

BMB Reports – Manuscript Submission

Manuscript Draft

DOI: [10.5483/BMBRep.2023-0050](https://doi.org/10.5483/BMBRep.2023-0050)

Manuscript Number: BMB-23-050

Title: Recording RNA by CRISPR-Cas adaptation: a brief review

Article Type: Mini Review

Keywords: CRISPR-Cas; CRISPR adaptation; Cas1-Cas2; RT-fused Cas1; RNA recording

Corresponding Author: Sungchul Kim

Authors: Sungchul Kim^{1,*}, Gyeong-Seok Oh¹, Seongjin An^{1,2}

Institution: ¹Center for RNA Research, Institute for Basic Science, Seoul 08826, Korea,

²Department of Life Sciences, School of Life Sciences and Biotechnology, Korea University, Seoul, 02841, Korea,

Manuscript Type: Mini Review

Title: Recording RNA by CRISPR-Cas adaptation: a brief review

Author's name: Gyeong-Seok Oh^{1,#}, Seongjin An^{1,2,#}, and Sungchul Kim^{1,*}

[#] These authors contributed equally to this work.

^{*} Correspondence

Affiliation:

1 Center for RNA Research, Institute for Basic Science, Seoul 08826, Korea

2 Department of Life Sciences, School of Life Sciences and Biotechnology, Korea University, Seoul, 02841, Korea

Running Title: The mechanism and application of RT-Cas1-Cas2

Keywords: CRISPR-Cas, CRISPR adaptation, Cas1-Cas2, RT-fused Cas1, RNA recording

Corresponding Author's Information: Tel. +82-2-880-6278, Fax. +82-2-887-0244, sungchulkim.kr@gmail.com

ABSTRACT

Prokaryotes encode clustered regularly interspaced short palindromic repeat (CRISPR) arrays and CRISPR-associated (Cas) genes as an adaptive immune machinery. CRISPR-Cas systems effectively protect hosts from the invasion of foreign enemies, such as bacteriophages and plasmids. Non-self nucleic acid fragments are acquired as spacers between repeats in the host CRISPR array during a process called ‘adaptation’ to establish immunological memory. The highly conserved Cas1-Cas2 complexes function as molecular recorders to integrate spacers in a time course manner, which can subsequently be expressed as crRNAs complexed with Cas effector proteins for the RNA-guided interference pathways. In some of RNA-targeting type III systems, Cas1 proteins are fused with reverse transcriptase (RT), indicating that RT-Cas1-Cas2 complexes can acquire RNA transcripts for spacer acquisition. In this review, we summarize current studies focused on the molecular structure and function of the RT-fused Cas1-Cas2 integrase and its potential applications as a directional RNA-recording tool in

cells. Furthermore, we highlight outstanding questions for RT-Cas1-Cas2 studies and future directions for RNA-recording CRISPR technologies.

Classification of CRISPR-Cas systems

CRISPR-Cas systems are diverse prokaryotic RNA-guided adaptive immune machineries that provide protection against invasions by ‘mobile genetic elements’(MGEs), such as plasmids, viruses, and transposons in ~40% of bacteria and ~90% of archaea (1-4). All CRISPR-Cas systems follow three basic steps for the inheritable immunity against harmful MGEs (Figure 1A). First, during the ‘adaptation’ (also referred to as spacer acquisition) step, foreign DNA or RNA fragments are captured and integrated between repeats in host CRISPR arrays, updating an inheritable memory, called a spacer, for future encounters (3, 5-8). In the ‘expression’ step, integrated spacers are transcribed into a single, long precursor CRISPR RNA (pre-crRNA) and further processed into mature crRNAs. The crRNAs are assembled into a Cas effector, becoming a surveillance ribonucleoprotein complex, such as type I Cascade, type II Cas9, type V Cas12 and type VI Cas13 effectors. During the final ‘interference’ step, target binding with an effector-crRNA complex through R-loop formation results in target cleavage and degradation of the invading threat (Fig 1A) (1, 9-12).

CRISPR–Cas systems are divided into two major classes that distinctly differ in the architectures of their effector modules related to crRNA processing and interference (Figure 1B). Class 1 CRISPR-Cas systems are comprised of types I, III and IV, and are further divided into 16 subtypes, whereas class 2 CRISPR-Cas systems include types II, V and VI, and 17 subtypes (13). Class 1 systems encode effector Cas modules as multi-subunit proteins, such as Cascade in type I, Csm complex in type III-A and Cmr complex in type III-B systems. Additional Cas proteins often contribute to pre-crRNA processing or interference steps in many subtypes. In contrast, class 2 systems consist of a single, multi-domain, and large crRNA-binding protein, such as Cas9 in type II systems, Cas12 in type V systems, and Cas13 in type VI systems. These effector complexes usually exhibit an all-in-one activity for target interference as well as pre-crRNA processing in some variants.

Likewise, two strategies exist for abortive infection by MGEs and are utilized depending on the type of an effector complex, (14). In types I, II, IV and V CRISPR–Cas systems, DNA-targeting effector complexes directly destroy invading DNA via crRNA-guided cleavage. Target DNA binding in these types leads to the activation of the DNase activity of the effector complex, resulting in specific degradation of the target DNA to

circumscribe infection. In types III and VI CRISPR-Cas systems, RNA-targeting effector complexes are utilized. Here, the effector modules are activated upon RNA binding, leading to the RNase activity that cleaves the target RNA. Other accessory Cas proteins can also be involved in combating the invader via the activation of protease-mediated cascade pathways (e.g. TPR-CHAT/Csx29 in type III-E) (15-22) or collateral RNase activity (e.g. Csm6 in type III-A) (Figure 1B) (23, 24). This can result in indirect inhibition of infection through cellular signaling pathways that leads to the activation of downstream defense genes.

RNA-targeting CRISPR-Cas systems

Type III and VI CRISPR-Cas systems, despite their evolutionary distance and structural differences, exhibit the common feature of acquiring, sensing, and cleaving target RNA molecules for (adaptive) immunity. Type III CRISPR-Cas systems, believed to be the oldest member of the CRISPR-Cas family (25, 26), are classified further into six different annotated subtypes : III-A to III-F (13). Type III CRISPR-Cas systems account for 25% of total CRISPR-Cas loci in bacteria and 34% in archaea (27). Class 1 Type III CRISPR-Cas effector complexes generally consist of multiple subunits, just to name a few, Csm2, Csm3, Csm4 and Csm5 in types III-A and III-D, and Cmr1, Cmr3, Cmr4, Cmr5 and Cmr6 in types III-B and III-C (11, 28-31). Of these, the signature Cas10 subunit, also referred to as Csm1 in types III-A and III-D, and Cmr2 in types III-B and III-C is typically complexed with the main effector complex. Upon target RNA binding, the conformational change of the Cas10 subunit provides a DNase activity for proximal ssDNA cleavage in co-transcriptional R-loops and an ATP cyclase activity for the generation of cyclic oligoadenylates as secondary messengers that activate Csm6 protein for collateral RNA degradation. In comparison, the type III-E and III-F systems lack Cas10 subunit. Of these, a recently reported type III-E effector, known as gRAMP (also called Cas7-11) has a unique architecture that comprises subunits fused together as single protein, resembling effectors in class 2 systems (15, 16, 21, 32-35). The gRAMP effector is complexed with a TPR-CHAT (also known as Csx29) subunit, which is a caspase-like peptidase that can cleave the Csx30 upon RNA binding to activate the CRISPR-associated sigma factor RpoE for cell cycle retardation or cell death.

