


Recording Biological Information with CRISPR–Cas Systems

Book Chapter**Author(s):**

Cherepkova, Mariia Y.; Tanna, Tanmay; [Platt, Randall](#) 

Publication date:

2022

Permanent link:

<https://doi.org/https://doi.org/10.3929/ethz-b-000584226>

Rights / license:

[In Copyright - Non-Commercial Use Permitted](#)

Originally published in:

<https://doi.org/10.1002/9781683673798.ch16>

Recording Biological Information with CRISPR-Cas Systems

Mariia Y. Cherepkova¹, Tanmay Tanna¹, and Randall J. Platt^{1,2}

¹Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland

²Department of Chemistry, University of Basel, 4003, Basel, Switzerland

Introduction

Biological systems—in all their diverse forms, ranging from macromolecules to multicellular organisms—are highly complex and dynamic. For instance, biological states for single cells are determined by an interplay of genetic and epigenetic regulation, transcriptional, translational, and posttranslational events, sensing mechanisms for external stimuli, signaling cascades, and microenvironments within the cell and cell niches. With such complexity, a stable yet evolvable medium for encoding, transmitting, and utilizing biological information is essential, and DNA has emerged as the evolutionarily preferred solution. Recent advances in genome editing, coupled with substantial reductions in DNA sequencing and synthesis costs, have opened new opportunities for harnessing this potential of DNA as an information storage system for biological applications (1).

One of the main challenges of biology is to understand how molecular events give rise to complex cellular behaviors and states and guide the development of multicellular systems. Most current technologies for studying these processes, however, are disruptive, therefore limiting observations to a few snapshots in time and resulting in a readout lacking in information that associates one state with the following. One potential solution to this challenge is to convert cells into biological recorders that continuously capture information about dynamic cellular processes and stably encode it in DNA (2, 3). This memorized information can then be read out from populations of cells to provide longitudinal insights into cellular events, such as the abundance of biomolecules over time or cellular lineage of multicellular systems such as organisms, organoids, or tumors.

A major step taken in this direction was the development of DNA writing, a process by which biological or artificial information is encoded through modifications in DNA using natural or engineered enzymes termed DNA writers. DNA modifications, in the form of mutations such as insertions, deletions, inversions, or base substitutions, serve as memory states corresponding to input signals (2, 4). The first DNA writers relied on synthetic gene circuits comprised of toggle switches (5) or site-specific DNA recombinases (6). While these advances established the principles of DNA writing in living cells, these systems lacked scalability and orthogonality, precluding their widespread use for interrogating biological systems.

CRISPR: Biology and Applications, First Edition. Edited by Rodolphe Barrangou, Erik J. Sontheimer, and Luciano A. Marraffini.

© 2022 American Society for Microbiology. DOI: 10.1128/9781683673798.ch16

Due to recent advancements in genome editing, fueled by the discovery and exploration of the microbial adaptive immune system CRISPR-Cas, a new generation of DNA writing tools was created and is enabling a diversity of novel applications for investigating and engineering biology in basic research, biotechnology, and medicine (3). In this chapter, we review the development and existing applications of CRISPR-Cas systems for DNA writing and discuss features and current limitations of these technologies for recording biological information in living cells.

DNA Writing with CRISPR-Cas Systems

The natural diversity of CRISPR-Cas systems provides us with a plethora of opportunities for developing DNA writing devices with diverse mutational signatures and response dynamics. The CRISPR-Cas immune response consists of three stages: adaptation, expression/processing, and interference. During the adaptation stage, a CRISPR acquisition complex comprised of Cas1 and Cas2 proteins recognizes foreign genetic elements and integrates short sequence fragments from these nucleic acids as new “spacers” into the host CRISPR array (7). In the subsequent stages, the acquired spacers are transcribed and processed into small CRISPR RNAs, which can guide effector Cas proteins to the complementary sequence in hostile genetic elements during future invasions, triggering cleavage, degradation, or mutation of the target sequence. Thus, during the natural CRISPR-Cas immune response, both the adaptation and interference stages involve modification of DNA, which can be leveraged for DNA writing.

The molecular machinery of the adaptation module from several microorganisms has been harnessed for DNA writing via spacer acquisition from DNA or RNA (Fig. 16.1) (8–10). The information encoded in the form of spacers can subsequently be retrieved by sequencing CRISPR arrays, either genomic or plasmid borne. Since CRISPR arrays can store multiple spacers, such systems enable multiple rounds of writing at a unique locus and can therefore provide greater information storage capacity. Additionally, new spacers are almost always inserted between a leader sequence and the first direct repeat sequence of the CRISPR array, which allows for resolution of the order of molecular events during recording (7, 11).

DNA writing by one of the most widely used Cas effector proteins, Cas9 from *Streptococcus pyogenes*, has provided a flexible platform for engineering DNA writing tools. The RNA-guided DNA endonuclease Cas9 can be utilized to generate random mutational outcomes at a specific target locus (Fig. 16.1). It recognizes target loci via short RNAs (single guide RNAs [sgRNAs]) and generates double-strand breaks (DSBs). In eukaryotic cells, these DSBs are then repaired through either nonhomologous end joining (NHEJ), which results in the formation of insertions and deletions (together called “indels”), or, less often, through homology-directed repair. Cas9 target sites must contain the recognition protospacer sequence (complementary to sgRNA) and a protospacer adjacent motif (PAM) (12, 13). Thus, a DNA writing module based on Cas9 should include a target locus with the PAM and enable the expression of sgRNA and Cas9.

