



Review

Toward DNA-Based Recording of Biological Processes

Hyeri Jang¹ and Sung Sun Yim^{1,2,3,4,*}

- ¹ Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea
² Graduate School of Engineering Biology, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea
³ KAIST Institute for BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea
⁴ Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 34141, Republic of Korea
* Correspondence: sunsunnyim@kaist.ac.kr

Abstract: Exploiting the inherent compatibility of DNA-based data storage with living cells, various cellular recording approaches have been developed for recording and retrieving biologically relevant signals in otherwise inaccessible locations, such as inside the body. This review provides an overview of the current state of engineered cellular memory systems, highlighting their design principles, advantages, and limitations. We examine various technologies, including CRISPR-Cas systems, recombinases, retrons, and DNA methylation, that enable these recording systems. Additionally, we discuss potential strategies for improving recording accuracy, scalability, and durability to address current limitations in the field. This emerging modality of biological measurement will be key to gaining novel insights into diverse biological processes and fostering the development of various biotechnological applications, from environmental sensing to disease monitoring and beyond.

Keywords: DNA-based data storage; molecular recording; cellular memory; synthetic biology



Citation: Jang, H.; Yim, S.S. Toward DNA-Based Recording of Biological Processes. *Int. J. Mol. Sci.* **2024**, *25*, 9233. <https://doi.org/10.3390/ijms25179233>

Academic Editor: Alicja Wegrzyn

Received: 2 July 2024

Revised: 21 August 2024

Accepted: 24 August 2024

Published: 26 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Biological processes are inherently complex and dynamic. Living organisms interact with each other and their environments by generating diverse biomolecules and metabolites, and these interactions continuously change over time. For example, microbial cells in the gut microbiome constantly sense environmental changes and respond by regulating the expression of specific genes necessary for their survival [1]. In multicellular organisms, the intricate regulation of numerous genes controls the differentiation of multiple cell types throughout development [2]. However, many of these dynamics remain poorly understood since native biological environments are often inaccessible, and tracking multiple biological events over time is still challenging [3].

While several approaches such as temporal RNA-seq [4–6] and biosensors [7–9] have been devised to address these challenges in biological measurement, they are still constrained by their temporal resolution and the number of channels available for data acquisition. Utilizing DNA as a data storage medium provides high-capacity storage, high density, and long-term stability to encode various types of data [10,11]. Advanced next-generation sequencing (NGS) technologies have facilitated convenient, cost-effective, and high-throughput decoding of information stored in DNA [12]. Furthermore, the inherent compatibility between DNA data storage and living systems has spurred the development of various DNA-based cellular recording techniques, which have the potential to acquire multiple and temporal biological information without disrupting cells (Figure 1a) [3]. Many different applications of DNA-based cellular recording have been demonstrated, such as diagnosing disease biomarkers [13,14], capturing horizontal gene transfer (HGT)

events [15,16], tracking cellular lineages throughout embryonic development [17,18], storing digital data [19,20], and constructing genetic circuits for therapeutic applications [21] (Figure 1b).

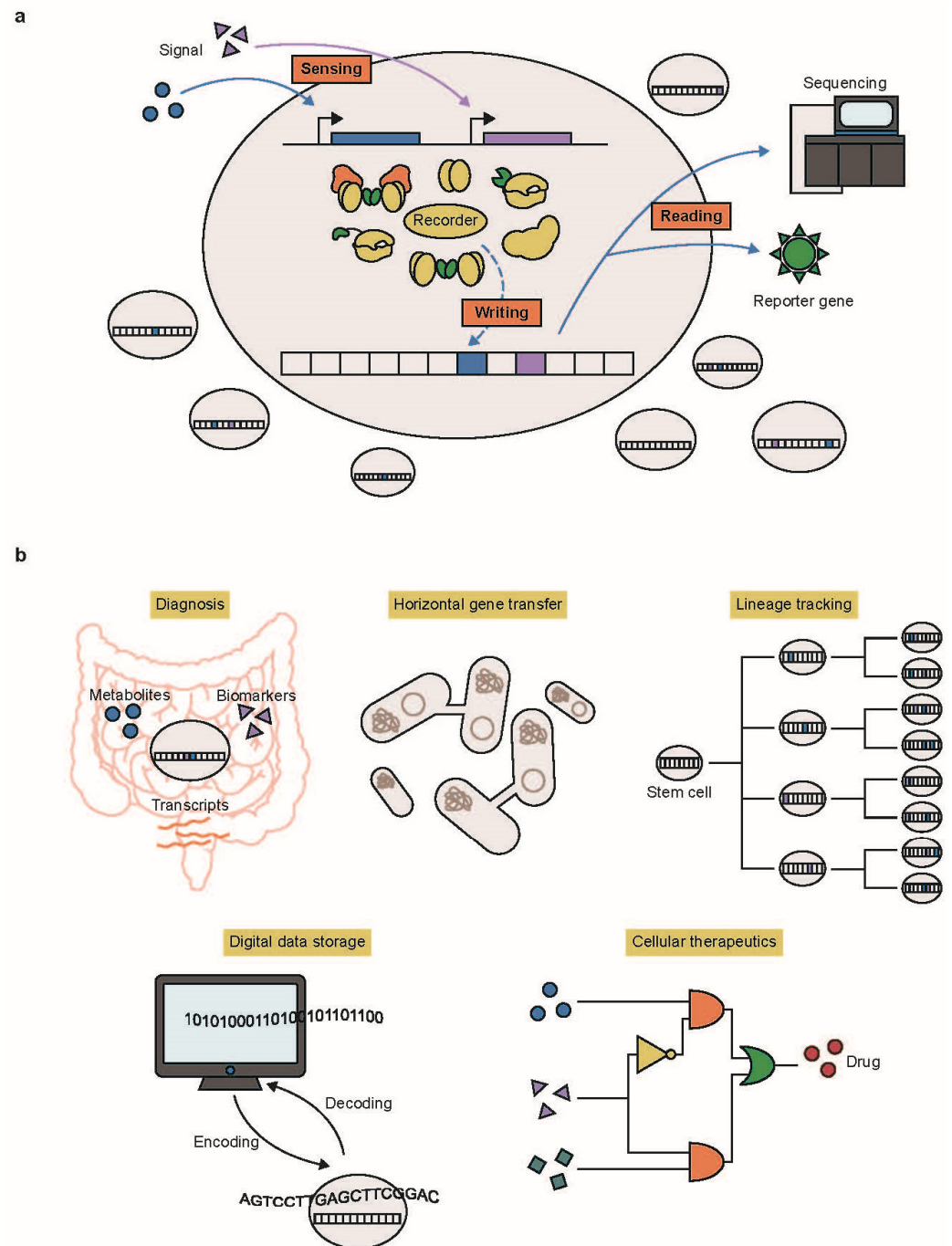


Figure 1. (a) In DNA-based cellular recording, various environmental or cellular signals activate molecular recorders. Once activated, these recorders alter the DNA sequence or epigenetic states to store the data. The recorded data can be retrieved through sequencing or reporter gene expression. (b) Examples of DNA-based cellular recording applications include diagnosing cellular states, understanding horizontal gene transfer (HGT) events within the microbiome, tracking cellular lineages, storing digital data, and constructing genetic circuits for therapeutic purposes.

In this review, we explore the principles of genome editing-based cellular recording systems, highlighting their benefits and applications across various fields. We also examine

potential strategies to overcome current limitations in this area. This emerging method of biological measurement is crucial for obtaining new insights into diverse biological processes and advancing various biotechnological applications.

2. Recombination-Based Cellular Recording

Recombinases are enzymes that mediate site-specific recombination by catalyzing excision, inversion, and integration of specific target DNA sequences, depending on the orientation of flanking homologous regions. These site-specific DNA recombinases have been utilized to construct various genetic circuits for cellular recording, such as permanent genetic memories and reversible genetic switches, which can be analyzed by recombination site sequences or reporter gene expression [22–27]. For the development of the recombination-based genetic memory system with >1-byte capacity, Yang et al. bioinformatically identified orthogonal phage integrases with their cognate recognition (attB-attP) sites and constructed a ‘memory array’ by linearly concatenating the recognition sites for each integrase [23]. With the orthogonal recombinase–recombination site pairs, the recording of temporally ordered signals could also be demonstrated as ‘recombinase-based state machines’ (RSMs) (Figure 2a) [28]. In the RSM concept, sequence states could be generated on DNA registers (memory arrays) made up of overlapping and orthogonal recombinase recognition sites. Depending on the order of a set of chemical inputs, corresponding recombination events could result in expected sequence states of the two-input five-state and three-input 16-state registers and cellular behaviors.

Beyond recording the occurrence (presence or absence) of events, the recombinase-based approach could also encode the duration and intensity of biological events. The ‘synthetic cellular recorders integrating biological events’ (SCRIBE) system was developed to record analog information, such as the magnitude and time course of inputs, within living cell populations by converting transcriptional signals into the production of single-stranded DNA (ssDNA), followed by ssDNA-based genome editing [29]. In the SCRIBE system, retrons, composed of a non-coding RNA (ncRNA) region with multicopy single-stranded RNA (msr) and multicopy single-stranded DNA (msd), as well as retron reverse transcriptase (retron RT) [30], are utilized to produce the ssDNAs [31–33]. And the ssDNAs write information at specific genomic loci as recombination frequencies within cell populations when single-strand annealing proteins (SSAPs) are co-expressed. Recently, the recombination efficiency of ssDNA retrons was improved by knocking out or knocking down cellular ssDNA-specific exonucleases, which affect the intracellular stability of ssDNA, enabling a broader range of applications for the system [16].

