Published in final edited form as:

Curr Protoc Mol Biol. ; 115: 31.4.1–31.4.6. doi:10.1002/cpmb.14.

Introduction to Gene Editing and Manipulation Using CRISPR/Cas9 Technology

Martin Newman and Frederick M. Ausubel

Department of Genetics, Harvard Medical School. Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114

Abstract

Until very recently, the prospect of introducing mutations or exogenous DNA sequences at precise locations in the genomes of plants and animals was difficult, if not impossible. This rapidly changed with the demonstration that the type II CRISPR-Cas complex, a bacterial anti-viral surveillance system, could be engineered into a simple and robust platform for introducing double-stranded DNA breaks at nearly any position of plant and animal genomes. The prospect of efficiently creating tailored changes to a gene of interest is revolutionizing biomedical research, allowing exciting new questions to be asked. This overview introduces CRISPR-Cas technology as a tool for molecular biology and briefly discusses the advantages of this method over earlier techniques, as well as unique opportunities to create new avenues of research.

INTRODUCTION

Major advances in the field of Molecular Biology have been catalyzed by the discovery of technologies to manipulate DNA both *in vitro* and *in vivo*. The modern era of DNA manipulation began in the early 1970s with the purification of bacterial restriction endonucleases that cleave DNA at specific nucleotide sequences. This was followed closely by the development of recombinant DNA techniques involving the use of restriction enzymes in conjunction with DNA ligases and polymerases (1971–1972), DNA sequencing (1977), the polymerase chain reaction (1985), and the commercialization of next generation high throughput DNA sequencing technologies (2006). Combined with classic genetic techniques, it has been possible since the mid 1980s to readily make site-specific changes in the genes of bacteria, yeast, and a variety of fungi, where specific DNA sequences or a set of random mutations, generated *in vitro*, can be inserted into the genomes of the target organisms by homologous recombination at the locus of interest.

In contrast to prokaryotes and fungi, making comparable site-specific changes in the genomes of most plants and animals is far more tedious. Generating random mutations at a particular locus can be achieved if it is possible to generate a double stranded break (DSB) at the locus; highly conserved non-homologous end joining (NHEJ) DNA repair mechanisms repair the DSB with a high likelihood of introducing small insertions or deletions (indels),

* Author for correspondence: Professor of Genetics, Harvard Medical School, Department of Molecular Biology, Massachusetts General Hospital, Richard B. Simches Research Building, 185 Cambridge Street, Boston, MA 02114, Phone: 617 726-5969, Fax: 617-726-5949, ausubel@molbio.mgh.harvard.edu.

which can lead to frame-shift mutations. Alternatively, supplying a DNA template *in vivo* with a desired mutation or more complex construct (e.g. a gene fusion) can result in the substitution of the desired sequences at the site of the DSB by homology-directed repair (HDR) mechanisms (i.e., homologous recombination). Making random mutations in a specific gene has been limited, however, by the ability to introduce DSBs at a particular locus *in vivo*.

Until about three years ago, improvements in the *in vivo* genetic engineering of specific genes in plants and animals have relied on technologies involving the combination of a protein that has been engineered to bind to a specific DNA sequence with a nuclease that introduces a double-strand break at the target site. These technologies rely on zinc finger nucleases (ZFNs) (Carroll, 2011) and transcription activator-like effector nucleases (TALENs) (Boch et al., 2009) to create the necessary DNA binding specificity.

A new era in our ability to manipulate the genomes of plants and animals has been made possible by the discovery of RNA-guided site-specific nucleases (RGNs) in bacteria and archaea. These nucleases utilize RNA molecules (guide RNAs) homologous to a specific target sequence to “guide” the RGN to specifically introduce a DSB in a gene or locus of interest. RNA-guided nucleases were discovered in the course of basic research studies in a wide variety of bacteria as a key component of an immune system that protects prokaryotes from viral attack [see (Marraffini, 2015) and (Lander, 2016) for reviews of the early history of CRISPR research]. These immune systems consist of loci that encode an RGN adjacent to a series of repeat sequences interspersed with short fragments of viral genomes that have previously infected the particular strain of bacteria. The repeat sequences were called CRISPRs (clustered, regularly interspaced, short, palindromic repeats) before their function was known, and the RGNs were named CRISPR-associated (CAS) because they were encoded by sequences directly adjacent to the CRISPR sequence. CRISPR loci were first described in 1995 (Mojica et al., 1995). A major advance in understanding their function was the inference, based upon bio-informatic analysis, that the CAS proteins include nucleases (Marafini, 2015). Work in several laboratories subsequently demonstrated that the CRISPR repeats are an important component of a prokaryotic immune system, where they use nuclease activity to cleave foreign DNA, and had significant potential for gene editing and biotechnology (Lander, 2016; Marraffini, 2015).

