CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes

Alexis C. Komor,1,2,3 Ahmed H. Badran,1,2,3 and David R. Liu1,2,3,*
1Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA
2Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138, USA
3Broad Institute of MIT and Harvard, Cambridge, MA 02141, USA
*Correspondence: drlu@fas.harvard.edu
http://dx.doi.org/10.1016/j.cell.2016.10.044

The CRISPR-Cas9 RNA-guided DNA endonuclease has contributed to an explosion of advances in the life sciences that have grown from the ability to edit genomes within living cells. In this Review, we summarize CRISPR-based technologies that enable mammalian genome editing and their various applications. We describe recent developments that extend the generality, DNA specificity, product selectivity, and fundamental capabilities of natural CRISPR systems, and we highlight some of the remarkable advancements in basic research, biotechnology, and therapeutics science that these developments have facilitated.

Introduction

Genome editing, the introduction of a desired change to the sequence of genomic DNA, is driving a revolution in the biomedical sciences and has the potential to provide future treatments for many human diseases with a genetic component. The ideal genome-editing tool would edit any genomic locus with high efficiency, high DNA sequence specificity, and little or no undesired byproducts. Such an ideal agent has not yet been developed and is unlikely to exist in nature, given that naturally occurring forms of genome-editing proteins evolved to achieve only partially related functions such as modulation of gene expression or protection from viral infection. Researchers have therefore recognized the need to develop new tools that increase the scope and effectiveness of genome editing, especially in eukaryotic cells and animal models of human disease. Recent efforts have resulted in remarkable advances toward this goal in a relatively short time period.

Early genome-editing efforts were enabled by the discovery that the endogenous cellular homologous recombination repair pathway could be used to replace a small portion of the genome of a living cell with an exogenous donor DNA sequence. To use this strategy for genome editing, the exogenous DNA sequence must have homology to the target genomic DNA site. Following transfection of the donor DNA, incorporation at the desired locus spontaneously occurs very inefficiently, at rates of 1 in every $10^3$ to $10^9$ cells, depending on the cell type and cell state (Smithies et al., 1985; Thomas et al., 1986). This technique of spontaneous homologous recombination was used in mouse embryo-derived stem cells, allowing researchers to generate mice with a desired genotype (Capecchi, 1989).

In addition to the low efficiency of editing using spontaneous homologous recombination, this approach could also induce undesired genome-editing events in which the exogenous DNA sequence was incorporated into the genome at random sites more frequently than at the desired locus (Lin et al., 1985).

A key advance in overcoming this limitation was the observation, first in yeast and then in mammalian cells, that the introduction of a double-stranded break (DSB) into the genomic locus using a meganuclease, an endonuclease that recognizes and cleaves a long DNA sequence, would stimulate homology-driven DNA incorporation (Figure 1A) (Rudin et al., 1989; Rouet et al., 1994). This “homology-directed repair” (HDR) strategy enhanced the efficiency of the desired genome-editing event by two to three orders of magnitude, resulting in targeted incorporation typically being much more efficient than incorporation at random sites in the genome (Choulika et al., 1995; Jasin, 1996).

Despite this breakthrough, genome editing still suffered from two major drawbacks. First, non-homologous end joining (NHEJ) also occurs at the site of DSBs, typically more efficiently than HDR, resulting in stochastic insertions and deletions (indels) of nucleotides at the target locus (Figure 1A) (Jeggo, 1998). While NHEJ-mediated genome editing is useful for gene disruption, indels are unwanted byproducts when precise genome editing is desired. Second, since the probability that a known meganuclease cleaves a particular target locus of interest is extremely small, either a meganuclease recognition site must be incorporated into the genomic locus of interest (Jasin, 1996) or a meganuclease must be engineered to cleave the target locus (Sussman et al., 2004; Rosen et al., 2006; Grizot et al., 2009).

Overcoming the first drawback is a focus of current research and will be discussed later in this Review. To address the second drawback, researchers turned to zinc-finger nucleases (ZFNs) (Bibikova et al., 2002, 2003; Porteus and Baltimore, 2003; Urnov et al., 2010) and transcription-activator-like effector nucleases (TALENs) (Figure 1B) (Li et al., 2011a, 2011b; Joung and Sander, 2013), engineered nucleases based on arrays of naturally occurring DNA-binding domains fused to the nonspecific DNA cleavage domain from FokI. Because the amino acid sequences of zinc-finger arrays and TALE repeat arrays, unlike most
DNA-binding proteins, can be readily designed to bind to virtually any target DNA sequence, ZFNs and TALENs can be engineered to cleave a target genomic loci with fairly high specificity (Carroll, 2008; Boch et al., 2009; Moscou and Bogdanove, 2009; Miller et al., 2011; Zhang et al., 2011; Joung and Sander, 2013). The design of ZFNs is complicated by their extensive protein-DNA contacts, however, and the cloning of TALEN genes is impeded by their highly repetitive nature. In addition, each new target locus requires the design, gene synthesis, expression, and validation of a new ZFN or TALEN protein (Figure 1B) (Urnov et al., 2010; Miller et al., 2011).

