

## REVIEW SUMMARY

## GENOME EDITING

# The new frontier of genome engineering with CRISPR-Cas9

Jennifer A. Doudna\* and Emmanuelle Charpentier\*

**BACKGROUND:** Technologies for making and manipulating DNA have enabled advances in biology ever since the discovery of the DNA double helix. But introducing site-specific modifications in the genomes of cells and organisms remained elusive. Early approaches relied on the principle of site-specific recognition of DNA sequences by oligonucleotides, small molecules, or self-splicing introns. More recently, the site-directed zinc finger nucleases (ZFNs) and TAL effector nucleases (TALENs) using the principles of DNA-protein recognition were developed. However, difficulties of protein design, synthesis, and validation remained a barrier to

widespread adoption of these engineered nucleases for routine use.

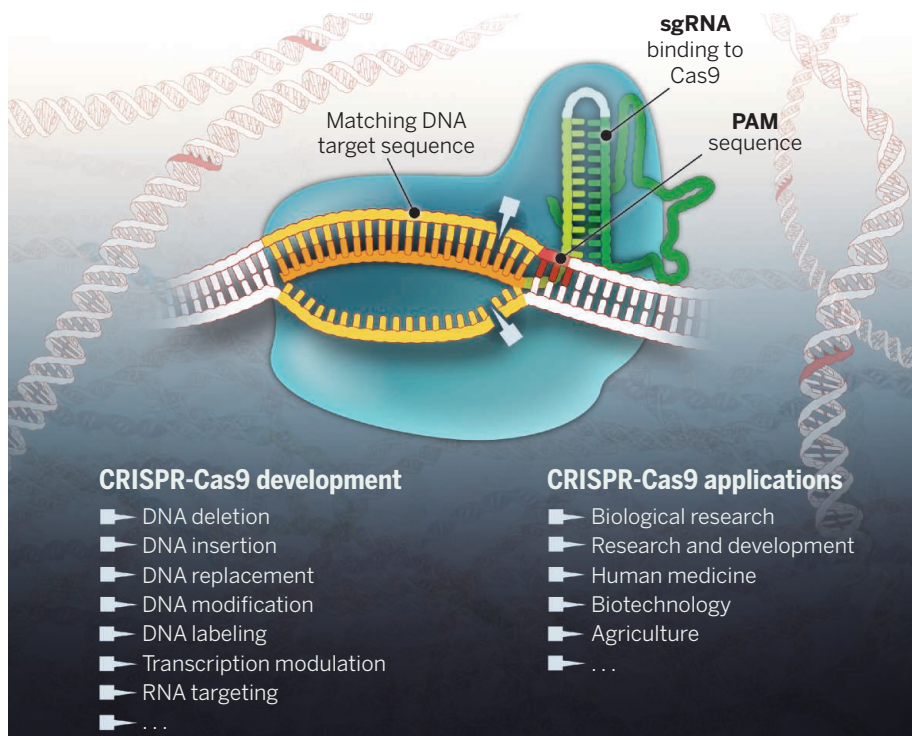
**ADVANCES:** The field of biology is now experiencing a transformative phase with the advent of facile genome engineering in animals and plants using RNA-programmable CRISPR-Cas9. The CRISPR-Cas9 technology originates from type II CRISPR-Cas systems, which provide bacteria with adaptive immunity to viruses and plasmids. The CRISPR-associated protein Cas9 is an endonuclease that uses a guide sequence within an RNA duplex, tracrRNA:crRNA, to form base pairs with DNA target sequences, enabling Cas9 to introduce a site-specific double-strand break

in the DNA. The dual tracrRNA:crRNA was engineered as a single guide RNA (sgRNA) that retains two critical features: a sequence at the 5' side that determines the DNA target site by Watson-Crick base-pairing and a duplex RNA structure at the 3' side that binds to Cas9. This finding created a simple two-component system in which changes in the guide sequence of the sgRNA program Cas9 to target any DNA sequence of interest. The simplicity of CRISPR-Cas9 programming, together with a unique DNA cleaving mechanism, the capacity for multiplexed target recognition, and the existence of many natural type II CRISPR-Cas system variants, has enabled remarkable developments using this cost-effective and easy-to-use technology to precisely and efficiently target, edit, modify, regulate, and mark genomic loci of a wide array of cells and organisms.

**OUTLOOK:** CRISPR-Cas9 has triggered a revolution in which laboratories around the world are using the technology for innovative applications in biology. This Review illustrates the power of the technology to systematically analyze gene functions in mammalian cells, study genomic rearrangements and the progression of cancers or other diseases, and potentially correct genetic mutations responsible for inherited disorders. CRISPR-Cas9 is having a major impact on functional genomics conducted in experimental systems. Its application in genome-wide studies will enable large-scale screening for drug targets and other phenotypes and will facilitate the generation of engineered animal models that will benefit pharmacological studies and the understanding of human diseases. CRISPR-Cas9 applications in plants and fungi also promise to change the pace and course of agricultural research. Future research directions to improve the technology will include engineering or identifying smaller Cas9 variants with distinct specificity that may be more amenable to delivery in human cells. Understanding the homology-directed repair mechanisms that follow Cas9-mediated DNA cleavage will enhance insertion of new or corrected sequences into genomes. The development of specific methods for efficient and safe delivery of Cas9 and its guide RNAs to cells and tissues will also be critical for applications of the technology in human gene therapy. ■

## ON OUR WEB SITE

Read the full article at <http://dx.doi.org/10.1126/science.1258096>



**The Cas9 enzyme (blue) generates breaks in double-stranded DNA by using its two catalytic centers (blades) to cleave each strand of a DNA target site (gold) next to a PAM sequence (red) and matching the 20-nucleotide sequence (orange) of the single guide RNA (sgRNA).** The sgRNA includes a dual-RNA sequence derived from CRISPR RNA (light green) and a separate transcript (tracrRNA, dark green) that binds and stabilizes the Cas9 protein. Cas9-sgRNA-mediated DNA cleavage produces a blunt double-stranded break that triggers repair enzymes to disrupt or replace DNA sequences at or near the cleavage site. Catalytically inactive forms of Cas9 can also be used for programmable regulation of transcription and visualization of genomic loci.

The list of author affiliations is available in the full article online.  
\*Corresponding author. E-mail: [doudna@berkeley.edu](mailto:doudna@berkeley.edu) (J.A.D.); [emmanuelle.charpentier@helmholtz-hzi.de](mailto:emmanuelle.charpentier@helmholtz-hzi.de) (E.C.)  
Cite this article as J. A. Doudna, E. Charpentier, *Science* **346**, 1258096 (2014). DOI:10.1126/science.1258096

## REVIEW

## GENOME EDITING

# The new frontier of genome engineering with CRISPR-Cas9

Jennifer A. Doudna<sup>1,2,3\*</sup> and Emmanuelle Charpentier<sup>4,5,6\*</sup>

The advent of facile genome engineering using the bacterial RNA-guided CRISPR-Cas9 system in animals and plants is transforming biology. We review the history of CRISPR (clustered regularly interspaced palindromic repeat) biology from its initial discovery through the elucidation of the CRISPR-Cas9 enzyme mechanism, which has set the stage for remarkable developments using this technology to modify, regulate, or mark genomic loci in a wide variety of cells and organisms from all three domains of life. These results highlight a new era in which genomic manipulation is no longer a bottleneck to experiments, paving the way toward fundamental discoveries in biology, with applications in all branches of biotechnology, as well as strategies for human therapeutics.

