# CRISPR/Cas9 and Targeted Genome Editing: A New Era in Molecular Biology

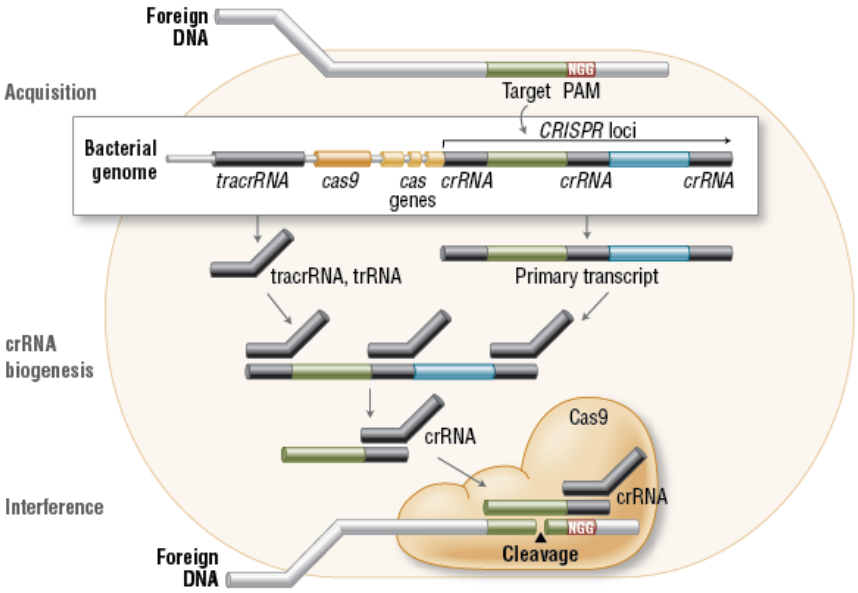
**The development of efficient and reliable ways to make precise, targeted changes to the genome of living cells is a long-standing goal for biomedical researchers. Recently, a new tool based on a bacterial CRISPR-associated protein-9 nuclease (Cas9) from *Streptococcus pyogenes* has generated considerable excitement (1). This follows several attempts over the years to manipulate gene function, including homologous recombination (2) and RNA interference (RNAi) (3). RNAi, in particular, became a laboratory staple enabling inexpensive and high-throughput interrogation of gene function (4, 5), but it is hampered by providing only temporary inhibition of gene function and unpredictable off-target effects (6). Other recent approaches to targeted genome modification – zinc-finger nucleases [ZFNs, (7)] and transcription-activator like effector nucleases [TALENs (8)]– enable researchers to generate permanent mutations by introducing doublestranded breaks to activate repair pathways. These approaches are costly and time-consuming to engineer, limiting their widespread use, particularly for large scale, high-throughput studies.**

## The Biology of Cas9

The functions of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and CRISPR-associated (Cas) genes are essential in adaptive immunity in select bacteria and archaea, enabling the organisms to respond to and eliminate invading genetic material. These repeats were initially discovered in the 1980s in *E. coli* (9), but their function wasn't confirmed until 2007 by Barrangou and colleagues, who demonstrated that *S. thermophilus* can acquire resistance against a bacteriophage by integrating a genome fragment of an infectious virus into its CRISPR locus (10).

Three types of CRISPR mechanisms have been identified, of which type II is the most studied. In this case, invading DNA from viruses or plasmids is cut into small fragments and incorporated into a CRISPR locus amidst a series of short repeats (around 20 bps). The loci are transcribed, and transcripts are then processed to generate small RNAs (crRNA – CRISPR RNA), which are used to guide effector endonucleases that target invading DNA based on sequence complementarity (Figure 1) (11).

**Figure 1. Cas9 *in vivo*: Bacterial Adaptive Immunity**




*In the acquisition phase, foreign DNA is incorporated into the bacterial genome at the CRISPR loci. CRISPR loci is then transcribed and processed into crRNA during crRNA biogenesis. During interference, Cas9 endonuclease complexed with a crRNA and separate tracrRNA cleaves foreign DNA containing a 20-nucleotide crRNA complementary sequence adjacent to the PAM sequence. (Figure not drawn to scale.)*

## Genome Editing Glossary

Cas = CRISPR-associated genes

Cas9, Csn1 = a CRISPR-associated protein containing two nuclease domains, that is programmed by small RNAs to cleave DNA

crRNA = CRISPR RNA

dCAS9 = nuclease-deficient Cas9

DSB = Double-Stranded Break

gRNA = guide RNA

HDR = Homology-Directed Repair

HNH = an endonuclease domain named for characteristic histidine and asparagine residues