This significant barrier to genome editing—that each new target site requires the design and construction of a new nuclease—was substantially lowered by the advent of CRISPR-Cas9 as an RNA-guided DNA endonuclease (Garneau et al., 2010; Jinek et al., 2012; Gasiuskas et al., 2012). In this system, a Cas endonuclease protein forms a complex with a “guide RNA” molecule and localizes to a target DNA sequence following simple guide RNA:genomic DNA base-pairing rules (Figure 1B) (Doudna and Charpentier, 2014; Hsu et al., 2014). The target DNA sequence (the protospacer) must be both complementary to the guide RNA and also contain a “protospacer-adjacent motif” (PAM), a short DNA sequence that is required for compatibility with the particular Cas protein being used (Deveau et al., 2008; Garneau et al., 2010; Sapraunaskas et al., 2011; Jinek et al., 2012; Gasiuskas et al., 2012). While this new technology places a modest limitation on the number of genomic sites amenable to genome editing due to the PAM requirement, it replaces the complex protein design and engineering tasks associated with ZFNs and TALENs with the much-simpler task of designing a new guide RNA for each genomic site of interest using simple Watson-Crick base-pairing (Cong et al., 2013; Mali et al., 2013b; Jinek et al., 2013).

The elucidation of the mechanics of CRISPR-Cas9 (Barrangou et al., 2007; Garneau et al., 2010; Deltcheva et al., 2011; Sapraunaskas et al., 2011; Jinek et al., 2012; Gasiuskas et al., 2012), and its adaptation for use in eukaryotic genome editing (Cong et al., 2013; Mali et al., 2013b; Wang et al., 2013; Cho et al., 2013; Hwang et al., 2013b; Jinek et al., 2013), has had a transformative impact on the life sciences. The ease with which new DNA sequences can be targeted for genome editing has enabled scientists to rapidly discover new gene functions, develop new cell and animal models of diseases, and make substantial progress toward human therapeutics. In this Review, we summarize some of the recently developed tools that use CRISPR-Cas9 for the manipulation of mammalian genomes and their applications in basic science, biotechnology, and medicine.

## New Natural CRISPR Enzymes

Several natural CRISPR nucleases have now been used for mammalian genome editing. Each CRISPR nuclease can vary in size, PAM requirement, and location of the introduced DSB within the protospacer (Table 1). The most commonly used variant is the 1,368-residue Cas9 protein from *Streptococcus pyogenes* (SpCas9) (Haft et al., 2005). Most known naturally occurring Cas9 nucleases, including SpCas9, natively use two different RNA molecules, the CRISPR-RNA (crRNA) and the trans-activating crRNA (tracrRNA), to form a functional guide RNA:Cas9 complex (Deltcheva et al., 2011). The discovery that a single-guide RNA (sgRNA) could take the place of the crRNA and the tracrRNA further simplified the use of the CRISPR-Cas9 system such that only one protein and one RNA molecule are needed to achieve RNA-programmed DNA cleavage (Jinek et al., 2012).

The relatively simple PAM requirement of NGG contributes to the popularity of SpCas9 for genome editing (Table 1). The *Staphylococcus aureus* (Sa) Cas9 analog (SaCas9) offers a smaller size (1,053 residues) that facilitates some of the applications described below but requires a more-complex PAM of NNGRRT (Ran et al., 2015; Friedland et al., 2015). Other Cas9 homologs with different PAM requirements have also been used for mammalian genome editing. For example, the *Streptococcus thermophilus* (St) Cas9 proteins St1Cas9 and St3Cas9 are 1,121 and 1,388 residues and require NNAGAAW and NGGNG PAMs, respectively (Table 1) (Gasiunas et al., 2012; Ran et al., 2013b; Hou et al., 2013). The Neisseria meningitides (Nm) Cas9 protein (NmCas9) is 1,082 residues and requires a NNNNGATT PAM (Table 1) (Gasiunas and Siksnys, 2013; Esvelt et al., 2013; Ran et al., 2013b; Müller et al., 2016).

## Figure 1. Genome Editing Using Double-Stranded Breaks

(A) A programmable nuclease incorporates a sequence-specific double-stranded break (DSB) in genomic DNA. In the absence of a repair template, the cell will process the DSB mostly by NHEJ, resulting in indels at the site of editing. In the presence of a separate DNA template containing sequences homologous to the regions flanking the DSB, HDR can result in incorporation of the repair template into the genomic DNA. (B) ZFNs, TALENs, and CRISPR-based nucleases have also been used to introduce programmable, sequence-specific DSBs. The ability of Cas9 to be reprogrammed to bind a new sequence (the protospacer and PAM) by designing a new sgRNA, rather than by engineering a new DNA-binding protein (such as ZF or TALE), has transformed the genome-editing field.
Expanding the Targeting Scope of Cas9

As genome-editing techniques using RNA-guided nucleases become more precise and diverse, the need for agents with different PAM requirements increases. The relatively simple NGG PAM sequence of SpCas9 occurs on average every 8–12 bp in the human genome (Cong et al., 2013; Hsu et al., 2013), a frequency that is not excessively limiting for classical HDR- and NHEJ-based genome editing, as multiple DNA cleavage locations can lead to the same desired HDR or NHEJ outcome. The discovery of additional naturally occurring RNA-guided nucleases such as those in Table 1 offer additional targeting flexibility. For other genome-editing techniques such as base editing (see below) or when it is necessary to distinguish between a wild-type (WT) and mutant allele, however, precise targeting of a locus with single-nucleotide resolution can be critical. In these cases, the PAM requirements can be a major restriction.

WT SpCas9 has been shown to have some activity on sites with NAG and NGA PAMs but typically with much lower efficiencies than on sites with canonical NGG PAMs (Jiang et al., 2013; Hsu et al., 2013; Zhang et al., 2014; Kleinstiver et al., 2015b). A recent study used a bacterial selection system to identify three new variants of SpCas9 that can target NGA, NGAG, and NGCG PAMs with high efficiencies and specificities (Table 1) (Kleinstiver et al., 2015b). This study is an exciting example of how a small number of mutations—in these cases, three to four—can substantially alter the PAM specificity of an RNA-guided nuclease.

