitudinal study of in situ biology is currently limited. Memory architectures that impose minimal fitness effects will be important for realizing the use of molecular recorders in challenging in vivo conditions.

Living biosensors—Nonbiological sensors are not optimized to interact with biological systems. Living cells, on the other hand, are useful chassis for hosting sensors that can respond to various biological cues. DNA-writing technologies can be used to create living biosensors for longitudinal health and environmental monitoring. For example, bacterial cells endowed with disease biomarker sensors coupled with DNA recorders could be consumed orally, transit through the gastrointestinal tract to record disease biomarkers, and report this information later when they exit the body (Fig. 1B, top right). Engineered human cells harboring molecular recorders could be deployed into the body to report on early signs of disease, such as cancer or neurodegeneration. Finally, engineered cells and animals equipped with recording capacities could be used to continuously monitor and record the levels and activities of biological and environmental cues (such as toxins, heavy metals, metabolites, and light) without requiring artificial power supplies and in conditions and places that are not readily accessible to nonbiological sensors. Similar to basic research purposes, biosensing applications will need memory architectures with minimized fitness effects and extended recording capacities to achieve continuous and robust recording.

Brain mapping—Mapping the activities and connectome of neural circuits in the brain is one of the greatest challenges of our time (Fig. 1B, bottom right). As an alternative to current imaging-based techniques, which suffer from trade-offs between resolution and throughput, DNA-based ticker tape circuits that allow for the dynamic logging of signals have been proposed for recording spatiotemporal neural activities (18). Although existing molecular-recording technologies offer temporal resolutions that are orders of magnitude longer than neural pulses (Box 1), they could potentially be used to study time-averaged neural activities. For example, neural activities can be linked to molecular recorders via neural activity-responsive regulatory elements, such as immediate early gene promoters (19). Live animals harboring these genetic recorders could then be subjected to different neural stimuli, and the resulting mutational signatures could be used to infer time-averaged activities across the entire animal brain. Alternatively, DNA writers encoded on mobilizable genetic elements that can pass through synapses, such as rabies or pseudorabies viruses, could be used to distinctively mark neural connections by DNA barcodes that could then be used to map the connectome in a high-resolution and high-throughput fashion (20). Despite many technical challenges, we envision that applying molecular-recording technologies to decipher the functional architecture of the brain will be a strong driving force for the advancement of these technologies, especially in terms of scalability, recording capacity, and temporal and spatial resolution.

Evolutionary cellular engineering

Continuous in vivo evolution—In vivo DNA-writing technologies could be used to recurrently mutate desired genomic segments and achieve targeted genetic diversification within a short period. Once coupled with continuous selection, this strategy could enable continuous rounds of evolution to improve cellular traits of interest or to quickly evolve protein and RNA scaffolds for biotechnological and therapeutic applications (Fig. 1C, left). Unlike in molecular-recording applications, where it is desirable to minimize fitness effects to achieve robust recording, in evolutionary engineering applications, a selective pressure is applied to direct evolutionary trajectories toward desired outcomes. DNA-writing technologies with relaxed (21) or obviated (2) requirements for cis-encoded elements and extended mutational spectra (22) could be especially useful for these applications.

Synthetic Lamarckian evolution—Living cells have evolved mechanisms to elevate their local mutation rate under certain conditions and in response to specific signals. For example, during antibody maturation, CRISPR-Cas9 spacer acquisition, and mutagenesis processes mediated by diversity-generating retroelements in phages and bacteria, a series of actively regulated molecular events lead to targeted mutagenesis in certain genomic loci. These Lamarckian evolutionary strategies can increase the overall fitness of cell populations in uncertain environments and help them to adapt to environmental changes at greater rates than are possible by random Darwinian mutations. DNA-writing technologies could be used to emulate Lamarckian evolution by increasing the local mutation rates of desired genetic loci in response to signals of interest and in the presence of suitable selective pressures (Fig. 1C, right). Cells engineered with such a capacity could evolve faster than possible by natural evolution and enable adaptive cell-based therapeutics that tune their responses to the conditions they encounter. Alternatively, engineered bacteriophages endowed with the capacity to target and mutagenize their own host-range determinants could be useful for the streamlined development of phage-based antimicrobials that could adapt to infect new hosts faster than natural phages.