Next, we will discuss the type VI CRISPR-Cas system. The signature effector protein of Type VI CRISPR-Cas systems is a single multidomain protein Cas13. It acts on in both crRNA processing and target RNA recognition and cleavage for the immune response (36-39). Six different type VI subtypes, types VI-A to VI-D, Cas13X and Cas13Y, have been

identified so far (13, 40). All Cas13 proteins in type VI subtypes contain two HEPN domains critical for RNA-mediated target surveillance. Upon target RNA loading, the conformational change activates the RNase activity in HEPN domains, resulting in target RNA cleavage as well as collateral subversion of bystander RNA hydrolysis, which establishes broad and nonspecific immunity (37, 41, 42).

The RNA targeting CRISPR-Cas systems described above require a mechanism to distinguish self from non-self. For discriminating between self and non-self, DNA targeting CRISPR-Cas effectors recognize a small RNA sequence motif called protospacer adjacent motif (PAM) to position a target sequence in MGEs (43). The absence of PAMs in the spacer flanking repeat sequences prevents self-recognition, thereby inhibiting autoimmunity. In contrast, the catalytic activity of type III and VI CRISPR-Cas effectors can be regulated by recognition of a small RNA sequence next to a target RNA sequence derived from a repeat portion, referred to as protospacer flanking site (PFS) (44-46). Whereas a PAM binding leads the activation of DNA targeting effectors, a mismatch in PFS with a crRNA primes RNA targeting effector for non-self RNA targeting, preventing toxic, nonspecific targeting of self-transcripts. In other words, autoimmune response can occur by signifying self-transcripts through complementarity of the crRNA to the PFS in the antisense CRISPR array transcript.

CRISPR adaptation by Cas1-Cas2

How do bacteria or archaea archive and remember previous invaders? This process occurs during the ‘adaptation’ step. In DNA-targeting CRISPR-Cas systems, the highly conserved Cas1-Cas2 complex mediates spacer acquisition derived from DNA and integrates them into the CRISPR array (Figure 1A) (47-49). Cas1-Cas2 complex form a heterohexameric Cas1₍₄₎-Cas2₍₂₎ complex (Figure 2A and 2B) (3, 5, 47, 48, 50-54). *In vivo* studies on spacer acquisition suggest that Cas1-Cas2 complex identifies suitable prespacers based on the PAM, which is also a prerequisite for the CRISPR-interference stage of immunity (55-57).

Structural studies on the *E. coli* type I-E Cas1-Cas2 complex have demonstrated that the C terminal tail of the Cas1 subunit is responsible for PAM recognition (47, 52). In contrast, Cas1-Cas2 complexes in some type I systems, except type I-E and I-F systems, use an additional adaptation factor called Cas4, for PAM recognition (58-61). Although the Cas1-Cas2 complex preferentially integrates partial duplex DNAs harboring single-stranded 3' overhangs *in vitro* (47, 52, 62), a single-molecule study showed that the Cas1-Cas2 complex actually binds to a single strand of DNA containing a PAM sequence in more favorable

manner, suggesting that annealing of complementary ssDNA facilitates the generation of a suitable substrate comprising a PS duplex with 3' overhangs (63). In most type II systems, Cas9 and Csn2, together with a Cas1-Cas2 heterohexamer, play a role in PAM recognition, although the precise mechanism remains to be investigated (64-67). During DNA adaptation, the Cas1-Cas2 complex controls the correct orientation of integrated spacers through asymmetric prespacer trimming in a delayed PAM trimming mode by DnaQ enzymes or Cas4 endonuclease itself (59, 63, 68).

Identification of RT-fused Cas1 in type III CRISPR-Cas systems

For type III and VI CRISPR-Cas systems, the possibility of spacer acquisition derived from RNA by type III and VI CRISPR systems has been raised, as these can target RNA in addition to DNA (49). Interestingly, some bacteria harboring type III CRISPR-Cas systems have Cas1 fused with reverse transcriptase (RT), and these RT domains have been thought to be related to RTs known as group II introns (Figure 3A) (49, 69-72). Several Cas1 and associated RTs in type III CRISPR systems even seem to have co-evolved (70, 72, 73). Specifically, 537 sequences were analyzed in a comprehensive analysis to elucidate the origin and relationship between RTs and the associated CRISPR-Cas system and found that cases of RTs related to Cas1 loci were more prevalent in bacteria (11 clades) than archaea (only 1 clade) (Figure 3B). Since Cas1 proteins mediate spacer acquisition together with Cas2 subunits in a heterohexameric complex, it has been suggested that an RT domain would be required for direct spacer acquisition from RNA (69). RT-Cas1 fusion proteins were also identified in type VI-A CRISPR-Cas systems (74). In this study, two variant type VI-A systems were assumed to be evolved independently to fuse Cas1 proteins with RTs possibly derived from type III-A and III-D systems. Moreover, in addition to the RT domain fused with Cas1, the Cas6 endoribonuclease domains are also N-terminally fused to RT-Cas1 parts in many variants (Figure 3B) (13, 73, 75, 76). The role of Cas6 protein has been well established as an effector protein directly required for processing pre-crRNA (77) and Mohr et al. proved that the Cas6 domain fused with RT-Cas1 is also involved in RNA spacer acquisition (76).

Structure of RT-fused Cas1-Cas2 complex

Determining the structure of a protein complex provides a detailed understanding of its molecular mechanism. To this end, Wang et al. recently revealed the entire cryo-EM structure

of type III Cas6-RT-Cas1-Cas2 complex in a naturally occurring *Thiomicrospira* (*Thio*) species (Figure 2C) (78). Despite the authors effort to characterize the DNA-bound complex, which combines purified Cas6-RT-Cas1- Cas2 proteins together with a DNA substrate designed to resemble a half-site integration intermediate, only the DNA-unbound structure (apo-Cas6-RT-Cas1-Cas2 complex) was solved. This structure of *Thio* apo-Cas6-RT-Cas1-Cas2 complex is heterohexameric and consists of two distal Cas6-RT-Cas1 dimers and a central Cas2 dimer, like typical Cas1-Cas2 integrases (Figure 2A–C) (48, 79-81). *Thio* Cas6-RT-Cas1-Cas2 complex are distinguishable from those of *E. coli* Cas1-Cas2 in some respects (or aspects of their structure). For example, in the *E. coli* Cas1-Cas2 complex, two positively charged regions in the DNA binding cleft and Cas2 dimer are critical for prespacer binding, enabling the intrinsic ruler mechanism (47, 52). Although similar charged regions are located on the Cas1 domains and Cas2 subunits in the Cas6-RT-Cas1-Cas2 complex, Cas2 dimers and one Cas1 dimers are rotated further away from another Cas1 dimer, resulting in an altered dimer interface. The authors also reported that there exists close contact between three active sites of Cas1-Cas2 integrase, RT and Cas6 maturase, implying that functional crosstalk is tightly coordinated (Figure 2C). Furthermore, the RT domain of Cas6-RT-Cas1 resembles other RTs, such as retroviral and group II intron RTs (Figure 2D) (82, 83), implying that the integration process by Cas6-RT-Cas1-Cas2 could be followed by the target-primed reverse transcription similar to retro-homing mechanism of group II introns.