Researchers have also engineered Cas9 enzymes to exhibit relaxed PAM specificities. In one approach, an unbiased selection system was used to relax the NNGRRT PAM requirement of SaCas9 to NNNRRT (Table 1). The engineered variant had three mutations and exhibited off-target editing comparable to that of the WT enzyme in human cells (Kleinstiver et al., 2015a). In a different study, the crystal structure of FnCas9 was used to guide the rational design of a variant with a relaxed PAM requirement. While the WT FnCas9 recognizes a NGG PAM, the engineered variant (which differs from WT at three residues) requires only a YG PAM and can be used to edit mammalian genomes when the protein is pre-complexed with sgRNA and directly injected into zygotes (Table 1) (Hirano et al., 2016). These important advances expand the number of target loci amenable to RNA-guided genome editing.

Improving the DNA Specificity of CRISPR-Based Agents

In addition to expanding the targeting scope of genome-editing agents, improving their DNA specificity has also been a major priority. Researchers have revealed the DNA-targeting specificities of CRISPR-based genome-editing agents using a variety of approaches. These methods include ChIP-seq (Cencic et al., 2014; Kuscu et al., 2014; Wu et al., 2014; O’Geen et al., 2015), targeted analysis of genomic sites identified through computational predictions (Fu et al., 2013; Hsu et al., 2013), in vitro high-throughput profiling methods (Pattanayak et al., 2013), whole-genome sequencing methods (Smith et al., 2014; Veres et al., 2014; Yang et al., 2014; Kim et al., 2015), the GUIDE-seq
Researchers have developed several strategies to substantially improve the specificity of SpCas9 (and, likely, other CRISPR agents) without making any changes to the Cas9 protein sequence. Off-target modification by SpCas9 can be decreased up to several orders of magnitude simply by truncating the sgRNA sequence. Off-target modification of SpCas9 variants appears to be higher than that of the SpCas9 variant (Figure 2D) (Kim et al., 2016; Kleinstiver et al., 2016b).

Researchers have also engineered variants of Cas9 that are activated by light or exogenous small molecules. These variants, including an intein-inactivated Cas9 system (Davis et al., 2015) and a small-molecule-dimerized split Cas9 system (Zetsche et al., 2015b), have been shown to substantially improve genome-editing specificity in mammalian cells compared with WT Cas9 by carefully controlling the temporal window within which active Cas9 is generated so that less active Cas9 is present after modification of the on-target loci is complete (Figures 2G and 2H). Similar systems, such as light-activated Cas9 variants (Nihongaki et al., 2015a; Hemphill et al., 2015; Jain et al., 2016), split Cas9 variants (Truong et al., 2015; Wright et al., 2015), small-molecule induction of Cas9 (Dow et al., 2015), and an engineered allosterically regulated Cas9 (Oakes et al., 2016), could also be used to reduce off-target genome editing following these same principles.

An additional strategy to reduce off-target genome editing through Cas9 engineering is to require that two separate Cas9 binding events take place at the same locus in order to result in DNA cleavage. Cas9 can be converted to a nickase enzyme (Cas9n) by inactivating either of its two catalytic residues (Mali et al., 2013a; Ran et al., 2013b). By designing two sgRNAs that bring separate Cas9n molecules to nick opposite DNA strands, double-stranded breaks only occur with simultaneous binding events (Figure 2E). This strategy reduces the theoretical likelihood of off-target events from 1/n to ~1/n^2; in practice, paired nicking reduced off-target activity up to several orders of magnitude in mammalian cells while retaining on-target activity (Ran et al., 2013a; Mali et al., 2013a). Inactivation of both catalytic residues results in dCas9, which cannot cleave either DNA strand but retains its ability to bind to a target DNA sequence. Fusion of the nonspecific restriction endonuclease FokI, which requires dimerization to become catalytically competent, to dCas9 results in an engineered variant that requires dual guide RNAs to coordinate FokI-dCas9 dimerization at a specific locus.
This approach results in up to two orders of magnitude of improved specificity in mammalian cells, although with somewhat reduced activity compared to WT Cas9. Because ChIP-seq experiments suggest that dCas9 binding is more promiscuous than Cas9 cleavage (Wu et al., 2014), however, a paired nickase approach may offer additional specificity advantages compared with the use of FokI-dCas9 dimers.

Recently, structure-guided engineering of SpCas9 has yielded variants with improved specificity. Several previous studies have implicated excess DNA-binding energy as a source of off-target genome-editing activity among ZFNs, TALENs, and Cas9 (Gupta et al., 2011; Pattanayak et al., 2011, 2013; Guilinger et al., 2014a). By introducing just three to four mutations into SpCas9 that neutralize nonspecific electrostatic interactions between the protein and the sugar-phosphate backbone of its target DNA, its DNA specificity increases dramatically (Figure 2C) (Slaymaker et al., 2016; Kleinstiver et al., 2016a).

Together, these advances in enhancing the specificity of CRISPR-based genome-editing agents in mammalian cells are key developments that increase their promise both as research tools and as potential future therapeutics.

**Improving the Product Selectivity of Genome-Editing Agents**

Because all of the above genome-editing tools are programmable nucleaseases that create a double-stranded break in DNA, a target locus is not desired, and precise editing in which the predominant product is the replacement of one allele for another is required. Although HDR is capable of affecting precise allele replacement, HDR efficiency is typically quite low (<5%); depends on many factors, including cell type and cell state; and is competitive with NHEJ. DNA nicks do not commonly result in NHEJ, and thus, the usage of a single Cas9 nickase with a donor DNA template usually results in lower indel frequency (Cong et al., 2013). This single-nickase strategy, however, also results in substantially decreased HDR-mediated editing efficiencies compared to that of WT Cas9 or double nicking strategies and is not amenable to all cell types.