Technologies for making and manipulating DNA have enabled many of the advances in biology over the past 60 years. This era began with the discovery of the DNA double helix and continued with the development of chemical methods for solid-phase DNA synthesis, enabling detection and exploration of genome organization. Enzymes (including polymerases, ligases, and restriction endonucleases) and the polymerase chain reaction (PCR) provided ways to isolate genes and gene fragments, as well as to introduce mutations into genes in vitro, in cells, and in model organisms. The advent of genomic sequencing technologies and the rapid generation of whole-genome sequencing data for large numbers and types of organisms, including humans, has been one of the singular advances of the past two decades. Now, the RNA-guided enzyme Cas9, which originates from the CRISPR-Cas adaptive bacterial immune system, is transforming biology by providing a genome engineering tool based on the principles of Watson-Crick base pairing. Ease of use and efficiency have led to rapid adoption by laboratories around the world. Below we discuss the history and biology of CRISPR systems, describe the molecular mechanisms underlying genome editing by Cas9, and review the rapid advances in applications of this technology since its initial publication in 2012.

<sup>1</sup>Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. <sup>2</sup>Department of Chemistry, University of California, Berkeley, CA 94720, USA. <sup>3</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. <sup>4</sup>Department of Regulation in Infection Biology, Helmholtz Centre for Infection Research, D-38124 Braunschweig, Germany. <sup>5</sup>Laboratory for Molecular Infection Medicine Sweden, Umeå Centre for Microbial Research, Department of Molecular Biology, Umeå University, S-90187 Umeå, Sweden. <sup>6</sup>Hannover Medical School, D-30625 Hannover, Germany.

\*Corresponding author. E-mail: doudna@berkeley.edu (J.A.D.); emmanuelle.charpentier@helmholtz-hzi.de (E.C.)

## Genome engineering—A decades-long goal

Ever since the discovery of the DNA double helix, researchers and clinicians have been contemplating the possibility of making site-specific changes to the genomes of cells and organisms. Many of the earliest approaches to what has been referred to as genome editing relied on the principle of site-specific recognition of DNA sequences (Fig. 1). The study of natural DNA repair pathways in bacteria and yeast, as well as the mechanisms of DNA recombination (1–5), revealed that cells have endogenous machinery to repair double-strand DNA breaks (DSBs) that would otherwise be lethal (6–9). Thus, methods for introducing precise breaks in the DNA at sites where changes are to be introduced was recognized as a valuable strategy for targeted genomic engineering.

Early approaches to such targeted DNA cleavage took advantage of DNA base pair recognition by oligonucleotides or small molecules. Building on the original description of triple helix formation by Rich and colleagues in the late 1950s (10, 11), oligonucleotides coupled to chemical cleavage or cross-linking reagents such as bleomycin and psoralen were shown to be useful for site-specific chromosome modification in yeast and mammalian cells (12–17). Other methods for chemical recognition of DNA sequences, such as peptide nucleic acids (PNAs) and polyamides, were shown to enable targeted binding of chromosomal loci that could be modified if the chemical recognition agent was coupled to a cleavage reagent such as bleomycin (18–20). Another strategy that relied on nucleic acid base pairing was the use of self-splicing introns to change sequences at the DNA (21, 22) or RNA (23) level. Although these approaches did not lead to robust methods, they demonstrated the utility of base pairing for site-specific genome modification.

The use of self-splicing introns for genome editing also suggested the possibility of using intron-encoded nucleases—homing endonucleases—that are capable of site-specific DNA cleavage and integration of the intron sequence. By inserting desired sequences into the intron first, researchers could incorporate selected genetic information into a genome at sites recognized by the homing endonuclease (24, 25). At around the same time, the initial reports of zinc finger-mediated DNA binding (26, 27) led to the creation of modular DNA recognition proteins that, when coupled to the sequence-independent nuclease domain of the restriction enzyme FokI, could function as site-specific nucleases (28). When designed to recognize a chromosomal sequence, such zinc finger nucleases (ZFNs) were found to be effective at inducing genomic sequence changes in *Drosophila* and mammalian cells (29, 30). Although ZFNs are effective genome editing reagents for some experiments, they were not widely adopted because of the difficulty inherent in designing and validating such proteins for a specific DNA locus of interest. Thus, the field was primed for the first reports of transcription activator-like (TAL) effectors, which occur naturally in bacteria that infect plants, enabling rapid creation of FokI-coupled versions that could be used similarly to ZFNs for site-directed genome editing (31–33). Such TAL effector nucleases (TALENs) were easier than ZFNs to produce and validate, generating widespread excitement about the possibility of facile genome editing that would be fast and inexpensive. But difficulties of protein design, synthesis, and validation remained a barrier to widespread adoption of these engineered nucleases for routine use.

## History and biology of CRISPR-Cas systems

In a parallel but completely separate area of research, a few microbiology and bioinformatics laboratories in the mid-2000s began investigating CRISPRs (clustered regularly interspaced palindromic repeats), which had been described in 1987 by Japanese researchers as a series of short direct repeats interspaced with short sequences in the genome of *Escherichia coli* (34) (Fig. 1). CRISPRs were later detected in numerous bacteria and archaea (35), and predictions were made about their possible roles in DNA repair or gene regulation (36, 37). A key insight came in 2005 with the observation that many spacer sequences within CRISPRs derive from plasmid and viral origins (38–40). Together with the finding that CRISPR loci are transcribed (41) and the observation that *cas* (CRISPR-associated) genes encode proteins with putative nuclease and helicase domains (38, 40, 42, 43), it was proposed that CRISPR-Cas is an adaptive defense system that might use antisense RNAs as memory signatures of past invasions (44). In 2007, infection experiments of the lactic acid bacterium *Streptococcus thermophilus* with lytic phages provided the first experimental evidence of CRISPR-Cas-mediated adaptive immunity (45). This finding led to the idea that natural CRISPR-Cas systems

existing in cultured bacteria used in the dairy industry could be harnessed for immunization against phages—a first successful application of CRISPR-Cas for biotechnological purposes (46). In 2008, mature CRISPR RNAs (crRNAs) were shown to serve as guides in a complex with Cas proteins to interfere with virus proliferation in *E. coli* (47). The same year, the DNA targeting activity of the CRISPR-Cas system was reported in the pathogen *Staphylococcus epidermidis* (48).

Functional CRISPR-Cas loci comprise a CRISPR array of identical repeats intercalated with invader DNA-targeting spacers that encode the crRNA components and an operon of *cas* genes encoding the Cas protein components. In natural environments, viruses can be matched to their bacterial or archaeal hosts by examining CRISPR spacers (49, 50). These studies showed that viruses are constantly evolving to evade CRISPR-mediated attenuation.

Adaptive immunity occurs in three stages [for recent reviews, see (51–53)]: (i) insertion of a short sequence of the invading DNA as a spacer sequence into the CRISPR array; (ii) transcription of precursor crRNA (pre-crRNA) that undergoes maturation to generate individual crRNAs, each