DNA writers based on a wild-type Cas9 protein are efficient but can generate only a limited number of mutational outcomes and are thus characterized by a short time of writing owing to saturation of target loci, requiring multiple target loci to achieve a high recording capacity. To overcome these limitations, alternative Cas9-based DNA writing modules utilize self-targeting guide RNAs (stgRNAs), also known as homing guide RNAs (hgRNAs) (14, 15) (Fig. 16.1). In these systems, the stgRNA scaffold sequence, which is necessary for Cas binding, carries a GUU-to-GGG mutation immediately downstream of

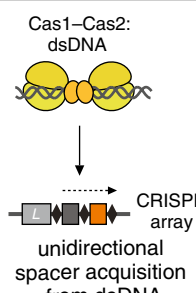
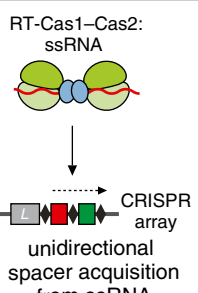
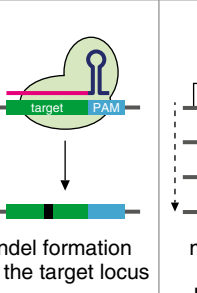
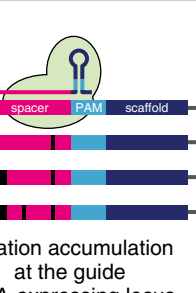
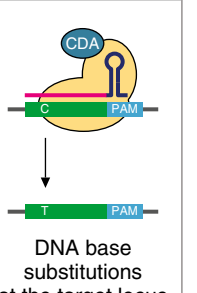
	spacer acquisition from DNA	spacer acquisition from RNA	wild-type Cas9	Cas9:hgRNA/stgRNA	nCas9/dCas9-CDA fusion
DNA writer	 <p>Cas1–Cas2: dsDNA</p> <p>CRISPR array</p> <p>unidirectional spacer acquisition from dsDNA</p>	 <p>RT-Cas1–Cas2: ssRNA</p> <p>CRISPR array</p> <p>unidirectional spacer acquisition from ssRNA</p>	 <p>target PAM</p> <p>indel formation at the target locus</p>	 <p>spacer PAM scaffold</p> <p>mutation accumulation at the guide RNA-expressing locus</p>	 <p>CDA PAM</p> <p>DNA base substitutions at the target locus</p>
method	TRACE, image recording	Record-seq	GESTALT, scGESTALT, MEMOIR	mSCRIBE, homing CRISPR	CAMERA, lineage tracing

Figure 16.1 DNA writing by CRISPR-Cas systems. Diversity of CRISPR-Cas-based DNA writers, their mode of action, and applications in molecular recording and lineage tracing.

the spacer sequence, so that it matches the requisite NGG PAM sequence of Cas9. These synthetic stgRNAs direct Cas9 to target the DNA locus of the stgRNA itself, enabling re-targeting and evolvability of the target locus and thus providing a higher diversity of mutations and extending the period of DNA writing.

Cas9-based DNA writers that generate mutational outcomes using DSBs are functional only in eukaryotic cells, since NHEJ repair pathways are rare in prokaryotes, and DSBs frequently cause cell death or a loss of extrachromosomal DNA (16).

By engineering Cas9 through the addition of different functional domains, the molecular arsenal of potential genetic modifications has been expanded. For instance, a “base editing” approach uses catalytically disabled (dCas9) or nicking Cas (nCas9) nucleases fused to nucleobase deaminases to insert mutations into DNA at single-base resolution without inducing DSBs, thus being functional in both eukaryotes and prokaryotes. As an example, a fusion of nCas9 to a cytidine deaminase domain (CDA) (Fig. 16.1) converts C:G base pairs to T:A base pairs in a 4- to 8-nucleotide window from the PAM site on the distal side of protospacer sequence (17).

In contrast to other DNA writers (such as site-specific DNA recombinases) that generally allow for only a limited number of predefined mutational outcomes, CRISPR-Cas systems provide a resource for the development of a new class of efficient, scalable, and easily programmable DNA writers. Furthermore, DNA writers that utilize CRISPR-Cas base editing or stgRNAs or leverage CRISPR spacer acquisition allow for multiple cycles of writing (4) onto a single DNA substrate, thus increasing recording capacity. The expanding toolbox of novel and engineered Cas proteins has found wide-ranging DNA writing applications, and here, we focus on two main applications: lineage tracing and molecular recording.

Lineage Tracing Using CRISPR-Cas9

In multicellular organisms, a single totipotent zygote gives rise to all cell types. This process is orchestrated by a plethora of cellular signals and involves multiple cell division and differentiation events that lead the zygote through cellular lineages resulting in different cell types. Lineage tracing methods aim to establish relationships between cell types by following the progeny of a single cell or groups of cells (Fig. 16.2A). Before the development of