While recombinase-based recording systems have primarily been established within model bacterial systems, their implementations have been successfully demonstrated in non-model bacteria and even in eukaryotes, including human and plant cells [34–36]. However, scalability remains challenging due to the limited number of available orthogonal recombinases. To address this, computational mining of efficient and orthogonal recombinases from microbial genomes could further expand the recombinase toolbox [37]. Alternatively, exploiting recombinases with orthogonal attachment sites and synthetic transcription factors together could increase memory capacity for each recombinase and enable much faster recombination [38].

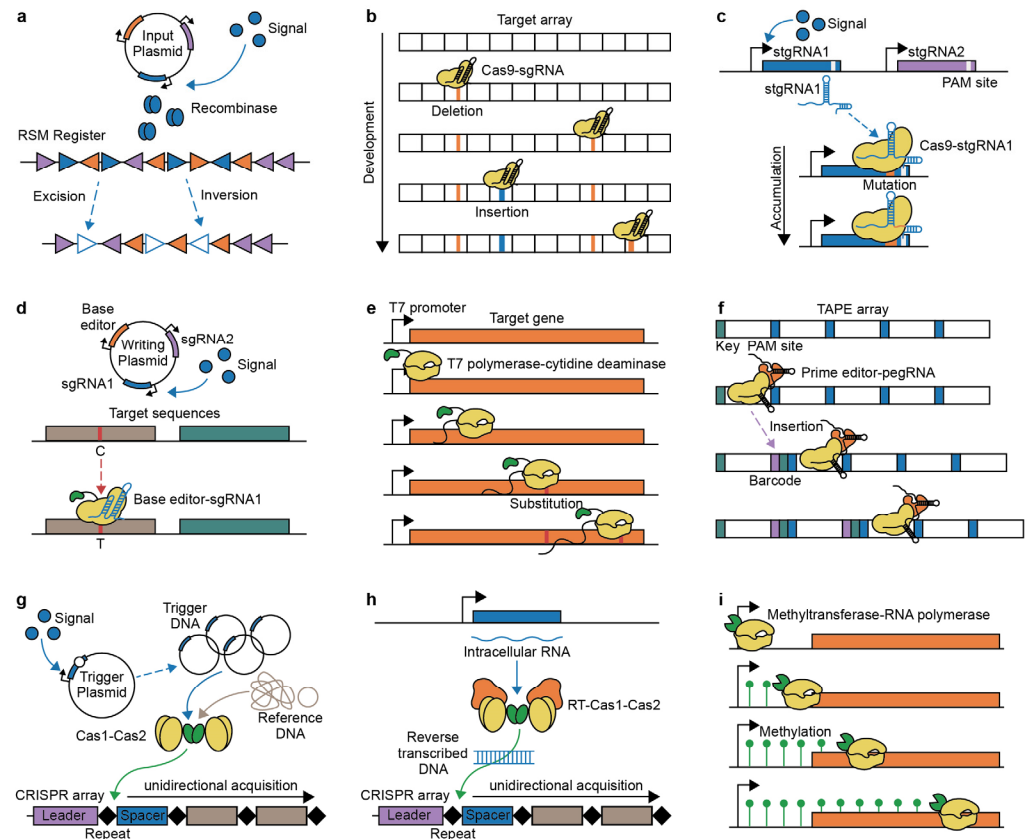


Figure 2. The principles of DNA-based recording systems are illustrated. (a) Recombinase-based state machines (RSMs): Orthogonal recombinases are activated in response to multiple signals. Depending on the order of signals, recombinases facilitate either excision or inversion of the RSM register, enabling the recording of the temporal order of multiple signals. (b) Genome editing of synthetic target arrays for lineage tracing (GESTALT): A contiguous array of target barcodes is edited by Cas9 nuclease-sgRNA throughout cell development. The accumulated patterns of deletions and insertions enable the reconstruction of lineage tree. (c) Mammalian synthetic cellular recorders integrating biological events (mSCRIBE): Multiple self-targeting guide RNAs (stgRNAs) and Cas9 nucleases are used to edit the stgRNA gene itself for monitoring biological signals. Within the cell population, self-targeting patterns correlate with either the duration or intensity of the signals. (d) CRISPR-mediated analog multi-event recording apparatus (CAMERA): Inducible base editors and sgRNAs generate C·G to T·A point mutations at recording sites. The editing frequencies depend on signal amplitude or duration, and the editing patterns indicate the order of events. (e) T7 polymerase-driven continuous editing system: T7 polymerase fused to cytidine deaminase transcribes a specific gene downstream, continuously generating substitution patterns. (f) DNA Typewriter: The pegRNA, consisting of key sequences, barcodes, and type guide sequences, is expressed under a promoter. The prime editor inserts the key and barcode sequences adjacent to the PAM site in a unidirectional manner, enabling temporal recording within cells. (g) Temporal recording in arrays by CRISPR expansion (TRACE): Biological signals activate replication proteins, facilitating the replication of the pTrigger plasmid. The Cas1–Cas2 complex integrates trigger DNA into the CRISPR array at a higher frequency compared to reference sequences. The unidirectionality of CRISPR acquisition allows for the temporal recording of multiple signals. (h) Record-seq: Expressed intracellular RNA is reverse transcribed into DNA sequences by RT. The resulting double-stranded DNA is then integrated into the CRISPR array by the Cas1–Cas2 complex. This system enables transcriptome-scale recording. (i) DCM-time machine (DCM-TM): The fusion protein of DCM methyltransferase and RNA polymerase is activated by an inducible signal. When the RNA polymerase acts on genes and active enhancers, DCM methyltransferase marks the methylation patterns along the sequences.

3. Implementation of Genome Editing for Molecular Recording

Genome editing involves the precise alteration of genomic sequences in living organisms by generating targeted insertions, deletions, and substitutions. While various genome-editing techniques, such as zinc-finger nucleases (ZFN) and transcription activator-like effector nuclease (TALEN), have demonstrated potential for effective genome engineering [39,40], the emergence of CRISPR technology has facilitated programmable genome engineering, leading to the development of diverse DNA-based recording systems [41–43].

3.1. CRISPR-Cas9 Barcoding-Based Lineage Tracing

The CRISPR-Cas9 system, a prokaryotic adaptive immune system, is composed of the Cas9 nuclease and single-guide RNA (sgRNA). CRISPR-Cas9 is a robust technology that facilitates genome engineering, screening, and transcription regulation by precisely recognizing and cleaving specific locations and editing target sequences within the genome [44–48]. The CRISPR-Cas9 nuclease causes DNA double-stranded breaks (DSBs) at specific locations, leading to irreversible insertions or deletions during the repair processes. The accumulation of these mutations could be utilized as unique barcodes for individual cells or cellular events in DNA-based cellular recording.

CRISPR-barcoding has been utilized for cellular recording, especially lineage tracing, by accumulating mutations such as deletions and insertions during cell division. For example, the ‘genome editing of synthetic target arrays for lineage tracing’ (GESTALT) strategy demonstrated this potential by applying CRISPR-Cas9 barcodes to fertilized zebrafish (*Danio rerio*) eggs for cumulative lineage barcoding (Figure 2b) [17]. Their lineage-informative barcodes were deciphered through DNA sequencing, allowing for the elucidation of lineage relationships based on mutation patterns. Similarly, the ‘memory by engineered mutagenesis with optical in situ readout’ (MEMOIR) system generates an irreversible collapse of a set of barcoded scratchpads by Cas9 targeted to the scratchpads during cell proliferation, enabling the recording of gene expression dynamics [49,50]. The states of these collapsed scratchpads were identified through multiplex single-molecule RNA fluorescence hybridization (smFISH) using sequential barcoding to multiplex different mRNAs by sequential hybridization [51].

To further improve CRISPR-Cas9 barcoding-based lineage tracing, combining CRISPR-Cas9 barcoding with single-cell RNA sequencing (scRNA-seq) allows for the acquisition of cellular transcriptomes and cell-type identification, facilitating robust lineage tracing of embryonic development [52–55] and tumor evolution [56]. While CRISPR-Cas9 barcoding is an effective method for cellular lineage tracing, the activity of the Cas9 nuclease can result in an off-target effect. Furthermore, scalability is restricted by the number of target arrays or barcodes, limiting its applications to early developmental processes [17].