In another alternative approach, Mohr et al. revealed a truncated mutant structure of the Cas6 maturase domain of *Marinomonas mediterranea* (*MMB-1*) type III-B Cas6-RT-Cas1-Cas2 (Figure 2E) (84). The structure of the Cas6 from MMB-1 was superimposed onto that from *Thio*, showing an overall good alignment, except for differences in the β -strand architecture of the C-terminal RRM fold. In addition, the overall architecture of Cas6 domains in Cas6-RT-Cas1 proteins looks quite distinct from other stand-alone Cas6 proteins (84-86). Thus, it suggests that the Cas6 domain in Cas6-RT-Cas1-Cas2 complex is not only critical to crRNA processing, but also may be functionally engaged in either RNA substrate capture and process during reverse transcription by RT domain or prespacer integration cooperated with Cas1-Cas2 integrase in the complex.

Functional features of RT-fused Cas1-Cas2 complex

Domain-fused proteins, including RT-fused Cas1-2 integrases, typically coordinate their series of actions. It has been suggested that RNA could be a suitable substrate for spacer

acquisition by Cas1-Cas2 integrase, with assistance from the fused RT. In vivo integration assay using RNA transcripts harboring self-splicing introns have shown that RNA transcripts can be integrated into CRISPR arrays by the *MMB-1* Cas6-RT-Cas1-Cas2 complex (87). This result was reproduced using the Cas6-RT-Cas1-Cas2 of *Fusicatenibacter saccharivorans* (*F. sac*) (88) and *Vibrio vulnificus* (*V. vul*) (89). *MMB-1* and *Thio* Cas6-RT-Cas1-Cas2 complexes show substrate versatility, as these have been shown to integrate dsDNA, ssDNA, and ssRNA oligonucleotides into the CRISPR DNA (78, 87). Full-site integration occurs only site-specifically when dsDNA substrates are provided. In *MMB-1*, both RT and Cas1-Cas2 integrase activities of the Cas6-RT-Cas1-Cas2 complex are indispensable for RNA integration, while RT activity is not required for DNA integration. In contrast, *Thio* Cas6-RT-Cas1 alone, regardless of the presence of Cas2 dimer, exhibits integration activity, although Cas2 improves integration efficiency. *Thio* Cas6-RT-Cas1-Cas2 is significantly more efficient at integrating dsDNA, ssDNA, and DNA/RNA hybrid substrates than ssRNA substrates. Collectively, these results suggest that the RT and Cas1 domains in the Cas6-RT-Cas1-Cas2 complex closely cooperate for RT process and integration functions (78, 87). Furthermore, mutations in the Cas6 active site affect both integration and RT activity, while mutations in the Cas1 and RT domains show no effect on Cas6 RNA processing activity. This suggests a unidirectional crosstalk between the Cas6 domain and the other two domains in *Thio* Cas6-RT-Cas1-Cas2 (78).

Next, the sequence and length specificity of spacer acquisition by Cas6-RT-Cas1-Cas2 has also been examined closely. In *MMB-1*, the majority of spacers (70–75%) were 34–36 base pairs (bp) long, with no significant sequence specificity observed (69). A small preference towards the sense strand spacer was also observed in an RT-independent manner in *MMB-1*, but this strand bias was not reconstructed in *E. coli*. In contrast, the median spacer length was 39 bp, with a distribution bias towards longer spacers in *F. sac*. (88). Strikingly, a strong bias towards AT-rich spacers was also observed. This apparent bias towards AT-rich spacers may be due to the AT-richness of RNA ends, but the bias persisted even when considering only spacers derived from the gene body. Furthermore, the *F. sac* Cas6-RT-Cas1-Cas2 complex did not exhibit any preference for PAM-like sequence motifs, but most spacers were acquired from the areas proximal to start and stop codons. Also, spacers were preferentially acquired toward an antisense orientation. Lastly, in *V. vul*, most of the spacers were between 34–38 bp with no PAM-like preference (89). However, there was an antisense bias for coding sequence orientation with a significant GC bias. Unlike *F. sac*, there was no

bias at the start and end codons. Taken together, the differential specificity for spacer length and sequence observed in previous independent studies implies significant variation among these systems in vivo, which needs to be characterized in future studies.

Application for RNA recording using RT-Cas1-Cas2

The Cas1-Cas2 adaptation complex can integrate foreign nucleic acids as spacers between the leader sequence and the first repeat sequence of its own CRISPR array (53, 55). Taking advantage of this nucleic acid induced CRISPR-Cas memory system, molecular recording using CRISPR-Cas systems has emerged as a prominent field of research with the potential to revolutionize various areas of biotechnology and life sciences. Cas1-Cas2-based DNA recording offers several advantages over other explored methods for storing information (90-92). Several recently published articles have described diverse approaches to molecular recording using CRISPR-Cas systems, each with its own advantages and limitations (93). In 2016, Shipman et al. developed the first DNA recording system using the type I-E Cas1-Cas2 complex in *E. coli* (94). One of key benefits of this recording system is that the Cas1-Cas2 integrase can target and capture specific DNA molecules as spacers and integrate them into the CRISPR array directionally and temporally, reflecting dynamic changes in real time within a cell (95-98). Since the recorded DNA was stored in the CRISPR array in a time-ordered way, it is possible to reconstruct the recorded information temporally and make a lineage of events by sequencing the CRISPR array later on (95, 99). By creating a permanent record of events within the genome, researchers can gain valuable insights into the underlying mechanisms driving cellular behavior and response to environmental stimuli at a single-cell level (95, 100). Moreover, since the CRISPR array is inherent within the genome of the bacterial cell, it offers a reliable and long-term record that can be easily retrieved and analyzed without the need for external devices or systems in the case of some prokaryotes. However, these molecular recording systems have several limitations that need to be addressed. For instance, these systems integrate spacers derived from DNA, not RNA and they do not co-ordinate the transcriptional dynamics of the cell. There is also a need to transform exogenous nucleic acids to verify the specificities and efficiencies of molecular recording (99).

Similarly, transcriptional molecular recorders are innovative tools that can capture various transcriptional events and incorporate them into the genome of the cell. This approach has the potential to reveal the precise timing, order, and intensity of transcriptional

activity at the cellular level, without requiring multiple destructive assays (Figure 4A). To this end, Schmidt et al. utilized *F. sac* Cas6-RT-Cas1-Cas2 for transcriptional recording instead of the DNA integrating *E. coli* Cas1-Cas2 complex (88), since Cas6-RT-Cas1-Cas2 has been used as a spacer integrase from cellular transcripts (Figure 4B) (69). This study showed that the tracking of transcriptional responses to specific stimuli within bacterial cells can be achieved via transcriptional recording (88). Furthermore, transcriptional events stored as RNA-derived spacers at the CRISPR array reflected temporal and global transcriptomic memories to various stresses (99, 101). In another recent study by Schmidt et al., they engineered multiplexed transcriptional recording to track transcriptional histories of the stress response in the physiological environment, enabling engineered microbiome as sentinel cells to record environmental changes in mouse guts by Record-seq (Figure 4B) (101) (Fig. 4). All together, these studies support the idea that RT-Cas1-Cas2-based recording technology can be a useful tool for studying the evolution of cells within various environmental contexts as well as for novel therapeutic development and diagnostics. Lastly, a retron RT, which can reverse transcribe non-coding RNA (ncRNA), has also been proposed to record temporal memories of transcriptional events in a live cell and emerged as an alternative to transcriptional recording using RT-Cas1-Cas2 (102).