To improve the product selectivity of genome editing to favor precise allele replacement, researchers have inhibited certain endogenous DNA repair components to favor HDR over NHEJ (Figure 3D). For example, the small molecule Scr7 is known to inhibit DNA ligase IV, a key component of NHEJ (Srivastava et al., 2012; Vartak and Raghavan, 2015). Administration of Scr7 in combination with Cas9, a sgRNA, and a donor DNA template has been shown to enhance HDR:NHEJ ratios substantially in some systems (Vartak and Raghavan, 2015; Chu et al., 2015; Maruyama et al., 2015), enhancers of HDR (Yu et al., 2015; Pinder et al., 2015; Song et al., 2016), or cell-cycle synchronizers (Lin et al., 2014b) can be used to increase the ratio of HDR:NHEJ genome editing products.
suppresses NHEJ-mediated indel formation and increases HDR-mediated genome editing when administered in combination with Cas9, a sgRNA, and a donor DNA template (Chu et al., 2015).

As an alternative strategy to suppressing NHEJ, a small molecule activator of Rad51, a protein involved in HDR, can augment HDR-mediated genome editing when added to cells treated with Cas9, a sgRNA, and a donor DNA template (Figure 3D) (Pinder et al., 2015; Song et al., 2016). A high-throughput screen designed to identify potential small molecules capable of enhancing HDR-mediated genome editing has been used to discover additional compounds that can be administered to cells to increase precision genome editing (Yu et al., 2015). HDR is known to be dependent on many variables, such as the cell and tissue type, the location of the target DNA within the chromosome, and the cell cycle (Biswas et al., 1992; Saleh-Gohari and Hellday, 2004; Heyer et al., 2010; Miyakoa et al., 2016). Chemically synchronizing cells to arrest them in G-phase by blocking M-phase before delivering Cas9 RNP and donor DNA template will increase HDR rates (Lin et al., 2014b).

The use of exogenous molecules to manipulate components of the NHEJ or HDR pathways therefore can be used to enhance HDR outcomes (Figure 3D), but such treatments are typically very perturbative to cells, are cell-type specific, and impair the cell’s ability to perform native DNA repair functions. These strategies therefore may have limited relevance in a therapeutic context or in research settings in which preserving native cell states is a priority.

Judicious design of the donor DNA template can also enhance HDR-mediated precision genome-editing efficiency. The observation that Cas9 dissociates from its cleaved DNA substrate asymmetrically prompted a study on the effects of donor template geometry on HDR rates. Researchers discovered that if the ssDNA template is designed such that it can anneal to the substrate for subsequent cleavage by Cas9 (Figure 3C). In some cases, this reprogramming of undesired HDR product can be avoided by installing silent (or acceptable) mutations into the donor template such that the HDR product contains mutations in the PAM or the PAM-proximal region of the protospacer, thereby blocking Cas9 from cutting the product (Figure 3C) (Paquet et al., 2016).

“Base editing” is a new strategy to introduce point mutations in an RNA-programmed manner that does not rely on HDR or double-stranded DNA breaks (Komor et al., 2016). The fusion to dCas9 of a cytidine deaminase enzyme that operates on single-stranded DNA (but not duplex DNA) allows C to U conversion within a small (~3- to 5-base) window of the protospacer, exploiting the presence of a short segment of accessible single-stranded DNA in the “R-loop” ternary complex (Jiang et al., 2016) between Cas9, the guide RNA, and the target DNA. To make base editing efficient and permanent in mammalian cells, the dCas9-cytidine deaminase fusion protein was further engineered to inhibit base excision repair at the site of the edit and to induce cellular mismatch repair to replace the original C:G base pair with a T:A base pair (Figure 3A) (Komor et al., 2016). Base editing efficiencies are typically much higher than the efficiency of HDR-mediated point mutation. In addition, because base editing avoids making double-stranded DNA breaks, indel formation is minimized (Nishida et al., 2016; Komor et al., 2016). Current limitations of this new strategy include off-target editing that parallels Cas9 off-target DNA cleavage, the requirement of an NGG PAM that is a specific distance from the target C, and the inability to distinguish among multiple Cs present within the editing window. Advances described above that expand the targeting scope and improve the DNA specificity of Cas9 may offer solutions to some of these limitations.

Epigenome Editing
In addition to changing the DNA sequence of a given genome, researchers have also begun to edit the epigenome in order to alter the regulation of a target gene, rather than its sequence identity (Thakore et al., 2016). One technique to increase expression of a specific gene is to tether the dCas9:sgRNA complex to a transcriptional activator and program it to bind near the transcriptional start site of a gene of interest. For example, the transcriptional activation domain VP64, which consists of four tandem copies of the Herpes Simplex Viral Protein 16 (VP16), can be fused directly to the C terminus of dCas9 and used to increase the expression of a wide variety of different genes (Figure 4A) (Mai et al., 2013a; Maeder et al., 2013; Perez-Pinera et al., 2013). This fusion has been shown to be target-sequence-specific and capable of opening chromatin (Polstein et al., 2015). A light-activated version of this system has been developed by fusing dCas9 and VP64 or p65, the transactivation domain of NF-κB, to each member of a light-activated heterodimerizing pair of proteins (Figure 4A) (Nihongaki et al., 2015b; Polstein and Gersbach, 2015). In addition, the small-molecule-dimerized split Cas9 system described above has been adapted to generate a small-molecule-activated dCas9-VP64 activator (Figure 4A) (Zetsche et al., 2015b). These systems allow dynamic spatial and temporal control of gene activation.