The best-described CRISPR-CAS system is from the bacterium *Streptococcus pyogenes*. This immune system consists of two RNAs that bind to a single RGN, the CAS9 RNA-guided nuclease. The first RNA molecule is called the crRNA and consists of a 20 nucleotide sequence corresponding to a viral DNA signature (so-called proto-spacer sequence) and part of the CRISPR repeat sequence. The second RNA molecule is called the transactivating CRISPR-RNA or tracrRNA and is also encoded at the CRISPR locus. The Cas9 nuclease when bound to these two RNAs target viral sequences complementary to the guide crRNA as long as the target sequences is directly 5' to a so-called protospacer adjacent motif (PAM), with the sequence NGG, where “N” can be any nucleotide. The absence of PAM (NGG) sequences at the CRISPR locus prevents the Cas9 RGN from cleaving the CRISPR locus itself.

In work that is revolutionizing our ability to carry out gene editing, the CRISPR-Cas9 system from *S. pyogenes* has been engineered to facilitate its use in carrying out a variety of gene editing and gene control functions [there are many comprehensive reviews; see, for example, (Doudna and Charpentier, 2014; Hsu et al., 2014; Sander and Joung, 2014; Selle and Barrangou, 2015)]. The *S. pyogenes* system was chosen for genetic engineering purposes for two major reasons. First it was shown that the crRNA and the tracrRNA, both of which are crucial for Cas9 targeting and function, are functional when fused as a single RNA molecule (Jinek et al., 2012). This engineered crRNA + tracrRNA RNA molecule is referred to as a single guide RNA (sgRNA) or more simply as a guide RNA (gRNA). Second, the RGN encoded by *S. pyogenes*, is a single protein, whereas the RGNs encoded by a variety of other bacteria are multi-subunit proteins. Thus, in order to use the Cas9 nuclease to introduce a DSB *in vivo* at the locus N₂₀-NGG, it is only necessary to engineer the *in vivo* production of the Cas9 protein and a gRNA molecule in which the first 20 nucleotides correspond to 20 nucleotides directly 5' of an NGG sequence. The 3' end of the gRNA allows it to bind to the Cas9 nuclease (Figure 1).

Once demonstration of the simplicity of the *S. pyogenes* Cas9 nuclease for gene editing was published, there has been an explosion of publications concerning the use of the Cas9-gRNA system for a wide variety of *in vivo* gene editing applications. In the past few years, the Cas9-gRNA system has been used to carry out genome editing in large variety of species and cell types, including human and mouse tissue culture cells and a variety of animals and plants, including several fungi, the metazoans *C. elegans*, fruit flies, salamanders, frogs, mice, rats, and monkeys, and the plants Arabidopsis, rice, wheat, tobacco, and sorghum. The Cas9-gRNA system can also be used to edit multiple genes simultaneously or to carry out genome-wide gene inactivation studies to determine the functions of specific genes. The specificity of the Cas9-gRNA system appears to be high with a relatively low level of off target effects, however there are many efforts to use alternative Cas9 proteins that improve specificity of cleavage to further enhance the utility of this system. These improved specificity reagents will be important in mammals, where potential therapeutic uses necessitates increased specificity.

Implementing Cas9-gRNA gene editing requires the engineering of the bacterial Cas9 protein to function in a particular eukaryotic cell type, for example by codon optimization and the addition of a nuclear localization signal, as well as the engineering of a gRNA, including its expression from a suitable promoter. Delivery of Cas9 and gRNAs into cells and organisms can be accomplished using a variety of techniques including electroporation, transfection, or the use of viral vectors. In many cases, it is also possible to deliver *in vitro* transcribed gRNA directly into cells or embryos, including *C. elegans*, fruit flies, zebra fish, and mice. The CAS9-gRNA complex can be formed *in vitro*, and that complex can be delivered into mammalian cells using nucleofection technologies or into other organisms using micro-injection. In the case of plants, Cas9 and gRNAs can be transfected into protoplasts or delivered into plant tissues by *Agrobacterium tumefaciens*-mediated transfer.