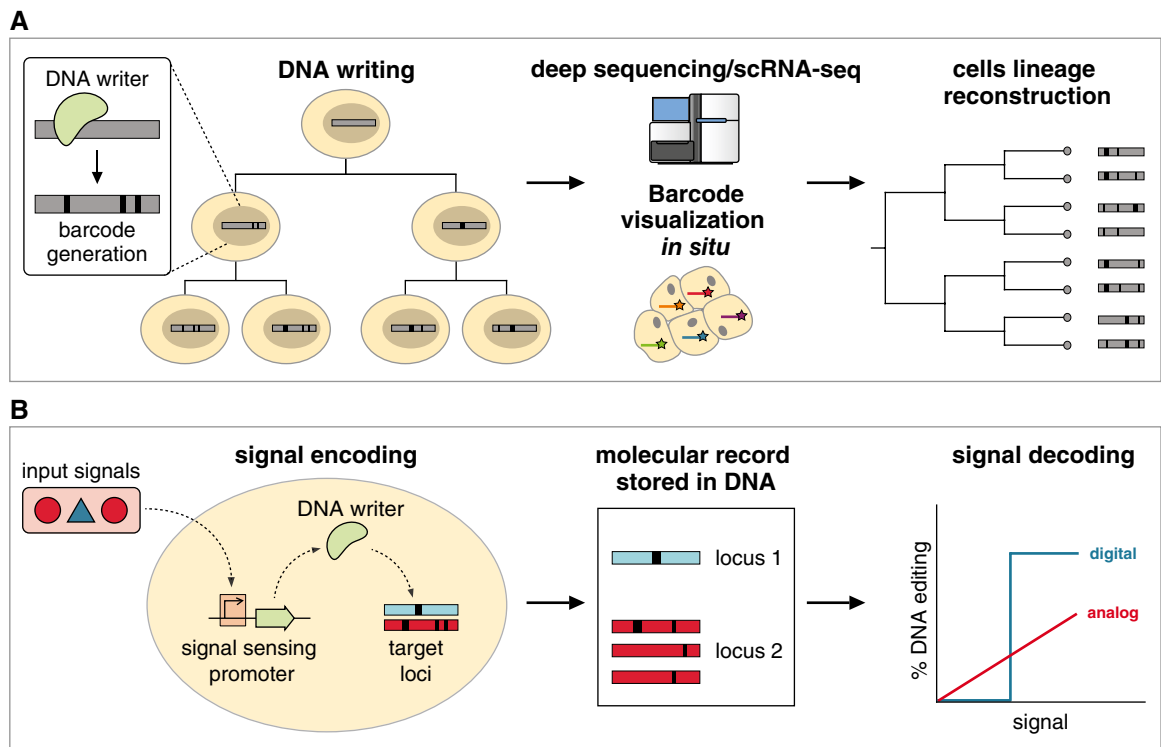


Figure 16.2 Applications of DNA writing. (A) Lineage tracing aims to establish relationships between cell types by following the progeny of a single cell or a group of cells. By introducing mutations at target loci, DNA writers generate barcodes over the course of cell division and differentiation. These barcodes can be subsequently retrieved by sequencing or using *in situ* hybridization with multiple fluorescently labeled probes. Subsequently, a lineage tree can be reconstructed using phylogenetic computational approaches. (B) Molecular recording aims to capture transient signals and stably encode them into DNA within cells. One example is depicted, wherein input signals activate transcription of a DNA writer from a signal-sensing promoter, leading to the accumulation of mutations at target loci. These molecular records can be sequenced and used to decode information about input signals. In a case of a digital recording regime (blue), molecular records represent the information about presence or absence of input signals, whereas in analog recording (red), mutations gradually accumulate reflecting differing magnitudes or durations of input signals.

CRISPR-Cas genome editing tools, lineage tracing techniques included dye-based markers, transplantation, nucleotide pulse-chase analysis, Cre-Lox, FLP-FRT recombination-based mutagenesis, and sequencing of somatic mutations (18). However, these methods do not provide a detailed lineage tree over time but rather allow cell labeling at single time points. Over the past few years, Cas9-mediated genome editing has proven to be a powerful tool for lineage tracing via the dynamic introduction of stochastic mutations in the genome, thus barcoding individual cells. These barcodes, continuously generated during cellular division and differentiation, facilitate reconstruction of the original lineage relationships. The lineage tree is built based on the diversity of the barcode sequences: the more distant these mutational signatures, the less related the cells.

Leveraging this idea, McKenna and colleagues developed genome editing of synthetic target arrays for lineage tracing (GESTALT), a lineage tracing method based on genome editing of a transgenic reporter sequence in the genome (19). By transfecting cells with a plasmid expressing Cas9 and sgRNAs, or injecting Cas9 ribonucleoproteins (RNPs) containing sgRNAs that target a reporter sequence into embryos, they were able to generate barcodes to record lineage information and demonstrate the reconstruction of lineage relationships in early embryonic development of the zebrafish (*Danio rerio*). The transgenic

reporter in GESTALT contains 10 distinct sites in tandem, each of which is targeted by a different sgRNA, resulting in the generation of irreversible edits, subsequently retrieved by DNA sequencing of the GESTALT reporter barcode in adult tissues. Lineage trees are generated using patterns of shared mutations between barcodes, decoding relationships between different cell populations. Although GESTALT can provide the developmental history of a whole organism, this method has important limitations. First, the transient presence of Cas9 and sgRNAs restricts editing to a short time window, allowing for tracking of only a part of embryogenesis. Second, due to destruction of the cells during sample preparation, lineage trees generated using GESTALT lack information about specific cell types associated with identified barcodes.

To overcome these limitations, single-cell GESTALT (scGESTALT) (20) was developed, which combines the lineage recording capabilities of GESTALT with cell type identification by single-cell RNA sequencing (scRNA-seq). In scGESTALT, the edited barcodes are expressed from a reporter transgene consisting of nine distinct CRISPR loci in tandem, enabling detection at the RNA level. To extend the time frame of barcoding, different loci of the reporter are edited at different developmental stages. For barcoding cells during early embryogenesis, Cas9 RNPs containing sgRNAs targeting the first four loci of the reporter are injected at the one-cell stage. At later stages of development, expression of Cas9 is induced by heat shock, while sgRNAs targeting the other five loci are constitutively expressed. scRNA-seq of ~60,000 cells from the developing zebrafish brain identified >100 cell types. Using these data, lineage trees were generated with hundreds of branches that helped uncover relationships at the level of cell types, brain regions, and gene expression cascades during differentiation.

Recently, Chan and colleagues used a similar approach to map cell fates during early development in mouse embryos, from fertilization through gastrulation (21). In their system, a target locus with three Cas9 target sites and a unique barcode, as well as complementary sgRNAs, is delivered to oocytes using a single piggyBAC transposon vector, with multiple target sites being integrated into the genome of each cell. These three target loci are embedded into the 3' untranslated region of a constitutively transcribed gene encoding a fluorescent protein to enable detection through RNA-seq. Further, the oocytes are fertilized with constitutive Cas9-green fluorescent protein-encoding sperm to initiate DNA writing. After gastrulation, the resulting mutational outcomes at target sites and cell's transcriptional phenotype are simultaneously detected using scRNA-seq. Applying this system, the authors reconstructed lineage identities and also identified phenotypic convergence in cells originating from different embryonic layers. Despite providing cell type information, the diversity of barcodes and time of editing provided by this approach, as well as by scGESTALT, are not sufficient to cover the number of cells during mammalian embryogenesis, limiting observations to a short period.

To increase the number of possible barcodes for lineage tracing, Kalhor and colleagues used an evolvable hgRNA, which, similarly to stgRNAs (Fig. 16.1), contains an NGG PAM sequence in the scaffold, enabling continuous mutagenesis of the hgRNA expression locus (14). hgRNAs generate barcode diversity more than eight times higher than that obtained with a conventional sgRNA design and do not require the integration of paired sgRNA and target elements into model organisms. This system can go through multiple rounds of mutagenesis until indels generated during DSB repair result in either the deletion of the PAM sequence or the formation of truncated nonfunctional hgRNAs. The recording capability of this system can be expanded by introducing multiple orthogonal hgRNA loci. Based on this approach, a mouse line carrying 60 independent hgRNA loci (22) has been created for

in vivo barcoding and used to generate developmentally barcoded animals. Using this mouse line, lineage trees of early embryonic development and germ layers, neuroectoderm, and neural tube were reconstructed. Another substantial improvement to hgRNA-based lineage tracing could be achieved by enabling barcode recovery from the transcriptomes of single cells, thus preserving information about cell types.