Applications specific to precise DNA writers

Layered molecular-recording, computation, and artificial-learning gene circuits—The precise and well-defined nature of the mutational outcomes generated by precise DNA writers allows them to be layered into more sophisticated genetic circuits in which the mutational outcome of one element can be used as inputs for other elements. By doing so, information regarding a series of input signals can be recorded in the form of well-defined transitions between multiple memory states. This strategy has been used to encode various forms of combinatorial, sequential, and temporal logic and other increasingly complex computing operations in living cells (Fig. 1D) (6–10). Additionally, because of the predictable nature of precise writers, their mutational output can be linked to functional genetic elements and used to control gene expression. These rationally designed genetic programs could be used, for example, to study or control the sequence and timing of developmental programs or to build gene circuits that classify disease conditions on the basis of multiple inputs. In addition, genetic programs could be created to endow cells with artificial learning capabilities such that specific circuit responses are gradually reinforced (or

degraded) in response to signals (6), much like the reinforcement of synaptic interconnections in neurons.

High-throughput interaction and activity mapping—Transient cellular events, such as protein-protein interactions, can be converted into transcriptional outputs and therefore captured into DNA memory. For example, a split DNA-writing system, where the N- and C-terminal domains of a precise DNA writer are fused to barcoded bait and prey, respectively, could be used to record protein-protein interactions (Fig. 1E, left). In cells harboring interacting partners, a functional DNA writer could be reconstituted and write a prey-specific barcode next to a bait-specific barcode. The joined barcode could then be retrieved by sequencing to identify interacting partners in pooled libraries in a high-throughput fashion. Analogous strategies could be used to study the activities of RNA and protein variant libraries in a high-throughput fashion (Fig. 1E, right).

Application specific to pseudorandom DNA writers: lineage tracing

Capturing cellular ancestry relationships during development and creating corresponding lineage maps, especially in larger animals, have been a long-standing challenge in developmental biology. Traditionally, various static genetic and non-genetic barcoding approaches have been used for lineage tracing (23–25). In these methods, once a cell receives a barcode, it passes the barcode to its progenies with no change. Therefore, lineages that are generated in later stages are not differentially barcoded and, as a result, only a low-resolution lineage tree can be constructed. DNA writers can be used to devise dynamic genetic barcoding schemes that continuously and distinctively mark cell lineages as they progress in vivo, thus enabling higher-resolution lineage maps (Fig. 1F). Lineage tracing can be considered a specific application of molecular recording, where, instead of a transient signal, the chronicle of transient events (e.g., cell divisions) is recorded in DNA and later retrieved by sequencing. Pseudorandom writers are especially useful for lineage-tracing applications because they can generate many distinct mutational signatures in an initially clonal population.

Precise DNA writers

Three classes of precise DNA writers have been described to date (Table 1), each featuring a different DNA-writing efficiency and thus a different recording regime (Box 1). Site-specific recombinases are the most efficient and well-established class of precise DNA writers. Depending on the orientation of their DNA recognition sites, these enzymes can either flip or excise a piece of DNA that lies between their cognate sites, thus memorizing the history of exposure to a signal in the form of defined and permanent DNA reconfiguration (Fig. 2A, transition from S0 to S1). Because of their relatively high efficiency, these DNA writers have been used mainly in digital recording (Box 1 and Fig. 2D) and building layered synthetic gene circuits for digital computation (7–9).