A variety of second-generation dCas9-activator fusions have been subsequently engineered that incorporate varying copies of VP16 (Cheng et al., 2013; Chakraborty et al., 2014), the tripartite activator VPR (VP64-p65-Rta, where Rta is the transcriptional activation domain of the Epstein-Barr virus) (Chavez et al., 2015), or the repeating peptide array SunTag that subsequently recruits multiple copies of an antibody-VP64 fusion (Tangenbaum et al., 2014) (Figure 4A). These various constructs offer strong gene activation in a variety of mammalian cell types (Chavez et al., 2016). Alternately, transcriptional activators can be attached to the sgRNA through an RNA hairpin that binds with very high affinity to the MS2 bacteriophage coat protein (Peabody, 1993). Coexpression of WT dCas9 or dCas9-VP64, the sgRNA-hairpin construct, and a MS2-VP64 (Zalatan et al., 2015; Konermann et al., 2015) or MS2-p65-HSF1 (where HSF1 is the activation domain from the human heat-shock transcription factor 1) (Konermann et al., 2015) fusion protein results in assembly of a RNA-guided transcriptional activator (Figure 4A). The combination of dCas9-VP64 and MS2-p65-HSF1, termed the synergistic activation mediator (SAM), exhibits particularly robust transcriptional activation (Konermann et al., 2015).
Conversely, dCas9 can be used as a transcriptional repressor on its own, simply exploiting its high affinity for target DNA and ability to block components of the transcriptional machinery (Qi et al., 2013) or as a fusion to the Kruppel-associated box (KRAB) transcriptional repressor (Figure 4B) (Gilbert et al., 2013; Lawhorn et al., 2014; Thakore et al., 2015). Together, these techniques of activating (CRISPRa) or repressing (CRISPRi) genes at will can be combined to allow researchers to reversibly modulate gene expression with a dynamic range of several orders of magnitude (Gilbert et al., 2014). A public tool is available to facilitate the design of appropriate sgRNAs to enable efficient CRISPRi and CRISPRa (Liu et al., 2015a). Researchers have also used genome-wide sgRNA libraries in high-throughput CRISPRi and CRISPRa screens to identify genes that result in phenotypes of interest when upregulated or downregulated (Gilbert et al., 2014). Orthogonal RNA-protein-binding modules have been incorporated into dCas9 transcriptional regulation complexes that allow for simultaneous gene activation and repression (Zalatan et al., 2015).

The packaging of DNA into chromatin is a key feature of eukaryotic DNA, and researchers have begun to develop tools that allow chromatin manipulation in a sequence-programmed manner (Keung et al., 2015). The fusion of the DNA methyltransferase enzyme DNMT3A to dCas9 results in an epigenetic modifier that methylates CpG islands within ~100 bp of the sgRNA-programmed genomic loci (Figure 4B) (McDonald et al., 2016; Liu et al., 2016; Vojta et al., 2016). Tethering the catalytic domain of the DNA demethylase Tet1 to dCas9 has been used to induce targeted DNA demethylation (Figure 4A). These fusions can induce over 90% demethylation of CpG islands within a 200-bp range of the target site (Xu et al., 2016; Morita et al., 2016; Liu et al., 2016). Researchers have also developed a histone-modifying CRISPR-based tool in which the catalytic domain of human acetyltransferase p300 was fused to the C terminus of dCas9. This fusion catalyzes histone H3 lysine 27 (H3K27) acetylation at loci up to thousands of base pairs from the sgRNA-specified locus and results in transcriptional activation of genes (Figure 4A) (Hilton et al., 2015). Alternately, the histone demethylase LSD1 has been fused to dCas9, allowing for demethylation of dimethylated histone H3 lysine 4 (H3K4me2) at sites > 350 bp from the sgRNA binding site (Figure 4B) (Kearns et al., 2015). The much-larger activity window of the methylation tools, unlike the small window of base editing, arises from the use of domains that operate on double-stranded DNA in the reported epigenome editing agents, in contrast with the use of single-stranded DNA-specific enzymes in base editors.

Figure 4. CRISPR-Based Epigenome Editing
(A) RNA-programmed gene activators can be assembled through the direct fusion of dCas9 with the transcriptional activators VP64 (Mali et al., 2013a; Maeder et al., 2013; Perez-Pinera et al., 2013; Tanenbaum et al., 2014) and VPR (Chavez et al., 2015), the histone acetyltransferase enzyme p300 (Hilton et al., 2015), or the DNA demethylase Tet1 (Xu et al., 2016; Morita et al., 2016; Liu et al., 2016). Alternately, the transcriptional activators VP64 and p65-HSF1 can be attached to the sgRNA (Zalatan et al., 2015; Konemann et al., 2019). Light-activated and small-molecule-activated variants can also be used (Nihongaki et al., 2015b; Zetsche et al., 2015b; Polstein and Gersbach, 2015).
(B) RNA-programmed gene repressors can be assembled by attaching the transcriptional repressor domain KRAB to dCas9 by either a direct fusion between the two proteins (Gilbert et al., 2013; Lawhorn et al., 2014; Thakore et al., 2015) or via the sgRNA. Alternatively, dCas9 can be fused to the DNA methyltransferase enzyme DNMT3a (McDonald et al., 2016; Liu et al., 2016; Vojta et al., 2016) or the histone demethylase LSD1 (Kearns et al., 2015) to result in a transcriptional repressor.
Delivery of Genome-Editing and Epigenome-Editing Agents

Although their substrates are intracellular, the genome-editing and epigenome-editing agents described above are all macro-molecules and, therefore, do not spontaneously enter cells. The delivery of genome-editing agents into cells has, therefore, been the subject of intense research over the past several decades and remains a significant barrier to some applications of genome editing (Bartus et al., 1998; Gaj et al., 2013). For many research applications, the transfection of plasmid DNA-expressing genome-editing proteins and guide RNAs is sufficient. In other cases, including in vivo therapeutic applications, however, DNA transfection is not possible, and alternative methods to deliver genome-editing agents are needed.