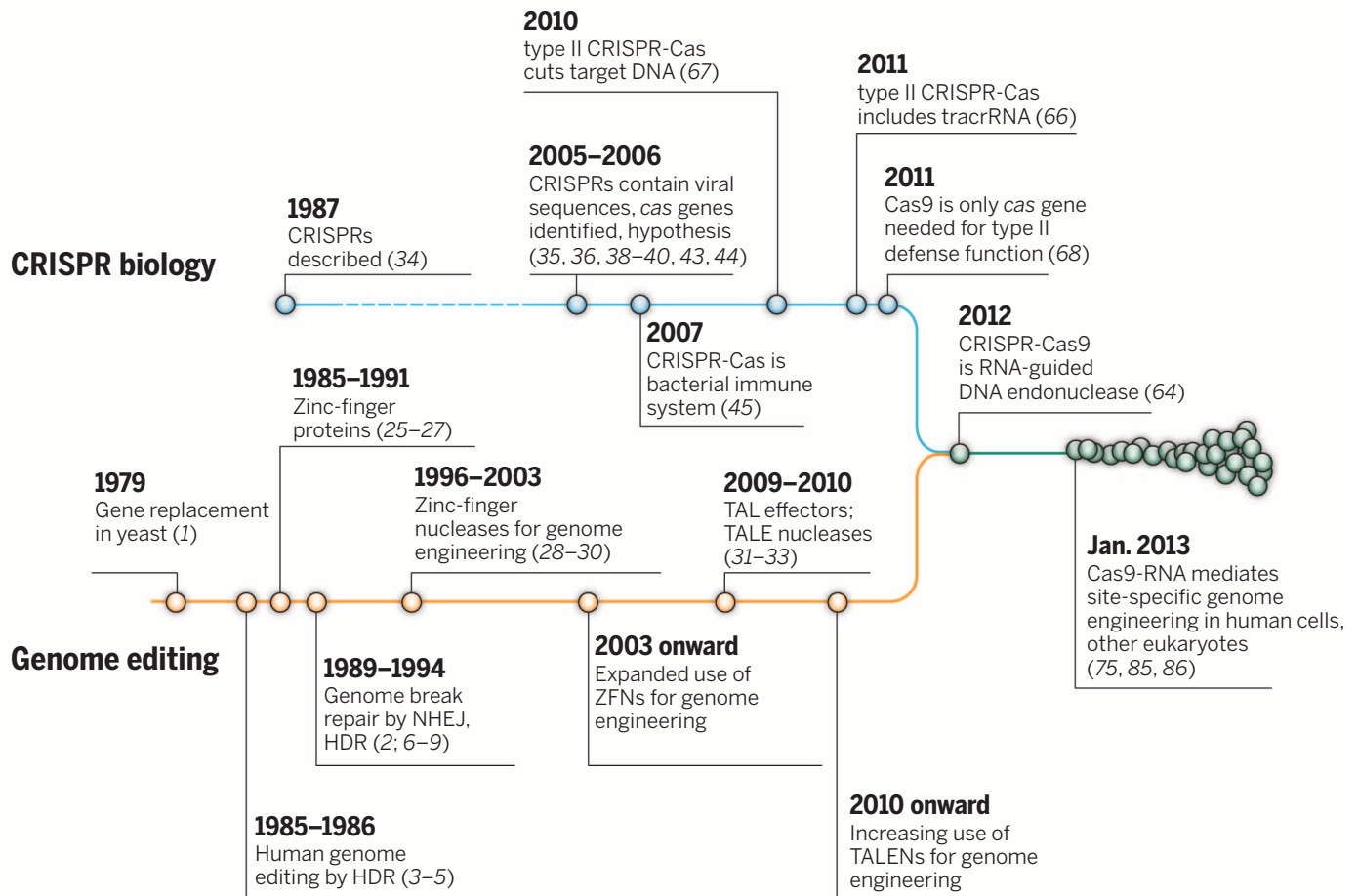
composed of a repeat portion and an invader-targeting spacer portion; and (iii) crRNA-directed cleavage of foreign nucleic acid by Cas proteins at sites complementary to the crRNA spacer sequence. Within this overall theme, three CRISPR-Cas system types (I, II, and III) use distinct molecular mechanisms to achieve nucleic acid recognition and cleavage (54, 55). The protospacer adjacent motif (PAM), a short sequence motif adjacent to the crRNA-targeted sequence on the invading DNA, plays an essential role in the stages of adaptation and interference in type I and type II systems (39, 56–58). The type I and type III systems use a large complex of Cas proteins for crRNA-guided targeting (47, 59–63). However, the type II system requires only a single protein for RNA-guided DNA recognition and cleavage (64, 65)—a property that proved to be extremely useful for genome engineering applications (see below).

### Functionality of CRISPR-Cas9

Bioinformatic analyses first identified Cas9 (formerly COG3513, Csx12, Cas5, or Csn1) as a large multifunctional protein (36) with two putative nuclease domains, HNH (38, 43, 44) and RuvC-like (44). Genetic studies showed that *S. thermophilus*

Cas9 is essential for defense against viral invasion (45, 66), might be responsible for introducing DSBs into invading plasmids and phages (67), enables in vivo targeting of temperate phages and plasmids in bacteria (66, 68), and requires the HNH and RuvC domains to interfere with plasmid transformation efficiency (68).

In 2011 (66), trans-activating crRNA (tracrRNA)—a small RNA that is trans-encoded upstream of the type II CRISPR-Cas locus in *Streptococcus pyogenes*—was reported to be essential for crRNA maturation by ribonuclease III and Cas9, and tracrRNA-mediated activation of crRNA maturation was found to confer sequence-specific immunity against parasite genomes. In 2012 (64), the *S. pyogenes* CRISPR-Cas9 protein was shown to be a dual-RNA-guided DNA endonuclease that uses the tracrRNA:crRNA duplex (66) to direct DNA cleavage (64) (Fig. 2). Cas9 uses its HNH domain to cleave the DNA strand that is complementary to the 20-nucleotide sequence of the crRNA; the RuvC-like domain of Cas9 cleaves the DNA strand opposite the complementary strand (64, 65) (Fig. 2). Mutating either the HNH or the RuvC-like domain in Cas9 generates a variant protein with single-stranded DNA cleavage (nickase)



**Fig. 1. Timeline of CRISPR-Cas and genome engineering research fields.** Key developments in both fields are shown. These two fields merged in 2012 with the discovery that Cas9 is an RNA-programmable DNA endonuclease, leading to the explosion of papers beginning in 2013 in which Cas9 has been used to modify genes in human cells as well as many other cell types and organisms.

activity, whereas mutating both domains (dCas9; Asp<sup>10</sup> → Ala, His<sup>840</sup> → Ala) results in an RNA-guided DNA binding protein (64, 65). DNA target recognition requires both base pairing to the crRNA sequence and the presence of a short sequence (PAM) adjacent to the targeted sequence in the DNA (64, 65) (Fig. 2).

The dual tracrRNA:crRNA was then engineered as a single guide RNA (sgRNA) that retains two critical features: the 20-nucleotide sequence at the 5' end of the sgRNA that determines the DNA target site by Watson-Crick base pairing, and the double-stranded structure at the 3' side of the guide sequence that binds to Cas9 (64) (Fig. 2). This created a simple two-component system in which changes to the guide sequence (20 nucleotides in the native RNA) of the sgRNA can be used to program CRISPR-Cas9 to target any DNA sequence of interest as long as it is adjacent to a PAM (64). In contrast to ZFNs and TALENs, which require substantial protein engineering for each DNA target site to be modified, the CRISPR-Cas9 system requires only a change in the guide RNA sequence. For this reason, the CRISPR-Cas9

technology using the *S. pyogenes* system has been rapidly and widely adopted by the scientific community to target, edit, or modify the genomes of a vast array of cells and organisms. Phylogenetic studies (69–71) as well as in vitro and in vivo experiments (64, 71, 72) show that naturally occurring Cas9 orthologs use distinct tracrRNA:crRNA transcripts as guides, defined by the specificity to the dual-RNA structures (69–71) (Fig. 3). The reported collection of Cas9 orthologs constitutes a large source of CRISPR-Cas9 systems for multiplex gene targeting, and several orthologous CRISPR-Cas9 systems have already been applied successfully for genome editing in human cells [*Neisseria meningitidis* (73, 74), *S. thermophilus* (73, 75), and *Treponema denticola* (73)].