In the case of *C. elegans*, researchers have introduced Cas9 and sgRNAs into the adult germline by micro-injection of DNA plasmids, mRNA, as well as *in vitro* translated Cas9 protein complexed with a synthetic sgRNA (Cho et al., 2013; Friedland et al., 2013; Lo et

al., 2013). Considering the many CRISPR/Cas9 protocols published in the *C. elegans* field, an average experiment to generate a specific edit or non-specific disruption of a target gene takes roughly 2 weeks (from engineering a specific sgRNA construct for the gene of interest to obtaining animals homozygous for the desired alteration).

Interestingly, Cas9 contains 2 nuclease domains, one of which cleaves the DNA strand complementary to the gRNA, whereas the other nuclease domain cleaves the opposite strand. Mutating one of these domains turns Cas9 into a DNA nickase, which can be used in some HDR-mediated gene editing applications (Doudna and Charpentier, 2014; Sander and Joung, 2014). Mutating both domains creates a nuclease-dead Cas9, which nevertheless binds to DNA and which can be engineered by the addition of suitable domains to function either as a transcriptional activator or repressor. Such constructs can be used in genome-wide studies to study the phenotypic effects of gene knockdowns, analogously to genome-wide RNAi (Doudna and Charpentier, 2014). In addition to altering transcription, GFP-tagged nuclease dead Cas9 can be used in conjunction with multiple sgRNAs that tile along a genomic locus to directly visualize chromosome conformational dynamics in human cells (Chen et al., 2013). Adaptation of this method to model organisms could potentially be an exciting new approach for examining DNA dynamics across some or all cells of a live multi-cellular organism.

Although most RGN-mediated gene-editing studies to date have been carried out with the Cas9-gRNA system from *S. pyogenes*, many other CRISPR-CAS systems are available for potential genetic engineering purposes, including RGNs that utilize different PAM sequences. Currently, the requirement for having the “NGG” PAM motif directly 3' to the 20mer protospacer sequence of choice constrains experiment design. The emergence of engineered variant Cas9 genes with expanded PAM specificities (Kleinstiver et al., 2015) should alleviate some of this design rigidity. Recently, variant PAMs were demonstrated to work in *C. elegans* (Bell et al., 2015). The widely-used crispr.mit.edu website provided by the Zhang Lab at the Broad Institute does not currently output sgRNA sequences with alternative PAMs, but there are other online resources, including “Chop Chop” (<https://chopchop.rc.fas.harvard.edu/index.php>), which can factor in variant PAMs.

While antibodies are readily available for mammalian proteins, there is a dearth of commercially available antibodies for model organisms including *C. elegans* and *D. melanogaster*. CRISPR/Cas9 technology allows researchers to append small epitope tags (Flag, HA, Myc, V5, etc...) in-frame to endogenous gene loci providing a potential solution to this lack of antibodies; tagging endogenous gene loci also eliminates any concern of position effects of single copy transgenes (Dickinson et al., 2013; Paix et al., 2014; Zhao et al., 2014). It should be noted that epitope tagging may also disrupt the function of certain genes. In the future, it will be important to collect information on successful CRISPR-tagging experiments that result in an unexpected disruption of a particular gene's function; information might be gleaned about a given protein's function when tagging either the N- or C- terminus (or both) disrupts that protein's function. For an organism like *C. elegans*, in which many interesting non-null alleles of genes have previously been isolated, epitope-tagging an endogenous gene locus carrying such a mutation of interest should greatly facilitate the investigation of how that mutation changes the gene's function.

As of June 2016, Current Protocols in Molecular Biology has published [FILL IN NUMBER] detailed protocols on the Cas9-gRNA system: “CRISPR/Cas9-Directed Genome Editing of Cultured Cells” (Unit 31.1); “CRISPR-Cas9 Genome Editing in *Drosophila*” (Unit 31.2); and “Genome Editing in Human Cells Using CRISPR/Cas Nucleases (Unit 31.3). Additional Cas9-gRNA protocols for *C. elegans* and *Arabidopsis* are in press and other will be commissioned in the future.