A number of effective ex vivo methods have been used to deliver proteins or their encoding genes into cultured mammalian cells. These methods include electroporation or nucleofection, lipid-based transfection, viruses, cationic peptides, and other approaches (Luo and Saltzman, 2000; Maasho et al., 2004; Zeitelhofer et al., 2007; Cockrell and Kafri, 2007; Yin et al., 2014a). For some cell types, including many cancer cell lines and certain blood cells, ex vivo delivery methods, when applied to genome-editing proteins such as Cas9, can be very effective, resulting in the exposure of the vast majority of treated cells to the genome-editing agent (Heckl et al., 2014). For other cell types of interest, including hematopoietic stem cells and some primary cells (i.e., cells taken directly from tissue, rather than replicated in culture), even ex vivo delivery using a wide variety of methods has proven challenging (Amsellem et al., 2003; Lombardo et al., 2007).

Viral delivery of genome-editing agents has been explored using lentivirus, adenovirus, and adeno-associated virus (AAV) (Gori et al., 2015) (Figure 5A). Lentiviruses are able to infect non-dividing cells and have been used in vivo to efficiently transduce a variety of specific target organs (Cockrell and Kafri, 2007). Furthermore, the packaging limit of lentivirus is ~8.5 kb (although inserts larger than ~3 kb are packaged less efficiently), sufficient to package most Cas9 genes, guide RNA expression constructs, and required promoter and regulatory sequences (Kumar et al., 2001; al Yadoub et al., 2007). Lentiviruses have been successfully used to deliver Cas9 and sgRNA genes into mice to characterize the contributions of a panel of tumor-suppressor genes to the progression of lung cancer (Sánchez-Rivera et al., 2014).

Adenoviruses are also capable of infecting both replicating and non-replicating cells but do not integrate their DNA into the host cell genome and can elicit a strong immune response in animals (Wang et al., 2004). Adenovirus-mediated delivery of Cas9 has been used to achieve in vivo genome editing in mouse lungs (Maddalo et al., 2014) and livers (Cheng et al., 2014; Ding et al., 2014; Wang et al., 2015a). Finally, AAV variants engineered for gene therapy can infect both dividing and non-dividing cells, do not integrate its DNA into the host genome, and do not elicit a significant immune response in the host (Wang et al., 2004). A variety of serotypes of AAV are known, offering delivery into different tissue types. However, AAV has a packaging limit of ~4.5 kb of foreign DNA (Wu et al., 2010). Thus, packaging into AAV genes encoding SpCas9 (4.2 kb), a sgRNA, a donor DNA template, and associated promoters and regulatory sequences is generally not possible. The gene encoding SaCas9 (3.2 kb) is significantly smaller than that encoding SpCas9 and can be packaged along with an sgRNA and associated promoters into a single AAV vector (Ran et al., 2015). Alternatively, genes encoding SpCas9 and
its sgRNA have been packaged into separate AAV vectors for in vivo genome editing in mouse brains (Swiech et al., 2015) and livers (Yang et al., 2016). An additional option is to use a split-intein Cas9 system and package each half into separate AAV vectors. Upon coinfection, the Cas9-intein halves associate and undergo protein splicing to yield a complete, covalently intact Cas9 protein. This system has successfully mediated genome editing in cultured human and mouse cells (Truong et al., 2015).

To circumvent the challenges associated with delivery of a Cas9-encoding gene, researchers have developed a mouse that expresses Cas9. A Cre-recombinase-inducible Cas9 transgene was inserted into the Rosa26 locus of the mouse. The resulting mouse can be crossed with mice expressing Cre recombinase from a tissue-specific promoter to generate mice that express Cas9 in specific tissues. An sgRNA and donor DNA can then be delivered using AAV or lentivirus to initiate genome editing (Platt et al., 2014).

Hydrodynamic injection (HDI) of plasmid DNA is a convenient and efficient non-viral delivery option that is particularly well suited for DNA delivery into hepatocytes in rodents (Suda and Liu, 2007) (Figure 5A). Researchers have shown that DNA plasmids encoding Cas9 and sgRNA can be delivered to the livers of mice infected with hepatitis B virus (HBV) via a hydrodynamic tail vein injection. The resulting Cas9:sgRNA complexes cleave HBV viral DNA, resulting in a gradual decrease in HBV expression in the animals (Lin et al., 2014a; Zhen et al., 2015; Ramanan et al., 2015). This technique has also been used to generate mouse liver cancer models. In this study, a plasmid encoding Cas9 and sgRNAs was delivered via HDI to disrupt tumor-suppressor genes. By co-injecting with a donor template for HDR, precise mutations can be introduced to genes of interest, as well (Xue et al., 2014).