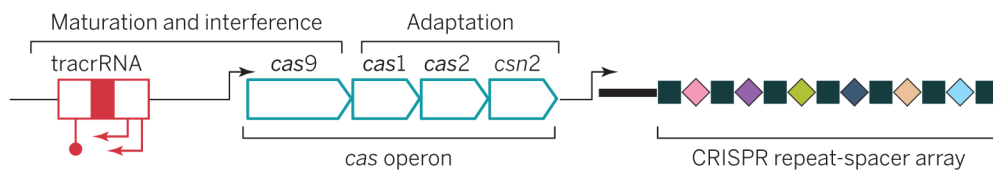
Although the CRISPR acronym has attracted media attention and is widely used in the scientific and popular literature, nearly all genome editing applications are based on the use of the protein Cas9 together with suitable sgRNAs. As discussed above, CRISPR refers to the repetitive nature of the repeats in the CRISPR arrays that encode crRNAs, and the term does not relate

directly to genome engineering. Nonetheless we prefer to use “CRISPR-Cas9” in a way that is less restrictive than other nomenclatures that have been used in the field (76).

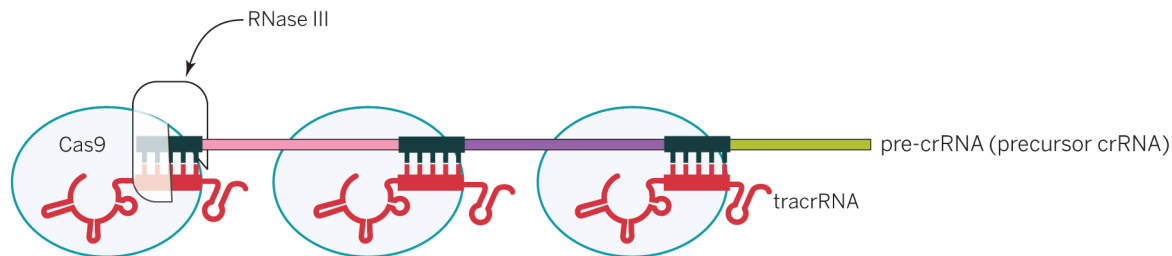
### Mechanism of CRISPR-Cas9-mediated genome targeting

Structural analysis of *S. pyogenes* Cas9 has revealed additional insights into the mechanism of CRISPR-Cas9 (Fig. 3). Molecular structures of Cas9 determined by electron microscopy and x-ray crystallography show that the protein undergoes large conformational rearrangement upon binding to the guide RNA, with a further change upon association with a target double-stranded DNA (dsDNA). This change creates a channel, running between the two structural lobes of the protein, that binds to the RNA-DNA hybrid as well as to the coaxially stacked dual-RNA structure of the guide corresponding to the crRNA repeat-tracrRNA antirepeat interaction (77, 78). An arginine-rich  $\alpha$  helix (77–79) bridges the two structural lobes of Cas9 and appears to be the hinge between them, in addition

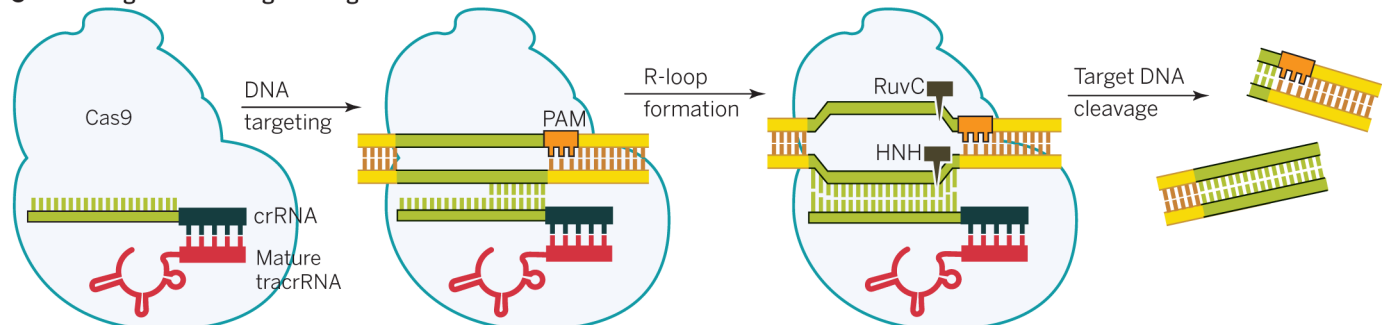
#### A Genomic CRISPR locus



#### B tracrRNA:crRNA co-maturation and Cas9 co-complex formation



#### C RNA-guided cleavage of target DNA



**Fig. 2. Biology of the type II-A CRISPR-Cas system.** The type II-A system from *S. pyogenes* is shown as an example. **(A)** The *cas* gene operon with tracrRNA and the CRISPR array. **(B)** The natural pathway of antiviral defense involves association of Cas9 with the antirepeat-repeat RNA (tracrRNA:crRNA) duplexes, RNA co-processing by ribonuclease III, further trimming, R-loop formation, and target DNA cleavage. **(C)** Details of the natural DNA cleavage with the duplex tracrRNA:crRNA.

to playing a central role in binding the guide RNA–target DNA hybrid as shown by mutagenesis (77, 78). The conformational change in Cas9 may be part of the mechanism of target dsDNA unwinding and guide RNA strand invasion, although this idea remains to be tested. Mechanistic studies also show that the PAM is critical for initial DNA binding; in the absence of the PAM, even target sequences fully complementary to the guide RNA sequence are not recognized by Cas9 (80). A crystal structure of Cas9 in complex with a guide RNA and a partially dsDNA target demonstrates that the PAM lies within a base-paired DNA structure (81). Arginine motifs in the C-terminal domain of Cas9 interact with the PAM on the noncomplementary strand within the major groove. The phosphodiester group at position +1 in the target DNA strand interacts with the minor groove of the duplexed PAM, possibly resulting in local strand separation, the so-called R-loop, immediately upstream of the PAM (81). Single-molecule experiments also suggest that R-loop association rates are affected primarily by the PAM, whereas R-loop stability is influenced mainly by protospacer elements distal to the PAM (82). Together with single-molecule and bulk biochemical experiments using mutated target DNAs, a mechanism can be proposed whereby target DNA melting starts at the level of PAM recognition, resulting in directional R-loop formation expanding toward the distal protospacer end and concomitant RNA strand invasion and RNA–DNA hybrid formation (80–82).

To assess the target-binding behavior of Cas9 in cells, researchers used chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq) to determine the numbers and types of Cas9 binding sites on the chromosome. Results showed that in both human embryonic kidney (HEK293) cells (83) and mouse embryonic stem cells (mESCs) (84), a catalytically inactive version of Cas9 bound to many more sites than those matching the sequence of the sgRNA used in each case. Such off-target interactions with DNA, typically at sites bearing a PAM and partially complementary to the guide RNA sequence, are consistent with established modes

of DNA interrogation by Cas9 (80). Active Cas9 rarely cleaves the DNA at off-target binding sites, implying decoupled binding and cleavage events in which nearly perfect complementarity between the guide RNA and the target site are necessary for efficient DNA cleavage. These observations are consistent with results obtained for Cas9–guide RNA complexes in single-molecule experiments (80). Furthermore, Cas9 binding events occur more densely in areas of open chromatin as compared to regions of compact, transcriptionally inactive chromatin. However, because the method involves cross-linking cells for ~10 min before quenching the reaction, transient and long-lived binding interactions cannot be distinguished. It is possible that many of the apparent off-target DNA interactions in fact reflect brief encounters that would not normally trigger strand invasion by the guide RNA.