3.2. Applications of Self-Targeting gRNA

Self-targeting CRISPR, also known as homing CRISPR, is a modified CRISPR-Cas9 system where the Cas9-gRNA complex directs its activity to the gRNA locus itself [57]. As self-targeting guide RNA (stgRNA) or homing guide RNA (hgRNA) contains a protospacer-adjacent motif (PAM) directly recognized by the Cas9 nuclease, it provides both guiding ability and target sites. When the stgRNA barcoding elements detect their target sequences to trigger mutations, the diversity of stgRNAs can be generated for barcoding and lineage tracing purposes [18]. While canonical CRISPR-Cas9 barcoding approaches capture only specific trajectories or moments due to their dependency on barcode sequences, stgRNA approaches establish a more independent barcoding system and produce substantially diverse barcodes.

The stgRNA approaches have shown potential for mapping cell development. For barcoding and recording cell lineages in mice, the Mouse for Actively Recording Cells 1 (MARC1) line carried multiple stgRNAs in its genome sequences and was crossed with Cas9 knock-in mouse [18,58]. In their offspring, the activation of stgRNAs generated diverse mutation patterns, which were passed to daughter cells with additional mutations.

This MARC1 system could construct a stable mouse line for barcoding and minimize the unwanted loss of patterns from large deletions. Additionally, self-targeting CRISPR approaches have been utilized to record biological events. For example, the ‘mammalian synthetic cellular recorders integrating biological events’ (mSCRIBE) system accumulates mutations in their stgRNA containing PAM sequences by linking the expression of stgRNA or Cas9 to specific biological events (Figure 2c) [59]. The frequency of accumulated stgRNA mutations within cell populations is correlated with the duration or magnitude of the biological signals. Moreover, beyond the relative duration of signals, the elapsed time of biological signals could also be gauged using stgRNAs that decay the intact target sequence frequency [60]. Most self-targeting CRISPR approaches have provoked deletions for their marking but face the risk of erasing existing records. Instead, terminal deoxynucleotidyl transferase (TdT) has been introduced to add new DNA sequences, thereby avoiding progressive erasure [13]. However, the increased lengths of stgRNA mediated by the insertions could also decrease editing efficiencies, limiting the scalability of the systems.

3.3. Base Editing-Based Cellular Recording

Deletions or insertions formed through the DSB repair pathway, including non-homologous end joining (NHEJ) or homologous recombination (HR) [61], may lead to cellular toxicity and the risk of overwriting new barcodes in existing recordings. Base editing, a CRISPR-based genome editing technique, differs from others by not relying on Cas9 nuclease, instead employing dead Cas9 (dCas9) or nickase Cas9 (nCAs9). Both lose the ability to cleave double-stranded DNA, reducing cellular toxicity but retaining the ability to bind target sequences guided by gRNAs. These modified Cas9 nucleases have been fused with base editors such as cytidine deaminase or adenine deaminase to modulate point mutations [62–64].

CRISPR-based base editing has facilitated the cellular recording of extracellular signals, especially effective for long-term analog recording due to its substantial storage capacity. Base editing-based recording has been demonstrated in both bacteria and mammalian cells. For example, in the ‘CRISPR-mediated analog multi-event recording apparatus’ (CAMERA) system, engineered bacteria demonstrated their recording ability in response to various stimuli, such as chemical signals, viral infections, and light exposure, by activating multiple gRNAs in response to these stimuli (Figure 2d) [65]. Simultaneously, the system was applied in mammalian cells, enabling the recording of chemical signals and Wnt signals. In CAMERA, expressed gRNAs direct a base editor composed of dCas9 and cytidine deaminase to targeted DNA sequences, facilitating C·G to T·A mutations. These mutation frequencies within populations indicate the magnitude or duration of specific signals. The sequential and temporal logics of multiple signals could also be constructed by more complex circuits [66]. In addition, dead Cas12a (dCas12a) has also been fused with a base editor for single nucleotide editing [67]. To enhance the efficiency of multiplex modulation, dCas12a was engineered through structure-guided protein engineering [68]. With an adenine base editor, it could effectively record much information in human cells [21]. With base editing-based recording approaches, analog characteristics such as the magnitude and duration of exogenous signals were reconstructed by the frequency of specific mutations at target sites within populations [21,65,66]. However, simultaneously distinguishing both remains challenging. Furthermore, most base editing approaches have focused on cellular recording at the population level. To address this, cellular recording at the single-cell level has been demonstrated through long-read sequencing of a ‘canvas’ with multiple target sites for base editing [69] or editing endogenous interspersed repeat regions for lineage tracing [70]. Additionally, recording multiple endogenous transcripts at the single-cell level could be performed by sensing transcripts with reprogrammed tracrRNAs (Rptrs) to convert the target endogenous mRNAs into gRNAs and mediate base editing to target DNA [71].

Base-editing-based approaches have also adopted other DNA-binding proteins to guide the target sequences. For example, the T7 polymerase-driven continuous editing

system demonstrated that transcriptional activities under the T7 promoter can be recorded through continuous nucleotide substitution mutations by exploiting T7 RNA polymerase (T7 RNAP) fused to cytidine deaminase (Figure 2e) [72]. The T7 promoter was integrated into genomic loci of specific genes, allowing the T7 RNAP-cytidine deaminase complex to constitutively access the T7 promoter and its downstream region, facilitating transcription and sequence editing. Furthermore, base-editing is not limited to DNA; it also enables transcriptional and temporal recording in RNA by utilizing RNA-specific adenosine deaminase with an RNA-binding domain [73,74].

3.4. Prime Editing-Based Recording Methods

Prime editing is a genome editing technique where target DNA is replaced by new genetic sequences [75,76]. It has the advantage of excluding bystander editing and Cas-independent off-target effects, which are challenges of base editing. The prime editor, comprising nCas9 fused to reverse transcriptase (RT), induces single-stranded breaks (SSBs) at specific locations directed by prime editor guide RNA (pegRNA). The pegRNA carries an editing sequence adjacent to the binding sequence as a template for reverse transcription. Specific sequences generated by RT are encoded by the prime editor, allowing for precise editing, such as DNA substitutions, insertions, and deletions, at targeted sites without requiring DSBs or donor DNA templates

Prime editing has been employed for robust temporally resolved cellular recording by producing sequential arrays with incorporated barcodes in the edited sequences. Individual pegRNAs with unique barcodes are inserted sequentially into specific genomic loci [20,77]. The 'prime editing cell history recording by ordered insertion' (peCHYRON) inserts 20 bp sequences, consisting of 3 bp signature mutations as the barcode and 17 bp constant propagator sequences adjacent to the PAM site [77]. With each cycle of insertion, the previous binding sequences are inactivated by being moved away from the PAM site. Another prime editing-based approach, DNA Typewriter, accomplished sequential recording by inserting short key sequences and barcodes into a tandem array of monomers containing the PAM sequence, subsequently shifting the position of the type of guide sequence (Figure 2f) [20,78]. Exploiting this sequential barcoding in the array mediated the encoding and decoding of short text messages within cell populations, collecting diverse encoded single cells. Within single cells, 3 bp barcodes were assigned to characters among alphabets, numbers, and symbols, and the barcode position in the tandem array encoded the order in sets of four characters.

For further multiplex recording, the 'enhancer-derived genomic recording of transcriptional activity in multiplex' (ENGRAM) integrates multiple signals and enhancer-specific barcodes into pegRNA [79]. This allowed for the scalable insertion of specific barcodes, capturing multiple transcriptional activities simultaneously. Despite their multiplexing and order dependency, the low efficiency of prime editing-based recording remains challenging. To improve the efficiency and precision of prime editing, engineered prime editors have been developed. For example, pegRNAs were modified to include structured 3' motif sequences that enhance RNA stability and prevent degradation, thereby increasing prime editing efficiencies [80]. Additionally, engineered RT and Cas9 nuclease were developed through phage-assisted evolution to further enhance prime editing efficiency [81]. These advancements in the prime editing approach can facilitate the incorporation of barcodes for rare events, thereby enhancing the reliability and accuracy of temporal recording.

4. CRISPR Adaptation for Temporal Recording

The CRISPR-Cas system functions as an adaptive immune response in prokaryotes, encompassing three main stages: adaptation or acquisition, expression and maturation, and interference. The CRISPR adaptation process involves recognizing foreign DNA sequences and integrating them into the CRISPR array to establish a genetic memory of viral infections. These CRISPR arrays consist of a leader sequence, short repeat sequences, and spacers derived from foreign DNA. These arrays are transcribed into CRISPR RNA (crRNA) and

subsequently processed to facilitate interference activity. The CRISPR integrases and Cas1–Cas2 complex incorporate DNA sequences, typically ranging from 30 to 40 bp, as new spacer sequences into the CRISPR array [82,83]. The new spacer sequences are integrated at the leader end of the CRISPR array, positioning the newest spacer ahead of older spacers [84].

Unidirectional CRISPR adaptation has facilitated the temporal recording of cellular events. Arbitrary DNA sequences of a specific size can be acquired as spacers in the CRISPR arrays by expressing CRISPR integrases Cas1 and Cas2 [85,86]. Recently, methods for capturing biological events have been developed by integrating intracellular DNA sequences. For example, the ‘temporal recording in arrays by CRISPR expansion’ (TRACE) system records temporal environmental signals into the CRISPR arrays by utilizing a copy number-inducible trigger plasmid (pTrig), which contains the phage P1 lytic replication initiation protein coding gene downstream of an inducible promoter (Figure 2g) [87]. In response to environmental signals, the increase in pTrig copy number led to a higher frequency of trigger DNA acquisition in the CRISPR array compared to reference sequences such as genomic and plasmid DNA. Furthermore, in the TRACE system, multiplex recording of three environmental signals was demonstrated by using a three-barcoded sensor population. This further enabled the encoding of arbitrary digital data in the CRISPR array by electronic stimulation of the trigger plasmid, maintaining robust long-term records in living cells [19].

