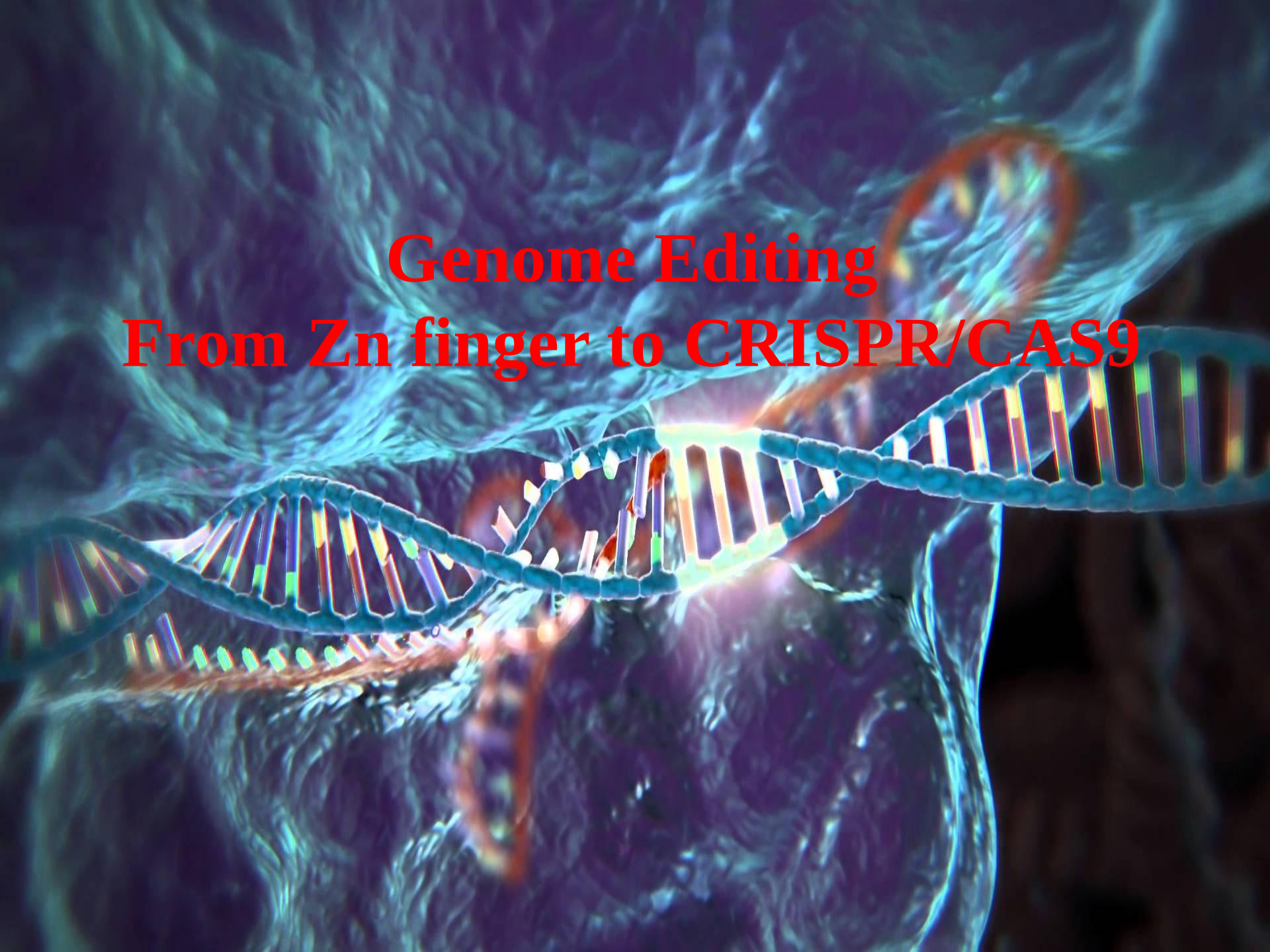

Genome Editing
From Zn finger to CRISPR/CAS9

NOBELPRISET I KEMI 2020
THE NOBEL PRIZE IN CHEMISTRY 2020

THE ROYAL SWEDIS ES

Vista

Photo: Hallbauer&Foretti

Emmanuelle Charpentier

Photo: UC Berkeley/Doudna Lab

Jennifer A. Doudna

”för utveckling av en metod för genomeditering”

“for the development of a method for genome editing”

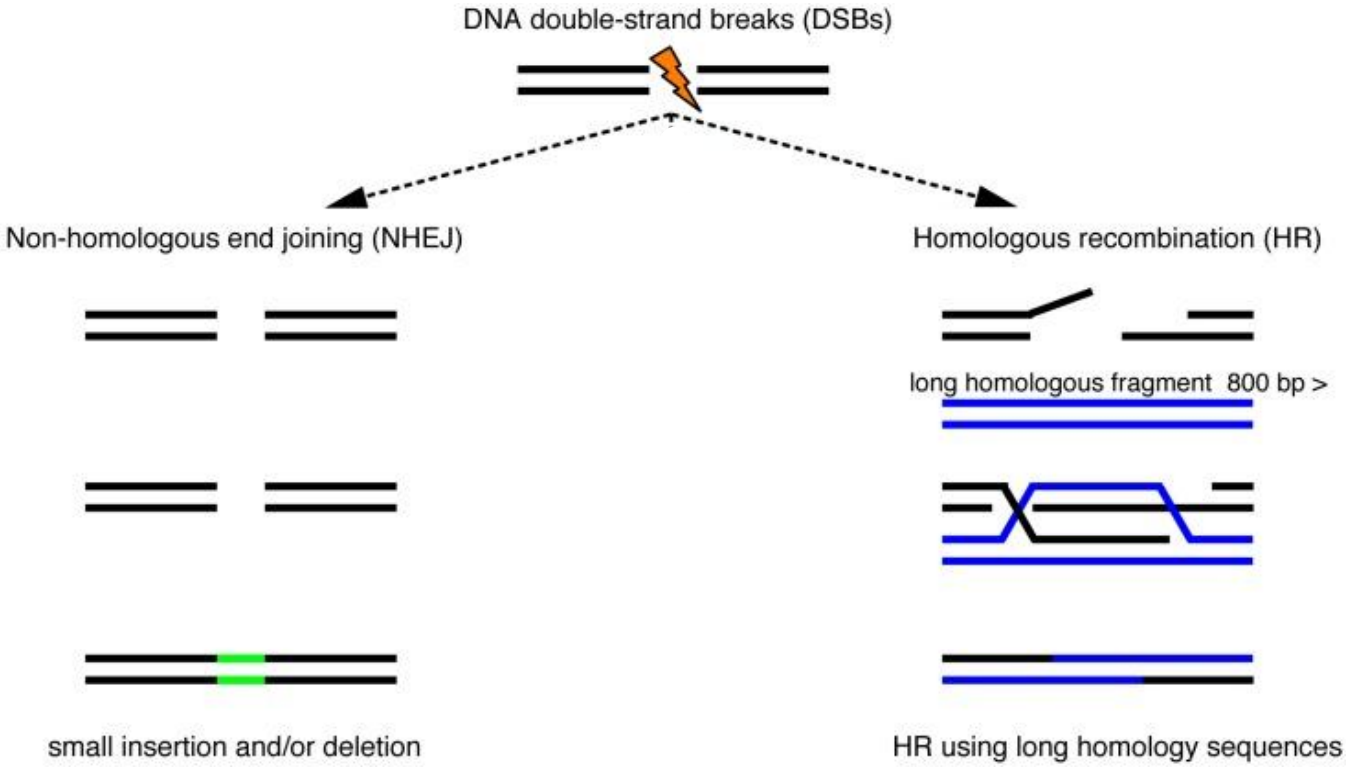
#nobelprize

THE NOBEL PRIZE

Registrato con Debut Home Edition. www.nchsoftware.com/capture/it

Targeted genome modification (TGM)