### Engineering cells and model organisms

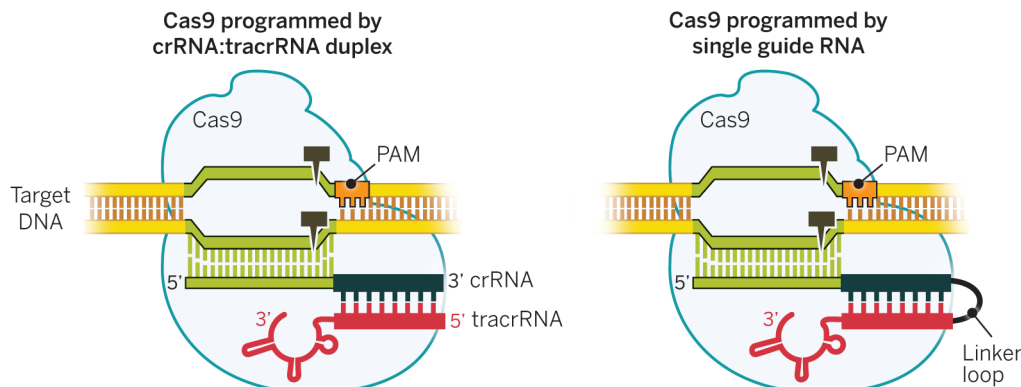
Following the 2012 publication of Jinek *et al.* (64), three studies in January 2013 demonstrated that CRISPR–Cas9 represents an efficient tool to edit the genomes of human cells (75, 85, 86). The “humanized” versions of *S. pyogenes* Cas9 (75, 85, 86) and *S. thermophilus* Cas9 (75) were coexpressed with custom-designed sgRNAs (75, 85, 86) or with tracrRNA coexpressed with custom-designed crRNAs (75) in human embryonic kidney, chronic myelogenous leukemia, or induced pluripotent stem cells (75, 85, 86) as well as in mouse cells (75). The expected alterations in the target DNA were observed, indicating that site-specific DSBs by RNA-guided Cas9 had stimulated gene editing by nonhomologous end joining repair or gene replacement by homology-directed repair (Fig. 4). Targeting with multiple sgRNAs—referred to as multiplexing—was also successfully achieved (75, 86). RNA-programmable *S. pyogenes* Cas9-mediated editing has now been applied to various human cells and embryonic stem cells [(87–90); for reviews, see (91–93)]. Although direct comparisons can be difficult to assess because of differences in target sites and protein expression levels, some analyses show that CRISPR–Cas9–mediated editing efficiencies can reach 80% or more depending on the target,

which is as high as or higher than levels observed using ZFNs or TALENs (89, 94).

These initial studies were only the beginning of what has become an incredibly fast-paced field in which laboratories around the world have used CRISPR–Cas9 to edit genomes of a wide range of cell types and organisms (summarized in Fig. 5). As of this writing, more than 1000 papers have been published that include the CRISPR acronym in the title or abstract, with the majority of these published since the beginning of 2013. Many of these applications have been discussed in recent reviews (91–93). Here we highlight a few examples that illustrate the power of the technology (Fig. 6). The first example is the precise reproduction of tumor-associated chromosomal translocations, which come about during carcinogenesis through illegitimate nonhomologous joining of two chromosomes. The ability of CRISPR–Cas9 to introduce DSBs at defined positions has made it possible to generate human cell lines and primary cells bearing chromosomal translations resembling those described in cancers such as lung cancer (95), acute myeloid leukemia, and Ewing’s sarcoma (96, 97). An improved method to generate liver cancer or myeloid malignancy models in mice facilitated by CRISPR–Cas9 was recently reported (98, 99). CRISPR–Cas9 thus provides a robust technology for studying genomic rearrangements and the development and progression of cancers or other diseases.

A second example is the systematic analysis of gene functions in mammalian cells. A genome-scale lentiviral sgRNA library was developed to generate a pooled loss-of-function genetic screening approach suitable for both positive and negative selection (100, 101). This approach was also used to identify genes essential for cell viability in cancer and pluripotent stem cells (102). Although such studies have been attempted using RNA interference (RNAi) to reduce the expression of genes, this strategy does not allow the generation of gene knockouts and can suffer from substantial off-target effects. The use of CRISPR–Cas9 for genome-wide studies will enable large-scale screening for drug targets and other phenotypes and thus will expand the nature and utility of genetic screens in human and other nonmodel cell types and organisms.

Other pertinent examples of CRISPR–Cas9 applications with relevance to human health include the ability to correct genetic mutations responsible for inherited disorders. A dominant mutation in the *Crygc* gene responsible for cataracts was successfully corrected in mice (103). Using cultured primary adult intestinal stem cells derived from cystic fibrosis patients, the CFTR locus responsible for cystic fibrosis was corrected by homologous recombination, resulting in the clonal expansion of miniature organlike cell cultures (organoids) harboring the desired, exact genetic change (104). These



**Fig. 3. Evolution and structure of Cas9.** The structure of *S. pyogenes* Cas9 in the unliganded and RNA–DNA-bound forms [from (77, 81)].

studies underscore the potential for this technology to be used for human gene therapy to treat genetic disorders.

A last example of CRISPR-Cas9 as a genome engineering technology is its application to plants and fungi. Since its demonstration as a genome editing tool in *Arabidopsis thaliana* and *Nicotiana benthamiana* (105, 106), editing has been demonstrated in crop plants including rice, wheat, and sorghum as well as sweet orange and liverwort

(107–111). This technology promises to change the pace and course of agricultural research. For example, a recent study in rice found that target genes were edited in nearly 50% of the embryogenic cells that received the Cas9–guide RNA constructs, and editing occurred before the first cell division (112). Furthermore, these genetic changes were passed to the next generation of plants without new mutation or reversion, and whole-genome sequencing did not reveal substantial off-target

editing. Such findings suggest that modification of plant genomes to provide protection from disease and resistance to pests may be much easier than has been the case with other technologies. The regulatory implications of CRISPR-Cas9 technology for use in plants are not yet clear and will certainly depend on the type of mutation(s) to be introduced.

In general, the lack of efficient, inexpensive, fast-to-design, and easy-to-use precision genetic