The second class of precise DNA writers relies on reverse transcriptase (Remediated in vivo single-stranded DNA (ssDNA) expression followed by recombineering to achieve cis element-independent DNA writing in bacteria (Fig. 2B) (2). The moderate writing efficiency of this system offers wider-dynamic-range molecular recording in which the analog

properties of biological signals, such as signal intensity and exposure duration, are recorded into the overall genomic DNA of cell populations (Box 1 and Fig. 2D). Because these DNA writers do not require cis-encoded elements on the target, they are desirable for evolutionary engineering applications.

The third class of precise DNA writers performs nucleotide-resolution manipulation of DNA via base editing (26). In this system, a base editor, such as a cytidine deaminase domain fused to dead Cas9 (dCas9), is addressed to a desired target site by expression of a complementary guide RNA (gRNA), generating deoxycytidine (dC)-to-deoxythymidine (dT) mutations within a narrow window in the target vicinity (Fig. 2C). As these memory operators are CRISPR-Cas9 based, they are more scalable than other precise writers. Additionally, they can be functionalized with regulatory modules (such as CRISPR interference and activation) to achieve complex recording and computation operations in living cells (6). Recently, an adenosine deaminase base editor, which writes deoxyadenosine (dA)-to-deoxyguanosine (dG) mutations, was developed (22), further expanding the mutation spectrum and utility of this class of DNA writers and paving the way toward bidirectional DNA-writing systems that could be used for advanced computation and evolutionary engineering applications.

Pseudorandom DNA writers

Two main classes of pseudorandom DNA writers have been described to date. The first class relies on targeted double-stranded DNA (dsDNA) breaks generated by site-specific nucleases, such as CRISPR-Cas9, followed by error-prone repair of the breaks by the nonhomologous end-joining (NHEJ) pathway (11). During this process, each individual cell can acquire a pseudorandom mutational signature (i.e., indel mutations) in the target locus. Several studies have used these mutational signatures as barcodes to trace cellular lineages during embryo development in zebrafish and other small animals or in situ in cell cultures (Fig. 3A) (11–13). Efforts to extend the write cycles (see Box 1) of these molecular recorders led to the development of evolving barcodes (4,14). This memory architecture was built by engineering a protospacer adjacent motif (PAM) into the gRNA-encoding locus, resulting in self-targeting gRNAs (stgRNAs) that undergo iterative barcoding cycles during which the stgRNA locus is repeatedly diversified (Fig. 3B). This memory architecture was leveraged to build a population-level analog recorder (4) and to dynamically barcode cellular lineages in mammalian cells (14).

Despite these improvements, the reliance of these memory architectures on dsDNA breaks and NHEJ still limits their write cycles and makes them unsuitable for usage in organisms that lack an efficient NHEJ pathway, such as most prokaryotes. The prevalence of deletions in NHEJ can result in shortening of the stgRNA and loss of the PAM, thereby rendering the recorder nonfunctional over time. Moreover, the stochastic and deletion-based nature of the mutations generated by these strategies can result in nonpersistent encoding, where new memory states overwrite previous ones, thus making it difficult to infer ancestral relationships. Furthermore, encoding multiple stgRNAs in the same cell could result in unwanted chromosomal rearrangements and cellular toxicity. To extend the use of DNA writers for high-resolution lineage-tracing applications, particularly in larger animals, new

memory architectures with improved efficiency, extended write cycles, reduced toxicity, and persistent (i.e., non-deletion-based) barcoding are desired. Alternative strategies, for example, using C- or A-rich stgRNAs in combination with base editors, could be devised to combine the high storage capacity of pseudorandom memory architectures with the well-defined and persistent memory states offered by precise DNA writers to address some of the above-mentioned limitations.