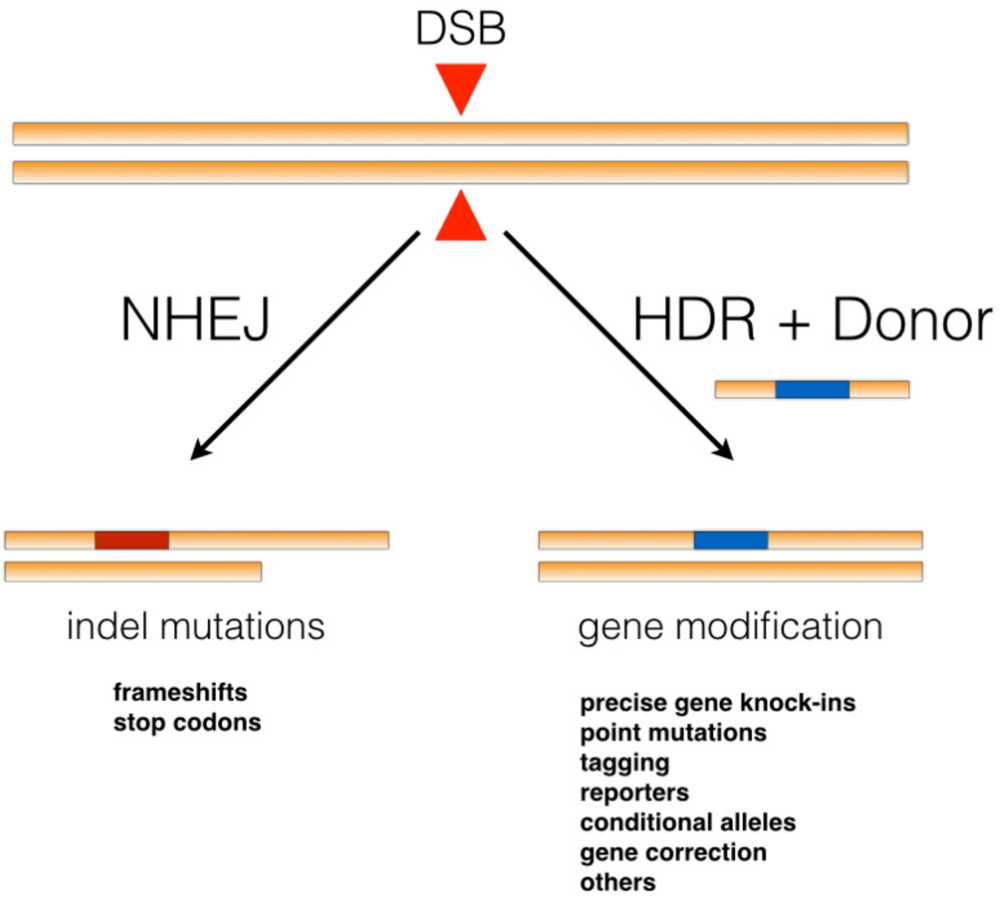
The deployment of engineered nucleases (Zinc finger nucleases, TALENs and the CRISPR/cas system) enabled a significant breakthrough in targeted genome editing and offers an unprecedented way of modifying the genome of organisms in an efficient and cost-effective way.



ZFNs: artificial nucleases that consist of a **synthetic ZFN** domain fused to a **Fok I cleavage domain**

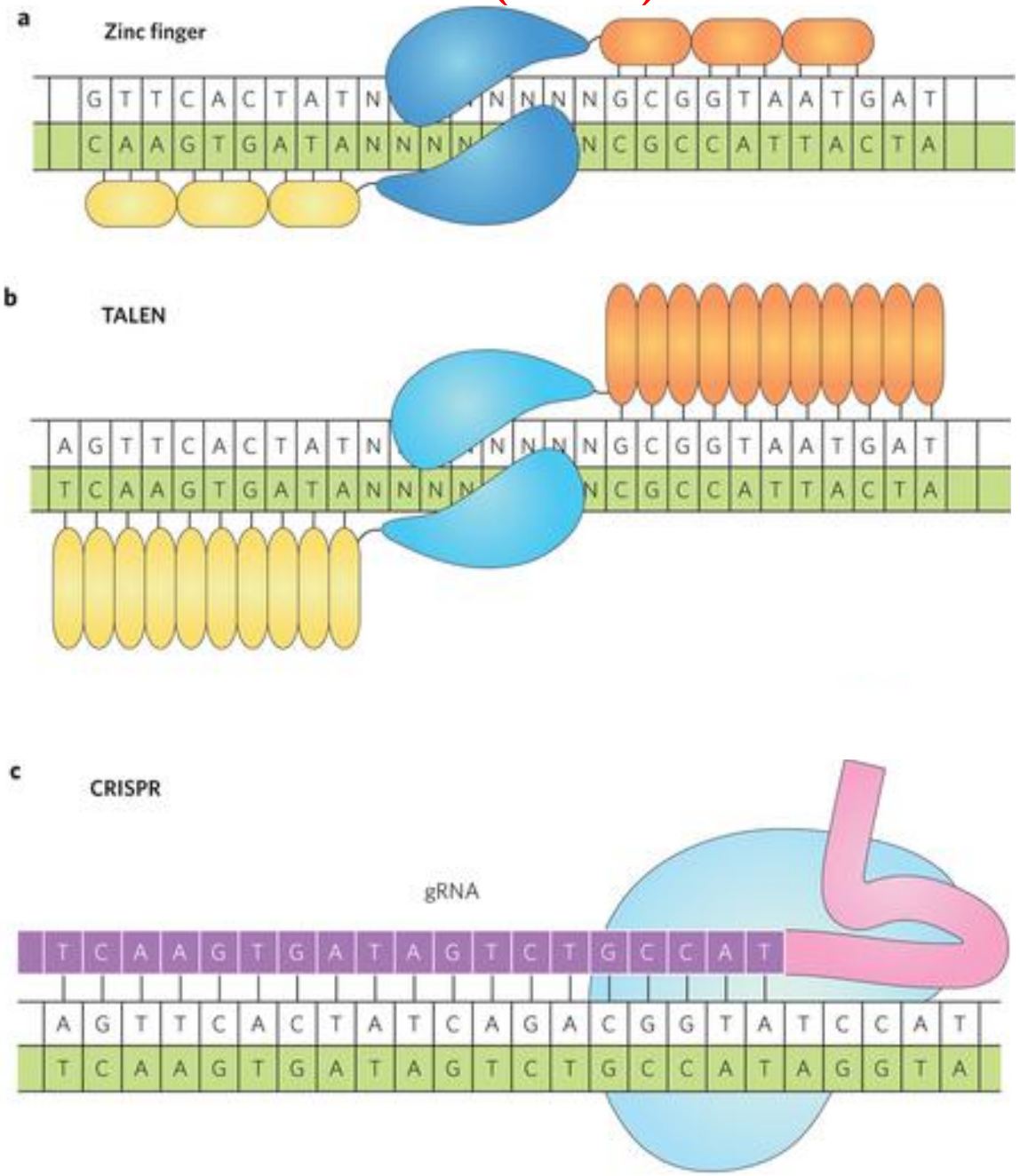
TALENs: transcription activator-like effector nucleases, consisting of an **engineered specific** (TALE) DNA binding domain and a **Fok I cleavage domain**

A crucial first step for performing targeted genome editing is **the creation of a DNA double-stranded break (DSB) at the genomic locus to be modified.** Nuclease-induced DSBs can be repaired by: **Non-Homologous End-Joining (NHEJ) and Homology-Directed Repair (HDR)**. **NHEJ can lead to the efficient introduction of insertion/deletion mutations** (indels) of various lengths. **HDR-mediated repair can be used to introduce specific point mutations or to insert desired sequences through recombination of the target locus with exogenously supplied DNA 'donor templates'**. With targeted nuclease-induced DSBs, the frequencies of these alterations are typically **greater than 1% and, in some cases, over 50%**.



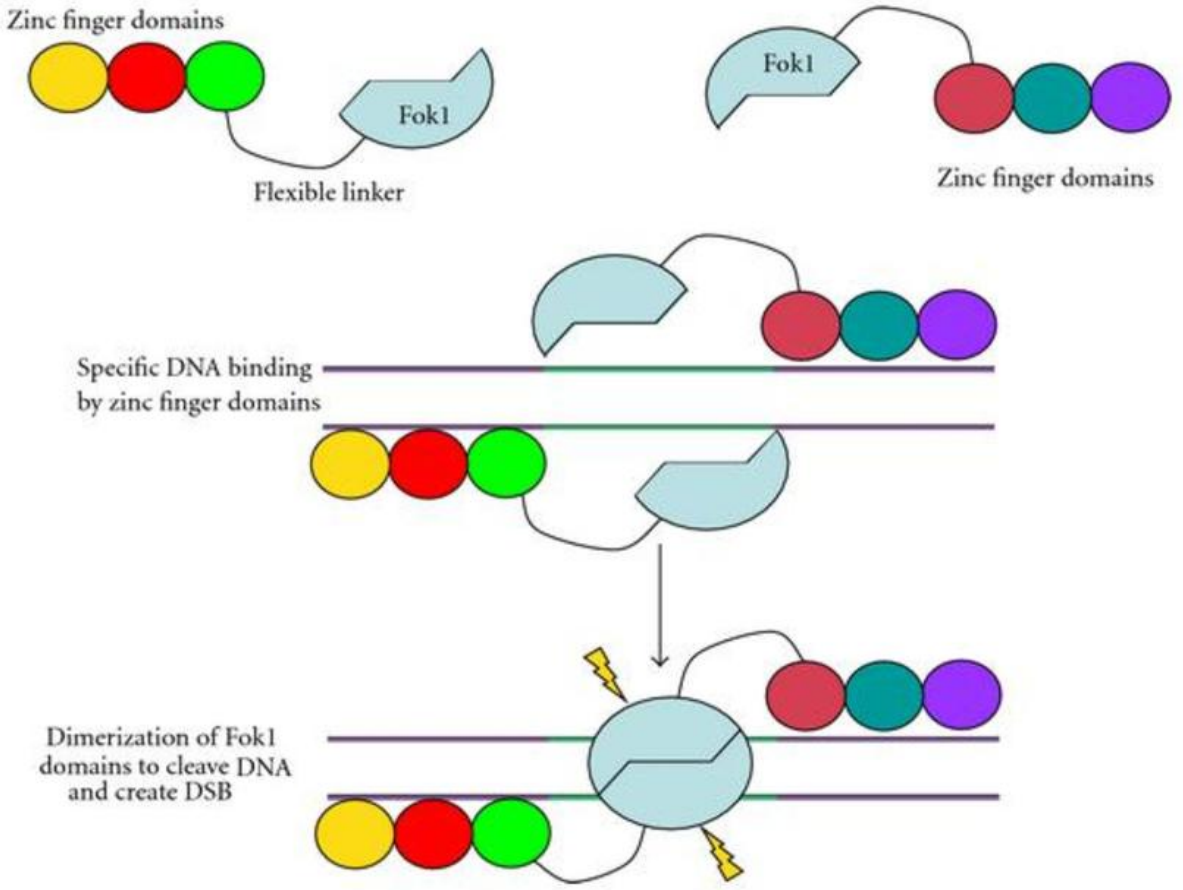
Targeted genome modification (TGM)

Site-directed nucleases. **a–c**, DNA nucleases bind to and cut DNA at specific locations. Each nuclease comprises a DNA-cutting domain (depicted in blue) and a DNA-targeting domain. **Zinc finger nucleases and Transcription Activator-Like Effector Nucleases** (TALENs) possess **protein-based DNA recognition domains** (depicted by yellow and orange ovals). **Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) nucleases** (**c**) rely on short-guide RNAs (gRNA) to locate the target DNA. The DNA recognition domain of all three nucleases can be engineered to target pre-determined sites in the genome for the purposes of genome editing.