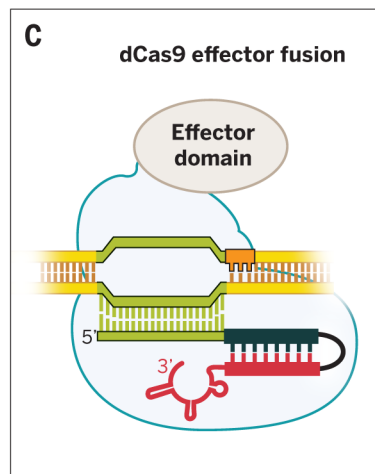
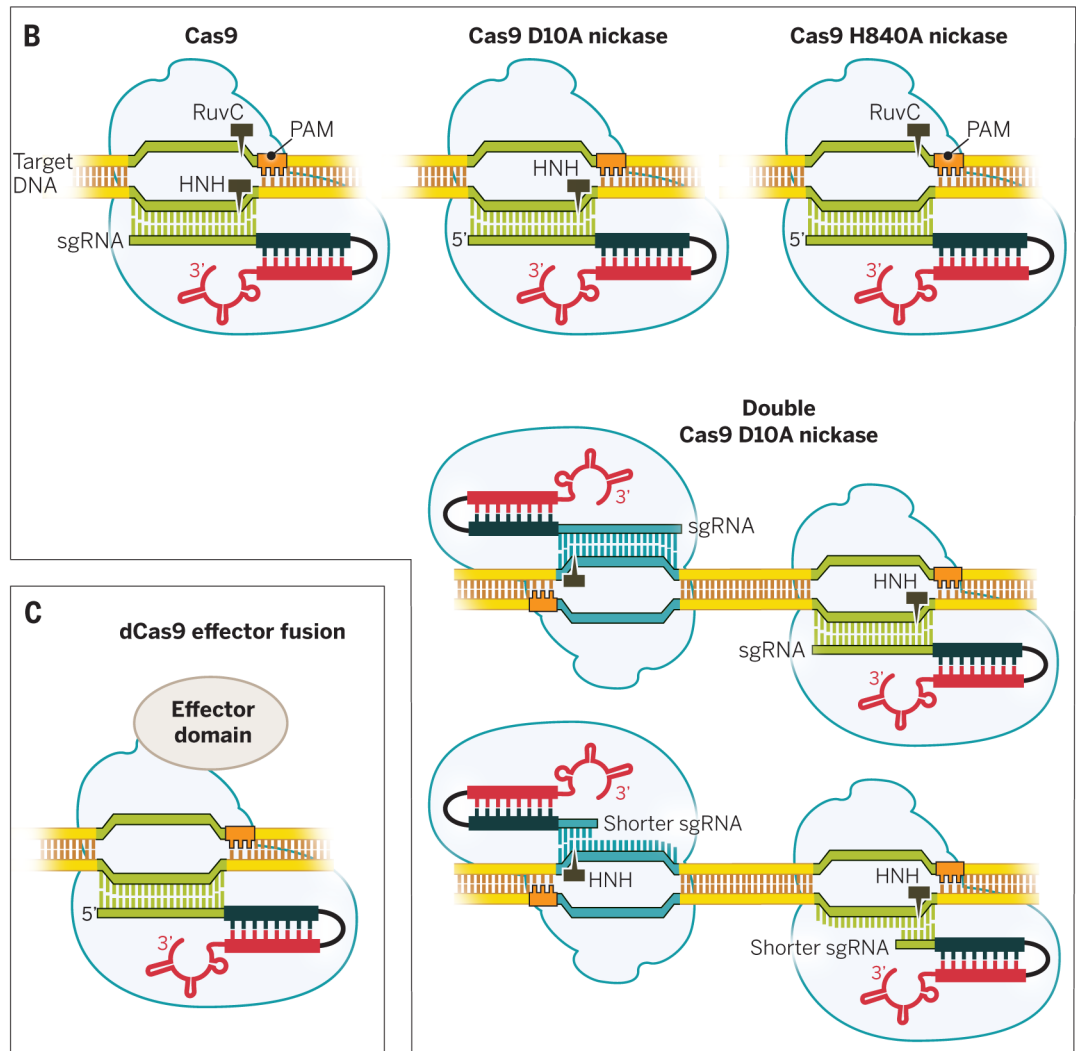
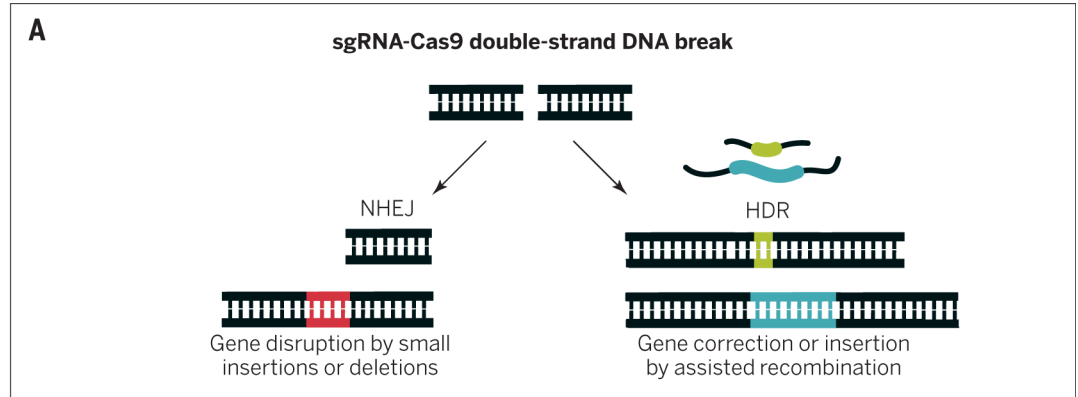
**Fig. 4. CRISPR-Cas9 as a genome engineering tool. (A)**

Different strategies for introducing blunt double-stranded DNA breaks into genomic loci, which become substrates for endogenous cellular DNA repair machinery that catalyze non-homologous end joining (NHEJ) or homology-directed repair (HDR).

(B) Cas9 can function as a nickase (nCas9) when engineered to contain an inactivating mutation in either the HNH domain or RuvC domain active sites. When nCas9 is used with two sgRNAs that recognize offset target sites in DNA, a staggered double-strand break is created.

(C) Cas9 functions as an RNA-guided DNA binding protein when engineered to contain inactivating mutations in both of its active sites. This catalytically inactive or dead Cas9 (dCas9) can mediate transcriptional down-regulation or activation, particularly when fused to activator or repressor domains.

In addition, dCas9 can be fused to fluorescent domains, such as green fluorescent protein (GFP), for live-cell imaging of chromosomal loci. Other dCas9 fusions, such as those including chromatin or DNA modification domains, may enable targeted epigenetic changes to genomic DNA.



tools has also been a limiting factor for the analysis of gene functions in model organisms of developmental and regenerative biology. Efficient genome engineering to allow targeted genome modifications in the germ lines of animal models such as fruit flies (113, 114), zebrafish (94, 115), nematodes (116), salamanders (117), and frogs (118, 119) is now possible with the development of the CRISPR-Cas9 technology. The technology can also facilitate the generation of mouse (120–122) and rat (123, 124) models better suited to pharmacological studies and the understanding of human diseases, as well as pigs (125) and monkeys (126). Overall, CRISPR-Cas9 is already having a major impact on functional genomic experiments that can be conducted in these model systems, which will advance the field of experimental biology in ways not imagined even a few years ago.

### Further development of the technology

A key property of Cas9 is its ability to bind to DNA at sites defined by the guide RNA sequence and the PAM, allowing applications beyond permanent modification of DNA. In particular, a catalytically deactivated version of Cas9 (dCas9) has been repurposed for targeted gene regulation on a genome-wide scale. Referred to as CRISPR interference (CRISPRi), this strategy was shown to block transcriptional elongation, RNA polymerase binding, or transcription factor binding, depending on the site(s) recognized by the dCas9–guide RNA complex. Demonstrated first in *E. coli*, whole-genome sequencing showed that there were no detectable off-target effects (127). CRISPRi has been used to repress multiple target genes simultaneously, and its effects are reversible (127–130).

By generating chimeric versions of dCas9 that are fused to regulatory domains, it has been possible to use CRISPRi for efficient gene regulation in mammalian cells. Specifically, fusion of dCas9 to effector domains including VP64 or KRAB allowed stable and efficient transcriptional activation or repression, respectively, in human and yeast cells (129). As observed in bacteria, site(s) of regulation were defined solely by the coexpressed guide RNA(s) for dCas9. RNA-seq analysis showed that CRISPRi-directed transcriptional repression is highly specific. More broadly, these results demonstrated that dCas9 can be used as a modular and flexible DNA-binding platform for

the recruitment of proteins to a target DNA sequence in a genome, laying the foundation for future experiments involving genome-wide screening similar to those performed using RNAi. The lack of CRISPR-Cas systems in eukaryotes is an important advantage of CRISPRi over RNAi for various applications in which competition with the endogenous pathways is problematic. For example, using RNAi to silence genes that are part of the RNAi pathway itself (i.e., Dicer, Argonaute) can lead to results that are difficult to interpret due to multiple direct and indirect effects. In addition, any RNAs used to silence specific genes may compete with endogenous RNA-mediated gene regulation in cells. With its ability to permanently change the genetic code and to up- or down-regulate gene expression at the transcriptional or posttranscriptional level, CRISPR-Cas9 offers a large versatility in harnessing alternatives, whereas RNAi is mostly restricted to knocking down gene expression. Although RNAi has been improving over the years, incomplete knockdowns or unpredictable off-targeting are still reported bottlenecks of this technology, and future comparative analyses should address the superiority of CRISPRi over RNAi in these aspects.

The programmable binding capability of dCas9 can also be used for imaging of specific loci in live cells. An enhanced green fluorescent protein-tagged dCas9 protein and a structurally optimized sgRNA were shown to produce robust imaging of repetitive and nonrepetitive elements in telomeres and coding genes in living cells (131). This CRISPR imaging tool has the potential to improve the current technologies for studying conformational dynamics of native chromosomes in living cells, particularly if multicolor imaging can be developed using multiple distinct Cas9 proteins. It may also be possible to couple fluorescent proteins or small molecules to the guide RNA, providing an orthogonal strategy for multicolor imaging using Cas9.