The complex of RT and Cas1–Cas2 has been employed to record transcriptional events through CRISPR adaptation. The Record-seq strategy showed transcriptome-scale molecular recording by leveraging RT-Cas1 and Cas2 to directly capture transcripts into the CRISPR array (Figure 2h) [88]. As the acquisition frequencies of spacers depend on the source RNA abundance, highly expressed genes were captured more frequently in the CRISPR arrays. To detect rarely acquired spacers, the ‘selective amplification of expanded CRISPR arrays’ (SENECA) method was developed to specifically amplify the acquired spacers for deep sequencing [89]. Record-seq demonstrated its ability to noninvasively assess cellular transcriptional events in the intestines of mice under different dietary or environmental conditions [14]. More recently, the Retro-Cascorder system utilized retrons, previously mentioned in the SCRIBE system, to reverse transcribe engineered ncRNA barcodes into ssDNA. Then, two generated ssDNA hybridized to form duplex DNA for CRISPR acquisition [90,91]. The expression of distinct barcoded ncRNA under different inducible promoters enabled CRISPR acquisition of different duplex sequences, mediating multiplex temporal recording.

CRISPR adaptation-based approaches are powerful for temporal information recording; however, their recording efficiencies and applicable host range remain constrained. Enhancing CRISPR adaptation efficiency by utilizing internal nucleases or evolved CRISPR integrases holds promise for expanding the recording capacity and applicability of these systems, making them more versatile and effective across diverse biological contexts. For example, Cas4 nucleases or endonucleases such as DnaQ and ExoT inherently control the size and orientation of integrated spacers via asymmetric trimming [92–94]. These nucleases coordinate with CRISPR integrases, facilitating efficient CRISPR adaptation. Furthermore, evolving CRISPR integrases through directed evolution and enriching the mutant integrases by perpetual DNA packaging and transduction (PeDPaT) offer the potential to improve CRISPR-adaptation-based recording [95,96].

5. Using DNA Methylation for Biological Recording

DNA methylation is a major epigenetic process characterized by the addition of a methyl group to nucleic acid bases, such as cytosine and adenine, without altering the original sequences. This reversible modification mediates the regulation of gene expression in development and disease [97,98]. DNA methyltransferases also play a role in the prokaryotic defense system associated with the restriction-modification (RM) system [99,100]. Three prevalent methylation patterns, including 5-methylcytosine (5mC), N4-methylcytosine (4mC), and N6-methyladenine (6mA), are controlled by their catalytic writer, reader, and

eraser enzymes. Recent advances in DNA methylome mapping technologies have enabled the analysis of these methylation profiles [101–103].

Synthetic epigenetic circuits, especially those involving targeted DNA methylation, have regulated specific gene expression levels and durably retained cellular epigenetic memory [104–107]. While CRISPRa and CRISPRi methods transiently manipulate gene function, targeted DNA methylation can provide long-term regulations. For example, an engineered bacterial 6mA regulatory system could be utilized to record biological events and control transcriptional events in mammalian cells, since 6mA modification is not common in eukaryotes [108]. In response to environmental signals, a 6mA writer, a fusion of an engineered Dam methylase, and an engineered zinc finger for DNA binding mediated targeted methylation at GATC motifs to construct epigenetic memory, recording the presence of environmental signals [109].

Genome-wide transcriptome recording could also be demonstrated using the DCM-time machine (DCM-TM) system through epigenome editing (Figure 2i) [110]. This system analyzed methylation patterns by methylated DNA sequencing (MeD-seq) based on Lp-nPI digestion of DCM methylated position [111]. An inducible fusion protein of DCM methyltransferase and the RNA polymerase 2 subunit b labeled methylation patterns on transcribed genes and active enhancers when the gene was transcribed by RNA polymerase. This strategy was utilized to understand the genetic activity and temporal dynamics of intestinal stem cells (ISCs) during their differentiation into enterocytes. DNA methylation-based approaches could further increase their utility by using methyltransferase and demethylase for reversible epigenetic modification, as demonstrated in the CRISPRoff and CRISPRon systems [112].

While methylation-based recording approaches offer extensive scalability for recording transcriptomes by using the whole genome sequence as a recording site, they still have certain limitations. Notably, methylation-based techniques for recording the temporal order of various signals and analog characteristics have not been demonstrated. Additionally, the requirement for specific recognition sites for each methyltransferase may limit their applications.

6. Outlook and Discussion

DNA-based cellular recordings using DNA recombination, CRISPR systems, and DNA methylation have enabled the generation of permanent memories of environmental and biological events in living cells. In this review, we examined various DNA-based cellular recording systems, focusing on their principles, advantages, and limitations (Table 1). Unlike existing reviews on cellular recording [3,113–116], we covered the most recent cellular recording techniques, such as prime editing-based multiplexed temporal recording systems. Additionally, we introduced methylation-based cellular recording strategies alongside the commonly discussed recombinase, CRISPR nuclease, and CRISPR integrase systems.

Molecular recording of cellular events can be applied to diagnosing cellular states, capturing HGT events, tracking cell lineage, storing digital data in DNA, and developing cellular therapeutics. Selecting an appropriate recording system will be necessary for specific applications because each strategy has different advantages and scalability. For instance, the CRISPR-Cas spacer acquisition strategy possesses a distinctive ability to record horizontal gene transfer (HGT) across a cell population by directly capturing mobile DNA from complex environments [15]. When combined with genetic logic computation or sophisticated computational algorithms, DNA-based cellular recording approaches have the potential to mediate the control of cellular functions based on cellular memory [117] and to reconstruct cellular lineages [118].

Table 1. Summary of major DNA-based cellular recording systems.

System	Approach	Information Type	Sensitivity (Timescales of Cellular Recording)	Scalability	Durability	Temporal Information	Storage Place	Features	Citation
RSM	Recombination of DNA register	Chemical	Hour scale (Fast)	Medium	Short	Yes	Plasmid	Applied to build state-dependent gene regulation programs	[28]
SCRIBE	Recombination of retron RT-DNA into genomic DNA	Chemical, Light	Day scale (Slow)	Medium	Long	No	Genome	Encoding of analog memory, reversible system	[29]
GESTALT	CRISPR-Cas9 targeted to synthetic target arrays	Cell differentiation	Day scale (Slow)	Low	Long	Yes	Genome	Mapping cell lineage information, vulnerable to off-target effect	[17]
mSCRIBE	CRISPR-Cas9 and self-targeting guide RNA targeted itself	Inflammation, Chemical	Day scale (Slow)	Medium	Long	Yes	Genome	Encoding of analog memory, vulnerable to off-target effect	[59]
CAMERA	Cas9 nuclease or base editor targeted to recording site	Chemical, Phage infection, Light, Cellular state	Hour scale (Fast)	Medium	Long	Yes	Plasmid, Genome	Encoding of analog memory, reversible system, universal system between bacteria and mammalian cells	[65]
HyperCas12a base editor system	Cas12a base editor targeted to recording circuit	Chemical, Cellular state	Hour scale (Fast)	Medium	Long	No	Plasmid, Genome	Encoding of analog memory, applied to sense-and-respond circuits	[21]
T7 polymerase-driven base editing	Base editor fused to T7 RNA polymerase targeted to T7 promoter-controlled gene sequence	Transcription	Hour scale (Fast)	Low	Long	No	Plasmid, Genome	Accompanied continuous mutagenesis to target site	[72]
DNA Typewriter	Sequential prime editing of target sequences	Transfection, Cell differentiation	Day scale (Slow)	High	Long	Yes	Genome	Applied to record complex event histories and short digital data	[20]

Table 1. Cont.

System	Approach	Information Type	Sensitivity (Timescales of Cellular Recording)	Scalability	Durability	Temporal Information	Storage Place	Features	Citation
ENGRAM	Prime editor programmed to insert CRE-specific barcode sequence	Enhancer activity, Cellular state	Day scale (Slow)	High	Long	Yes	Genome	Could be coupled with DNA Typewriter system for temporal recording	[79]
TRACE	CRISPR adaptation of copy inducible plasmid into CRISPR array	Chemical	Hour scale (Fast)	Medium	Long	Yes	Genome	Applied to record temporal biological/digital data	[87]
Record-seq	RT-Cas1 and Cas2-based acquisition of RNA transcripts into CRISPR array	Transcription	Hour scale (Fast)	High	Long	Yes	Genome	Genome-wide transcriptional information	[88]
Retro-Cascorder	CRISPR adaptation of retron RT-DNA into CRISPR array	Chemical	Hour scale (Fast)	Medium	Long	Yes	Genome	Identification of molecular events and their orders in individual cells	[90]
DCM-TM	Methyltransferase DCM fused to RNA polymerase targeted to transcript	Transcription, Enhancer activity	Day scale (Slow)	High	Long	No	Genome	Applied to track cellular state in mouse intestine	[110]

We anticipate that improving DNA-based cellular recorders by enhancing their sensitivity, scalability, and durability will be key to utilizing molecular recording across various applications. While the sensitivity of most molecular recorders is limited to an hour or day scale, it is important to address stimuli that occur on a second or minute scale for responding to instant signals. Developing methods to increase the sensitivity for cellular recording at such high temporal resolution will provide real-time monitoring capabilities, which are essential for applications such as detecting rapid changes in cellular states or environmental conditions. A recent study with minute resolution demonstrated the potential for highly sensitive encoding of environmental signals [119]. Engineered TdT transduced these signals by incorporating specific nucleotides in response to cation concentrations, such as Co^{2+} , Ca^{2+} , and Zn^{2+} , and temperature changes within 1 min in vitro. Similar to this engineered TdT, exploiting highly sensitive enzymes could improve the resolution of cellular recording.

