CRISPR-Cas systems for editing, regulating and targeting genomes

Jeffry D Sander1,2 & J Keith Joung1,2

Targeted genome editing using engineered nucleases has rapidly gone from being a niche technology to a mainstream method used by many biological researchers. This widespread adoption has been largely fueled by the emergence of the clustered, regularly interspaced, short palindromic repeat (CRISPR) technology, an important new approach for generating RNA-guided nucleases, such as Cas9, with customizable specificities. Genome editing mediated by these nucleases has been used to rapidly, easily and efficiently modify endogenous genes in a wide variety of biomedically important cell types and in organisms that have traditionally been challenging to manipulate genetically. Furthermore, a modified version of the CRISPR-Cas9 system has been developed to recruit heterologous domains that can regulate endogenous gene expression or label specific genomic loci in living cells. Although the genome-wide specificities of CRISPR-Cas9 systems remain to be fully defined, the power of these systems to perform targeted, highly efficient alterations of genome sequence and gene expression will undoubtedly transform biological research and spur the development of novel molecular therapeutics for human disease.

The introduction of targeted genomic sequence changes into living cells and organisms has become a powerful tool for biological research and is a potential avenue for therapy of genetic diseases. Framedhift knockout mutations enable reverse genetics and identification of gene functions; sequence insertions can fuse epitope tags or other functional domains, such as fluorescent proteins, to endogenous gene products; and specific sequence alterations can induce amino acid substitutions for disease modeling, transfer traits in agricultural crops and livestock, and correct defective genes for therapeutic applications. For many years, strategies for efficiently inducing precise, targeted genome alterations were limited to certain organisms (e.g., homologous recombination in yeast or recombiningneering in mice) and often required drug-selectable markers or left behind 'scar' sequences associated with the modification method (e.g., residual loxP sites from Cre recombinase-mediated excision). Targeted genome editing using customized nucleases provides a general method for inducing targeted deletions, insertions and precise sequence changes in a broad range of organisms and cell types. The high efficiency of genome editing obviates the need for additional sequences, such as drug-resistance marker genes, and therefore the need for additional manipulations to remove them.

A crucial first step for performing targeted genome editing is the creation of a DNA double-stranded break (DSB) at the genomic locus to be modified1. Nuclease-induced DSBs can be repaired by one of at least two different pathways that operate in nearly all cell types and organisms: nonhomologous end-joining (NHEJ) and homology-directed repair (HDR) (Fig. 1). NHEJ can lead to the efficient introduction of insertion/deletion mutations (indels) of various lengths, which can disrupt the translational reading frame of a coding sequence or the binding sites of trans-acting factors in promoters or enhancers. HDR-mediated repair can be used to introduce specific point mutations or to insert desired sequences through recombination of the target locus with exogenously supplied DNA 'donor templates'. With targeted nuclease-induced DSBs, the frequencies of these alterations are typically greater than 1% and, in some cases, over 50%; at these rates, desired mutations can be identified by simple screening, without drug-resistance marker selection.

Early methods for targeting DSB-inducing nucleases to specific genomic sites relied on protein-based systems with customizable DNA-binding specificities, such as meganucleases, zinc finger nucleases (ZFNs) and transcription activator–like effector nucleases (TALENs). These platforms have made possible important advances, but each has its own set of associated advantages and disadvantages (Box 1). More recently, a platform based on a bacterial CRISPR-associated protein 9 nuclease from Streptococcus pyogenes (hereafter referred to as Cas9) has been developed; it is unique and flexible owing to its dependence on RNA as the moiety that targets the nuclease to a desired DNA sequence. In contrast to ZFN and TALEN methods, which use protein-DNA interactions for targeting, RNA-guided nucleases (RGNs) use simple, base-pairing rules between an engineered RNA and the target DNA site.