Novel technologies aiming to disrupt proviruses may be an attractive approach to eliminating viral genomes from infected individuals and thus curing viral infections. An appeal of this strategy is that it takes advantage of the primary native functions of CRISPR-Cas systems as antiviral adaptive immune systems in bacteria. The targeted CRISPR-Cas9 technique was shown to efficiently cleave and mutate the long terminal repeat sites of HIV-1 and also to remove internal

viral genes from the chromosome of infected cells (132, 133).

CRISPR-Cas9 is also a promising technology in the field of engineering and synthetic biology. A multiplex CRISPR approach referred to as CRISPRm was developed to facilitate directed evolution of biomolecules (134). CRISPRm consists of the optimization of CRISPR-Cas9 to generate quantitative gene assembly and DNA library insertion into the fungal genomes, providing a strategy to improve the activity of biomolecules. In addition, it has been possible to induce Cas9 to bind single-stranded RNA in a programmable fashion by using short DNA oligonucleotides containing PAM sequences (PAMmers) to activate the enzyme, suggesting new ways to target transcripts without prior affinity tagging (135).

A series of studies have reported the efficiency with which the RNA-programmable *S. pyogenes* Cas9 targets and cleaves DNA and have also addressed the level of its specificity by monitoring the ratio of off-site targeting (136–140). Off-site targeting is defined by the tolerance of Cas9 to mismatches in the RNA guide sequence and is dependent on the number, position, and distribution of mismatches throughout the entire guide sequence (136–140) beyond the initial seed sequence originally defined as the first 8 to 12 nucleotides of the guide sequence proximal to the PAM (64) (Fig. 2). The amount of Cas9 enzyme expressed in the cell is an important factor in tolerance to mismatches (138). High concentrations of the enzyme were reported to increase off-site targeting, whereas lowering the concentration of Cas9 increases specificity while diminishing on-target cleavage activity (137). Several groups have developed algorithmic tools that predict the sequence of an optimal sgRNA with minimized off-target effects (for example, <http://tools.genome-engineering.org>, <http://zifit.partners.org>, and [www.e-crisp.org](http://www.e-crisp.org)) (141–145). The development of alternative genome-wide approaches that would also consider other features of the reaction, such as the thermodynamic properties of the sgRNA, may also increase the specificity of the design.

Several studies of the CRISPR-Cas9 technology relate to the specificity of DNA targeting (Fig. 4): a double-nicking approach consisting of using the nickase variant of Cas9 with a pair of offset sgRNAs properly positioned on the target DNA (146–148); an sgRNA-guided dCas9 fused to the FokI nuclease where two fused dCas9-FokI monomers can simultaneously bind target sites at a defined distance apart (149, 150); and shorter sgRNAs truncated by two or three nucleotides at the distal end relative to the PAM that can be used with the double nicking strategy to further reduce off-target activity (151). The first two methods rely on Cas9 dimerization similar to the engineered dimeric ZFNs and TALENs, with the principle that two adjacent off-target binding events and subsequent cleavage

## Biology

Cell lines	Model organisms
HEK293	Mice
U2OS	Rats
K562	Fruit flies
	Nematodes
	<i>Arabidopsis</i>
	Salamanders
	Frogs
	Monkeys

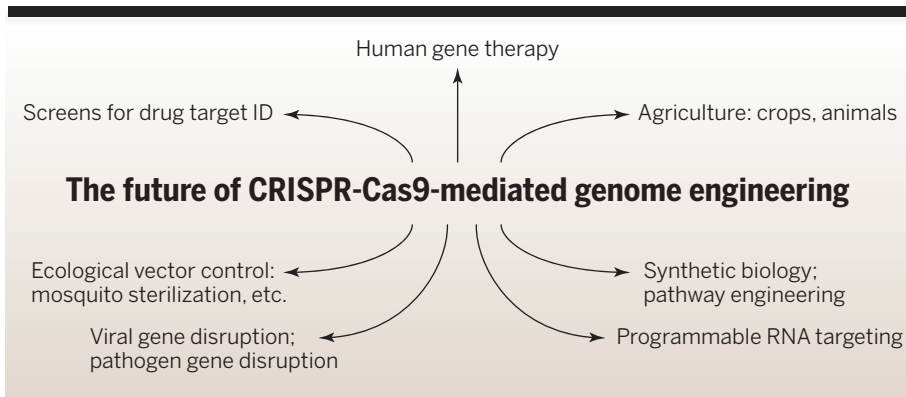
## Biotechnology

Crop plants	Fungi
Rice	<i>Kluyveromyces</i>
Wheat	<i>Chlamydomonas</i>
Sorghum	
Tobacco	

## Biomedicine

Organoids
hESCs
iPSCs

Fig. 5. Examples of cell types and organisms that have been engineered using Cas9.



**Fig. 6. Future applications in biomedicine and biotechnology.** Potential developments include establishment of screens for target identification, human gene therapy by gene repair and gene disruption, gene disruption of viral sequences, and programmable RNA targeting.

are less likely to occur than a single off-target cleavage (146–150). The latter method follows the reasoning according to which the 5'-end nucleotides of the sgRNAs are not necessary for their full activity; however, they may compensate for mismatches at other positions along the guide RNA–target DNA interface, and thus shorter sgRNAs may be more specific (151). Future efforts will focus on further developing the precision of the technology, as well as increasing the frequency of homology-directed repair relative to nonhomologous end joining in order to favor site-specific insertion of new genetic information.

### Conclusions and perspectives

Our understanding of how genomes direct development, normal physiology, and disease in higher organisms has been hindered by a lack of suitable tools for precise and efficient gene engineering. The simple two-component CRISPR-Cas9 system, using Watson-Crick base pairing by a guide RNA to identify target DNA sequences, is a versatile technology that has already stimulated innovative applications in biology. Understanding the CRISPR-Cas9 system at the biochemical and structural level allows the engineering of tailored Cas9 variants with smaller size and increased specificity. A crystal structure of the smaller Cas9 protein from *Actinomyces*, for example, showed how natural variation created a streamlined enzyme, setting the stage for future engineered Cas9 variants (77). A deeper analysis of the large panel of naturally evolving bacterial Cas9 enzymes may also reveal orthologs with distinct DNA binding specificity, will broaden the choice of PAMs, and will certainly reveal shorter variants more amenable for delivery in human cells.

Furthermore, specific methods for delivering Cas9 and its guide RNA to cells and tissues should benefit the field of human gene therapy. For example, recent experiments confirmed that the Cas9 protein-RNA complex can be introduced directly into cells using nucleofection or cell-penetrating peptides to enable rapid and timed editing (89, 152), and transgenic organisms

that express Cas9 from inducible promoters are being tested. An exciting harbinger of future research in this area is the recent demonstration that Cas9–guide RNA complexes, when injected into adult mice, provided sufficient editing in the liver to alleviate a genetic disorder (153). Understanding the rates of homology-directed repair after Cas9-mediated DNA cutting will advance the field by enabling efficient insertion of new or corrected sequences into cells and organisms. In addition, the rapid advance of the field has raised excitement about commercial applications of CRISPR-Cas9.

The era of straightforward genome editing raises ethical questions that will need to be addressed by scientists and society at large. How can we use this powerful tool in such a way as to ensure maximum benefit while minimizing risks? It will be imperative that nonscientists understand the basics of this technology sufficiently well to facilitate rational public discourse. Regulatory agencies will also need to consider how best to foster responsible use of CRISPR-Cas9 technology without inhibiting appropriate research and development.