One key limitation of DNA writers based on random mutagenesis using Cas9 is that DSBs are preferentially repaired with a deletion, rather than an insertion, which results in quick saturation or loss of target sites and barcodes. A recently described lineage tracing method named *cell history recording by ordered insertion* (CHYRON) addresses this by combining a Cas9-hgRNA system with a terminal deoxynucleotidyl transferase, which incorporates random nucleotides into Cas9-induced DSBs (23). This system allows the generation of continuous and ordered insertional mutations at the target locus and thus increases information encoding capacity. As a lineage tracer, this system proved powerful for reconstructing lineage relationships in human cell culture, setting the stage for future applications *in vivo*. Future improvements of CHYRON writing efficiency can potentially lead to a single-cell resolution of lineage tracing and molecular recording.

Base editing by cytidine deaminase-nCas9 fusion proteins can also be leveraged to address the problem of rapid saturation of barcodes due to deletions (17). The mutational activity of nCas9-CDA results in a high barcode diversity with single nucleotide resolution, without loss of barcodes due to deletions. This base editing approach was used to specifically target endogenous interspersed repeat regions in mammalian cell culture (24). The resulting genetic editing patterns in endogenous sites were used as cellular barcodes to reconstruct lineage trees. This method generates endogenous barcodes without making DSBs and also does not rely on the creation and insertion of complex target arrays. The base editing approach enables the editing of targets slowly but continuously, and it theoretically provides a diversity of cell barcodes sufficient to cover the approximately 37 trillion cells of the human body.

In contrast to deep sequencing, optical *in situ* detection of barcodes can provide spatial information about cell lineages. Another lineage tracing application of Cas9-based DNA writers called *memory by engineered mutagenesis with optical in situ readout* (MEMOIR) utilizes multiplexed single-molecule RNA fluorescence *in situ* hybridization to detect barcodes and cell-type-specific mRNA (25). Combining this optical readout with highly efficient DNA writers would be especially beneficial for defining developmental trajectories in embryos, tumors, and other systems.

Introduction to Molecular Recording

Current technologies to investigate cellular processes (such as transcriptomics, proteomics, and metabolomics) can only provide snapshots of molecular information, and therefore, the readout lacks information about a continuous sequence of biological events. Further, it is difficult to detect and monitor transient events using these snapshots. For addressing these limitations, DNA writers have been applied to capture biological, environmental, or artificial information within cells and encode it stably into DNA; this emergent application of DNA writing is termed molecular recording (Fig. 16.2B).

In general, the mechanism of molecular recorders involves three steps (2): (i) sensing of input stimuli such as specific cellular signals (e.g., presence or absence of small molecules) or stochastic information within the cell (e.g., abundance of mRNA transcripts), (ii) transformation of input signals or information into a standardized format for being encoded (e.g., through triggering the expression of DNA writing modules), and (iii) writing into a DNA medium (e.g., indel generation at predefined genomic loci) (4). Resulting

“molecular records” encoded into the DNA of cell populations can be retrieved by sequencing or hybridization methods long after transient input signals have disappeared.

Molecular recording can be classified into digital and analog based on the encoding regime. Digital recording involves the encoding of information about presence or absence of input signals, whereas in analog recording, mutations or modifications gradually accumulate in response to differing magnitudes or durations of input signals. Thus, both the presence and concentration of biological molecules can serve as input signals for a recording system. Potential inputs can also include environmental stimuli such as temperature, light, pH, electricity, and other physiological characteristics capable of being sensed by cells and transformed into cellular signals (Fig. 16.2B). Further, in addition to detecting predefined signals, analog recording can also be used to encode information about cellular processes and cell states. A range of biological signals can be used as proxies for cell states; for example, transcription levels of mRNA and noncoding RNAs from different genes represent the transcriptional state of cells. Similarly, proteins and metabolites represent other markers of cell states.

One approach to sense input stimuli is to directly couple signals of interest to the expression or posttranscriptional/posttranslational activation of a DNA writer. Transcriptional activation can be achieved by using naturally occurring or engineered signal-responsive promoters (e.g., cAMP response element in mammalian cells), whereas signal-dependent conformation changes or modifications of DNA writers can be used for posttranslational activation (e.g., split-Cas9 nuclease) (4). Alternatively, engineered biosensor modules can be used to specifically detect the signal of interest and transform it into altered activation levels of the DNA writer (2). Another approach is to stochastically sample a broad range of input information, such as the pool of mRNA transcripts present within a cell, and continuously encode this sampled information using DNA writers.

The type of signal transformation, DNA writing module efficiency, and response time are other important features of molecular recorders, which determine whether an analog or digital recording regime can be achieved. Using synthetic biology and genome editing techniques, the final characteristics of the recorder can be optimized to match the desired application. For instance, transcription-based activation demonstrates slow signal transduction ($>10^2$ s), whereas enzyme-based posttranslational sensors respond to a signal much quicker ($<10^2$ s), making them suitable for tracking fast biological processes (26).

Molecular Recording Using CRISPR-Cas9

In a recent approach, named mammalian synthetic cellular recorders integrating biological events (mSCRIBE), Perli et al. leveraged stgRNAs and Cas9 for molecular recording in mammalian cells (15) (Fig. 16.3). By linking the expression of stgRNA or Cas9 to biological signals, the authors found an accumulation of stgRNA mutations upon increasing the magnitude and duration of input signals, thereby demonstrating analog memory being stored in a population of human cells. To demonstrate the power of this approach, they developed an inflammation recording cell line, where Cas9 expression is linked to the activation of NF- κ B signaling that plays an essential role in coordinating a response to inflammation. These sentinel cells enabled the recording of different concentrations of tumor necrosis factor alpha (a potent activator of the NF- κ B pathway) and upon implantation into mice were able to detect systemic inflammation caused by lipopolysaccharide injections, demonstrating *in vivo* recording of physiologically relevant biological signals. In the future, coupling stgRNA modules with the response elements of other signaling pathways could allow simultaneous tracking of multiple cellular processes. Although mSCRIBE enables