Another delivery option that can result in germline genome editing in animals is the direct injection of Cas9 mRNA and purified sgRNA into oocytes or zygotes (Figure 5B). This technique has been used to knock out specific genes in mice (Wang et al., 2013; Li et al., 2013a; Yang et al., 2013a), rats (Li et al., 2013b), zebrafish (Hruscha et al., 2013; Hvlang et al., 2013; Chang et al., 2013; Hu et al., 2013), flies (Bassett et al., 2013), frogs (Nakayama et al., 2013), pigs (Hai et al., 2014), and monkeys (Niu et al., 2014). By co-injecting with a donor DNA, precise HDR-mediated genome editing has been achieved in mice (Wang et al., 2013; Yang et al., 2013a; Wu et al., 2013), nematodes (Lo et al., 2013), and zebrafish (Hruscha et al., 2013; Chang et al., 2013; Auer et al., 2014).

A DNA- and mRNA-free in vivo delivery method directly delivers purified RNPs using cationic lipids (Zuris et al., 2015) (Figure 5A). Due to the polyanionic charge of the sgRNA and its tight association with Cas9 protein, commercially available cationic lipid nucleic acid transfection reagents can efficiently deliver Cas9:sgRNA complexes into mammalian cells both in cell culture and in vivo. This strategy has been successfully used to perform genome editing in mouse inner ear hair cells and in neurons (Zuris et al., 2015; Wang et al., 2016). An additional advantage of the direct Cas9:sgRNA protein:RNA complex delivery over mRNA or DNA delivery is the more transient nature of protein delivery, which results in substantially higher DNA specificity and less off-target editing, for the reasons discussed above (Kim et al., 2014; Zuris et al., 2015; Liu et al., 2015b).

**Applications of CRISPR-Based Genome Editing**

CRISPR components have been used for a variety of creative applications. Here, we summarize some recent uses of CRISPR-based genome editing in eukaryotic cells and organisms for basic research, biotechnology, and therapeutics development. Although a major and important application of Cas9 is the generation of animal and cell models of diseases, including target gene knockouts, these have been reviewed recently and are not covered in this Review (Sander and Joung, 2014; Mou et al., 2015).

**Genome Editing in Human Primary Cells**

Pluripotent stem cells are undifferentiated primary cells that are able to differentiate into virtually any cell type of the human body. Induced pluripotent stem cells (iPSCs) are pluripotent stem cells that can be directly generated from adult somatic cells and have proven useful for regenerative medicine research (Takahashi and Yamanaka, 2006; Yu et al., 2007). Genome editing in iPSCs allows the precise study of human genetic variants in a wide variety of tissues in cell culture. The use of Cas9 for mammalian genome editing advanced this capability by allowing scientists to perform genetic manipulations in iPSCs at efficiencies that are difficult to attain with TALENs or ZFNs (Ding et al., 2013; Yang et al., 2013b; Byrne et al., 2014; Hockenmeyer and Jaenisch, 2016).

Cells isolated from an individual can be genetically modified in a specific way using Cas9, then differentiated alongside identical but unmodified cells. These otherwise isogenic cells allow researchers to directly determine the impact of a given genetic variation on a disease phenotype or to simply study gene function. Cas9-mediated iPSC-derived knockout cell lines for a variety of different genes have been reported for use in loss-of-function studies (Yang et al., 2013b; González et al., 2014; Smith et al., 2015; Liao et al., 2015; Chen et al., 2015b; Liang et al., 2015; Wang et al., 2015b). This technique can also be combined with CRISPRi to specifically, rapidly, and reversibly downregulate specific genes in iPSCs and iPSC-derived cell types (Mandelgeri et al., 2016). Similarly, specific mutations have been introduced into iPSCs and iPSC-derived cells using Cas9 and HDR-mediated genome editing for the study of genetic diseases (Yang et al., 2013b; Hou et al., 2013; Rong et al., 2014; Smith et al., 2015).

A central goal of regenerative medicine is to replace unhealthy or diseased cells with healthy ones. One approach to this goal is cell therapy, in which primary cells are genetically manipulated then implanted into patients in order to cure or treat genetic diseases (Mironov et al., 2004). Several studies have used Cas9 to correct genetic mutations in patient-derived primary cells, including Duchenne muscular dystrophy (MDM) (Li et al., 2015b; Ousterout et al., 2015), Fanconian anemia (Osborn et al., 2015), hemophilia (Park et al., 2015), cystic fibrosis (Schwank et al., 2013), and beta thalassemia (Xie et al., 2014). Additionally, primary immune cells have been made resistant to HIV infection by Cas9-mediated knockout of CCR5 or CXCR4, receptors for HIV entry (Wang et al., 2014b; Li et al., 2015a; Hou et al., 2015; Schumann et al., 2015).
Together, these studies highlight how CRISPR-based genome-editing technologies have accelerated biological studies in primary cells. The development of more precise, specific, and capable genome-editing technologies will likely further augment our understanding of diseased cells as these new technologies are applied to primary cell editing.

**CRISPR-Based Treatment of Animal Models of Human Genetic Disease**

Cas9-mediated genome editing in vivo has been used to correct disease-associated alleles in animal models of genetic diseases. The zygotes of mice heterozygous for a dominant-negative cataract-causing mutation in the CRYGC gene were injected with Cas9 mRNA and an sgRNA targeting only the mutant allele. HDR-mediated correction resulted in cataract-free progeny (Wu et al., 2013). This technique was also applied to mdx mice, a model of DMD harboring a mutation in the gene encoding dystrophin. Cas9, sgRNA, and a donor template were injected into mouse zygotes, resulting in genetically mosaic progeny with 2%-100% gene correction and varying degrees of phenotypic rescue (Long et al., 2014). Disease correction, both genotypically and phenotypically, in post-natal mdx mice has also been reported following AAV-mediated delivery of SaCas9 or SpCas9 and sgRNA to skeletal and cardiac muscle cells to delete the mutated exon from the dystrophin gene (Long et al., 2016; Nelson et al., 2016; Tabebordar et al., 2016). Finally, a mouse model of hereditary tyrosinemia type I (HT1) was corrected in vivo using Cas9. These mice have a homozygous G to A point mutation in the fumarylacetoacetate hydrolase (FAH) gene, which results in severe liver damage. Plasmid encoding Cas9 and sgRNA and a ssDNA oligo were delivered via HDI into adult mice. WT Fah protein was detected in ~1/250 liver cells, and significantly less liver damage was observed as compared to untreated controls (Yin et al., 2014b). Correction of this point mutation was also achieved using the concomitant lipid-mediated delivery of Cas9 mRNA and AAV-mediated delivery of sgRNA and a repair template. This combined viral and non-viral delivery method yielded point mutation correction in >6% of hepatocytes (Yin et al., 2016).