The enzyme Fok1, naturally found in *Flavobacterium okeanokoites*, is a bacterial restriction endonuclease consisting of an N-terminal DNA-binding domain and a non-specific DNA cleavage domain at the C-terminal.

Fok1 domain requires dimerization to function.



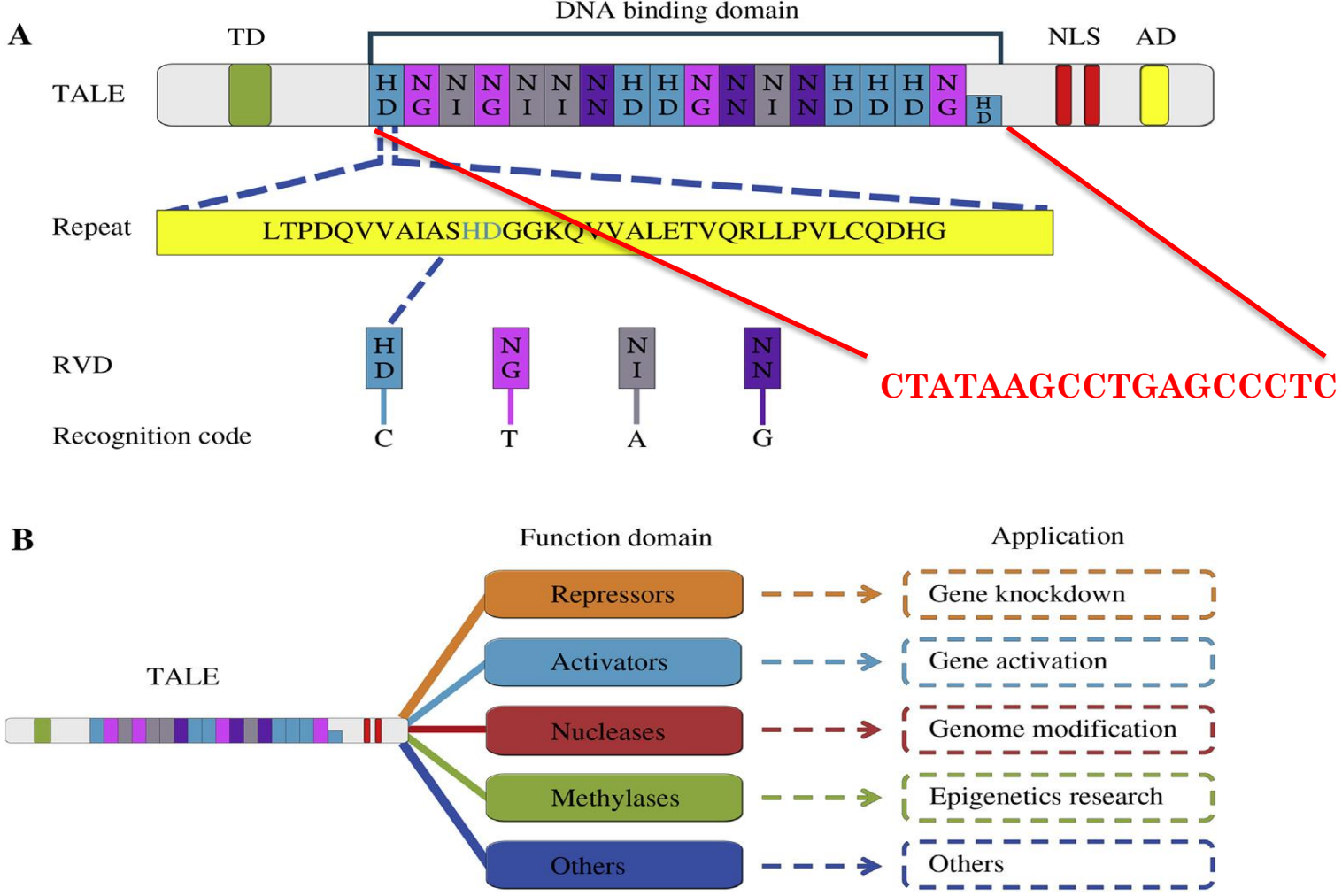
The customizable TALE DNA binding domain, composed of several nearly identical tandem repeat arrays, can target any given sequence according to a simple repeat variable di-residue (RVD)-nucleotide recognition code

- TALEs are major virulence factors secreted by the plant pathogenic bacterial genus *Xanthomonas*. They are injected into host cells through the secretion system and interfere with cellular activities by activating the transcription of specific target genes

- They have specific structural features, including secretion and translocation signals in the N-terminal region, nuclear localization signals (NLS) and an acidic transcription-activation domain (AD) in the C-terminal region and a central DNA binding domain (DBD) with 33-35 nearly identical long amino acid repeats, followed by the last module which contains only 20 amino acids ("half repeat")

- The Repeat Variable Di-residue (RVD) at positions 12 and 13 of each repeat dictates the specificity of repeat binding to a nucleotide in the DNA target

- The DNA binding specificity of a TALE is determined by its repeat number and the sequence of the RVD: the repeat number determines the length of the target sequence, while the RVD corresponds directly to the nucleotide in the target site.



A: TALE structure and DNA recognition code: a typical TALE structure (top) comprises an N-terminal translocation domain (TD), a central DNA binding domain (DBD), two nuclear localization signals (NLS) and a transcriptional activation domain (AD) in the C-terminal region. The DBD is composed of several tandem repeats which end with a half repeat. Each repeat consists of 34 nearly identical amino acids except for the central repeat variable di-residue (RVD) in positions 12 and 13. The one-to-one RVD-nucleotide code of the four common RVDs used for DNA specific targeting in TALENs is shown below. B: TALE proteins: an optimal TALE scaffold including the central DBD can be fused to some functional domain, such as repressor, activator, nuclease or methylase, to generate TALE-based designer proteins for site-specific modification of the genome.

1987 a Japanese group identifies **clustered regularly interspaced palindromic repeats** (CRISPRs) as a series of short direct repeats interspaced with short sequences in the genome of *Escherichia coli*

CRISPRs were later detected in numerous bacteria and archaea, and predictions were made about their possible roles in DNA repair or gene regulation

TGAAAATGGGAGGGAGTTCTACCGCAGAGGCGGGGGAACTCCAAGTGATATCCATCATCGCATCCAGTGCGCC (1,451)
(1,452) CGGTTTATCCCCGCTGATGCGGGGAACACCAGCGTCAGGCGTGAAATCTCACCGTCGTTGC (1,512)
(1,513) CGGTTTATCCCTGCTGGCGCGGGGAACTCTCGGTTCAGGCGTTGCAAACCTGGCTACCGGG (1,573)
(1,574) CGGTTTATCCCCGCTAACGCGGGGAACTCGTAGTCCATCATTCCACCTATGTCTGAACTCC (1,634)
(1,635) CGGTTTATCCCCGCTGGCGCGGGGAACTCG (1,664)

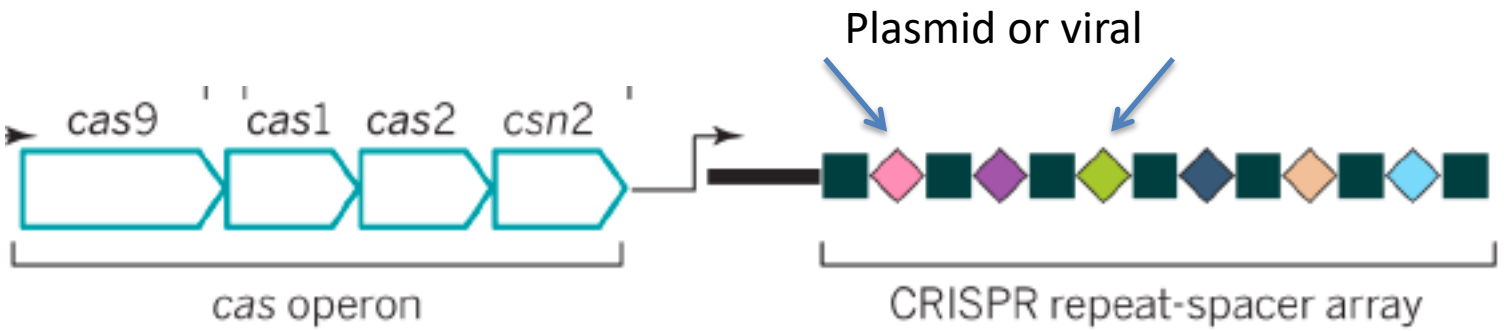
consensus: CGGTTTATCCCCGCT$^{GG}_{AA}$CGCGGGGAACTC

FIG. 5. Comparison of direct-repeat sequences consisting of 61 base pairs in the 3′-end flanking region of *iap*. The 29 highly conserved nucleotides, which contain a dyad symmetry of 14 base pairs (underlined), are shown at the bottom. Homologous nucleotides found in at least two DNA segments are shown in boldface type. The second translational termination codon is boxed. The nucleotide numbers are in parentheses.