The identification of the CRISPR-Cas9 technology underscores the way in which many inventions that have advanced molecular biology and medicine emanated, through basic research on natural mechanisms of DNA replication, repair, and defense against viruses. In many cases, key methodologies emerged from the study of bacteria. The CRISPR-Cas9 technology originated through a similar process: Once the mechanism underlying how the CRISPR-Cas9 system works was understood, it could be harnessed for applications in molecular biology and genetics that were not previously envisioned.

### REFERENCES AND NOTES

- S. Scherer, R. W. Davis, Replacement of chromosome segments with altered DNA sequences constructed in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4951–4955 (1979). doi: [10.1073/pnas.76.10.4951](https://doi.org/10.1073/pnas.76.10.4951); pmid: 388424
- Y. S. Rong, K. G. Golic, Gene targeting by homologous recombination in *Drosophila*. *Science* **288**, 2013–2018 (2000). doi: [10.1126/science.288.5473.2013](https://doi.org/10.1126/science.288.5473.2013); pmid: 10856208
- O. Smithies, R. G. Gregg, S. S. Boggs, M. A. Koralewski, R. S. Kucherlapati, Insertion of DNA sequences into the

- human chromosomal beta-globin locus by homologous recombination. *Nature* **317**, 230–234 (1985). doi: [10.1038/317230a0](https://doi.org/10.1038/317230a0); pmid: 2995814
- K. R. Thomas, K. R. Folger, M. R. Capecchi, High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44**, 419–428 (1986). doi: [10.1016/0092-8674\(86\)90463-0](https://doi.org/10.1016/0092-8674(86)90463-0); pmid: 3002636
- S. L. Mansour, K. R. Thomas, M. R. Capecchi, Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: A general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348–352 (1988). doi: [10.1038/336348a0](https://doi.org/10.1038/336348a0); pmid: 3194019
- N. Rudin, E. Sugarman, J. E. Haber, Genetic and physical analysis of double-strand break repair and recombination in *Saccharomyces cerevisiae*. *Genetics* **122**, 519–534 (1989). pmid: 2668114
- A. Plessis, A. Perrin, J. E. Haber, B. Dujon, Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics* **130**, 451–460 (1992). pmid: 1551570
- P. Rouet, F. Smih, M. Jasin, Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**, 8096–8106 (1994). pmid: 7969147
- A. Choulika, A. Perrin, B. Dujon, J. F. Nicolas, Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**, 1968–1973 (1995). pmid: 7891691
- G. Felsenfeld, D. R. Davies, A. Rich, Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.* **79**, 2023–2024 (1957). doi: [10.1021/ja01565a074](https://doi.org/10.1021/ja01565a074)
- A. Varshavsky, Discovering the RNA double helix and hybridization. *Cell* **127**, 1295–1297 (2006). doi: [10.1016/j.cell.2006.12.008](https://doi.org/10.1016/j.cell.2006.12.008); pmid: 17190591
- S. A. Strobel, L. A. Doucette-Stamm, L. Riba, D. E. Housman, P. B. Dervan, Site-specific cleavage of human chromosome 4 mediated by triple-helix formation. *Science* **254**, 1639–1642 (1991). doi: [10.1126/science.1836279](https://doi.org/10.1126/science.1836279); pmid: 1836279
- S. A. Strobel, P. B. Dervan, Single-site enzymatic cleavage of yeast genomic DNA mediated by triple helix formation. *Nature* **350**, 172–174 (1991). doi: [10.1038/350172a0](https://doi.org/10.1038/350172a0); pmid: 1848684
- S. A. Strobel, P. B. Dervan, Site-specific cleavage of a yeast chromosome by oligonucleotide-directed triple-helix formation. *Science* **249**, 73–75 (1990). doi: [10.1126/science.2195655](https://doi.org/10.1126/science.2195655); pmid: 2195655
- A. F. Faruqi, M. M. Seidman, D. J. Segal, D. Carroll, P. M. Glazer, Recombination induced by triple-helix-targeted DNA damage in mammalian cells. *Mol. Cell. Biol.* **16**, 6820–6828 (1996). pmid: 8943337
- G. Wang, D. D. Levy, M. M. Seidman, P. M. Glazer, Targeted mutagenesis in mammalian cells mediated by intracellular triple helix formation. *Mol. Cell. Biol.* **15**, 1759–1768 (1995). pmid: 7862165
- Z. Sandor, A. Bredberg, Deficient DNA repair of triple helix-directed double psoralen damage in human cells. *FEBS Lett.* **374**, 287–291 (1995). doi: [10.1016/0014-5793\(95\)01133-Y](https://doi.org/10.1016/0014-5793(95)01133-Y); pmid: 7589555
- J. Cho, M. E. Parks, P. B. Dervan, Cyclic polyamides for recognition in the minor groove of DNA. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10389–10392 (1995). doi: [10.1073/pnas.92.22.10389](https://doi.org/10.1073/pnas.92.22.10389); pmid: 7479790
- A. F. Faruqi, M. Egholm, P. M. Glazer, Peptide nucleic acid-targeted mutagenesis of a chromosomal gene in mouse cells. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1398–1403 (1998). doi: [10.1073/pnas.95.4.1398](https://doi.org/10.1073/pnas.95.4.1398); pmid: 9465026
- J. M. Gottesfeld, L. Neely, J. W. Trauger, E. E. Baird, P. B. Dervan, Regulation of gene expression by small molecules. *Nature* **387**, 202–205 (1997). doi: [10.1038/387202a0](https://doi.org/10.1038/387202a0); pmid: 9144294
- J. Yang, S. Zimmerly, P. S. Perlman, A. M. Lambowitz, Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature* **381**, 332–335 (1996). doi: [10.1038/381332a0](https://doi.org/10.1038/381332a0); pmid: 8692273
- S. Zimmerly *et al.*, A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* **83**, 529–538 (1995). doi: [10.1016/0092-8674\(95\)90092-6](https://doi.org/10.1016/0092-8674(95)90092-6); pmid: 7585955
- B. A. Sullenger, T. R. Cech, Ribozyme-mediated repair of defective mRNA by targeted, trans-splicing. *Nature* **371**, 619–622 (1994). doi: [10.1038/371619a0](https://doi.org/10.1038/371619a0); pmid: 7935797
- A. Jacquier, B. Dujon, An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* **41**, 383–394 (1985). doi: [10.1016/S0092-8674\(85\)80011-8](https://doi.org/10.1016/S0092-8674(85)80011-8); pmid: 3886163







EXTENDED PDF FORMAT  
SPONSORED BY



**The new frontier of genome engineering with CRISPR-Cas9**  
Jennifer A. Doudna and Emmanuelle Charpentier (November 27, 2014)  
*Science* **346** (6213), . [doi: 10.1126/science.1258096]

Editor's Summary

### CRISPR-cas: A revolution in genome engineering

The ability to engineer genomic DNA in cells and organisms easily and precisely will have major implications for basic biology research, medicine, and biotechnology. Doudna and Charpentier review the history of genome editing technologies, including oligonucleotide coupled to genome cleaving agents that rely on endogenous repair and recombination systems to complete the targeted changes, self-splicing introns, and zinc-finger nucleases and TAL effector nucleases. They then describe how clustered regularly interspaced palindromic repeats (CRISPRs), and their associated (Cas) nucleases, were discovered to constitute an adaptive immune system in bacteria. They document development of the CRISPR-Cas system into a facile genome engineering tool that is revolutionizing all areas of molecular biology.

*Science*, this issue 10.1126/science.1258096

---

This copy is for your personal, non-commercial use only.

---

- Article Tools** Visit the online version of this article to access the personalization and article tools:  
<http://science.sciencemag.org/content/346/6213/1258096>
- Permissions** Obtain information about reproducing this article:  
<http://www.sciencemag.org/about/permissions.dtl>

*Science* (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.