These studies collectively demonstrate significant progress toward developing treatments for genetic diseases. Nearly all genetic diseases, including DMD, currently have no cure, and in many cases, those affected suffer from low quality of life and a shortened life expectancy. These significant advances suggest potential treatments and even potential cures of human genetic diseases.

**High-Throughput Genetic Screens**

The ease with which CRISPR-based tools can be reprogrammed simply by constructing a new guide RNA has facilitated genomewide screening. CRISPR-based knockout screens are reported to be more efficient and specific than RNAi, enabling researchers to obtain more reliable and robust results (Sanjana, 2016). In these experiments, sgRNA libraries and Cas9 are delivered into cells. Researchers then screen the treated cells for a phenotype of interest (Sanjana et al., 2014; Shalem et al., 2015). CRISPR-mediated knockout screens have been used to identify genes involved in cancer progression (Shalem et al., 2014; Chen et al., 2015a; Shi et al., 2015), drug resistance (Shalem et al., 2014; Wang et al., 2014a), the immune response (Parnas et al., 2015), vulnerability to bacterial toxins (Zhou et al., 2014), and other biomedically relevant phenotypes.

Alternatively, Cas9-mediated saturation mutagenesis, in which an sgRNA library is targeted to thoroughly mutate a targeted region of the genome, can be used in a high-throughput manner to obtain high-resolution information on genetic loci. For example, a sgRNA library targeting the BCL11A enhancer was used to dissect the importance of each portion of the enhancer on HbF levels. Similarly, a single sgRNA was combined with a donor DNA library to replace a 6-bp portion of the BRCA1 gene with all possible hexamers and to replace exon 18 of BRCA1 with all possible SNP variations to dissect the mutations’ effects on transcript processing (Kiani et al., 2015).

Together, these studies demonstrate how Cas9 has enhanced the ability to screen the modification of many loci within the human genome in a single experiment. These types of studies provide researchers with powerful tools to dissect complex cellular signaling pathways, determine gene function, identify targets for therapeutic intervention, and predict drug side effects.

**Gene Drives**

Gene drives are a particularly powerful application of CRISPR technology. Gene drives, genetic elements that insert themselves into target sites lacking that element, convert heterozygous alleles to homozygous alleles within an organism. In organisms that support sexual reproduction, gene drives enable non-Mendelian inheritance of alleles that can spread throughout a population. This spread can be rapid for species with short reproductive generation times. While gene drives can be implemented in other ways, Cas9-based gene drives are very efficient. CRISPR gene drives were first demonstrated by inserting into the Drosophila genome a construct consisting of Cas9 and an sgRNA flanked by two “homology arms” that match the genomic sequence surrounding the sgRNA-programmed target site. Cas9:sgRNA complex expression results in cleavage of the WT allele and incorporation of the Cas9:sgRNA cassette via HDR to produce a homozygous mutant organism (Gantz and Bier, 2015). In mosquitos, researchers developed a Cas9-based gene drive that inserts a gene inducing a parasite-resistance phenotype (Gantz et al., 2015; Hammond et al., 2016). Gene drives have the potential to allow the facile generation of animal models with recessive mutations, as well as the genome editing of whole populations of rapidly reproducing organisms—a remarkably powerful capability that should be considered only with great thoughtfulness and caution (Oye et al., 2014; Esvelt et al., 2014).

**Conclusion**

The discovery and characterization of CRISPR systems have transformed genome editing and the life sciences. The development of new CRISPR technologies that extend the generality, DNA specificity, product selectivity, and even the fundamental capabilities of natural CRISPR systems has driven and accelerated this transformation. These technologies have armed researchers with powerful new tools to study living systems and human disease. Incredibly, they have also made it possible to imagine a near-term future in which the treatment of genetic diseases is within our reach. As these technologies grow in scope and capability, ethical and regulatory guidelines must also be
thoughtfully developed (Baltimore et al., 2015) to ensure a balance between realizing the enormous potential of these tools to benefit mankind and minimizing the risk of their misuse.

ACKNOWLEDGMENTS

We are grateful to Jennifer Doudna and her coworkers Mitchell O’Connell and Alexandra Seletsky, Keith Joung, Gerald Joyce, and Feng Zhang for their helpful comments. This work was supported by U.S. NIH R01 EB022376 (formerly R01 GM065400), NIH R53GM118062, F-Prime Biomedical Research Initiative (A28161), and the Howard Hughes Medical Institute. A.C.K. is a Ruth L. Kirkstein National Research Service Awards Postdoctoral Fellow (F32 GM 112366-2). A.H.B. was supported by the Harvard Chemical Biology Program and a National Science Foundation Graduate Research Fellowship. A.C.K. and D.R.L. have filed patent applications on base editing. D.R.L. is a consultant and co-founder of Editas Medicine, a company that seeks to develop genome-editing therapeutics.

REFERENCES