The second class of pseudorandom DNA writers was built upon Cas1 and Cas2 proteins, which naturally mediate spacer acquisition in the CRISPR bacterial immune system. In this system, the Cas1-Cas2 complex samples the intracellular ssDNA pool, which can originate from various intracellular or extra-cellular sources, and integrates short (~20 to 30 base pairs) ssDNA fragments from this pool into a preexisting CRISPR array, resulting in extension of the array over time (Fig. 3C). New spacers are added to the leader-proximal site, so the chronological order of spacer addition events is preserved within the array configuration. By placing the expression of the Cas1-Cas2 cassette under the control of signal-inducible promoters and introducing exogenous oligonucleotides into *Escherichia coli* cell populations, Shipman *et al.* (27) demonstrated that the signal intensity and duration can be inferred from array extensions and that the temporal order of the addition of the oligonucleotide pools can be inferred from the array composition. In a follow-up study, the authors demonstrated that artificial digital information, such as small pictures and movies, could be encoded into oligonucleotide pools and recorded into the distributed genomic DNA of a cell population (28). Building on these results, Sheth *et al.* (3) showed that instead of providing exogenous oligonucleotides, the intracellular ssDNA pool could be dynamically modulated by using template plasmids with tunable copy numbers. Using this strategy and multiple template plasmids, the authors demonstrated that the temporal order of multiple signals and lineage information of bacterial populations could be recorded in the CRISPR array composition (Fig. 3D). Though the Cas1-Cas2 writing system offers a relatively persistent memory architecture with desirable features, such as sequentially and temporally resolved recording and extended write cycles, it is currently limited to bacteria. The system could offer an attractive strategy for lineage tracing if it could be adapted to eukaryotes and function stably over multiple generations.

Conclusion and future prospects

In the past few years, we have witnessed the transition from the read-only genomic era to the read-and-write era. DNA-writing technologies have transformed genomic DNA into a dynamic medium for processing and storing biological and artificial information in living cells. These advances herald a new generation of powerful approaches for investigating and engineering in situ biology in basic research, biotechnology, and medicine. Although substantial progress has been made, there is plenty of room for improving existing memory architectures or developing new ones with desirable features, especially in terms of recording capacity, scalability, robustness, fitness effects, cellular resource consumption, write cycles, temporal resolution, and recording kinetics. These technologies promise to further advance our ability to manipulate life's natural memory storage media in a dynamic, longitudinal, and multiplexed fashion.

ACKNOWLEDGMENTS

Funding: This work was supported by the National Institutes of Health (P50 GM098792), the Office of Naval Research (N00014-13-1-0424), the NSF (MCB-1350625), the Defense Advanced Research Projects Agency, and NSF Expeditions in Computing program award 1522074. F.F. thanks the Schmidt Science Fellows Program, in partnership with the Rhodes Trust, for support.