Indel = insertion and/or deletion

NHEJ = Non-Homologous End Joining

PAM = Protospacer-Adjacent Motif

RuvC = an endonuclease domain named for an *E. coli* protein involved in DNA repair

sgRNA = single guide RNA

tracrRNA, trRNA = trans-activating crRNA

TALEN = Transcription-Activator Like Effector Nuclease

ZFN = Zinc-Finger Nuclease

One Cas protein, Cas9 (also known as Csn1), has been shown, through knockdown and rescue experiments to be a key player in certain CRISPR mechanisms (specifically type II CRISPR systems). The type II CRISPR mechanism is unique compared to other CRISPR systems, as only one Cas protein (Cas9) is required for gene silencing (12). In type II systems, Cas9 participates in the processing of crRNAs (12), and is responsible for the destruction of the target DNA (11). Cas9's function in both of these steps relies on the presence of two nuclease domains, a RuvC-like nuclease domain located at the amino terminus and a HNH-like nuclease domain that resides in the mid-region of the protein (13).

To achieve site-specific DNA recognition and cleavage, Cas9 must be complexed with both a crRNA and a separate trans-activating crRNA (tracrRNA or trRNA), that is partially complementary to the crRNA (11). The tracrRNA is required for crRNA maturation from a primary transcript encoding multiple pre-crRNAs. This occurs in the presence of RNase III and Cas9 (12).

During the destruction of target DNA, the HNH and RuvC-like nuclease domains cut both DNA strands, generating double-stranded breaks (DSBs) at sites defined by a 20-nucleotide target sequence within an associated crRNA transcript (11, 14). The HNH domain cleaves the complementary strand, while the RuvC domain cleaves the noncomplementary strand.

The double-stranded endonuclease activity of Cas9 also requires that a short conserved sequence, (2–5 nts) known as protospacer-associated motif (PAM), follows immediately 3´- of the crRNA complementary sequence (15). In fact, even fully complementary sequences are ignored by Cas9-RNA in the absence of a PAM sequence (16).

## Cas9 and CRISPR as a New Tool in Molecular Biology

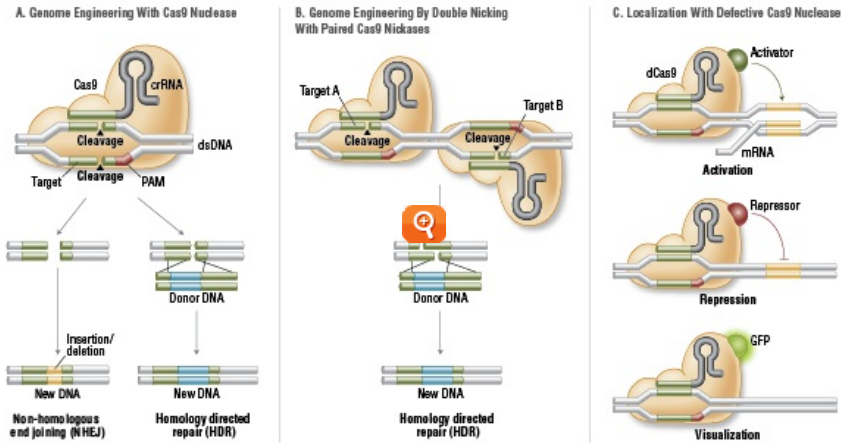
The simplicity of the type II CRISPR nuclease, with only three required components (Cas9 along with the crRNA and trRNA) makes this system amenable to adaptation for genome editing. This potential was realized in 2012 by the Doudna and Charpentier labs (11). Based on the type II CRISPR system described previously, the authors developed a simplified two-component system by combining trRNA and crRNA into a single synthetic single guide RNA (sgRNA). sgRNAprogrammed Cas9 was shown to be as effective as Cas9 programmed with separate trRNA and crRNA in guiding targeted gene alterations (Figure 2A).

To date, three different variants of the Cas9 nuclease have been adopted in genome-editing protocols. The first is wild-type Cas9, which can site-specifically cleave double-stranded DNA, resulting in the activation of the doublestrand break (DSB) repair machinery. DSBs can be repaired by the cellular Non-Homologous End Joining (NHEJ) pathway (17), resulting in insertions and/or deletions (indels) which disrupt the targeted locus. Alternatively, if a donor template with homology to the targeted locus is supplied, the DSB may be repaired by the homology-directed repair (HDR) pathway allowing for precise replacement mutations to be made (Figure 2A) (17, 18).