Ishino et al., 1987

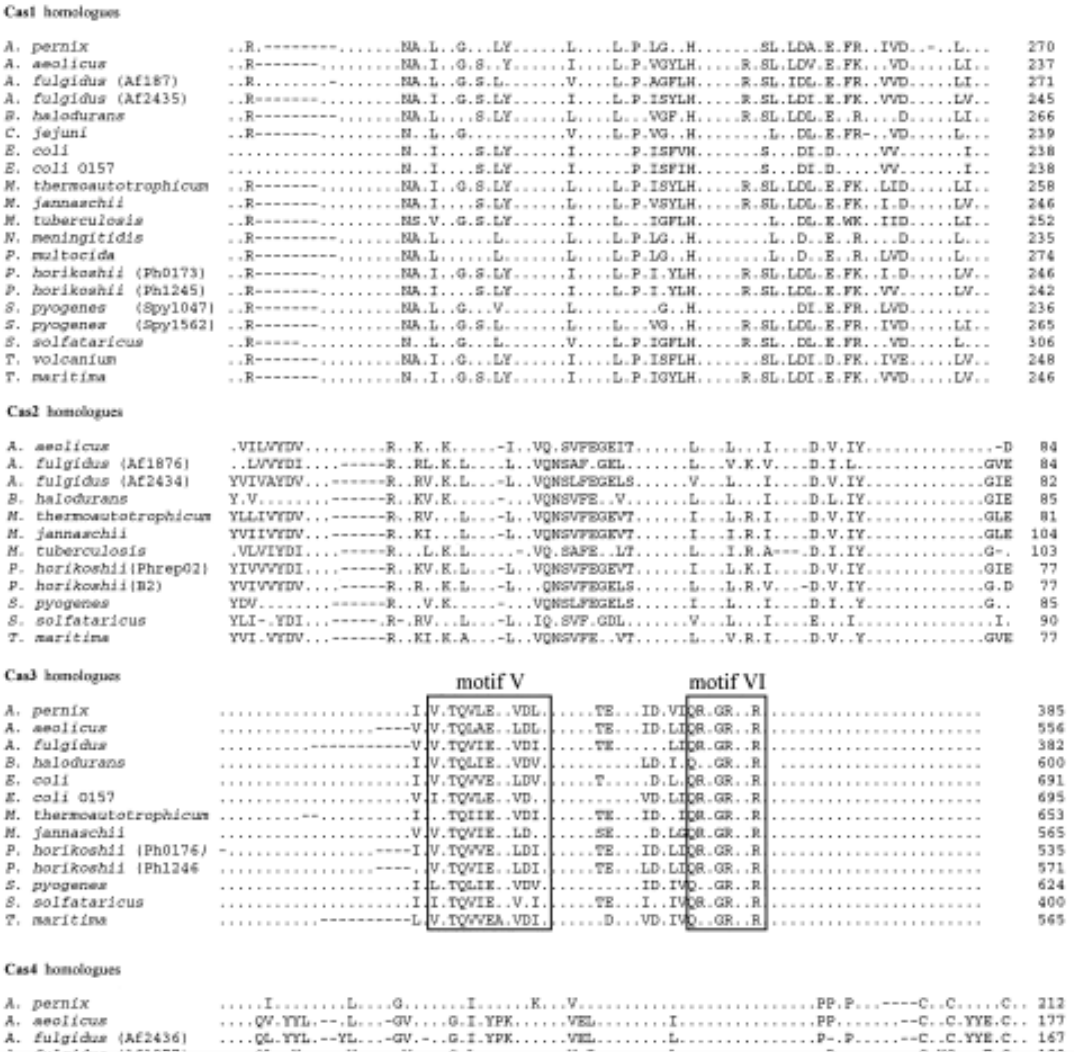
2005: sequence analyses indicate that many spacer sequences within CRISPRs derive from **plasmid and viral origins.**

CAS (CRISP-Associated) genes:



Cas (CRISPR-associated) genes encode proteins with putative nuclease and helicase domains

CRISPR-Cas is an adaptive defense system that might use antisense RNAs as memory signatures of past invasions. Alignment of the amino acid sequences of putative Cas proteins 1–4 from published genomes. The number indicates the last amino acid residue of the alignment. Only matching residues of identical or chemically related amino acid residues are given. The boxed regions in the Cas3 homologues are characteristic of helicases.



CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes

Rodolphe Barrangou,[1] Christophe Fremaux,[2] Hélène Deveau,[3] Melissa Richards,[1] Patrick Boyaval,[2] Sylvain Moineau,[3] Dennis A. Romero,[1] Philippe Horvath[2]*

Clustered regularly interspaced short palindromic repeats (CRISPR) are a distinctive feature of the genomes of most Bacteria and Archaea and are thought to be involved in resistance to bacteriophages. We found that, after viral challenge, bacteria integrated new spacers derived from phage genomic sequences. Removal or addition of particular spacers modified the phage-resistance phenotype of the cell. Thus, CRISPR, together with associated *cas* genes, provided resistance against phages, and resistance specificity is determined by spacer-phage sequence similarity.

CRISPR-Cas is an adaptive defense system
that might use antisense RNAs
as memory signatures of past invasions

Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes

Stan J. J. Brouns,[1*] Matthijs M. Jore,[1*] Magnus Lundgren,[1] Edze R. Westra,[1] Rik J. H. Slijkhuis,[1] Ambrosius P. L. Snijders,[2] Mark J. Dickman,[2] Kira S. Makarova,[3] Eugene V. Koonin,[3] John van der Oost[1†]

Prokaryotes acquire virus resistance by integrating short fragments of viral nucleic acid into clusters of regularly interspaced short palindromic repeats (CRISPRs). Here we show how virus-derived sequences contained in CRISPRs are used by CRISPR-associated (Cas) proteins from the host to mediate an antiviral response that counteracts infection. After transcription of the CRISPR, a complex of Cas proteins termed Cascade cleaves a CRISPR RNA precursor in each repeat and retains the cleavage products containing the virus-derived sequence. Assisted by the helicase Cas3, these mature CRISPR RNAs then serve as small guide RNAs that enable Cascade to interfere with virus proliferation. Our results demonstrate that the formation of mature guide RNAs by the CRISPR RNA endonuclease subunit of Cascade is a mechanistic requirement for antiviral defense.

Doudna, Charpentier, and Zhang

1. J. A. Hoffmann, *Nature* **426**, 33 (2003).
2. P. Schmid-Hempel, *Annu. Rev. Entomol.* **50**, 529 (2005).
3. Y. Moret, M. T. Siva-Jothy, *Proc. Biol. Sci.* **270**, 2475 (2003).
4. M. J. Brown, Y. Moret, P. Schmid-Hempel, *Parasitology* **126**, 253 (2003).

Materials and Methods
Figs. S1 to S8
Table S1

9 April 2010; accepted 22 July 2010
10.1126/science.1190689

Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease

Rachel E. Haurwitz,[1]* Martin Jinek,[1]* Blake Wiedenheft,[1,2] Kaihong Zhou,[1,2] Jennifer A. Doudna[1,2,3,4]†

Many bacteria and archaea contain clustered regularly interspaced short palindromic repeats (CRISPRs) that confer resistance to invasive genetic elements. Central to this immune system is the production of CRISPR-derived RNAs (crRNAs) after transcription of the CRISPR locus. Here, we identify the endoribonuclease (Csy4) responsible for CRISPR transcript (pre-crRNA) processing i *aeruginosa*. A 1.8 angstrom crystal structure of Csy4 bound to its cognate RNA rev sequence-specific interactions in the major groove of the crRNA repeat stem-loop. electrostatic contacts to the phosphate backbone, these enable Csy4 to bind selecti pre-crRNAs using phylogenetically conserved serine and histidine residues in the a recognition mechanism identified here explains sequence- and structure-specific p family of CRISPR-specific endoribonucleases.