REFERENCES AND NOTES

1. Church GM, Gao Y, Kosuri S, Science 337,1628–1628 (2012). [PubMed: 22903519]
2. Farzadfard F, Lu TK, Science 346, 1256272 (2014).
3. Sheth RU, Yim SS, Wu FL, Wang HH, Science 358, 1457–1461 (2017). [PubMed: 29170279]
4. Perli SD, Cui CH, Lu TK, Science 353, aag0511 (2016).
5. Tang W, Liu DR, Science 360, eaap8992 (2018).
6. Farzadfard F. et al., bioRxiv 263657 [Preprint]. 16 2 2018.
7. Roquet N, Soleimany AP, Ferris AC, Aaronson S, Lu TK, Science 353, aad8559 (2016).
8. Siuti P, Yazbek J, Lu TK, Nat. Biotechnol 31,448–452 (2013). [PubMed: 23396014]
9. Bonnet J, Yin P, Ortiz ME, Subsoontorn P, Endy D, Science 340, 599–603 (2013). [PubMed: 23539178]
10. Weinberg BH et al., Nat Biotechnol 35, 453–462 (2017). [PubMed: 28346402]
11. McKenna A. et al., Science 353, aaf7907 (2016).
12. Schmidt ST, Zimmerman SM, Wang J, Kim SK, Quake SR, ACS Synth. Biol 6, 936–942 (2017). [PubMed: 28264564]
13. Frieda KL et al., Nature 541, 107–111 (2017). [PubMed: 27869821]
14. Kalhor R, Mali P, Church GM, Nat. Methods 14,195–200 (2017). [PubMed: 27918539]
15. Farzadfard F, Perli SD, Lu TK, ACS Synth. Biol 2,604–613 (2013). [PubMed: 23977949]
16. Meyer AJ, Segall-Shapiro TH, Voigt CA, bioRxiv 285866 [Preprint]. 10 4 2018.
17. Chen Y. et al., Nat. Commun 9, 64 (2018). [PubMed: 29302024]
18. Kording KP, PLOS Comput. Biol 7, e1002291 (2011).
19. Kawashima T, Okuno H, Bito H, Front. Neural Circuits 8,37 (2014). [PubMed: 24795570]
20. Zador AM et al., PLOS Biol 10, e1001411 (2012).
21. Hu JH et al., Nature 556, 57–63 (2018). [PubMed: 29512652]
22. Gaudelli NM et al., Nature 551, 464–471 (2017). [PubMed: 29160308]
23. Kretschmar K, Watt FM, Cell 148, 33–45 (2012). [PubMed: 22265400]
24. Sulston JE, Schierenberg E, White JG, Thomson JN, Dev. Biol 100, 64–119 (1983). [PubMed: 6684600]
25. Livet J. et al., Nature 450, 56–62 (2007). [PubMed: 17972876]
26. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR, Nature 533, 420–424 (2016). [PubMed: 27096365]
27. Shipman SL, Nivala J, Macklis JD, Church GM, Science 353, aaf1175 (2016).
28. Shipman SL, Nivala J, Macklis JD, Church GM, Nature 547, 345–349 (2017). [PubMed: 28700573]

Box 1.**DNA memory features.****Population-distributed versus single-cell recording**

Because of the probabilistic nature of DNA writing at the single-molecule level, a statistically significant number of recording substrates (i.e., DNA molecules) are required to achieve robust recording. All the molecular recorders described so far have utilized the distributed genomic DNA of cell subpopulations to achieve robust recording. Developing efficient writers and/or using these together with high-copy-number recording substrates could pave the way toward single-cell recording.

Write cycles

We define write cycles as the number of iterations in which new information can be added to a memory register encoded on a single molecule of DNA by a single DNA writer or recorder complex before the memory register becomes nonresponsive to that complex. With the use of base editing (6), stgRNA (4,14), and Cas1-Cas2 (3,28) technologies, memory architectures with write cycles of >1 have been demonstrated.

Recording capacity

We define recording capacity as the number of distinct memory states that can be recorded (and practically retrieved) in the entire storage unit (the cell population for population-level recording or an individual cell for single-cell recording) by using a single DNA writer or recorder complex.

Digital versus analog recording

Depending on the writer efficiency and the potential memory states in the population (the number of memory states in each cell x the number of cells), two signal-recording regimes can be defined (Fig. 2D). Digital recording (a sharp, saturating increase in the mutation frequency in the population in response to an input) can be achieved when highly efficient DNA writers are used or when the number of potential memory states is limited. Analog recording (a gradual accumulation of mutations in the population in response to an input) can be implemented when moderately efficient writers are used or when there are many potential memory states. This extended dynamic range enables one to infer information regarding signal intensity and duration, which are analog properties, as opposed to the absence or presence of a signal, which is digital information.

Sequential and temporal resolution

Analog recorders integrate a signal over time but do not necessarily preserve information about the relative order or timing of multiple signals or the recurrence of a signal. Memory architectures with the capacity to record sequential and temporal information have been developed by using site-specific recombinases (7), base editing (6), and Cas1-Cas2 (3), although their resolution, write cycles, and recording capacity still need to be improved for demanding applications. Notably, ticker tape memory architectures (18) that record signaling dynamics in a temporally resolved fashion could enable one to infer signal intensity as a function of time.

“With the advent of genome-editing technologies, we can now dynamically change genetic information and harness the vast capacity of genomic DNA for information processing and storage in living cells.”