Cong and colleagues (1) took the Cas9 system a step further towards increased precision by developing a mutant form, known as Cas9D10A, with only nickase activity. This means it cleaves only one DNA strand, and does not activate NHEJ. Instead, when provided with a homologous repair template, DNA repairs are conducted via the high-fidelity HDR pathway only, resulting in reduced indel mutations (1, 11, 19). Cas9D10A is even more appealing in terms of target specificity when loci are targeted by paired Cas9 complexes designed to generate adjacent DNA nicks (20) (see further details about "paired nickases" in Figure 2B).

The third variant is a nuclease-deficient Cas9 (dCas9, Figure 2C) (21). Mutations H840A in the HNH domain and D10A in the RuvC domain inactivate cleavage activity, but do not prevent DNA binding (11, 22). Therefore, this variant can be used to sequence-specifically target any region of the genome without cleavage. Instead, by fusing with various effector domains, dCas9 can be used either as a gene silencing or activation tool (21, 23–26). Furthermore, it can be used as a visualization tool. For instance, Chen and colleagues used dCas9 fused to Enhanced Green Fluorescent Protein (EGFP) to visualize repetitive DNA sequences with a single sgRNA or nonrepetitive loci using multiple sgRNAs (27).

**Figure 2. CRISPR/Cas9 System Applications**



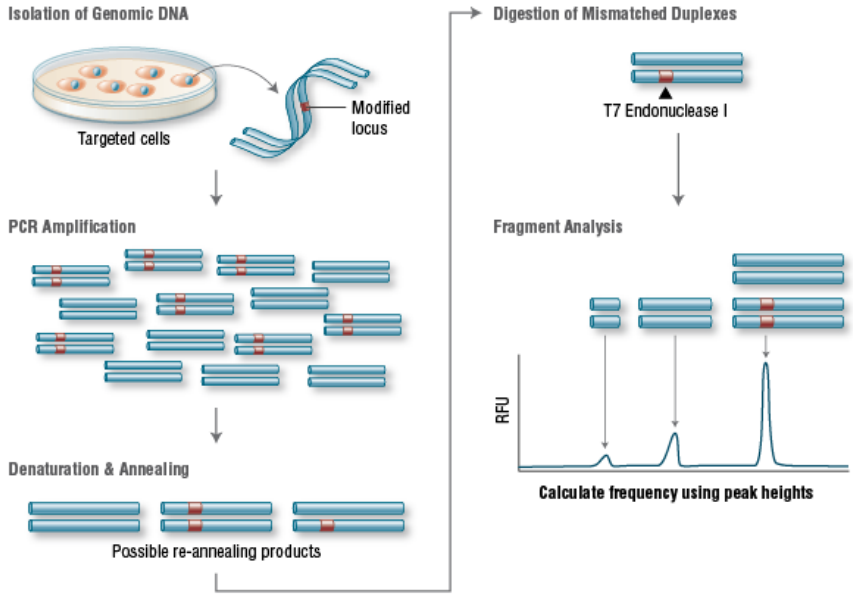

*A. Wild-type Cas9 nuclease site specifically cleaves double-stranded DNA activating double-strand break repair machinery. In the absence of a homologous repair template non-homologous end joining can result in indels disrupting the target sequence. Alternatively, precise mutations and knock-ins can be made by providing a homologous repair template and exploiting the homology directed repair pathway.*
*B. Mutated Cas9 makes a site specific single-strand nick. Two sgRNA can be used to introduce a staggered double-stranded break which can then undergo homology directed repair.*
*C. Nuclease-deficient Cas9 can be fused with various effector domains allowing specific localization. For example, transcriptional activators, repressors, and fluorescent proteins.*

## Targeting Efficiency and Off-target Mutations

Targeting efficiency, or the percentage of desired mutation achieved, is one of the most important parameters by which to assess a genome-editing tool. The targeting efficiency of Cas9 compares favorably with more established methods, such as TALENs or ZFNs (8). For example, in human cells, custom-designed ZFNs and TALENs could only achieve efficiencies ranging from 1% to 50% (29–31). In contrast, the Cas9 system has been reported to have efficiencies up to >70% in zebrafish (32) and plants (33), and ranging from 2–5% in induced pluripotent stem cells (34). In addition, Zhou and colleagues were able to improve genome targeting up to 78% in one-cell mouse embryos, and achieved effective germline transmission through the use of dual sgRNAs to simultaneously target an individual gene (35).