In prokaryotes, fragments of foreign DNA are integrated into clustered regularly interspaced short palindromic repeat (CRISPR) loci that are transcribed as long RNAs containing a repetitive sequence element derived from the host (*1*–*6*). These CRISPR transcripts posttranscriptionally processe that serve as homing oligon the propagation of invading harboring cognate sequence

ncemag.org/ on April 23, 2018

ARTICLE

doi:10.1038/nature09886

CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III

Elitza Deltcheva[1,2], Krzysztof Chylinski[1,2]*, Cynthia M. Sharma[3]*, Karine Gonzales[2], Yanjie Chao[3,4], Zaid A. Pirzada[2], Maria R. Eckert[2], Jörg Vogel[3,4] & Emmanuelle Charpentier[1,2]

CRISPR/Cas systems constitute a widespread class of immunity systems that protect bacteria and archaea against phages and plasmids, and commonly use repeat/spacer-derived short crRNAs to silence foreign nucleic acids in a sequence-specific manner. Although the maturation of crRNAs represents a key event in CRISPR activation, the responsible endoribonucleases (CasE, Cas6, Csy4) are missing in many CRISPR/Cas subtypes. Here, differential RNA sequencing of the human pathogen *Streptococcus pyogenes* uncovered tracrRNA, a *trans*-encoded small RNA with 24-nucleotide complementarity to the repeat regions of crRNA precursor transcripts. We show that tracrRNA directs the maturation of crRNAs by the activities of the widely conserved endogenous RNase III and the CRISPR-associated Csn1 protein; all these components are essential to protect *S. pyogenes* against prophage-derived DNA. Our study reveals a novel pathway of small guide RNA maturation and the first example of a host factor (RNase III) required for bacterial RNA-mediated immunity against invaders.

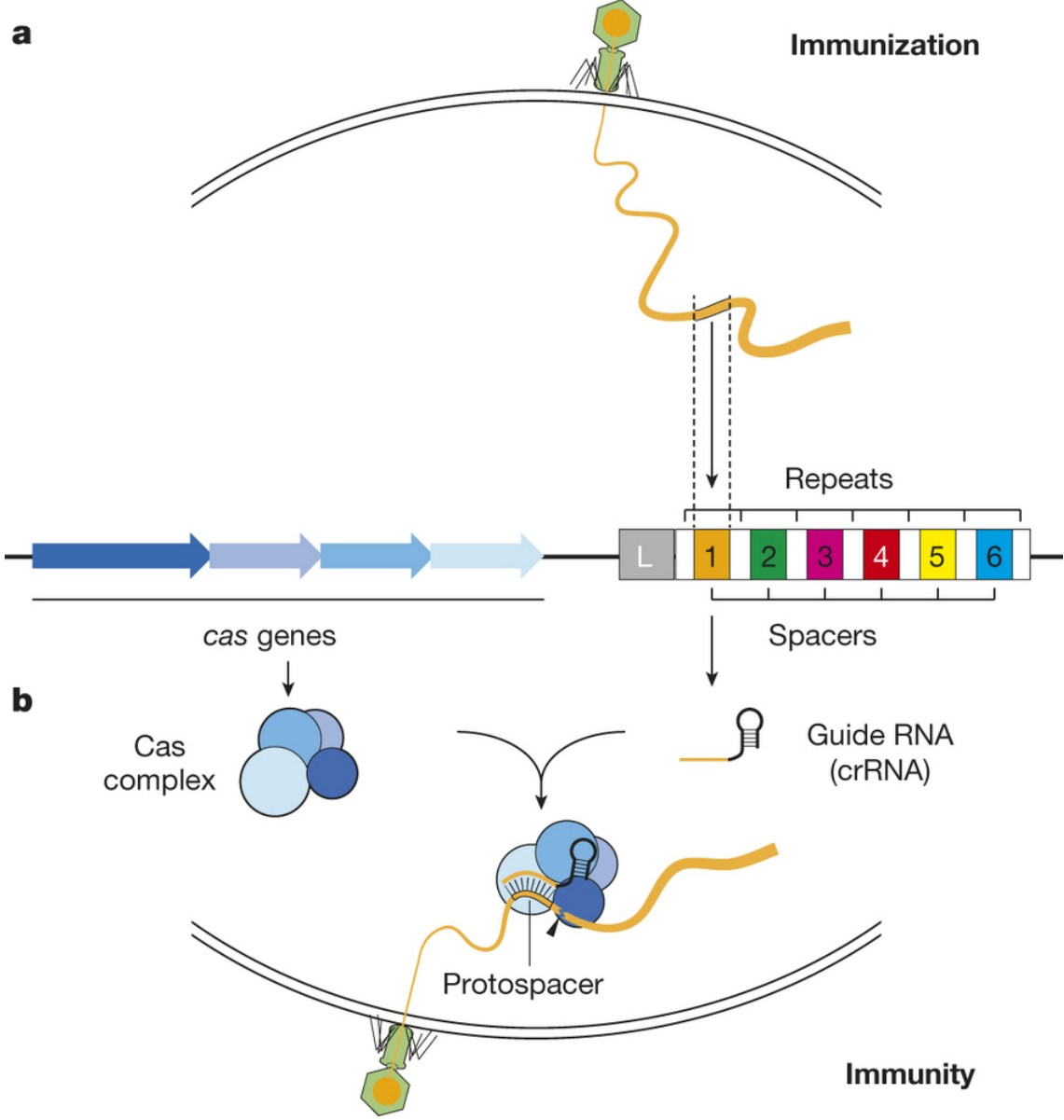
Organisms of all kingdoms of life have evolved RNA-guided immunity mechanisms to protect themselves against genome invaders[1–6]. In bacteria and archaea, CRISPR/Cas (clustered, regularly interspaced short palindromic repeats/CRISPR-associated proteins) constitutes

To examine the *in vivo* expression of CRISPR01 and CRISPR02, we analysed *S. pyogenes* strain SF370 (M1 serotype) by differential RNA sequencing (dRNA-seq)[39]. The most abundantly recovered small RNA species were CRISPR01 crRNAs originating from a ~511-nucleotide

CRISPR loci are a cluster of short DNA repeats (white boxes) separated by **equally short spacer sequences of phage and plasmid origin**. This repeat/spacer array is **flanked by an operon of CRISPR-associated (cas)** genes that encode the machinery for the immunization and immunity stages of the system. The CRISPR array is preceded by a leader sequence (grey box) containing the promoter for its expression.

***Immunization stage*:** spacer sequences are captured upon entry of the foreign DNA into the cell and integrated into the first position of the CRISPR array.

***Immunity stage*:** spacer used to target invading DNA that carries a cognate sequence for destruction. **Spacers are transcribed and processed into small CRISPR RNAs (crRNAs).** These small RNAs act as **antisense guides for Cas RNA-guided nucleases** that cleave the target sequence (black arrowhead) in the invader's genome

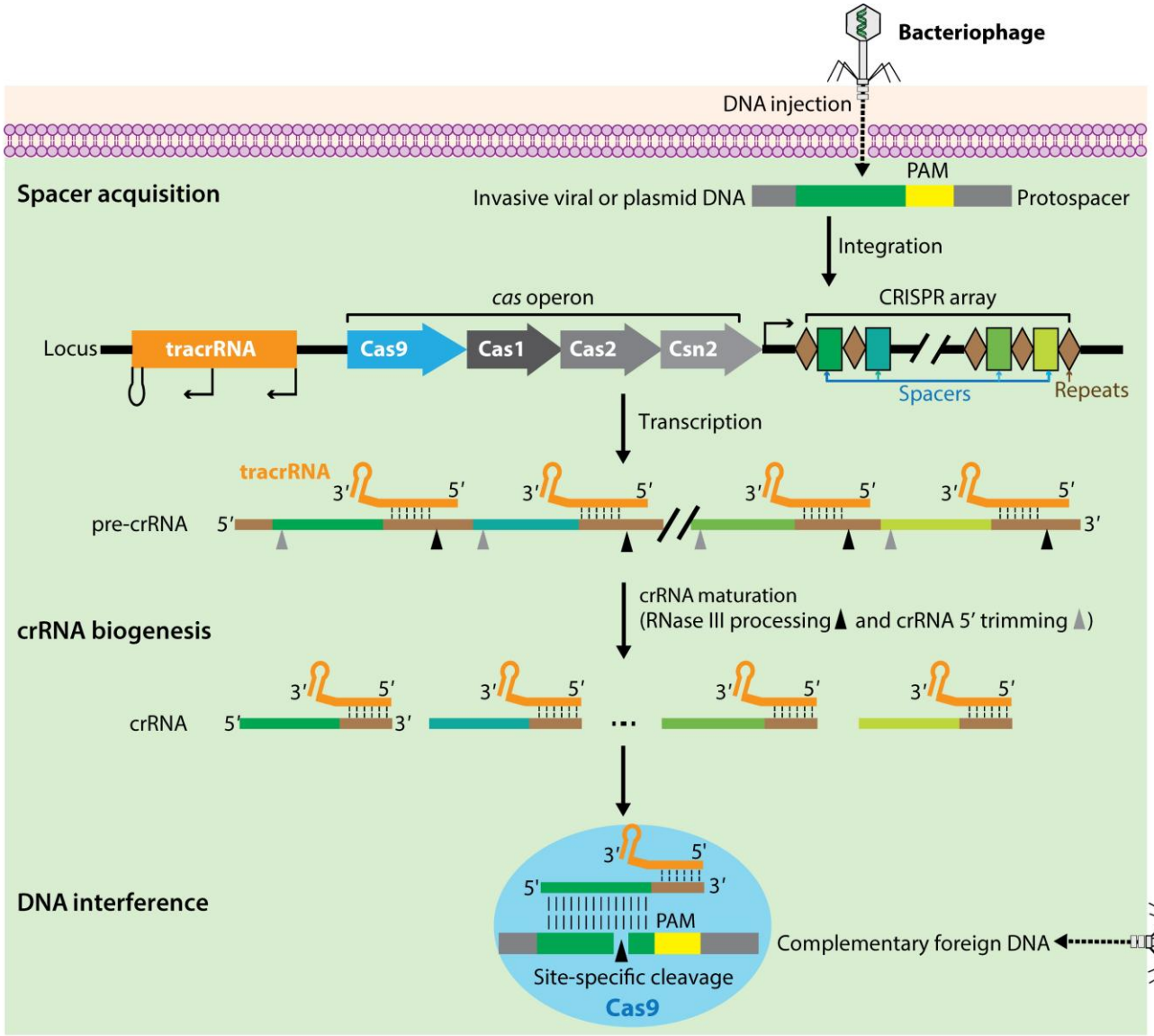


Preceding the cas operon is the **trans-activating CRISPR RNA (tracrRNA) gene, encoding a unique noncoding RNA homologous to the repeat sequences**.