Furthermore, expanding the capability for multiplexing and temporal recording will be necessary as these characteristics tend to be inversely proportional. Encoding temporal information of two or three environmental signals is relatively straightforward; however, managing temporal transcriptional recording on a genome-wide scale is still challenging and requires significant experimental and computational advancements. Expanded scalability to support long-term genome-wide transcriptional recording with high temporal resolution will allow for comprehensive monitoring of complex biological processes and interactions over time. A potential approach to enhance these capabilities is to combine the multiplexed and quantitative recording capacity of ENGRAM with the sequential recording capacity of DNA Typewriter [20,79]. The pegRNAs linked to signal-responsive cis-regulatory elements (CREs) that are targeted to a tandem array of partial target sites could potentially mediate unidirectional insertions of barcodes for temporal recording by shifting the editable positions.

Finally, enhancing the durability and robustness of recorded data by minimizing off-target effects and ensuring long-term stability will provide reliable data for extended studies and applications, such as longitudinal tracking of cellular changes. The off-target effects can occur not only in approaches using CRISPR-Cas nucleases but also in those using CRISPR adaptation machineries [120]. To overcome these challenges, better control of CRISPR off-target effects and the integration of robust memory maintenance mechanisms will be essential. To minimize off-target effects, engineered sgRNAs can increase the specificity of CRISPR activity by varying the hairpin structure of sgRNAs [121] or by delivering off-target-directed short gRNA while maintaining on-target efficiencies [122]. Furthermore, engineered enzymes can reduce the risks of off-target effects [123]. We expect this emerging DNA-based modality of biological measurement will be key to gaining novel insights into diverse biological processes and fostering the development of various biotechnological applications, from environmental sensing to disease monitoring and beyond.

Author Contributions: Conceptualization, H.J. and S.S.Y.; writing, H.J. and S.S.Y.; funding acquisition, S.S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) funded by the Korea government (MSIT) (RS-2024-00358175, RS-2024-00399424) and KAIST (G04220037, N10230105, N11230043, N10240028, N10240039).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zheng, D.; Liwinski, T.; Elinav, E. Interaction between microbiota and immunity in health and disease. *Cell Res.* **2020**, *30*, 492–506. [[CrossRef](#)]
2. Joung, J.; Ma, S.; Tay, T.; Geiger-Schuller, K.R.; Kirchgatterer, P.C.; Verdine, V.K.; Guo, B.; Arias-Garcia, M.A.; Allen, W.E.; Singh, A.; et al. A transcription factor atlas of directed differentiation. *Cell* **2023**, *186*, 209–229.e26. [[CrossRef](#)] [[PubMed](#)]
3. Sheth, R.U.; Wang, H.H. DNA-based memory devices for recording cellular events. *Nat. Rev. Genet.* **2018**, *19*, 718–732. [[CrossRef](#)]

4. Schofield, J.A.; Duffy, E.E.; Kiefer, L.; Sullivan, M.C.; Simon, M.D. TimeLapse-seq: Adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat. Methods* **2018**, *15*, 221–225. [[CrossRef](#)]
5. La Manno, G.; Soldatov, R.; Zeisel, A.; Braun, E.; Hochgerner, H.; Petukhov, V.; Lidschreiber, K.; Kastrioti, M.E.; Lönnerberg, P.; Furlan, A.; et al. RNA velocity of single cells. *Nature* **2018**, *560*, 494–498. [[CrossRef](#)] [[PubMed](#)]
6. Chen, W.; Guillaume-Gentil, O.; Rainer, P.Y.; Gäbelein, C.G.; Saelens, W.; Gardeux, V.; Klaeger, A.; Dainese, R.; Zachara, M.; Zambelli, T.; et al. Live-seq enables temporal transcriptomic recording of single cells. *Nature* **2022**, *608*, 733–740. [[CrossRef](#)] [[PubMed](#)]
7. Gootenberg, J.S.; Abudayyeh, O.O.; Lee, J.W.; Essletzbichler, P.; Dy, A.J.; Joung, J.; Verdine, V.; Donghia, N.; Daringer, N.M.; Freije, C.A.; et al. Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **2017**, *356*, 438–442. [[CrossRef](#)]
8. Kim, J.; Lee, S.; Jung, K.; Oh, W.C.; Kim, N.; Son, S.; Jo, Y.; Kwon, H.-B.; Heo, W.D. Intensiometric biosensors visualize the activity of multiple small GTPases in vivo. *Nat. Commun.* **2019**, *10*, 211. [[CrossRef](#)]
9. Kaczmarczyk, A.; van Vliet, S.; Jakob, R.P.; Teixeira, R.D.; Scheidat, I.; Reinders, A.; Klotz, A.; Maier, T.; Jenal, U. A genetically encoded biosensor to monitor dynamic changes of c-di-GMP with high temporal resolution. *Nat. Commun.* **2024**, *15*, 3920. [[CrossRef](#)]
10. Matange, K.; Tuck, J.M.; Keung, A.J. DNA stability: A central design consideration for DNA data storage systems. *Nat. Commun.* **2021**, *12*, 1358. [[CrossRef](#)]
11. Doricchi, A.; Platnich, C.M.; Gimpel, A.; Horn, F.; Earle, M.; Lanzavecchia, G.; Cortajarena, A.L.; Liz-Marzán, L.M.; Liu, N.; Heckel, R.; et al. Emerging Approaches to DNA Data Storage: Challenges and Prospects. *ACS Nano* **2022**, *16*, 17552–17571. [[CrossRef](#)]
12. Hu, T.; Chitnis, N.; Monos, D.; Dinh, A. Next-generation sequencing technologies: An overview. *Hum. Immunol.* **2021**, *82*, 801–811. [[CrossRef](#)] [[PubMed](#)]
13. Loveless, T.B.; Grotts, J.H.; Schechter, M.W.; Forouzmand, E.; Carlson, C.K.; Agahi, B.S.; Liang, G.; Ficht, M.; Liu, B.; Xie, X.; et al. Lineage tracing and analog recording in mammalian cells by single-site DNA writing. *Nat. Chem. Biol.* **2021**, *17*, 739–747. [[CrossRef](#)] [[PubMed](#)]
14. Schmidt, F.; Zimmermann, J.; Tanna, T.; Farouni, R.; Conway, T.; Macpherson, A.J.; Platt, R.J. Noninvasive assessment of gut function using transcriptional recording sentinel cells. *Science* **2022**, *376*, eabm6038. [[CrossRef](#)] [[PubMed](#)]
15. Munck, C.; Sheth, R.U.; Freedberg, D.E.; Wang, H.H. Recording mobile DNA in the gut microbiota using an Escherichia coli CRISPR-Cas spacer acquisition platform. *Nat. Commun.* **2020**, *11*, 95. [[CrossRef](#)]
16. Farzadfard, F.; Gharaei, N.; Citorik, R.J.; Lu, T.K. Efficient retroelement-mediated DNA writing in bacteria. *Cell Syst.* **2021**, *12*, 860–872.e5. [[CrossRef](#)]
17. McKenna, A.; Findlay, G.M.; Gagnon, J.A.; Horwitz, M.S.; Schier, A.F.; Shendure, J. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **2016**, *353*, aaf7907. [[CrossRef](#)]
18. Kalhor, R.; Kalhor, K.; Mejia, L.; Leeper, K.; Graveline, A.; Mali, P.; Church, G.M. Developmental barcoding of whole mouse via homing CRISPR. *Science* **2018**, *361*, eaat9804. [[CrossRef](#)]
19. Yim, S.S.; McBee, R.M.; Song, A.M.; Huang, Y.; Sheth, R.U.; Wang, H.H. Robust direct digital-to-biological data storage in living cells. *Nat. Chem. Biol.* **2021**, *17*, 246–253. [[CrossRef](#)]
20. Choi, J.; Chen, W.; Minkina, A.; Chardon, F.M.; Suiter, C.C.; Regalado, S.G.; Domcke, S.; Hamazaki, N.; Lee, C.; Martin, B.; et al. A time-resolved, multi-symbol molecular recorder via sequential genome editing. *Nature* **2022**, *608*, 98–107. [[CrossRef](#)]
21. Kempton, H.R.; Love, K.S.; Guo, L.Y.; Qi, L.S. Scalable biological signal recording in mammalian cells using Cas12a base editors. *Nat. Chem. Biol.* **2022**, *18*, 742–750. [[CrossRef](#)] [[PubMed](#)]
22. Siuti, P.; Yazbek, J.; Lu, T.K. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **2013**, *31*, 448–452. [[CrossRef](#)]
23. Yang, L.; Nielsen, A.A.; Fernandez-Rodriguez, J.; McClune, C.J.; Laub, M.T.; Lu, T.K.; Voigt, C.A. Permanent genetic memory with >1-byte capacity. *Nat. Methods* **2014**, *11*, 1261–1266. [[CrossRef](#)] [[PubMed](#)]
24. Courbet, A.; Endy, D.; Renard, E.; Molina, F.; Bonnet, J. Detection of pathological biomarkers in human clinical samples via amplifying genetic switches and logic gates. *Sci. Transl. Med.* **2015**, *7*, 289ra83. [[CrossRef](#)]
25. Chiu, T.-Y.; Jiang, J.-H.R. Logic Synthesis of Recombinase-Based Genetic Circuits. *Sci. Rep.* **2017**, *7*, 12873. [[CrossRef](#)] [[PubMed](#)]
26. Kim, T.; Weinberg, B.; Wong, W.; Lu, T.K. Scalable recombinase-based gene expression cascades. *Nat. Commun.* **2021**, *12*, 2711. [[CrossRef](#)]
27. Huang, B.D.; Kim, D.; Yu, Y.; Wilson, C.J. Engineering intelligent chassis cells via recombinase-based MEMORY circuits. *Nat. Commun.* **2024**, *15*, 2418. [[CrossRef](#)] [[PubMed](#)]
28. Roquet, N.; Soleimany, A.P.; Ferris, A.C.; Aaronson, S.; Lu, T.K. Synthetic recombinase-based state machines in living cells. *Science* **2016**, *353*, aad8559. [[CrossRef](#)] [[PubMed](#)]
29. Farzadfard, F.; Lu, T.K. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* **2014**, *346*, 1256272. [[CrossRef](#)]
30. Millman, A.; Bernheim, A.; Stokar-Avihail, A.; Fedorenko, T.; Voichek, M.; Leavitt, A.; Oppenheimer-Shaanan, Y.; Sorek, R. Bacterial Retrons Function in Anti-Phage Defense. *Cell* **2020**, *183*, 1551–1561.e12. [[CrossRef](#)] [[PubMed](#)]