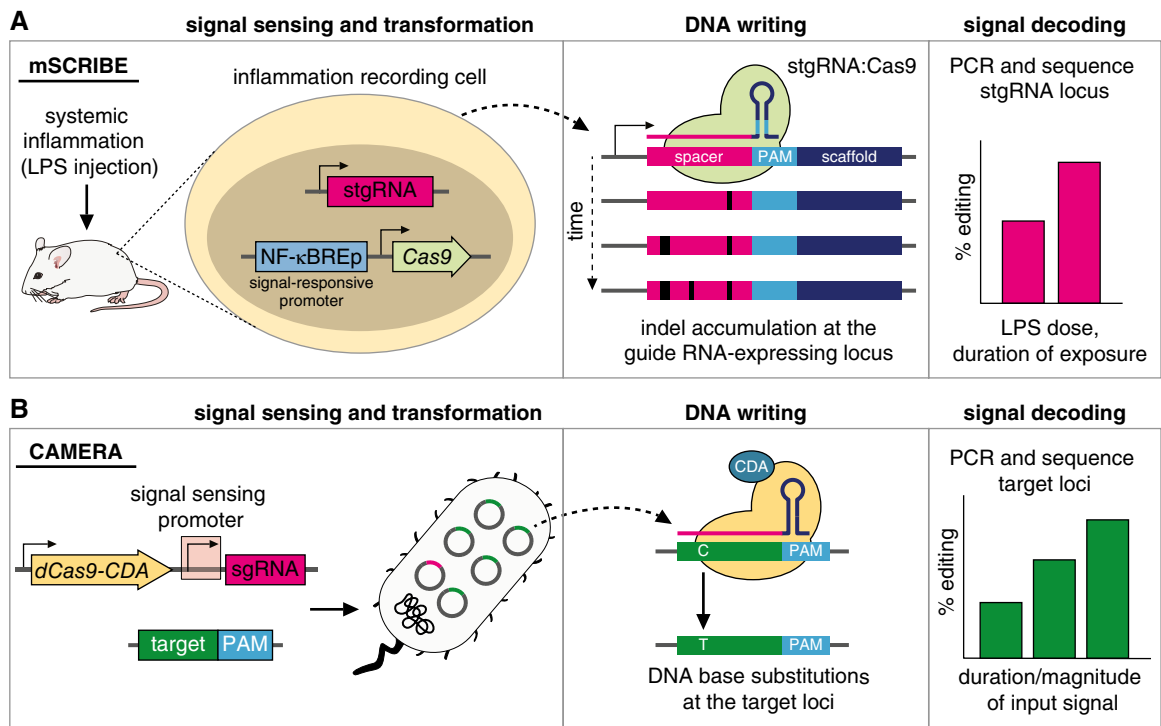


Figure 16.3 Examples of molecular recording by CRISPR-Cas9. (A) Leveraging Cas9 and self-targeting guide RNAs (stgRNAs), mSCRIBE enables detection of information about systemic inflammation. In this system, inflammation recording cells contain an NFκB-responsive promoter (NFκBEP) controlling the expression of Cas9. Systemic inflammation caused by lipopolysaccharide (LPS) injection into mice results in the accumulation of indels at the stgRNA loci. By quantifying mutations at the stgRNA loci, information about LPS dose and duration of exposure can be revealed. (B) Fusion of catalytically inactive dCas9 with a cytidine deaminase domain (dCas9-CDA) is utilized in a molecular recording approach named CAMERA. In this system, a signal sensing promoter triggers sgRNA expression and leads to base editing at target loci. The target loci are sequenced and editing efficiency is calculated to decode and quantify the input signals.

analog recording of biological signals, the recording is interrupted when the PAM sequence is deleted or the stgRNA is shortened below 16 bp, resulting in a limited recording time. Additionally, continuous self-mutation generates a diverse set of stgRNAs variants, leading to potential off-target effects and introducing noise into the recording process.

To avoid random DNA mutagenesis, base editing approaches utilizing dCas9 or nCas9 fusions with CDA have been utilized for molecular recording (27, 28). In a system called CRISPR-mediated analog multievent recording apparatus 2 (CAMERA 2), Tang and Liu used base editors to record various stimuli in bacteria and human cell cultures in the form of single-base mutations at defined genomic loci (27) (Fig. 16.3). In bacteria, CAMERA 2 utilizes a two-plasmid system with a “writing” plasmid expressing sgRNAs and dCas9-CDA fusion protein and multiple copies of a “recording” plasmid carrying different target sequences. The information is then recovered by sequencing the target loci within the recording plasmid. The percentage of edits per specific nucleotide position reflects the dosage, amplitude, and duration of input signals. Since a high-copy-number recording plasmid could have thousands of copies per cell, CAMERA 2 supports analog recording even in small populations (10 to 100 cells), in comparison to other recording technologies that enable recording only on a scale of 1,000 to 10,000 bacterial cells. Combining CAMERA 2 with different inducible promoters, the authors were able to record a wide range of signals, including exposure to antibiotics, nutrients, viruses, and light in bacteria. In a similar manner, CAMERA 2m exploits human safe

harbor gene *CCR5* as a recording locus in mammalian cell culture. Controlling the expression of sgRNAs targeting *CCR5* by a synthetic inducible promoter allowed simultaneous recording of multiple stimuli in a population of cells. Furthermore, leveraging a Wnt-inducible T-cell factor/lymphoid enhancer factor promoter (LEF-TCF) to control the expression of the base editor, the authors were able to detect the activation of the Wnt pathway in human cells. Base editors provide predictable mutational outcomes with single-nucleotide resolution, which enables the direct association of input stimuli with mutations and allows for a high recording capacity. Recent works also leverage adenine base editors to introduce A:T-to-G:C mutations in a programmable manner (29). Combining cytidine and adenine base editors could be used to develop new powerful rewritable tools for molecular recording in the future.

Molecular Recording Using CRISPR Spacer Acquisition

CRISPR spacer acquisition evolved in bacteria and archaea as a natural mechanism to capture molecular records of genetic encounters and horizontal gene transfer events in the form of spacers stored within CRISPR arrays. Inspired by this, recent molecular recording approaches apply the molecular machinery of the CRISPR adaptation module as a tool for recording biological and arbitrary information into a DNA medium (Fig. 16.4). In addition to recording a diversity of input information, this approach can provide temporal resolution of recordings via the sequential and ordered spacer acquisition. This approach is also capable of directly sensing and encoding input information such as presence and abundance of transcripts or DNA sequences, precluding the need for signal transformation.