Concluding remark and future perspectives

In the last decade, the CRISPR-based genome engineering field has been successfully revolutionized along with our understanding for mechanisms of the CRISPR-Cas adaptive immune systems. DNA cutting CRISPR-Cas effectors, such as type II Cas9 and type V Cas12a (also called as Cpf1), have been mainly focused and engineered to develop diverse genome editing technologies (103). Additional need for advancing RNA editing tools has enabled RNA-targeting type VI CRISPR Cas13 effectors to be reconfigured for the use in gene silencing (104) and RNA diagnostics (105). Beyond the gene editing, an endeavor is still being made to discover novel CRISPR-based technologies, such as CRISPR-mediated imaging, epigenome manipulation, and molecular recording tools. In this review, we focus on recent advances and understandings regarding molecular recording techniques based on CRISPR adaptation modules. Throughout recent genetic, structural and biochemical approaches, we can now explain how proteins related to the DNA CRISPR adaptation function to ensure the efficiency and fidelity for the precise spacer acquisition (8). However, our knowledge about the mechanism of CRISPR RNA adaptation is very limited.

To extend our knowledge and to improve technical limitations of current RNA recording tools, further studies are required to be addressed to more understand the basic mechanism of RT-Cas1-Cas2-mediated CRISPR adaptation. Outstanding questions are like below. How the ssRNA or RNA/DNA or dsDNA bound RT-Cas1-Cas2 structures are configured? How the cooperation between domains of RT-Cas1-Cas2 can be achieved for CRISPR RNA acquisition, and which role of Cas6 is critical in this process? How the RT domain in this complex reverse transcribes RNA substrates? How the actual process from ssRNA capture to full site integration can be coordinated? How is the suitable spacer size determined during RNA adaptation? And how the orientation of integrated spacers ensures for immunity? It is also not known whether RNA spacer acquisition can contribute to immunity against RNA bacteriophages and other foreign RNA elements beside targeting nascent viral transcripts.

While these molecular recording systems using RT-Cas1-Cas2 have several promising advantages warranting their future use, RT-Cas1-Cas2-based recorders still have inherent limitations that need to be addressed. It remains to be solved how these recorders can function the same way in the eukaryotic system. The process of implementing RT-Cas1-Cas2-based recorders into mammalian cells may depend on non-Cas bacterial factors (51, 106), or may be established by a combination with epigenome manipulating techniques in eukaryotic environments. Furthermore, the capacity and efficiency to detect CRISPR arrays, which are intended to integrate transcript-derived spacers of interest, could be very slim. Given that not all spacers of interest will be integrated as expected, this inefficient recording possibility may further reduce the chance of detecting the spacer of interest. Therefore, more optimized sequencing techniques are needed to be developed to overcome the detection limit in the future.

ACKNOWLEDGMENTS

This work was supported by IBS-R008-D1, Young Scientist Fellowship program of the Institute for Basic Science from the Ministry of Science and ICT of Korea.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FIGURE LEGENDS

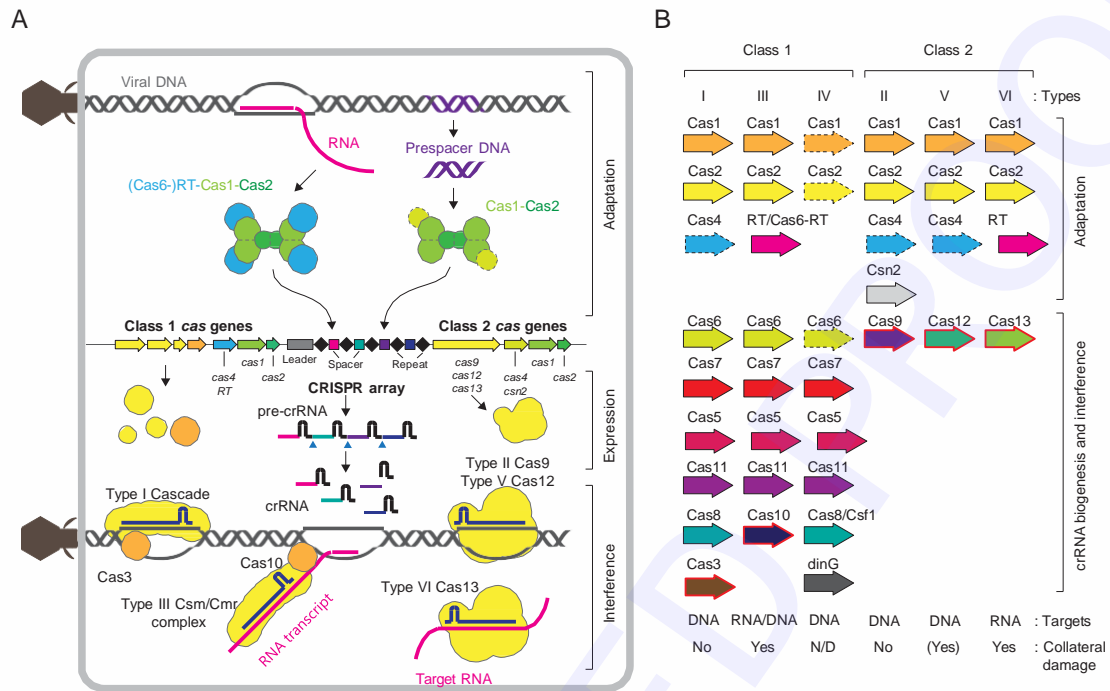


Figure 1. Overview of the mechanism and classification of CRISPR-Cas systems

(A) Schematic of the CRISPR-Cas adaptive immunity. During the adaptation stage, foreign DNA (gray helical lines) or RNA (pink lines) can be captured and integrated into CRISPR array, by Cas1-Cas2 and (Cas6-)-RT-Cas1-Cas2 complexes, respectively. Acquired nucleic acids are converted to new spacers and recorded in a leader-proximal order, creating immunological memories. In the expression stage, spacers in CRISPR array are transcribed to pre-crRNA, and processed into crRNAs (crRNA biogenesis). The effector RNP complex are assembled crRNAs with expressed effector Cas proteins. At the final interference stage, effector complexes target against revisiting foreign (viral) DNAs and their transcripts through sequence-specific cleavage complementary to the sequence of crRNA. The image is adopted and modified from Lee et al. Trends in Biochemical Sciences, 2022 (8). (B) Classification and gene structure of CRISPR-Cas systems. In adaptation associated *cas* genes, there are well conserved genes, Cas1 and Cas2. Some accessory genes exist like Cas4 in type I, II and V, or reverse transcriptase (RT) in type III and VI, or Csn2 in type II. The classification of each type primarily based on effector genes. Class 1 has multiple interfering and crRNA biogenesis genes. Meanwhile, Class 2 has single effector genes, e.g. Cas9, Cas12 or Cas13. The signature proteins in each type are represented with red outline. Target nucleic acids and the existing collateral damage effect are indicated below.