31. Schubert, M.G.; Goodman, D.B.; Wannier, T.M.; Kaur, D.; Farzadfard, F.; Lu, T.K.; Shipman, S.L.; Church, G.M. High-throughput functional variant screens via in vivo production of single-stranded DNA. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2018181118. [[CrossRef](#)] [[PubMed](#)]
32. Lopez, S.C.; Crawford, K.D.; Lear, S.K.; Bhattarai-Kline, S.; Shipman, S.L. Precise genome editing across kingdoms of life using retron-derived DNA. *Nat. Chem. Biol.* **2022**, *18*, 199–206. [[CrossRef](#)] [[PubMed](#)]
33. Liu, W.; Zuo, S.; Shao, Y.; Bi, K.; Zhao, J.; Huang, L.; Xu, Z.; Lian, J. Retron-mediated multiplex genome editing and continuous evolution in *Escherichia coli*. *Nucleic Acids Res.* **2023**, *51*, 8293–8307. [[CrossRef](#)]
34. Weinberg, B.H.; Pham, N.T.H.; Caraballo, L.D.; Lozanoski, T.; Engel, A.; Bhatia, S.; Wong, W.W. Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat. Biotechnol.* **2017**, *35*, 453–462. [[CrossRef](#)]
35. Guiziou, S.; Maranas, C.J.; Chu, J.C.; Nemhauser, J.L. An integrase toolbox to record gene-expression during plant development. *Nat. Commun.* **2023**, *14*, 1844. [[CrossRef](#)]
36. Kalvapalle, P.B.; Sridhar, S.; Silberg, J.J.; Stadler, L.B. Long-duration environmental biosensing by recording analyte detection in DNA using recombinase memory. *Appl. Environ. Microbiol.* **2024**, *90*, e02363-23. [[CrossRef](#)] [[PubMed](#)]
37. Durrant, M.G.; Fanton, A.; Tycko, J.; Hinks, M.; Chandrasekaran, S.S.; Perry, N.T.; Schaepe, J.; Du, P.P.; Lotfy, P.; Bassik, M.C.; et al. Systematic discovery of recombinases for efficient integration of large DNA sequences into the human genome. *Nat. Biotechnol.* **2023**, *41*, 488–499. [[CrossRef](#)] [[PubMed](#)]
38. Short, A.E.; Kim, D.; Milner, P.T.; Wilson, C.J. Next generation synthetic memory via intercepting recombinase function. *Nat. Commun.* **2023**, *14*, 5255. [[CrossRef](#)]
39. Urnov, F.D.; Rebar, E.J.; Holmes, M.C.; Zhang, H.S.; Gregory, P.D. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **2010**, *11*, 636–646. [[CrossRef](#)]
40. Joung, J.K.; Sander, J.D. TALENs: A widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 49–55. [[CrossRef](#)] [[PubMed](#)]
41. Li, H.; Yang, Y.; Hong, W.; Huang, M.; Wu, M.; Zhao, X. Applications of genome editing technology in the targeted therapy of human diseases: Mechanisms, advances and prospects. *Signal Transduct. Target. Ther.* **2020**, *5*, 1. [[CrossRef](#)] [[PubMed](#)]
42. Doudna, J.A. The promise and challenge of therapeutic genome editing. *Nature* **2020**, *578*, 229–236. [[CrossRef](#)]
43. Anzalone, A.V.; Koblan, L.W.; Liu, D.R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **2020**, *38*, 824–844. [[CrossRef](#)]
44. Cong, L.; Ran, F.A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P.D.; Wu, X.; Jiang, W.; Marraffini, L.A.; et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **2013**, *339*, 819–823. [[CrossRef](#)]
45. Ran, F.A.; Hsu, P.D.; Wright, J.; Agarwala, V.; Scott, D.A.; Zhang, F. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **2013**, *8*, 2281–2308. [[CrossRef](#)] [[PubMed](#)]
46. Sander, J.D.; Joung, J.K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **2014**, *32*, 347–355. [[CrossRef](#)] [[PubMed](#)]
47. Wang, J.Y.; Doudna, J.A. CRISPR technology: A decade of genome editing is only the beginning. *Science* **2023**, *379*, eadd8643. [[CrossRef](#)]
48. Celli, L.; Gasparini, P.; Biino, G.; Zannini, L.; Cardano, M. CRISPR/Cas9 mediated Y-chromosome elimination affects human cells transcriptome. *Cell Biosci.* **2024**, *14*, 15. [[CrossRef](#)]
49. Frieda, K.L.; Linton, J.M.; Hormoz, S.; Choi, J.; Chow, K.-H.K.; Singer, Z.S.; Budde, M.W.; Elowitz, M.B.; Cai, L. Synthetic recording and in situ readout of lineage information in single cells. *Nature* **2017**, *541*, 107–111. [[CrossRef](#)]
50. Wang, Z.; Zhu, J. MEMOIR: A Novel System for Neural Lineage Tracing. *Neurosci. Bull.* **2017**, *33*, 763–765. [[CrossRef](#)]
51. Lubeck, E.; Coskun, A.F.; Zhiyentayev, T.; Ahmad, M.; Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **2014**, *11*, 360–361.
52. Alemany, A.; Florescu, M.; Baron, C.S.; Peterson-Maduro, J.; van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. *Nature* **2018**, *556*, 108–112. [[PubMed](#)]
53. Raj, B.; Wagner, D.E.; McKenna, A.; Pandey, S.; Klein, A.M.; Shendure, J.; Gagnon, J.A.; Schier, A.F. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **2018**, *36*, 442–450. [[CrossRef](#)] [[PubMed](#)]
54. Chan, M.M.; Smith, Z.D.; Grosswendt, S.; Kretzmer, H.; Norman, T.M.; Adamson, B.; Jost, M.; Quinn, J.J.; Yang, D.; Jones, M.G.; et al. Molecular recording of mammalian embryogenesis. *Nature* **2019**, *570*, 77–82. [[CrossRef](#)] [[PubMed](#)]
55. Wagner, D.E.; Klein, A.M. Lineage tracing meets single-cell omics: Opportunities and challenges. *Nat. Rev. Genet.* **2020**, *21*, 410–427. [[CrossRef](#)]
56. Yang, D.; Jones, M.G.; Naranjo, S.; Rideout, W.M.; Min, K.H.; Ho, R.; Wu, W.; Replogle, J.M.; Page, J.L.; Quinn, J.J.; et al. Lineage tracing reveals the phylogenetics, plasticity, and paths of tumor evolution. *Cell* **2022**, *185*, 1905–1923.e25.
57. Kalhor, R.; Mali, P.; Church, G.M. Rapidly evolving homing CRISPR barcodes. *Nat. Methods* **2017**, *14*, 195–200. [[CrossRef](#)]
58. Leeper, K.; Kalhor, K.; Vernet, A.; Graveline, A.; Church, G.M.; Mali, P.; Kalhor, R. Lineage barcoding in mice with homing CRISPR. *Nat. Protoc.* **2021**, *16*, 2088–2108.
59. Perli, S.D.; Cui, C.H.; Lu, T.K. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* **2016**, *353*, aag0511. [[CrossRef](#)]
60. Park, J.; Lim, J.M.; Jung, I.; Heo, S.J.; Park, J.; Chang, Y.; Kim, H.K.; Jung, D.; Yu, J.H.; Min, S.; et al. Recording of elapsed time and temporal information about biological events using Cas9. *Cell* **2021**, *184*, 1047–1063.e23. [[CrossRef](#)]