The pioneering example of applying CRISPR spacer acquisition for molecular recording utilizes the Cas1-Cas2 acquisition complex from *Escherichia coli* to integrate spacers directly into genomic CRISPR arrays. By overexpressing *E. coli* Cas1 and Cas2 proteins, Shipman and colleagues were able to generate records of specific DNA sequences in a population of bacterial genomes (8). Spacer acquisition was demonstrated from genomic and plasmid sources, as well as from synthetic double-stranded DNA sequences electroporated into the cell, and shown to be dependent on spacer sequence and the presence of a 5' PAM. Since the spacers are integrated between a leader sequence and the first direct repeat sequence of the CRISPR array, this system was also shown to reconstruct the temporal sequence of events based on the order of acquisition. Further, the PAM recognition of this acquisition system can be modified through directed evolution, leading to the possibility of using multiple systems recognizing different PAMs orthogonally to enhance the range of input information. In a follow-up study, they encoded pixel information in the form of short fragments of synthetic DNA, which were then electroporated into recorder *E. coli* cells (30). These fragments served as artificial protospacers, being recognized by the CRISPR acquisition complex and integrated into CRISPR arrays in the bacterial genome. The images were thus permanently recorded into a population of cells and could subsequently be retrieved by sequencing the respective CRISPR arrays. Leveraging the unidirectional nature of CRISPR spacer acquisition, the authors also encoded multiple images over time within a population of bacteria, generating a short movie (Fig. 16.4A).

Besides recording synthetic inputs such as oligonucleotides, CRISPR spacer acquisition can be used to detect horizontal gene transfer events. Munck and colleagues leveraged the *E. coli* Cas1-Cas2 spacer acquisition complex to identify horizontally transferred elements in the human gut microbiome via exposing a recording *E. coli* strain to human clinical fecal samples (31). In the recording strain, upon overexpression of the Cas1-Cas2 acquisition complex, spacers were acquired from horizontally transferred elements. These exogenous spacers, permanently stored in genomic CRISPR arrays, were

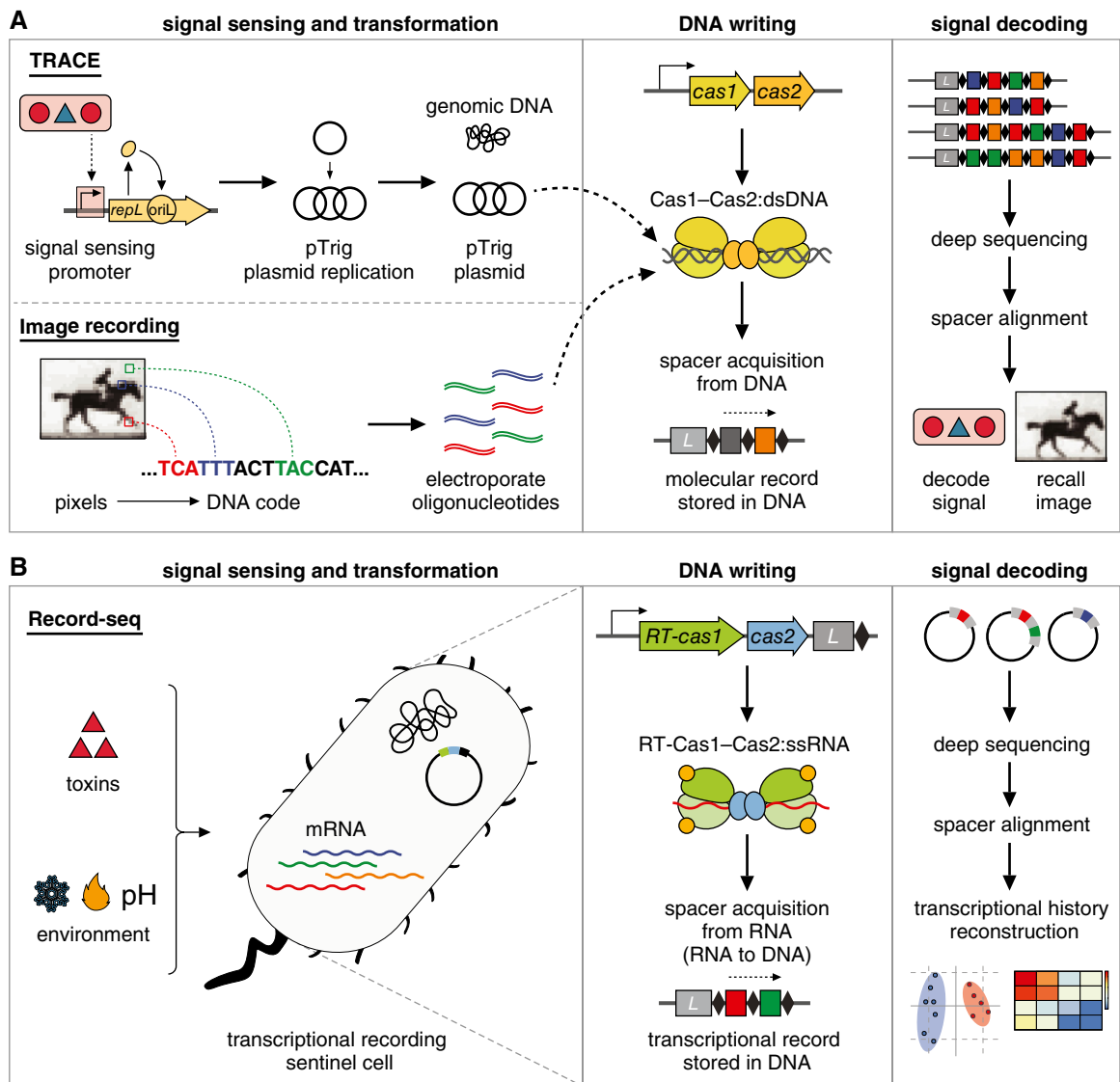


Figure 16.4 Molecular recording by CRISPR spacer acquisition. (A) Spacer acquisition from double-stranded DNA (dsDNA) by the Cas1-Cas2 protein complex from *E. coli* can be used for recording biological information or encoding digital information in a population of bacteria. Input digital information, such as an image or sequence of images, is encoded in DNA oligonucleotides (image recording), whereas in TRACE, an input biological signal is converted to a pTrig plasmid DNA abundance. The Cas1-Cas2 DNA writing module acquires spacers from input DNA, thus recording this information permanently in the form of spacers stored within CRISPR arrays. Deep sequencing of CRISPR arrays and aligning of the newly acquired spacers to reference sequences (i.e., *E. coli* genome, plasmid, or oligonucleotides) allows the original input signals to be decoded. (B) Record-seq leverages spacer acquisition from single-stranded RNA (ssRNA) for recording transcriptional information. Upon heterologous expression in *E. coli*, RT-Cas1-Cas2 acquires new spacers from RNA, based on transcript abundance. Thus, changes in the transcriptional state of a cell caused by a variety of input signals are reflected in the number of spacers aligning to the genes differentially expressed in response to the stimuli. Deep sequencing of CRISPR arrays and alignment of the newly acquired spacers to reference sequences allow the transcriptional history of a population of bacterial cells to be decoded.

then sequenced and used to assess their origins and transfer rates. This approach allows for detection and quantification of horizontal gene transfer events and demonstrates a potential application of molecular recording by CRISPR spacer acquisition to study the dynamics of complex microbial communities.