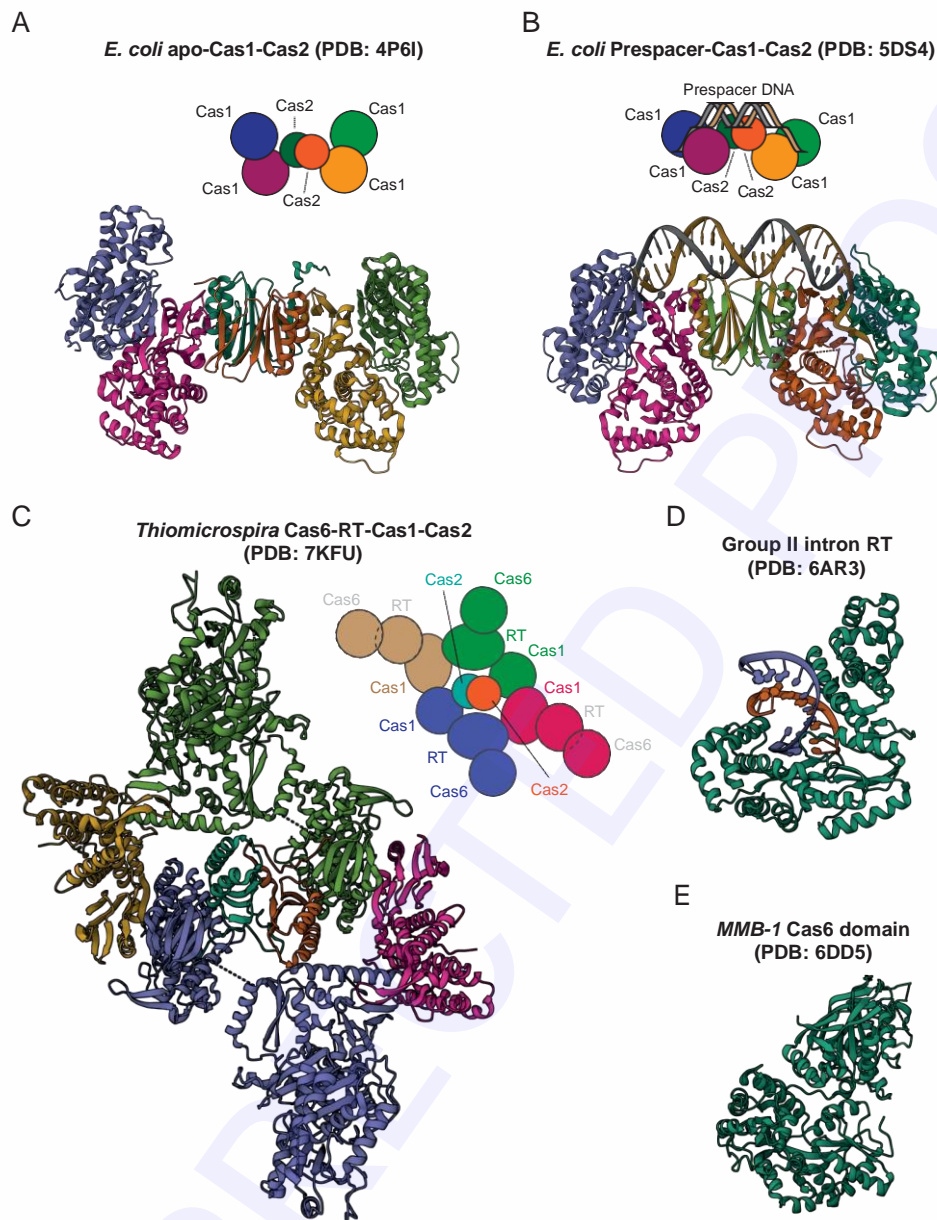


Figure 2. Structural architectures of CRISPR integrases and related proteins

(A) Structure of *E. coli* type I-E DNA-unbound (apo) Cas1-Cas2 complex (Protein Data Bank, PDB, ID: 4P6I) (50). (B) Structure of *E. coli* type I-E prespacer DNA-bound Cas1-Cas2 complex (PDB ID: 5DS4) (52). (C) Structure of *Thiomicrospira* Cas6-RT-Cas1-Cas2 complex (PDB ID: 7KFU) (75). Two of the four Cas6-RT domains are invisible. (A–C) The arrangement models of the color-coded subunits and DNA (in B) are represented above the crystal (in A and B) and cryo-EM (in C) structures of the complexes. (D) Structure of *Marinomonas mediterranea* (*MMB-1*) truncated Cas6 domain (PDB ID: 6DD5) (76). (E) Structure of group II intron reverse transcriptase bound to an RNA template base paired with a DNA primer (PDB ID: 6AR3) (82).

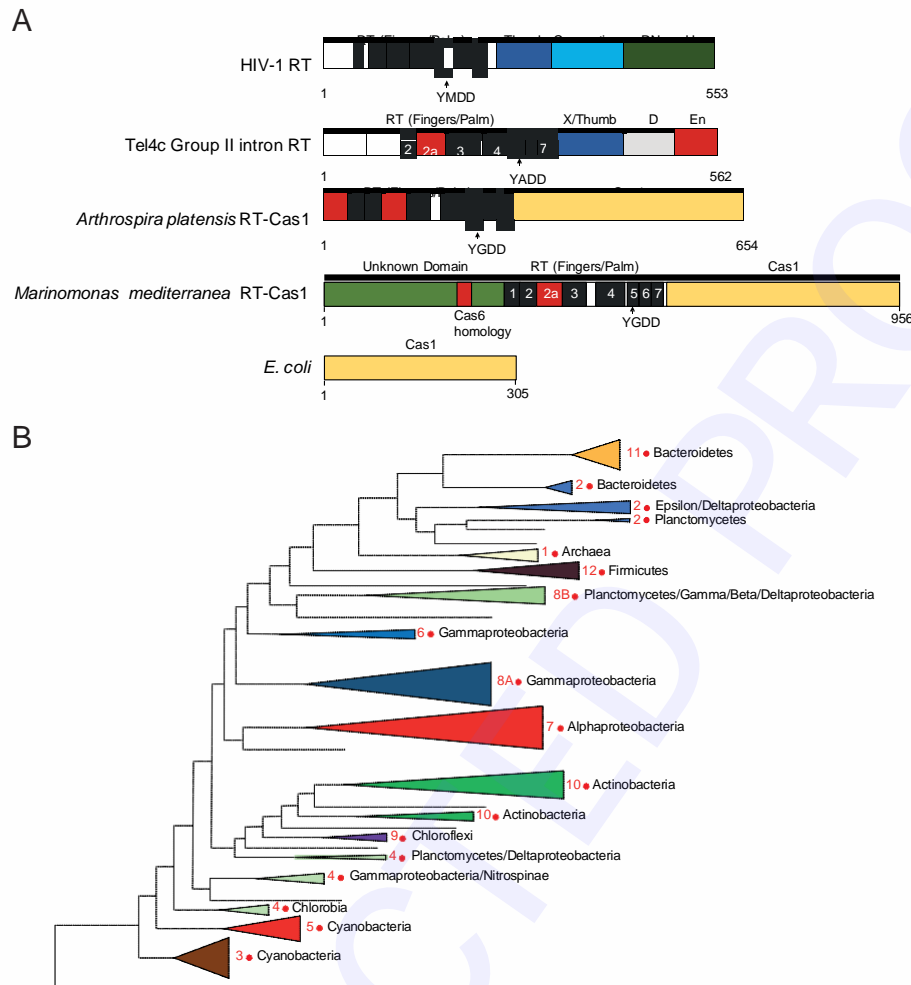


Figure 3. Domain structure and Phylogenetic distribution of the RT-fused Cas1 and its related proteins