61. Xue, C.; Greene, E.C. DNA Repair Pathway Choices in CRISPR-Cas9-Mediated Genome Editing. *Trends Genet.* **2021**, *37*, 639–656. [[CrossRef](#)]
62. Komor, A.C.; Kim, Y.B.; Packer, M.S.; Zuris, J.A.; Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **2016**, *533*, 420–424. [[CrossRef](#)] [[PubMed](#)]
63. Nishida, K.; Arazoe, T.; Yachie, N.; Banno, S.; Kakimoto, M.; Tabata, M.; Mochizuki, M.; Miyabe, A.; Araki, M.; Hara, K.Y.; et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **2016**, *353*, aaf8729. [[CrossRef](#)]
64. Gaudelli, N.M.; Komor, A.C.; Rees, H.A.; Packer, M.S.; Badran, A.H.; Bryson, D.I.; Liu, D.R. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **2017**, *551*, 464–471. [[CrossRef](#)]
65. Tang, W.; Liu, D.R. Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* **2018**, *360*, eaap8992. [[CrossRef](#)] [[PubMed](#)]
66. Farzadfard, F.; Gharaei, N.; Higashikuni, Y.; Jung, G.; Cao, J.; Lu, T.K. Single-Nucleotide-Resolution Computing and Memory in Living Cells. *Mol. Cell* **2019**, *75*, 769–780.e4. [[CrossRef](#)] [[PubMed](#)]
67. Li, X.; Wang, Y.; Liu, Y.; Yang, B.; Wang, X.; Wei, J.; Lu, Z.; Zhang, Y.; Wu, J.; Huang, X.; et al. Base editing with a Cpf1–cytidine deaminase fusion. *Nat. Biotechnol.* **2018**, *36*, 324–327. [[CrossRef](#)]
68. Guo, L.Y.; Bian, J.; Davis, A.E.; Liu, P.; Kempton, H.R.; Zhang, X.; Chemparathy, A.; Gu, B.; Lin, X.; Rane, D.A.; et al. Multiplexed genome regulation in vivo with hyper-efficient Cas12a. *Nat. Cell Biol.* **2022**, *24*, 590–600. [[CrossRef](#)]
69. Tu, B.; Sundar, V.; Esvelt, K.M. An ultra-high-throughput method for measuring biomolecular activities. *bioRxiv* 2024. [[CrossRef](#)]
70. Hwang, B.; Lee, W.; Yum, S.-Y.; Jeon, Y.; Cho, N.; Jang, G.; Bang, D. Lineage tracing using a Cas9-deaminase barcoding system targeting endogenous L1 elements. *Nat. Commun.* **2019**, *10*, 1234. [[CrossRef](#)]
71. Jiao, C.; Reckstadt, C.; König, F.; Homberger, C.; Yu, J.; Vogel, J.; Westermann, A.J.; Sharma, C.M.; Beisel, C.L. RNA recording in single bacterial cells using reprogrammed tracrRNAs. *Nat. Biotechnol.* **2023**, *41*, 1107–1116. [[CrossRef](#)]
72. Chen, H.; Liu, S.; Padula, S.; Lesman, D.; Griswold, K.; Lin, A.; Zhao, T.; Marshall, J.L.; Chen, F. Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor. *Nat. Biotechnol.* **2020**, *38*, 165–168. [[CrossRef](#)]
73. Rodrigues, S.G.; Chen, L.M.; Liu, S.; Zhong, E.D.; Scherrer, J.R.; Boyden, E.S.; Chen, F. RNA timestamps identify the age of single molecules in RNA sequencing. *Nat. Biotechnol.* **2021**, *39*, 320–325. [[CrossRef](#)]
74. Lin, Y.; Kwok, S.; Hein, A.E.; Thai, B.Q.; Alabi, Y.; Ostrowski, M.S.; Wu, K.; Floor, S.N. RNA molecular recording with an engineered RNA deaminase. *Nat. Methods* **2023**, *20*, 1887–1899. [[CrossRef](#)]
75. Anzalone, A.V.; Randolph, P.B.; Davis, J.R.; Sousa, A.A.; Koblan, L.W.; Levy, J.M.; Chen, P.J.; Wilson, C.; Newby, G.A.; Raguram, A.; et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **2019**, *576*, 149–157. [[CrossRef](#)]
76. Chen, P.J.; Liu, D.R. Prime editing for precise and highly versatile genome manipulation. *Nat. Rev. Genet.* **2023**, *24*, 161–177. [[CrossRef](#)] [[PubMed](#)]
77. Loveless, T.B.; Carlson, C.K.; Dentzel Helmy, C.A.; Hu, V.J.; Ross, S.K.; Demelo, M.C.; Murtaza, A.; Liang, G.; Ficht, M.; Singhai, A.; et al. Open-ended molecular recording of sequential cellular events into DNA. *bioRxiv* **2024**. [[CrossRef](#)]
78. Liao, H.; Choi, J.; Shendure, J. Molecular recording using DNA Typewriter. *Nat. Protoc.* **2024**. [[CrossRef](#)] [[PubMed](#)]
79. Chen, W.; Choi, J.; Li, X.; Nathans, J.F.; Martin, B.; Yang, W.; Hamazaki, N.; Qiu, C.; Lalanne, J.-B.; Regalado, S.; et al. Symbolic recording of signalling and cis-regulatory element activity to DNA. *Nature* **2024**. [[CrossRef](#)] [[PubMed](#)]
80. Nelson, J.W.; Randolph, P.B.; Shen, S.P.; Everette, K.A.; Chen, P.J.; Anzalone, A.V.; An, M.; Newby, G.A.; Chen, J.C.; Hsu, A.; et al. Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **2022**, *40*, 402–410. [[CrossRef](#)]
81. Doman, J.L.; Pandey, S.; Neugebauer, M.E.; An, M.; Davis, J.R.; Randolph, P.B.; McElroy, A.; Gao, X.D.; Raguram, A.; Richter, M.F.; et al. Phage-assisted evolution and protein engineering yield compact, efficient prime editors. *Cell* **2023**, *186*, 3983–4002.e26. [[CrossRef](#)] [[PubMed](#)]
82. Nuñez, J.K.; Lee, A.S.Y.; Engelman, A.; Doudna, J.A. Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature* **2015**, *519*, 193–198. [[CrossRef](#)]
83. Amitai, G.; Sorek, R. CRISPR–Cas adaptation: Insights into the mechanism of action. *Nat. Rev. Microbiol.* **2016**, *14*, 67–76. [[CrossRef](#)] [[PubMed](#)]
84. McGinn, J.; Marraffini, L.A. Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nat. Rev. Microbiol.* **2019**, *17*, 7–12. [[CrossRef](#)]
85. Shipman, S.L.; Nivala, J.; Macklis, J.D.; Church, G.M. Molecular recordings by directed CRISPR spacer acquisition. *Science* **2016**, *353*, aaf1175. [[CrossRef](#)]
86. Shipman, S.L.; Nivala, J.; Macklis, J.D.; Church, G.M. CRISPR–Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature* **2017**, *547*, 345–349. [[CrossRef](#)]
87. Sheth, R.U.; Yim, S.S.; Wu, F.L.; Wang, H.H. Multiplex recording of cellular events over time on CRISPR biological tape. *Science* **2017**, *358*, 1457–1461. [[CrossRef](#)]
88. Schmidt, F.; Cherepkova, M.Y.; Platt, R.J. Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* **2018**, *562*, 380–385. [[CrossRef](#)] [[PubMed](#)]
89. Tanna, T.; Schmidt, F.; Cherepkova, M.Y.; Okoniewski, M.; Platt, R.J. Recording transcriptional histories using Record-seq. *Nat. Protoc.* **2020**, *15*, 513–539. [[CrossRef](#)]