Upon phage infection, a new spacer derived from the phage is integrated into the CRISPR array by the acquisition machinery (Cas1, Cas2, and Csn2). It is cotranscribed into a precursor CRISPR RNA (pre-crRNA). **The tracrRNA is transcribed separately and anneals to the pre-crRNA repeats for crRNA maturation by RNase III.**

During interference, the mature **crRNA–tracrRNA structure engages Cas9 endonuclease and further directs it to cleave foreign DNA containing a 20-nt crRNA complementary sequence preceding the PAM sequence (protospacer adjacent motif)**.

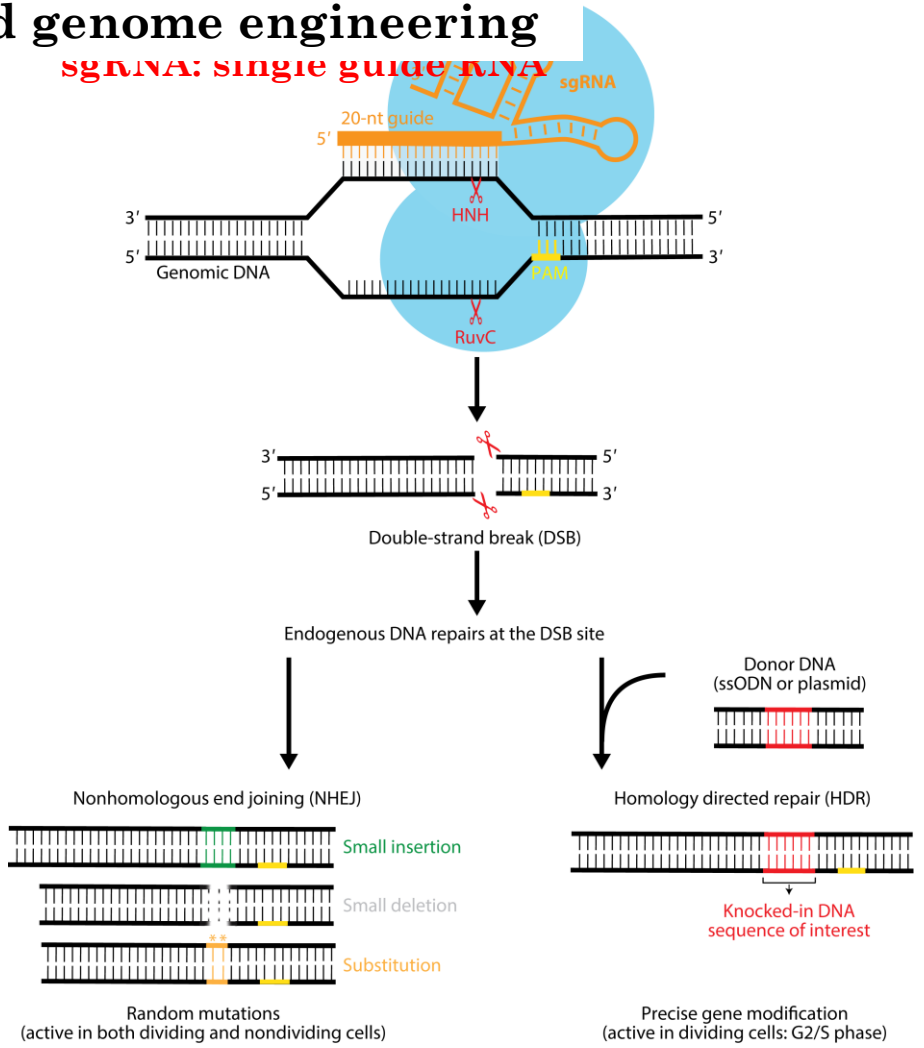
CRISPR–Cas system: an array of repeats interspaced by short stretches of nonrepetitive sequences (spacers), and a set of CRISPR-associated (cas) genes.



CRISPR–Cas9–mediated genome engineering

The sgRNA or crRNA: single guide RNA directs Cas9 to DNA sequence in the genome through a user-defined 20-nt guide RNA sequence guiding Cas9 to introduce a double-strand break (DSB) in targeted genomic DNA. The DSB generated by two distinct Cas9 nuclease domains is repaired by host-mediated DNA repair mechanisms. In the absence of a repair template, the prevalent error-prone nonhomologous end joining (NHEJ) pathway is activated and causes random indels or even substitutions at the DSB site, frequently resulting in the disruption of gene function. *In the presence of a donor template* containing a sequence of interest flanked by homology arms, the error-free homology directed repair (HDR) pathway can be initiated to create desired mutations through homologous recombination.

CRISPR–Cas9 DNA targeting can be uncoupled from cleavage activity by mutating the catalytic residues in the HNH and RuvC nuclease domains, making it a versatile platform for many other applications beyond genome editing.



Error-prone NHEJ
- Indel mutations
- Frameshift mutations
- Large deletions or inversions using two adjacent DSBs
- Loss-of-function screens
- Genomic rearrangements
- NHEJ-mediated homology-independent knock-in

Other applications
- Gene targeting (dCas9–effector):
  - transcriptional regulation
  - epigenetic modification
  - live-cell imaging
  - nucleotide editing
- Genetic screens/drug screens
- Ligation-mediated gene editing by double Cas9 nickases (D10A)

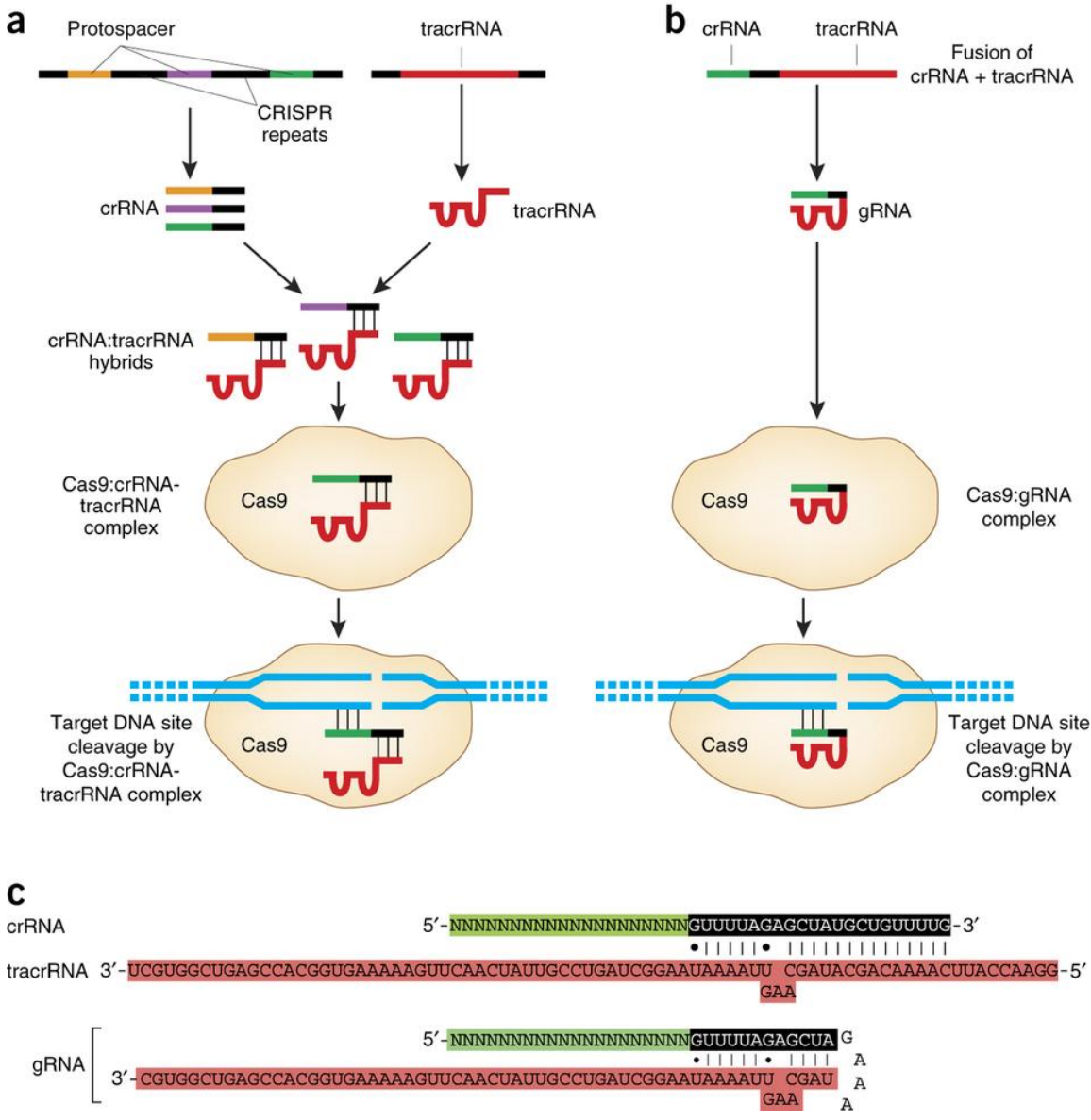
High-fidelity HDR
- Gene insertion/correction/replacement
- Precise point mutations
- Precise gene knockout
- Conditional alleles (Cre-*lox*P, etc.)
- Introduction of tags, reporters, etc.
- Gain-of-function mutations