(A) Schematic of the domain organization of HIV-1 RT, Tel4c group II intron RT, *A. platensis* type III-B RT-Cas1, *MMB-1* type III-B RT-Cas1, and *E. coli* type I-E Cas1. Conserved RT motifs (1 to 7) are indicated in black boxes. Conserved motifs in mobile group II intron and non-LTR-retrotransposon RTs (0 and 2a) are labeled in red boxes. The YXDD sequence found in motif 5 represented with two aspartic acid residues is indicated by arrow. The X/Thumb domains commonly found in HIV-1 and group II intron RTs are indicated. Amino acid numbers are indicated below the bars. D, DNA binding domain. En, endonuclease domain. The image is taken from Silas et al. Science, 2016 (69). (B) Phylogenetic tree of Cas1 associated with RTs. The analysis is reconstructed with 148 Cas1 proteins. The identified clades with numbers are named and colored according to the RT-associated clade. The figure is taken and modified from Toro et al. Scientific Reports, 2017 (73).

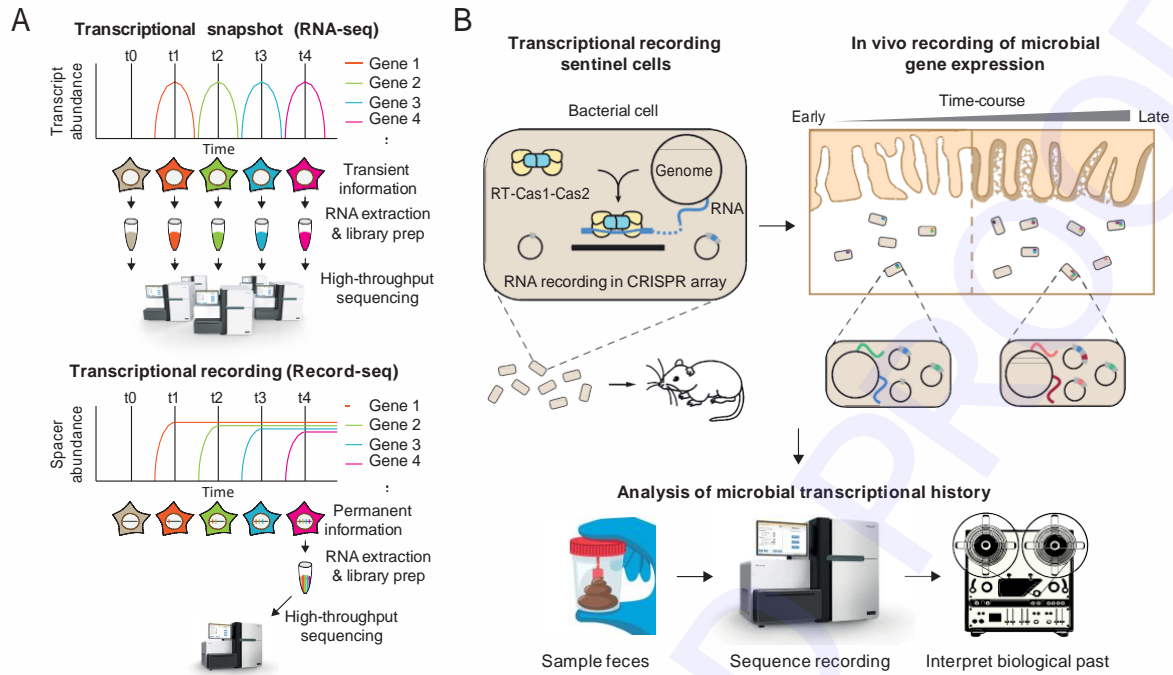


Figure 4. In vivo transcriptional recording by RT-Cas1-Cas2

(A) Comparison between RNA-seq and Record-seq. RNA-seq captures the transcriptomes from subject cells at each point in time, providing a transient snapshot of cellular events. By contrast, transcriptional records in CRISPR arrays provide molecular records that can be used to reconstruct transcriptional events that occurred over time only by a one-shot analysis. The image is adopted and revised from Schmidt et al. Nature, 2018 (88). (B) Schematic of experimental platform for the noninvasive interpretation of transcriptional recordings using *F. sac* RT-Cas1-Cas2 in bacterial cells. The *E. coli* cells transformed to express RT-Cas1-Cas2 proteins and harbor plasmids with CRISPR arrays are orally gavaged into mice. Stress response patterns by bacterial strains as sentinel cells are stored in CRISPR arrays in an inheritable and time-ordered way. Eventually, the recorded information can be reconstructed by sequencing the CRISPR arrays from microbial samples without any effort for the time-course harvest. The figure is taken and modified from Schmidt et al. Science, 2022 (101).

REFERENCES

1. Hille F, Richter H, Wong SP, Bratovic M, Ressel S and Charpentier E (2018) The Biology of CRISPR-Cas: Backward and Forward. *Cell* 172, 1239-1259
2. Marraffini LA (2015) CRISPR-Cas immunity in prokaryotes. *Nature* 526, 55-61
3. McGinn J and Marraffini LA (2019) Molecular mechanisms of CRISPR-Cas spacer acquisition. *Nat Rev Microbiol* 17, 7-12
4. Frost LS, Leplae R, Summers AO and Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 3, 722-732
5. Jackson SA, McKenzie RE, Fagerlund RD, Kieper SN, Fineran PC and Brouns SJ (2017) CRISPR-Cas: Adapting to change. *Science* 356
6. Barrangou R, Fremaux C, Deveau H et al (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709-1712
7. Fineran PC and Charpentier E (2012) Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* 434, 202-209
8. Lee H and Sashital DG (2022) Creating memories: molecular mechanisms of CRISPR adaptation. *Trends Biochem Sci* 47, 464-476
9. Deveau H, Garneau JE and Moineau S (2010) CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol* 64, 475-493
10. Garneau JE, Dupuis ME, Villion M et al (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67-71
11. Hale CR, Zhao P, Olson S et al (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139, 945-956
12. Marraffini LA and Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843-1845
13. Makarova KS, Wolf YI, Iranzo J et al (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol* 18, 67-83