A widely used method to identify mutations is the T7 Endonuclease I mutation detection assay (36, 37) (Figure 3). This assay detects heteroduplex DNA that results from the annealing of a DNA strand, including desired mutations, with a wildtype DNA strand (37).

**Figure 3. T7 Endonuclease I Targeting Efficiency Assay**




*Genomic DNA is amplified with primers bracketing the modified locus. PCR products are then denatured and re-annealed yielding 3 possible structures. Duplexes containing a mismatch are digested by T7 Endonuclease I. The DNA is then electrophoretically separated and fragment analysis is used to calculate targeting efficiency.*

Another important parameter is the incidence of off-target mutations. Such mutations are likely to appear in sites that have differences of only a few nucleotides compared to the original sequence, as long as they are adjacent to a PAM sequence. This occurs as Cas9 can tolerate up to 5 base mismatches within the protospacer region (36) or a single base difference in the PAM sequence (38). Off-target mutations are generally more difficult to detect, requiring whole-genome sequencing to rule them out completely.

Recent improvements to the CRISPR system for reducing off-target mutations have been made through the use of truncated gRNA (truncated within the crRNA-derived sequence) or by adding two extra guanine (G) nucleotides to the 5´ end (28, 37). Another way researchers have attempted to minimize off-target effects is with the use of "paired nickases" (20). This strategy uses D10A Cas9 and two sgRNAs complementary to the adjacent area on opposite strands of the target site (Figure 2B). While this induces DSBs in the target DNA, it is expected to create only single nicks in off-target locations and, therefore, result in minimal off-target mutations.

By leveraging computation to reduce off-target mutations, several groups have developed webbased tools to facilitate the identification of potential CRISPR target sites and assess their potential for off-target cleavage. Examples include the CRISPR Design Tool (38) and the ZiFiT Targeter, Version 4.2 (39, 40).

## Applications as a Genome-editing and Genome Targeting Tool

Following its initial demonstration in 2012 (9), the CRISPR/Cas9 system has been widely adopted. This has already been successfully used to target important genes in many cell lines and organisms, including human (34), bacteria (41), zebrafish (32), *C. elegans* (42), plants (34), *Xenopus tropicalis* (43), yeast (44), *Drosophila* (45), monkeys (46), rabbits (47), pigs (42), rats (48) and mice (49). Several groups have now taken advantage of this method to introduce single point mutations (deletions or insertions) in a particular target gene, via a single gRNA (14, 21, 29). Using a pair of gRNA-directed Cas9 nucleases instead, it is also possible to induce large deletions or genomic rearrangements, such as inversions or translocations (50). A recent exciting development is the use of the dCas9 version of the CRISPR/Cas9 system to target protein domains for transcriptional regulation (26, 51, 52), epigenetic modification (25), and microscopic visualization of specific genome loci (27).

The CRISPR/Cas9 system requires only the redesign of the crRNA to change target specificity. This contrasts with other genome editing tools, including zinc finger and TALENs, where redesign of the protein-DNA interface is required. Furthermore, CRISPR/Cas9 enables rapid genome-wide interrogation of gene function by generating large gRNA libraries (51, 53) for genomic screening.

## The future of CRISPR/Cas9

The rapid progress in developing Cas9 into a set of tools for cell and molecular biology research has been remarkable, likely due to the simplicity, high efficiency and versatility of the system. Of the designer nuclease systems currently available for precision genome engineering, the CRISPR/Cas system is by far the most user friendly. It is now also clear that Cas9's potential reaches beyond DNA cleavage, and its usefulness for genome locus-specific recruitment of proteins will likely only be limited by our imagination.