As of June 2016, Current Protocols in Molecular Biology has published six units of detailed protocols on the Cas9-gRNA system: “CRISPR/Cas9-Directed Genome Editing of Cultured Cells” (Unit 31.1); “CRISPR-Cas9 Genome Editing in *Drosophila*” (Unit 31.2); “Genome Editing in Human Cells Using CRISPR/Cas Nucleases (Unit 31.3); “Practical Considerations for Using Pooled Lentiviral CRISPR Libraries” (Unit 31.5); “CRISPR/Cas9-Based Multiplex Genome Editing in Monocot and Dicot Plants” (Unit 31.6); and “CRISPR-Cas9-guided Genome Engineering in *C. elegans*” (Unit 31.7). Additional Cas9-gRNA protocols for *E. coli* and zebrafish are in press and others will be commissioned in the future.

In closing, it is important to point out that the development of the Cas9-gRNA system of genome editing was a completely unintended consequence of basic research in prokaryotic species, demonstrating that support of fundamental research independent of its potential applications is a key driver of scientific and technological innovations. It is also important to stress that the Cas9-gRNA technology now affords scientists unprecedented ability to genetically engineer the genomes of all organisms on earth, including man. The potential of Cas9-mediated genomic editing on a population level was recently demonstrated by Gantz and Bier (Gantz and Bier, 2015) who converted the eye color of fruit flies using an autocatalytic CRISPR/Cas9 editing system. Genetically manipulating organisms on a population scale raises important ethical considerations that need to be addressed by scientists working in conjunction with appropriate policy planners.

Acknowledgments

FMA acknowledges support from the National Institutes of Health (5P01AI083214-08 (M. Gilmore, PI) and 5P30 DK040561-19 (W.A. Walker, PI)).

Literature Cited

- Bell RT, Fu BX, Fire AZ. Cas9 Variants Expand the Target Repertoire in *Caenorhabditis elegans*. *Genetics*. 2015
- Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*. 2009; 326:1509–1512. [PubMed: 19933107]
- Carroll D. Genome engineering with zinc-finger nucleases. *Genetics*. 2011; 188:773–782. [PubMed: 21828278]
- Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*. 2013; 155:1479–1491. [PubMed: 24360272]
- Cho SW, Lee J, Carroll D, Kim JS, Lee J. Heritable gene knockout in *Caenorhabditis elegans* by direct injection of Cas9-sgRNA ribonucleoproteins. *Genetics*. 2013; 195:1177–1180. [PubMed: 23979576]

- Dickinson DJ, Ward JD, Reiner DJ, Goldstein B. Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. *Nat Methods*. 2013; 10:1028–1034. [PubMed: 23995389]
- Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 2014; 346:1258096. [PubMed: 25430774]
- Friedland AE, Tzur YB, Esvelt KM, Colaiacovo MP, Church GM, Calarco JA. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat Methods*. 2013; 10:741–743. [PubMed: 23817069]
- Gantz VM, Bier E. Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science*. 2015; 348:442–444. [PubMed: 25908821]
- Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. 2014; 157:1262–1278. [PubMed: 24906146]
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012; 337:816–821. [PubMed: 22745249]
- Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, Gonzales AP, Li Z, Peterson RT, Yeh JR, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*. 2015; 523:481–485. [PubMed: 26098369]
- Lander ES. The Heroes of CRISPR. *Cell*. 2016; 164:18–28. [PubMed: 26771483]
- Lo TW, Pickle CS, Lin S, Ralston EJ, Gurling M, Schartner CM, Bian Q, Doudna JA, Meyer BJ. Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions. *Genetics*. 2013; 195:331–348. [PubMed: 23934893]
- Marraffini LA. CRISPR-Cas immunity in prokaryotes. *Nature*. 2015; 526:55–61. [PubMed: 26432244]
- Mojica FJ, Ferrer C, Juez G, Rodriguez-Valera F. Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol Microbiol*. 1995; 17:85–93. [PubMed: 7476211]
- Paix A, Wang Y, Smith HE, Lee CY, Calidas D, Lu T, Smith J, Schmidt H, Krause MW, Seydoux G. Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 Sites in *Caenorhabditis elegans*. *Genetics*. 2014; 198:1347–1356. [PubMed: 25249454]
- Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*. 2014; 32:347–355. [PubMed: 24584096]
- Selle K, Barrangou R. Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends Microbiol*. 2015; 23:225–232. [PubMed: 25698413]
- Zhao P, Zhang Z, Ke H, Yue Y, Xue D. Oligonucleotide-based targeted gene editing in *C. elegans* via the CRISPR/Cas9 system. *Cell Res*. 2014; 24:247–250. [PubMed: 24418757]

Cas9 anti-viral immunity in *S. pyogenes*

Cas9 - sgRNA complex for genome engineering

Figure 1.