14. van Beljouw SPB, Sanders J, Rodriguez-Molina A and Brouns SJJ (2023) RNA-targeting CRISPR-Cas systems. *Nat Rev Microbiol* 21, 21-34
15. Ekundayo B, Torre D, Beckert B et al (2023) Structural insights into the regulation of Cas7-11 by TPR-CHAT. *Nat Struct Mol Biol* 30, 135-139
16. Wang S, Guo M, Zhu Y, Lin Z and Huang Z (2022) Cryo-EM structure of the type III-E CRISPR-Cas effector gRAMP in complex with TPR-CHAT. *Cell Res* 32, 1128-1131
17. Huo Y, Zhao H, Dong Q and Jiang T (2023) Cryo-EM structure and protease activity of the type III-E CRISPR-Cas effector. *Nat Microbiol* 8, 522-532
18. Wang X, Yu G, Wen Y et al (2022) Target RNA-guided protease activity in type III-E CRISPR-Cas system. *Nucleic Acids Res* 50, 12913-12923
19. Kato K, Okazaki S, Schmitt-Ulms C et al (2022) RNA-triggered protein cleavage and cell growth arrest by the type III-E CRISPR nuclease-protease. *Science* 378, 882-889
20. Strecker J, Demircioglu FE, Li D et al (2022) RNA-activated protein cleavage with a CRISPR-associated endopeptidase. *Science* 378, 874-881
21. Yu G, Wang X, Zhang Y et al (2022) Structure and function of a bacterial type III-E CRISPR-Cas7-11 complex. *Nat Microbiol* 7, 2078-2088
22. Hu C, van Beljouw SPB, Nam KH et al (2022) Craspase is a CRISPR RNA-guided, RNA-activated protease. *Science* 377, 1278-1285
23. Niewoehner O, Garcia-Doval C, Rostol JT et al (2017) Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature* 548, 543-548
24. Kazlauskienė M, Kostiuk G, Venclovas C, Tamulaitis G and Siksnyš V (2017) A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* 357, 605-609
25. Koonin EV and Makarova KS (2018) Discovery of Oligonucleotide Signaling Mediated by CRISPR-Associated Polymerases Solves Two Puzzles but Leaves an Enigma. *ACS Chem Biol* 13, 309-312

26. Coleman GA, Davin AA, Mahendrarajah TA et al (2021) A rooted phylogeny resolves early bacterial evolution. *Science* 372
27. Makarova KS, Wolf YI, Alkhnbashi OS et al (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* 13, 722-736
28. Molina R, Sofos N and Montoya G (2020) Structural basis of CRISPR-Cas Type III prokaryotic defence systems. *Curr Opin Struct Biol* 65, 119-129
29. Tamulaitis G, Venclovas C and Siksnys V (2017) Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. *Trends Microbiol* 25, 49-61
30. Rouillon C, Zhou M, Zhang J et al (2013) Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol Cell* 52, 124-134
31. Zhang J, Rouillon C, Kerou M et al (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* 45, 303-313
32. van Beljouw SPB, Haagsma AC, Rodriguez-Molina A, van den Berg DF, Vink JNA and Brouns SJJ (2021) The gRAMP CRISPR-Cas effector is an RNA endonuclease complexed with a caspase-like peptidase. *Science* 373, 1349-1353
33. Goswami HN, Rai J, Das A and Li H (2022) Molecular mechanism of active Cas7-11 in processing CRISPR RNA and interfering target RNA. *Elife* 11
34. Kato K, Zhou W, Okazaki S et al (2022) Structure and engineering of the type III-E CRISPR-Cas7-11 effector complex. *Cell* 185, 2324-2337 e2316
35. Ozcan A, Krajeski R, Ioannidi E et al (2021) Programmable RNA targeting with the single-protein CRISPR effector Cas7-11. *Nature* 597, 720-725
36. East-Seletsky A, O'Connell MR, Knight SC et al (2016) Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* 538, 270-273
37. Liu L, Li X, Wang J et al (2017) Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities. *Cell* 168, 121-134 e112

38. Liu L, Li X, Ma J et al (2017) The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell* 170, 714-726 e710
39. Abudayyeh OO, Gootenberg JS, Konermann S et al (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353, aaf5573
40. Xu C, Zhou Y, Xiao Q et al (2021) Programmable RNA editing with compact CRISPR-Cas13 systems from uncultivated microbes. *Nat Methods* 18, 499-506
41. Tambe A, East-Seletsky A, Knott GJ, Doudna JA and O'Connell MR (2018) RNA Binding and HEPN-Nuclease Activation Are Decoupled in CRISPR-Cas13a. *Cell Rep* 24, 1025-1036
42. Meeske AJ, Nakandakari-Higa S and Marraffini LA (2019) Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature* 570, 241-245
43. Leenay RT and Beisel CL (2017) Deciphering, Communicating, and Engineering the CRISPR PAM. *J Mol Biol* 429, 177-191
44. Meeske AJ and Marraffini LA (2018) RNA Guide Complementarity Prevents Self-Targeting in Type VI CRISPR Systems. *Mol Cell* 71, 791-801 e793
45. Marraffini LA and Sontheimer EJ (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 463, 568-571
46. Elmore JR, Sheppard NF, Ramia N et al (2016) Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. *Genes Dev* 30, 447-459
47. Wang J, Li J, Zhao H et al (2015) Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell* 163, 840-853
48. Xiao Y, Ng S, Nam KH and Ke A (2017) How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. *Nature* 550, 137-141
49. Makarova KS and Koonin EV (2015) Annotation and Classification of CRISPR-Cas Systems. *Methods Mol Biol* 1311, 47-75

50. Nunez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW and Doudna JA (2014) Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* 21, 528-534
51. Nunez JK, Bai L, Harrington LB, Hinder TL and Doudna JA (2016) CRISPR Immunological Memory Requires a Host Factor for Specificity. *Mol Cell* 62, 824-833
52. Nunez JK, Harrington LB, Kranzusch PJ, Engelman AN and Doudna JA (2015) Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* 527, 535-538
53. Nunez JK, Lee AS, Engelman A and Doudna JA (2015) Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 519, 193-198
54. Wright AV, Nunez JK and Doudna JA (2016) Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* 164, 29-44
55. Yosef I, Goren MG and Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res* 40, 5569-5576
56. Diez-Villasenor C, Guzman NM, Almendros C, Garcia-Martinez J and Mojica FJ (2013) CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of *Escherichia coli*. *RNA Biol* 10, 792-802
57. Levy A, Goren MG, Yosef I et al (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* 520, 505-510
58. Dhingra Y, Suresh SK, Juneja P and Sashital DG (2022) PAM binding ensures orientational integration during Cas4-Cas1-Cas2-mediated CRISPR adaptation. *Mol Cell* 82, 4353-4367 e4356
59. Hu C, Almendros C, Nam KH et al (2021) Mechanism for Cas4-assisted directional spacer acquisition in CRISPR-Cas. *Nature* 598, 515-520
60. Kieper SN, Almendros C, Haagsma AC, Barendregt A, Heck AJR and Brouns SJJ (2021) Cas4-Cas1 Is a Protospacer Adjacent Motif-Processing Factor Mediating Half-Site Spacer Integration During CRISPR Adaptation. *CRISPR J* 4, 536-548