Jiang F, Doudna JA. 2017.
Annu. Rev. Biophys. 46:505–29

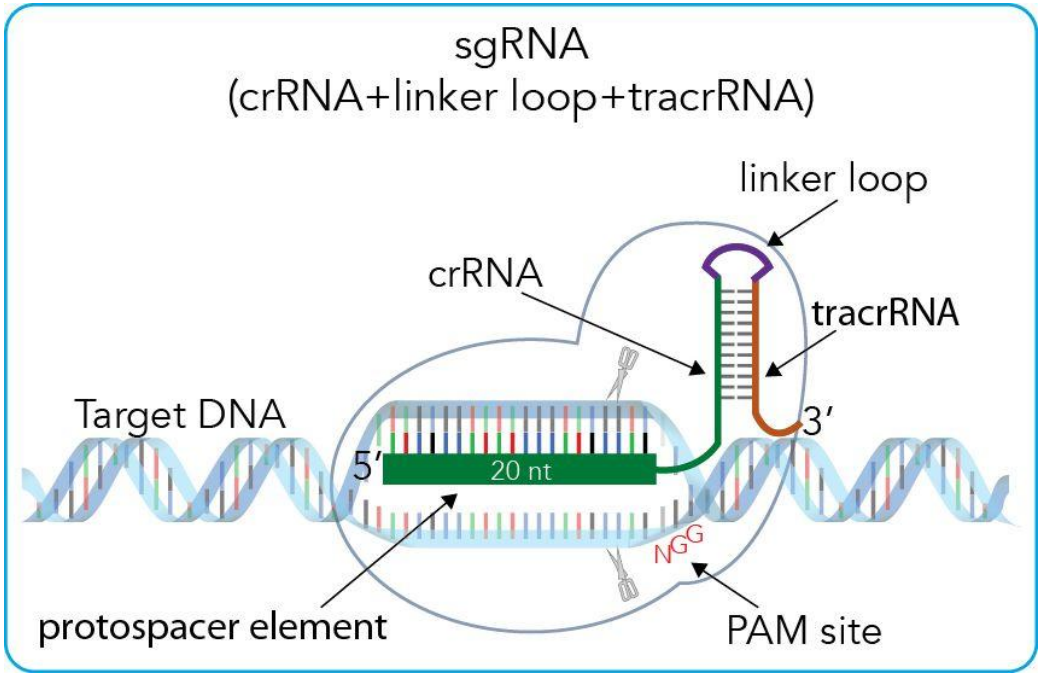
Naturally occurring and engineered CRISPR-Cas systems.

(**a**) Naturally occurring CRISPR systems incorporate foreign DNA sequences into CRISPR arrays, which then produce crRNAs bearing "protospacer" regions complementary to the foreign DNA site. crRNAs hybridize to tracrRNAs (also encoded by the CRISPR system) and this pair of RNAs can associate with the Cas9 nuclease. crRNA-tracrRNA:Cas9 complexes recognize and cleave foreign DNAs bearing the protospacer sequences.

(**b**) Engineered CRISPR-Cas system utilizes a fusion between a crRNA and part of the tracrRNA sequence. This single **gRNA** complexes with Cas9 to mediate cleavage of target DNA sites that are complementary to the 5′ 20 nt of the gRNA and that lie next to a PAM sequence. (**c**) Example sequences of a crRNA-tracrRNA hybrid and a gRNA.



A CRISPR/Cas9 nuclease system requires two components: a Cas enzyme for cutting the target sequence and a single guide RNA (sgRNA), which binds to the target sequence of 20-base pair (bp). The target sequence (complementary to the sgRNA sequence) is followed by two cytosine nucleotides because the sgRNA binds best when the opposite DNA strand is comprised of any nucleotide followed by two guanines (NGG). This sequence is called a Protospacer Adjacent Motif (PAM) sequence. The PAM varies depending on the origin of Cas9.



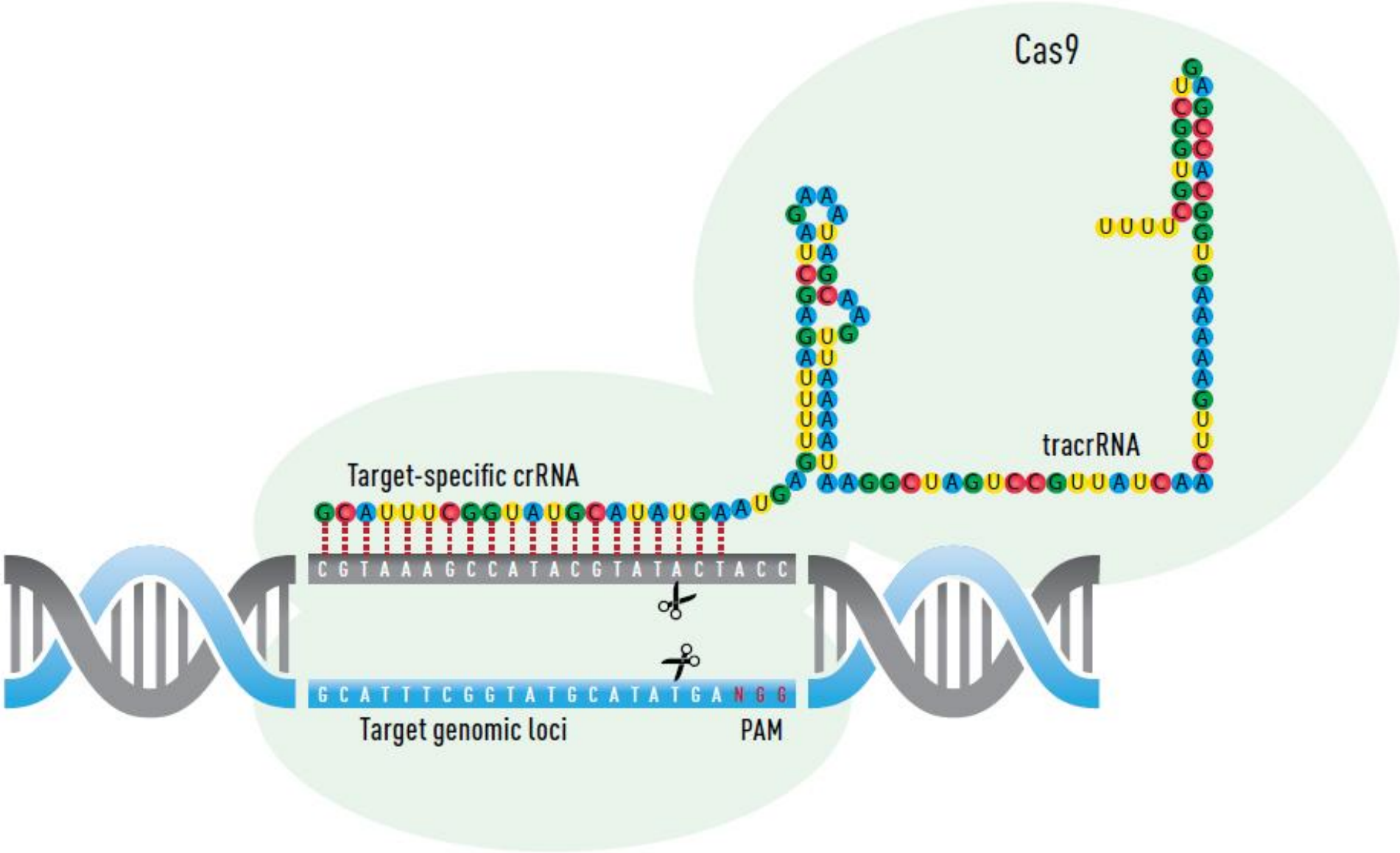
crRNA (CRISPR RNA)—contains both the 20 base protospacer element and additional nucleotides which are complementary to the tracrRNA.

tracrRNA (transactivating RNA)—hybridizes to the crRNA and binds to the CAS9 protein activating the complex to creating double-stranded breaks at specific sites within genomic sequence.