Analogously to Cas9-based molecular recording devices, the activity of CRISPR spacer acquisition can be coupled to a sensing module to detect input signals within cells. A technology termed temporal recording in arrays by CRISPR expansion (TRACE) employs *E. coli* spacer acquisition machinery to enable the recording of extracellular stimuli (Fig. 16.4A) (9). The input signal is recognized by an inducible synthetic promoter, which regulates the expression of RepL protein, essential for the replication of a copy number-inducible trigger plasmid. Consequently, a transcriptional signal can be converted to an altered abundance of the trigger plasmid in the intracellular DNA pool, providing a source of DNA protospacers for Cas1-Cas2 proteins expressed in the same cell. This signal can then be detected through an increase in spacers mapping to the corresponding trigger plasmid. TRACE has been successfully applied for recording multiple stimuli in a population of bacterial cells, such as the presence of isopropyl- β -D-thiogalactopyranoside (IPTG) or metabolites, and provided information regarding chronological order of input signals and magnitude of exposure. Despite providing a sequential resolution of recording, the input signals detected by TRACE are limited to a set of stimuli that need to be defined prior to the experiment, and further applications depend on the availability of inducible promoters that trigger spacer acquisition.

Although the prototypical Cas1-Cas2 integration complex from *E. coli* acquires spacers only from DNA, adaptation modules from other organisms are capable of acquiring spacers from RNA through naturally occurring fusions between a reverse transcriptase (RT) and Cas1 (RT-Cas1) (10, 32, 33). Such RT-Cas1-Cas2 protein complexes can be leveraged to directly record information about transcript abundance in the form of RNA-derived spacers, thus eliminating the need for coupling additional sensors to the adaptation module for recording biological stimuli.

Schmidt and colleagues recently demonstrated that an RT-Cas1-Cas2 complex derived from the human commensal bacterium *Fusicatenibacter saccharivorans* (FsRT-Cas1-Cas2), upon heterologous expression in *E. coli*, acquires new spacers from RNA transcripts based on their abundance. This feature was harnessed to develop Record-seq (Fig. 16.4B), a method for recording RNA on a transcriptome scale (10). In a Record-seq experiment, the FsRT-Cas1-Cas2 adaptation module records short spacer sequences from RNA templates into plasmid-borne CRISPR arrays, with spacer abundance serving as a proxy for transcript abundance. The plasmid DNA can then be extracted and deep sequenced to identify acquired spacers. By aligning acquired spacer sequences to the corresponding genes in the *E. coli* genome, cumulative gene expression can be quantified and transcriptional histories of the cells can be reconstructed. Thus, Record-seq allows for the direct recording of transcriptional responses without the need to create reporter strains dedicated to specific stimuli. Record-seq has been applied for detecting complex cellular responses to environmental stimuli, such as oxidative and acid stress, or exposure to different dosages of herbicides. Remarkably, in comparison to RNA sequencing, Record-seq is capable of capturing transcriptional response to transient stimuli and preserving this information even after the transcriptional state has changed. However, the extremely low efficiency of CRISPR spacer acquisition from RNA allows the recording of transcriptional events only on a population scale and hinders further applications of Record-seq, which would require higher sensitivity and better time resolution.

Outlook

CRISPR-Cas systems provide an advanced toolbox for the development of sophisticated DNA writing technologies, providing an entry point to recording multiple facets of dynamic biological processes through time. These CRISPR-based tools broaden the repertoire of

potential DNA modifications, providing a highly flexible and efficient means to store information in DNA. This, in turn, provides a technology suite enabling the sensing of diverse stimuli, recording of multiple signals, and enhanced resolution of cellular lineage relationships. In this chapter, we described the diversity of CRISPR-Cas-based DNA writers and their emerging applications for cell lineage reconstruction and molecular recording.

Genome editing technology development is a dynamic area of research, and the next generation of DNA writing technologies will be realized through both rational engineering approaches and fundamentally new discoveries. For instance, one of the main limitations of Cas9 base editors is their off-target effects and the narrow range in which they can edit DNA. However, recent work on the continuous evolution of base editors resulted in the development of new protein variants with expanded target compatibility and improved activity (34), and the implementation of this new generation of base editors for DNA writing is imminent. Other CRISPR effector proteins with novel properties could also provide unique opportunities to encode information in DNA, such as Cas12a (35), which generates staggered cuts (in contrast to the blunt ends generated by Cas9), or Cas12k (36), which acts as an RNA-guided transposase, offering fundamentally new opportunities for DNA writing.

In addition to DNA writing modules based on Cas effector proteins, CRISPR spacer acquisition complexes present a resource for DNA writing with enormous information capture capacity and built-in temporal resolution. Moreover, molecular recording using CRISPR spacer acquisition from RNA eliminates the need for specific input signals or coupled sensing modules and enables us to directly encode a wide range of signals of interest. However, DNA writing modules based on CRISPR spacer acquisition are less efficient than those based on Cas9 and have only been demonstrated for prokaryotes, necessitating further development in this area before their full potential can be realized.

The application of CRISPR-Cas DNA writers to lineage tracing has enabled dynamic mutagenesis at target loci, increasing the time frame and resolution of cell lineage reconstruction. However, further application to mammalian *in vivo* studies is impeded by the low diversity of cellular barcodes created by existing DNA writers. In the future, DNA writing modules with higher information storage capacity, such as CRISPR spacer acquisition protein complexes or the new generation of improved base editors, can be harnessed for cellular lineage tracing. Another limitation of existing lineage tracing technology is a quick saturation of target loci available for mutagenesis, since current DNA writing modules continually introduce mutations upon being expressed. This could potentially be addressed by coupling DNA writing activity to cell division events, for instance, using cell cycle regulatory machinery of mammalian cells.

Analogously, molecular recording techniques have advanced with the expanded CRISPR-Cas toolbox, enabling the encoding of specific cellular signals or biological information into cell populations. A crucial limitation of current molecular recording technologies, especially those based on CRISPR spacer acquisition, is the low activity of DNA writing modules, which limits the amount of information retrieved per cell and necessitates population-level analyses. Enhancing the efficiency of these systems could potentially enable recordings at a single-cell level, in addition to improving temporal resolution, and facilitate the development of sophisticated approaches to understand dynamic cellular processes.