### References


1. Cong L., et al. (2013) *Science*, 339, 819–823.
2. Capecchi, M.R. (2005) *Nat. Rev. Genet.* 6, 507–512.
3. Fire, A., et al. (1998) *Nature,* 391, 806–811.
4. Elbashir, S.Mm, et al. (2002) *Methods,* 26, 199–213.
5. Martinez, J., et al. (2003) *Nucleic Acids Res. Suppl*. 333.
6. Alic, N, et al. (2012) *PLoS One,* 7, e45367.
7. Miller, J., et al. (2005) *Mol. Ther.* 11, S35–S35.
8. Mussolino, C., et al. (2011) *Nucleic Acids Res.* 39, 9283–9293.
9. Ishino, Y., et al. (1987) *J. Bacteriol*. 169, 5429–5433.
10. Barrangou, R., et al. (2007). *Science,* 315, 1709–1712.
11. Jinek, M., et al. (2012) *Science,* 337, 816–821.
12. Deltcheva, E., et al. (2011) *Nature,* 471, 602–607.
13. Sapranauskas, R., et al. (2011) *Nucleic Acids Res.* 39, 9275–9282.
14. Nishimasu, H., et al. (2014) *Cell,* doi:10.1016/j.cell.2014.02.001
15. Swarts, D.C., et al. (2012) *PLoS One,* 7:e35888.
16. Sternberg, S.H., et al. (2014) *Nature,* doi:10.1038/nature13011.
17. Overballe-Petersen, S., et al. (2013) *Proc. Natl. Acad. Sci. U.S.A.* 110,19860–19865.
18. Gong, C., et al. (2005) *Nat. Struct. Mol. Biol.* 12, 304–312.
19. Davis, L., Maizels, N. (2014) *Proc. Natl. Acad. Sci. USA,* 111, E924–932.
20. Ran, F.A., et al. (2013) *Cell,* 154, 1380–1389.
21. Qi, L.S., et al. (2013) *Cell,* 152, 1173–1183.
22. Gasiunas, G., et al. (2012) *Proc. Natl. Acad. Sci. USA,* 109, E2579–2586.
23. Maede, M.L., et al. (2013) *Nat. Methods,* 10, 977–979.
24. Gilbert, L.A., et al. (2013) *Cell,* 154, 442–451.
25. Hu, J., et al. (2014) *Nucleic Acids Res.* doi:10.1093/nar/gku109.
26. Perez-Pinera, P., et al. (2013) *Nat. Methods,* 10, 239–242.
27. Chen, B., et al. (2013) *Cell,* 155, 1479–1491.
28. Seung, W., et al. (2014) *Genome Res.* 24, 132–141.
29. Miller, J.C., et al. (2011). *Nat. Biotechnol*. 29, 143–148.
30. Mussolino, C., et al. (2011). *Nucleic Acids Res.* 39, 9283–9293.
31. Maeder, M.L., et al. (2008) *Mol. Cell,* 31, 294–301.
32. Hwang, W.Y., et al. (2013) *PLoS One,* 8:e68708.
33. Feng, Z., et al. (2013) *Cell Res.* 23, 1229–1232.
34. Mali, P., et al. (2013) *Science,* 339, 823–826.
35. Zhou, J., et al. (2014) *FEBS J.* doi:10.1111/febs.12735.
36. Fu, Y., et al. (2013) *Nat. Biotechnol*. 31, 822–826.
37. Fu, Y., et al. (2014) *Nat Biotechnol*. doi: 10.1038/nbt.2808.
38. Hsu, P.D., et al. (2013) *Nat. Biotechnol*. 31, 827–832.
39. Sander, J.D., et al. (2007) *Nucleic Acids Res.* 35, W599-605.
40. Sander, J.D., et al (2010) *Nucleic Acids Res.* 38, W462–468.
41. Fabre, L., et al. (2014) *PLoS Negl. Trop. Dis.* 8:e2671.
42. Hai, T., et al. (2014) *Cell Res.* doi: 10.1038/cr.2014.11.
43. Guo, X., et al. (2014) *Development,* 141, 707–714.
44. DiCarlo, J.E., et al. (2013) *Nucleic Acids Res.* 41, 4336–4343.
45. Gratz, S.J., et al. (2014) *Genetics,* doi:10.1534/genetics.113.160713.
46. Niu, Y., et al. (2014) *Cell,* 156, 836–843.
47. Yang, D., et al. (2014) *J. Mol. Cell Biol*. 6, 97-99.
48. Ma, Y., et al. (2014) *Cell Res.* 24, 122–125.
49. Mashiko, D., et al. (2014) *Dev. Growth Differ.* 56, 122–129.
50. Gratz, S.J., et al. (2013) *Fly,* 249.
51. Mali, P., et al. (2013) *Nat. Biotechnol*. 31, 833–838.
52. Cheng, A.W., et al. (2013) *Cell Res.* 23, 1163–1171.
53. Koike-Yusa, H., et al. (2013) *Nat. Biotechnol*. doi: 10.1038/nbt.2800.
54. Sander, J.D., and Joung, J.K. (2014) *Nat Biotechnol*. doi:10.1038/nbt.2842.



*From NEB expressions Issue I, 2014*


*Article by Alex Reis, Ph.D., Bitesize Bio*
*Breton Hornblower, Ph.D., Brett Robb, Ph.D. and George Tzertzinis, Ph.D., New England*