90. Bhattarai-Kline, S.; Lear, S.K.; Fishman, C.B.; Lopez, S.C.; Lockshin, E.R.; Schubert, M.G.; Nivala, J.; Church, G.M.; Shipman, S.L. Recording gene expression order in DNA by CRISPR addition of retron barcodes. *Nature* **2022**, *608*, 217–225. [[CrossRef](#)]
91. Lear, S.K.; Lopez, S.C.; González-Delgado, A.; Bhattarai-Kline, S.; Shipman, S.L. Temporally resolved transcriptional recording in *E. coli* DNA using a Retro-Cascorder. *Nat. Protoc.* **2023**, *18*, 1866–1892. [[CrossRef](#)]
92. Ramachandran, A.; Summerville, L.; Learn, B.A.; DeBell, L.; Bailey, S. Processing and integration of functionally oriented prespacers in the *Escherichia coli* CRISPR system depends on bacterial host exonucleases. *J. Biol. Chem.* **2020**, *295*, 3403–3414. [[CrossRef](#)] [[PubMed](#)]
93. Wang, J.Y.; Tuck, O.T.; Skopintsev, P.; Soczek, K.M.; Li, G.; Al-Shayeb, B.; Zhou, J.; Doudna, J.A. Genome expansion by a CRISPR trimmer-integrase. *Nature* **2023**, *618*, 855–861. [[CrossRef](#)]
94. Hu, C.; Almendros, C.; Nam, K.H.; Costa, A.R.; Vink, J.N.A.; Haagsma, A.C.; Bagde, S.R.; Brouns, S.J.J.; Ke, A. Mechanism for Cas4-assisted directional spacer acquisition in CRISPR–Cas. *Nature* **2021**, *598*, 515–520. [[CrossRef](#)] [[PubMed](#)]
95. Heler, R.; Wright, A.V.; Vucelja, M.; Bikard, D.; Doudna, J.A.; Marraffini, L.A. Mutations in Cas9 Enhance the Rate of Acquisition of Viral Spacer Sequences during the CRISPR–Cas Immune Response. *Mol. Cell* **2017**, *65*, 168–175. [[CrossRef](#)] [[PubMed](#)]
96. Yosef, I.; Mahata, T.; Goren, M.G.; Degany, O.J.; Ben-Shem, A.; Qimron, U. Highly active CRISPR-adaptation proteins revealed by a robust enrichment technology. *Nucleic Acids Res.* **2023**, *51*, 7552–7562. [[CrossRef](#)] [[PubMed](#)]
97. Moore, L.D.; Le, T.; Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **2013**, *38*, 23–38. [[CrossRef](#)]
98. Greenberg, M.V.C.; Bourc’his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 590–607. [[CrossRef](#)]
99. Sánchez-Romero, M.A.; Casadesús, J. The bacterial epigenome. *Nat. Rev. Microbiol.* **2020**, *18*, 7–20. [[CrossRef](#)]
100. Seong, H.J.; Han, S.W.; Sul, W.J. Prokaryotic DNA methylation and its functional roles. *J. Microbiol.* **2021**, *59*, 242–248. [[CrossRef](#)]
101. Beaulaurier, J.; Schadt, E.E.; Fang, G. Deciphering bacterial epigenomes using modern sequencing technologies. *Nat. Rev. Genet.* **2019**, *20*, 157–172. [[CrossRef](#)] [[PubMed](#)]
102. Rauluseviute, I.; Drabløs, F.; Rye, M.B. DNA methylation data by sequencing: Experimental approaches and recommendations for tools and pipelines for data analysis. *Clin. Epigenet.* **2019**, *11*, 193. [[CrossRef](#)] [[PubMed](#)]
103. Zhou, Q.; Zhou, C.; Zhu, Z.; Sun, Y.; Li, G. DNA Methylation (DM) data format and DMtools for efficient DNA methylation data storage and analysis. *bioRxiv* **2024**. [[CrossRef](#)]
104. Maier, J.A.H.; Möhrle, R.; Jeltsch, A. Design of synthetic epigenetic circuits featuring memory effects and reversible switching based on DNA methylation. *Nat. Commun.* **2017**, *8*, 15336. [[CrossRef](#)]
105. Lei, Y.; Zhang, X.; Su, J.; Jeong, M.; Gundry, M.C.; Huang, Y.-H.; Zhou, Y.; Li, W.; Goodell, M.A. Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. *Nat. Commun.* **2017**, *8*, 16026. [[CrossRef](#)] [[PubMed](#)]
106. Van, M.V.; Fujimori, T.; Bintu, L. Nanobody-mediated control of gene expression and epigenetic memory. *Nat. Commun.* **2021**, *12*, 537. [[CrossRef](#)] [[PubMed](#)]
107. Sapozhnikov, D.M.; Szyf, M. Unraveling the functional role of DNA demethylation at specific promoters by targeted steric blockage of DNA methyltransferase with CRISPR/dCas9. *Nat. Commun.* **2021**, *12*, 5711. [[CrossRef](#)]
108. Heyn, H.; Esteller, M. An Adenine Code for DNA: A Second Life for N6-Methyladenine. *Cell* **2015**, *161*, 710–713. [[CrossRef](#)]
109. Park, M.; Patel, N.; Keung, A.J.; Khalil, A.S. Engineering Epigenetic Regulation Using Synthetic Read-Write Modules. *Cell* **2019**, *176*, 227–238.e20. [[CrossRef](#)]
110. Boers, R.; Boers, J.; Tan, B.; van Leeuwen, M.E.; Wassenaar, E.; Sanchez, E.G.; Sleddens, E.; Tenhagen, Y.; Mulugeta, E.; Laven, J.; et al. Retrospective analysis of enhancer activity and transcriptome history. *Nat. Biotechnol.* **2023**, *41*, 1582–1592. [[CrossRef](#)]
111. Boers, R.; Boers, J.; de Hoon, B.; Kockx, C.; Ozgur, Z.; Molijn, A.; van IJcken, W.; Laven, J.; Gribnau, J. Genome-wide DNA methylation profiling using the methylation-dependent restriction enzyme LpnPI. *Genome Res.* **2018**, *28*, 88–99. [[CrossRef](#)]
112. Nuñez, J.K.; Chen, J.; Pommier, G.C.; Cogan, J.Z.; Replogle, J.M.; Adriaens, C.; Ramadoss, G.N.; Shi, Q.; Hung, K.L.; Samelson, A.J.; et al. Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* **2021**, *184*, 2503–2519.e17. [[CrossRef](#)] [[PubMed](#)]
113. Schmidt, F.; Platt, R.J. Applications of CRISPR–Cas for synthetic biology and genetic recording. *Curr. Opin. Syst. Biol.* **2017**, *5*, 9–15. [[CrossRef](#)]
114. Ishiguro, S.; Mori, H.; Yachie, N. DNA event recorders send past information of cells to the time of observation. *Curr. Opin. Chem. Biol.* **2019**, *52*, 54–62. [[CrossRef](#)]
115. Masuyama, N.; Mori, H.; Yachie, N. DNA barcodes evolve for high-resolution cell lineage tracing. *Curr. Opin. Chem. Biol.* **2019**, *52*, 63–71. [[CrossRef](#)] [[PubMed](#)]
116. Lear, S.K.; Shipman, S.L. Molecular recording: Transcriptional data collection into the genome. *Curr. Opin. Biotechnol.* **2023**, *79*, 102855. [[CrossRef](#)] [[PubMed](#)]
117. Green, A.A.; Kim, J.; Ma, D.; Silver, P.A.; Collins, J.J.; Yin, P. Complex cellular logic computation using ribocomputing devices. *Nature* **2017**, *548*, 117–121. [[CrossRef](#)]
118. Konno, N.; Kijima, Y.; Watano, K.; Ishiguro, S.; Ono, K.; Tanaka, M.; Mori, H.; Masuyama, N.; Pratt, D.; Ideker, T.; et al. Deep distributed computing to reconstruct extremely large lineage trees. *Nat. Biotechnol.* **2022**, *40*, 566–575. [[CrossRef](#)] [[PubMed](#)]
119. Bhan, N.; Callisto, A.; Strutz, J.; Glaser, J.; Kalhor, R.; Boyden, E.S.; Church, G.; Kording, K.; Tyo, K.E.J. Recording Temporal Signals with Minutes Resolution Using Enzymatic DNA Synthesis. *J. Am. Chem. Soc.* **2021**, *143*, 16630–16640. [[CrossRef](#)] [[PubMed](#)]

120. Nivala, J.; Shipman, S.L.; Church, G.M. Spontaneous CRISPR loci generation in vivo by non-canonical spacer integration. *Nat. Microbiol.* **2018**, *3*, 310–318. [[CrossRef](#)]
121. Kocak, D.D.; Josephs, E.A.; Bhandarkar, V.; Adkar, S.S.; Kwon, J.B.; Gersbach, C.A. Increasing the specificity of CRISPR systems with engineered RNA secondary structures. *Nat. Biotechnol.* **2019**, *37*, 657–666. [[CrossRef](#)] [[PubMed](#)]
122. Coelho, M.A.; De Braekeleer, E.; Firth, M.; Bista, M.; Lukasiak, S.; Cuomo, M.E.; Taylor, B.J.M. CRISPR GUARD protects off-target sites from Cas9 nuclease activity using short guide RNAs. *Nat. Commun.* **2020**, *11*, 4132. [[CrossRef](#)] [[PubMed](#)]
123. Li, A.; Mitsunobu, H.; Yoshioka, S.; Suzuki, T.; Kondo, A.; Nishida, K. Cytosine base editing systems with minimized off-target effect and molecular size. *Nat. Commun.* **2022**, *13*, 4531. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.