61. Lee H, Dhingra Y and Sashital DG (2019) The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. *Elife* 8
62. Shiriaeva AA, Savitskaya E, Datsenko KA et al (2019) Detection of spacer precursors formed in vivo during primed CRISPR adaptation. *Nat Commun* 10, 4603
63. Kim S, Loeff L, Colombo S, Jergic S, Brouns SJJ and Joo C (2020) Selective loading and processing of prespacers for precise CRISPR adaptation. *Nature* 579, 141-145
64. Wei Y, Terns RM and Terns MP (2015) Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation. *Genes Dev* 29, 356-361
65. Heler R, Samai P, Modell JW et al (2015) Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* 519, 199-202
66. Jakhanwal S, Cress BF, Maguin P, Lobba MJ, Marraffini LA and Doudna JA (2021) A CRISPR-Cas9-integrase complex generates precise DNA fragments for genome integration. *Nucleic Acids Res* 49, 3546-3556
67. Wilkinson M, Drabavicius G, Silanskas A, Gasiunas G, Siksnys V and Wigley DB (2019) Structure of the DNA-Bound Spacer Capture Complex of a Type II CRISPR-Cas System. *Mol Cell* 75, 90-101 e105
68. Ramachandran A, Summerville L, Learn BA, DeBell L and Bailey S (2020) Processing and integration of functionally oriented prespacers in the *Escherichia coli* CRISPR system depends on bacterial host exonucleases. *J Biol Chem* 295, 3403-3414
69. Silas S, Mohr G, Sidote DJ et al (2016) Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* 351, aad4234
70. Silas S, Makarova KS, Shmakov S et al (2017) On the Origin of Reverse Transcriptase-Using CRISPR-Cas Systems and Their Hyperdiverse, Enigmatic Spacer Repertoires. *mBio* 8
71. Simon DM and Zimmerly S (2008) A diversity of uncharacterized reverse transcriptases in bacteria. *Nucleic Acids Res* 36, 7219-7229
72. Toro N and Nisa-Martinez R (2014) Comprehensive phylogenetic analysis of bacterial

reverse transcriptases. PLoS One 9, e114083

73. Toro N, Martinez-Abarca F and Gonzalez-Delgado A (2017) The Reverse Transcriptases Associated with CRISPR-Cas Systems. *Sci Rep* 7, 7089
74. Toro N, Mestre MR, Martinez-Abarca F and Gonzalez-Delgado A (2019) Recruitment of Reverse Transcriptase-Cas1 Fusion Proteins by Type VI-A CRISPR-Cas Systems. *Front Microbiol* 10, 2160
75. Wang JY, Hoel CM, Al-Shayeb B, Banfield JF, Brohawn SG and Doudna JA (2021) Structural coordination between active sites of a CRISPR reverse transcriptase-integrase complex. *Nat Commun* 12, 2571
76. Mohr G, Silas S, Stamos JL et al (2018) A Reverse Transcriptase-Cas1 Fusion Protein Contains a Cas6 Domain Required for Both CRISPR RNA Biogenesis and RNA Spacer Acquisition. *Mol Cell* 72, 700-714 e708
77. Carte J, Wang R, Li H, Terns RM and Terns MP (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev* 22, 3489-3496
78. Wang JY, Hoel CM, Al-Shayeb B, Banfield JF, Brohawn SG and Doudna JA (2021) Structural coordination between active sites of a CRISPR reverse transcriptase-integrase complex. *Nature communications* 12, 2571
79. Wright AV, Liu J-J, Knott GJ, Doxzen KW, Nogales E and Doudna JA (2017) Structures of the CRISPR genome integration complex. *Science* 357, 1113-1118
80. Nuñez JK, Harrington LB, Kranzusch PJ, Engelman AN and Doudna JA (2015) Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* 527, 535-538
81. Wang J, Li J, Zhao H et al (2015) Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. *Cell* 163, 840-853
82. Stamos JL, Lentzsch AM and Lambowitz AM (2017) Structure of a thermostable group II intron reverse transcriptase with template-primer and its functional and evolutionary implications. *Molecular cell* 68, 926-939. e924

83. Mitchell M, Gillis A, Futahashi M, Fujiwara H and Skordalakes E (2010) Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nature structural & molecular biology* 17, 513-518
84. Mohr G, Silas S, Stamos JL et al (2018) A reverse transcriptase-Cas1 fusion protein contains a Cas6 domain required for both CRISPR RNA biogenesis and RNA spacer acquisition. *Molecular cell* 72, 700-714. e708
85. Hochstrasser ML and Doudna JA (2015) Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends in biochemical sciences* 40, 58-66
86. Reeks J, Sokolowski RD, Graham S, Liu H, Naismith JH and White MF (2013) Structure of a dimeric crenarchaeal Cas6 enzyme with an atypical active site for CRISPR RNA processing. *Biochemical Journal* 452, 223-230
87. Silas S, Mohr G, Sidote DJ et al (2016) Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* 351, aad4234
88. Schmidt F, Cherepkova MY and Platt RJ (2018) Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* 562, 380-385
89. González-Delgado A, Mestre MR, Martínez-Abarca F and 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 Research* 47, 10202-10211
90. Burrill DR and Silver PA (2010) Making cellular memories. *Cell* 140, 13-18
91. Gardner TS, Cantor CR and Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339-342
92. Siuti P, Yazbek J and Lu TK (2013) Synthetic circuits integrating logic and memory in living cells. *Nat Biotechnol* 31, 448-452
93. Lear SK and Shipman SL (2023) Molecular recording: transcriptional data collection into the genome. *Curr Opin Biotechnol* 79, 102855
94. Shipman SL, Nivala J, Macklis JD and Church GM (2016) Molecular recordings by

directed CRISPR spacer acquisition. *Science* 353, aaf1175

95. Sheth RU, Yim SS, Wu FL and Wang HH (2017) Multiplex recording of cellular events over time on CRISPR biological tape. *Science* 358, 1457-1461
96. Shipman SL, Nivala J, Macklis JD and Church GM (2017) CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature* 547, 345-349
97. Matsoukas IG (2017) Commentary: CRISPR-Cas Encoding of a Digital Movie into the Genomes of a Population of Living Bacteria. *Front Bioeng Biotechnol* 5, 57
98. Sheth RU and Wang HH (2018) DNA-based memory devices for recording cellular events. *Nat Rev Genet* 19, 718-732
99. Tanna T, Schmidt F, Cherepkova MY, Okoniewski M and Platt RJ (2020) Recording transcriptional histories using Record-seq. *Nat Protoc* 15, 513-539
100. McKenna A and Gagnon JA (2019) Recording development with single cell dynamic lineage tracing. *Development* 146
101. Schmidt F, Zimmermann J, Tanna T et al (2022) Noninvasive assessment of gut function using transcriptional recording sentinel cells. *Science* 376, eabm6038
102. Bhattarai-Kline S, Lear SK, Fishman CB et al (2022) Recording gene expression order in DNA by CRISPR addition of retron barcodes. *Nature* 608, 217-225
103. Wang JY and Doudna JA (2023) CRISPR technology: A decade of genome editing is only the beginning. *Science* 379, eadd8643
104. Tang L (2020) Guiding Cas13 for RNA knockdown. *Nat Methods* 17, 461
105. Kaminski MM, Abudayyeh OO, Gootenberg JS, Zhang F and Collins JJ (2021) CRISPR-based diagnostics. *Nat Biomed Eng* 5, 643-656
106. Yoganand KN, Sivathanu R, Nimkar S and Anand B (2017) Asymmetric positioning of Cas1-2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. *Nucleic Acids*

632 Res 45, 367-381
633

UNCORRECTED PROOF