“...there is plenty of room for improving existing memory architectures or developing new ones with desirable features, especially in terms of recording capacity, scalability, robustness, fitness effects, cellular resource consumption, write cycles, temporal resolution, and recording kinetics.”

Fig. 1. DNA-writing technologies and their emerging applications.

(A) A schematic representation of a DNA writer (left) and mutation signatures generated by precise (middle) and pseudorandom (right) DNA writers. S0 and S1 indicate unmodified (memory state 0) and mutated (memory state 1) alleles, respectively. B1 to B5 indicate random memory states 1 to 5 that are generated by pseudorandom DNA writers and could serve as distinct barcodes. (B) Schematic representation of a molecular recorder and strategies that can be used to couple its activity to signals of interest (left), along with examples of applications in basic research (middle) and biotechnology (right). (C) Examples

of evolutionary cellular engineering applications enabled by DNA writers. **(D)** Various forms of logic and computation can be achieved by layering multiple precise recorders. **(E)** Examples of strategies for high-throughput mapping of interactions or activities of variant libraries by DNA writers. **(F)** Pseudorandom DNA writers can be used to develop dynamic in vivo genetic barcoding schemes that distinctively and progressively mark cellular lineages over time.

Fig. 2. Precise DNA-writing technologies.

(A) Site-specific recombinases. (B) Recombineering. (C) Base editing. CDA, cytidine deaminase; d/nCas9, dCas9 or nickase Cas9. (D) Digital versus analog recording (see Box 1).

Fig. 3. Pseudorandom DNA-writing technologies.

(A) A schematic representation of the Cas9 nuclease-based DNA-writing system used for lineage tracing in zebrafish in (11). S[R1], S[R2], and S[Rn] represent random memory states 1, 2, ..., n generated by the pseudorandom DNA writers. These random memory states individually, or in combination with other random memory states, can serve as distinct barcodes (e.g., B1 = S[R1] and B2 = S[R2][R1]). (B) A schematic illustration of the Cas9 + stgRNA recording system. (C) DNA writing by the Cas1-Cas2 spacer acquisition system. (D) The strategy used in (3) to record temporal information into a CRISPR array.

Table 1.
Features and demonstrated applications for the current DNA-writing technologies.

TBD, to be determined; RSM, recombinase-based state machine; BLADE, Boolean logic and arithmetic through DNA excision; SCRIBE, synthetic cellular recorders integrating biological events; CAMERA, CRISPR-mediated analog multi-event recording apparatus; DOMINO, DNA-based ordered memory and iteration network operator; GESTALT, genome editing of synthetic target arrays for lineage tracing; MEMOIR, memory by engineered mutagenesis with optical in situ readout; mSCRIBE, mammalian synthetic cellular recorders integrating biological events; TRACE, temporal recording in arrays by CRISPR expansion.

DNA-writing technology	Mutation type	Used in	Host-specific requirement	Example platform(s)	Application(s)
<i>Precise writing mode</i>					
Site-specific recombinases	Inversion and deletion	Bacteria and eukaryotes	–	RSM (7), BLADE (10)	Digital recording and computation
Recombineering	Small insertion, deletion, and base substitution	Bacteria	ssDNA-mediated recombineering	SCRIBE (2)	Analog recording
Base editing	dC to dT or dA to dG	Bacteria and eukaryotes	–	CAMERA (5), DOMINO (6)	Analog and digital recording and computation
<i>Pseudorandom writing mode</i>					
Cas9 nuclease + gRNA	Indel	Eukaryotes	NHEJ	GESTALT (11), MEMOIR (13)	Lineage tracing
Cas9 nuclease + stgRNA	Indel	Eukaryotes	NHEJ	mSCRIBE (4,14)	Analog recording, lineage tracing
Cas1-Cas2 (spacer acquisition)	Insertion	Bacteria	TBD	TRACE (3, 27, 28)	Analog recording, lineage tracing, digital information storage