Besides experimental innovations, advances in analytical methods are needed to fully and efficiently leverage biological information encoded by DNA writers for recapitulating complex and dynamic biological processes. For instance, computational approaches for inferring the order of transcriptional states based on temporal information stored by molecular recorders could help reconstruct the sequence of molecular processes within the cell. Ultimately, we envision that combinatorial strategies integrating temporal and lineage

information from DNA writing methods, with functional and spatial information from orthogonal technologies such as scRNA-seq, proteomics, metabolomics, topological mappings, and other multiplexed perturbation strategies, will enable single-cell-level analyses of the molecular mechanisms of development and disease. The continued development of novel DNA writing technologies will further expand our capacity to store information in DNA, and the application of these DNA writers will provide a powerful toolbox for unraveling the molecular events that orchestrate biology and disease.

References

- Church GM, Gao Y, Kosuri S. 2012. Next-generation digital information storage in DNA. *Science* 337:1628.
- Sheth RU, Wang HH. 2018. DNA-based memory devices for recording cellular events. *Nat Rev Genet* 19:718–732.
- Schmidt F, Platt RJ. 2017. Applications of CRISPR-Cas for synthetic biology and genetic recording. *Curr Opin Syst Biol* 5:9–15.
- Farzadfard F, Lu TK. 2018. Emerging applications for DNA writers and molecular recorders. *Science* 361:870–875.
- Gardner TS, Cantor CR, Collins JJ. 2000. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403:339–342.
- Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ. 2009. Synthetic gene networks that count. *Science* 324:1199–1202.
- Amitai G, Sorek R. 2016. CRISPR-Cas adaptation: insights into the mechanism of action. *Nat Rev Microbiol* 14:67–76.
- Shipman SL, Nivala J, Macklis JD, Church GM. 2016. Molecular recordings by directed CRISPR spacer acquisition. *Science* 353:aaf1175.
- Sheth RU, Yim SS, Wu FL, Wang HH. 2017. Multiplex recording of cellular events over time on CRISPR biological tape. *Science* 358:1457–1461.
- Schmidt F, Cherepkova MY, Platt RJ. 2018. Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* 562:380–385.
- Yosef I, Goren MG, Qimron U. 2012. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res* 40:5569–5576.
- Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F. 2015. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 520:186–191.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821.
- Kalhor R, Mali P, Church GM. 2017. Rapidly evolving homing CRISPR barcodes. *Nat Methods* 14:195–200.
- Perli SD, Cui CH, Lu TK. 2016. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* 353:aag0511.
- Pitcher RS, Wilson TE, Doherty AJ. 2005. New insights into NHEJ repair processes in prokaryotes. *Cell Cycle* 4:675–678.
- Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533:420–424.
- Kretzschmar K, Watt FM. 2012. Lineage tracing. *Cell* 148:33–45.
- McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J. 2016. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* 353:aaf7907.
- Raj B, Wagner DE, McKenna A, Pandey S, Klein AM, Shendure J, Gagnon JA, Schier AF. 2018. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat Biotechnol* 36:442–450.
- Chan MM, Smith ZD, Grosswendt S, Kretzmer H, Norman TM, Adamson B, Jost M, Quinn JJ, Yang D, Jones MG, Khodaverdian A, Yosef N, Meissner A, Weissman JS. 2019. Molecular recording of mammalian embryogenesis. *Nature* 570:77–82.
- Kalhor R, Kalhor K, Mejia L, Leeper K, Graveline A, Mali P, Church GM. 2018. Developmental barcoding of whole mouse via homing CRISPR. *Science* 361:eaat9804.
- Loveless TB, Grotts JH, Schechter MW, Forouzmand E, Carlson CK, Agahi BS, Liang G, Ficht M, Liu B, Xie X, Liu CC. 2021. Lineage tracing and analog recording in mammalian cells by single-site DNA writing. *Nat Chem Biol* 17:739–747.
- Hwang B, Lee W, Yum S-Y, Jeon Y, Cho N, Jang G, Bang D. 2019. Lineage tracing using a Cas9-deaminase barcoding system targeting endogenous L1 elements. *Nat Commun* 10:1234–1239.
- Frieda KL, Linton JM, Hormoz S, Choi J, Chow KK, Singer ZS, Budde MW, Elowitz MB, Cai L. 2017. Synthetic recording and in situ readout of lineage information in single cells. *Nature* 541:107–111.
- Olson EJ, Tabor JJ. 2012. Post-translational tools expand the scope of synthetic biology. *Curr Opin Chem Biol* 16:300–306.
- Tang W, Liu DR. 2018. Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* 360:eaap8992.
- Farzadfard F, Gharaei N, Higashikuni Y, Jung G, Cao J, Lu TK. 2019. Single-nucleotide-resolution computing and memory in living cells. *Mol Cell* 75:769–780.e4.
- Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. 2018. Publisher Correction: Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 559:E8.
- Shipman SL, Nivala J, Macklis JD, Church GM. 2017. CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature* 547:345–349.
- Munck C, Sheth RU, Freedberg DE, Wang HH. 2020. Recording mobile DNA in the gut microbiota using an *Escherichia coli* CRISPR-Cas spacer acquisition platform. *Nat Commun* 11:95.
- Silas S, Mohr G, Sidote DJ, Markham LM, Sanchez-Amat A, Bhaya D, Lambowitz AM, Fire AZ. 2016. Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* 351:aad4234.
- González-Delgado A, Mestre MR, Martínez-Abarca F, Toro N. 2019. Spacer acquisition from RNA mediated by a natural reverse transcriptase-Cas1 fusion protein associated with a type III-D CRISPR-Cas system in *Vibrio vulnificus*. *Nucleic Acids Res* 47:10202–10211.
- Thuronyi BW, Koblan LW, Levy JM, Yeh W-H, Zheng C, Newby GA, Wilson C, Bhaumik M, Shubina-Oleinik O, Holt JR, Liu DR. 2019. Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat Biotechnol* 37:1070–1079.
- Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin EV, Zhang F. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163:759–771.
- Strecker J, Ladha A, Gardner Z, Schmid-Burgk JL, Makarova KS, Koonin EV, Zhang F. 2019. RNA-guided DNA insertion with CRISPR-associated transposases. *Science* 365:48–53.