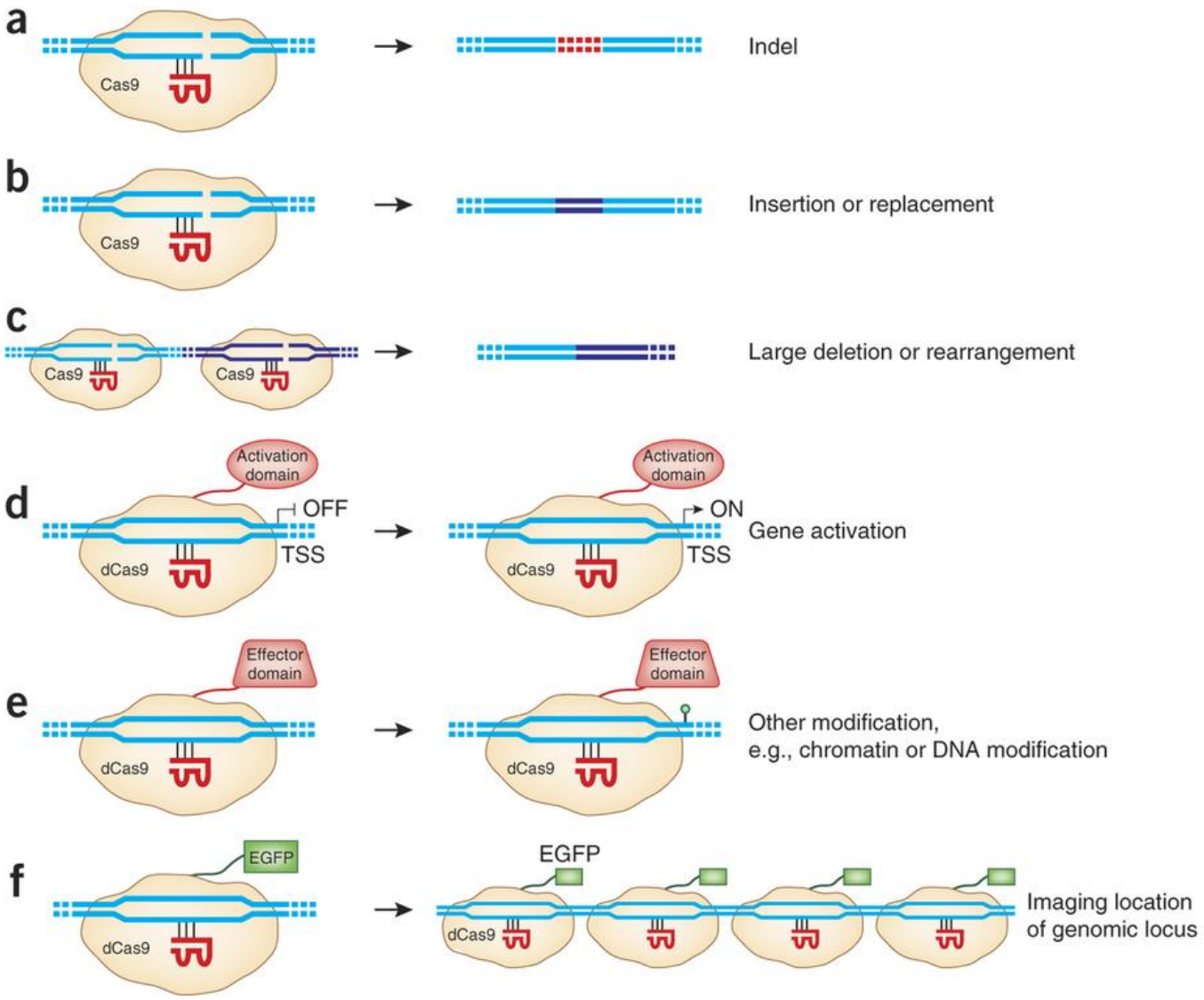
sgRNA (single-guide RNA)—combines the tracrRNA and crRNA, which are separate molecules in the native CRISPR/Cas9 system in *S. pyogenes*, into a single RNA construct, simplifying the components needed to use CRISPR/Cas9 for genome editing. A linker loop sequence is included between the two.

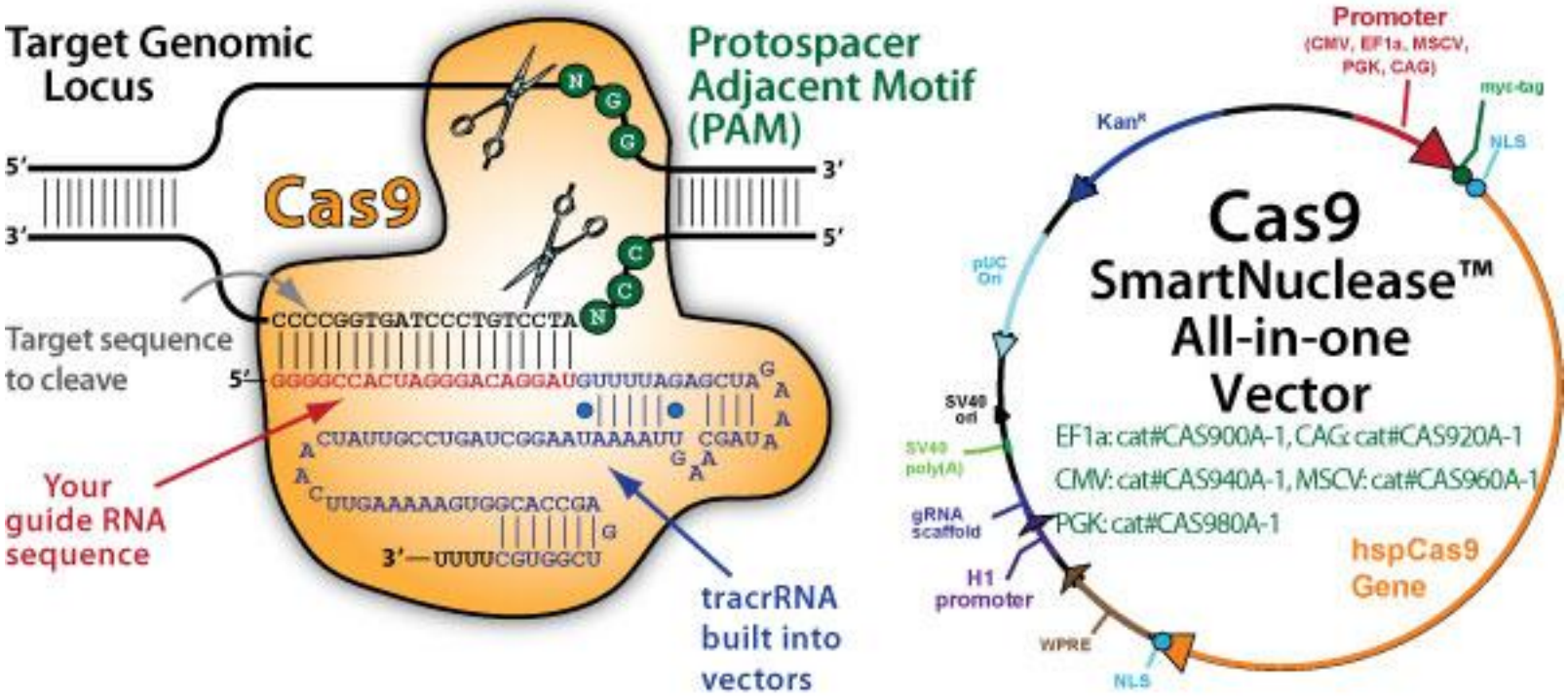
protospacer element — the portion of the crRNA (or sgRNA) that is complementary to the genomic DNA target sequence; usually 20 nucleotides in length.

RNA trigger—a generic term for the RNAs that activate the CRIPSR/CAS9 complex. They can be sgRNA or crRNA/tracrRNA.


Cas9
tracrRNA
Target-specific crRNA
Target genomic loci
PAM

Overview of various Cas9-based applications. (**a**,**b**) gRNA-directed Cas9 nuclease can induce indel mutations (**a**) or specific sequence replacement or insertion (**b**). (**c**) Pairs of gRNA-directed Cas9 nucleases can stimulate large deletions or genomic rearrangements (e.g., inversions or translocations). (**d**–**f**) gRNA-directed dCas9 can be fused to activation domains (**d**) to mediate upregulation of specific endogenous genes, heterologous effector domains (**e**) to alter histone modifications or DNA methylation, or fluorescent proteins (**f**) to enable imaging of specific genomic loci.




Target Genomic Locus
Protospacer Adjacent Motif (PAM)
Cas9
5′
3′
3′
5′
N G G
C C N
CCCCGGTGATCCCTGTCCTA
Target sequence to cleave
5′ GGGGCCACUAGGGACAGGAUGUUUUAGAGCUA
CUAUUGCCUGAUCGGAAUAAAAUU CGAUA
UUGAAAAAGUGGCACCGA
3′ UUUUCGUGGCU
Your guide RNA sequence
tracrRNA built into vectors
Promoter
(CMV, EF1a, MSCV, PGK, CAG)
myc-tag
NLS
Kan^R
Cas9
SmartNuclease™
All-in-one
Vector
pUC Ori
SV40 ori
SV40 poly(A)
gRNA scaffold
H1 promoter
WPRE
NLS
EF1a: cat#CAS900A-1, CAG: cat#CAS920A-1
CMV: cat#CAS940A-1, MSCV: cat#CAS960A-1
PGK: cat#CAS980A-1
hspCas9
Gene