In this Review, we describe how this RNA-guided system works and how it has been applied to perform genome editing across a wide variety of cell types and whole organisms. We also discuss the advantages and limitations of this system, and assess off-target effects and recent strategies for improving specificity and how the system can be repurposed for other applications, such as regulation of gene expression and selective labeling of the genome (Fig. 2 summarizes different applications). Finally, we consider the challenges that will need to be addressed for this emerging genome editing platform.

1Molecular Pathology Unit, Center for Computational and Integrative Biology, Center for Cancer Research, Massachusetts General Hospital, Charlestown, Massachusetts, USA. 2Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA, Massachusetts, USA. Correspondence should be addressed to J.K.J. (jjoung@mgh.harvard.edu).

Received 2 September 2013; accepted 31 January 2014; published online 2 March 2014; doi:10.1038/nbt.2842
From a bacterial CRISPR immune system to engineered RGNs

CRISPR systems are adaptable immune mechanisms used by many bacteria to protect themselves from foreign nucleic acids, such as viruses or plasmids. Type II CRISPR systems incorporate sequences from invading DNA between CRISPR repeat sequences encoded as arrays within the bacterial host genome. Transcripts from the CRISPR repeat arrays are processed into CRISPR RNAs (crRNAs), each harboring a variable sequence transcribed from the CRISPR repeat. Each crRNA hybridizes with a second RNA, known as the transactivating CRISPR RNA (tracrRNA), and these two RNAs complex with the Cas9 nuclease. The protospacer-encoded portion of the crRNA directs Cas9 to cleave complementary target-DNA sequences, if they are adjacent to short sequences known as protospacer adjacent motifs (PAMs). Protospacer sequences incorporated into the CRISPR locus are not cleaved presumably because they are not next to a PAM sequence.

The type II CRISPR system from S. pyogenes has been adapted for inducing sequence-specific DSBs and targeted genome editing. In the simplest and most widely used form of this system, two components must be introduced into and/or expressed in cells or an organism to perform genome editing: the Cas9 nuclease and a guide RNA (gRNA), consisting of a fusion of a crRNA and a fixed tracrRNA. Twenty nucleotides at the 5’ end of the gRNA (corresponding to the protospacer portion of the crRNA) direct Cas9 to a specific target DNA site using standard RNA:DNA complementarity base-pairing rules. These target sites must lie immediately 5’ of a PAM sequence that matches the canonical form 5’-NGG (although recognition at sites with alternate PAM sequences (e.g., 5’-NAG) has also been reported, albeit at less efficient rates). Thus, with this system, Cas9 nuclease activity can be directed to any DNA sequence of the form N20-NGG simply by altering the first 20 nt of the gRNA to correspond to the target DNA sequence. Type II CRISPR systems from other species of bacteria that recognize alternative PAM sequences and that utilize different crRNA and tracrRNA sequences have also been used for targeted genome editing. However, because the most commonly used and extensively characterized system is based on the S. pyogenes system, the remainder of this Review focuses on this particular platform and its components, unless otherwise noted.

Following the initial demonstrations in 2012 that Cas9 could be programmed to cut various DNA sites in vitro, a flurry of papers published in 2013 showed that this platform also functions efficiently in a variety of cells and organisms. Initial proof-of-principle studies showed that Cas9 could be targeted to endogenous genes in bacteria, cultured transformed human cancer cell lines and human pluripotent stem cells in culture, as well as in a whole organism, the zebrafish (J.K.J. and colleagues). Subsequently, Cas9 has been used to alter...
genes in yeast, tobacco, thale cress, rice, wheat, sorghum, mice, rats, rabbits, frogs, fruit flies, silkworms, and roundworms (see Table 1 for a list of these published reports). Cas9-induced DSBs have been used to introduce NHEJ-mediated indel mutations as well as to stimulate HDR with both double-stranded plasmid DNA and single-stranded oligonucleotide donor templates. Being able to introduce DSBs at multiple sites in parallel with Cas9 is a unique advantage of this RNA-guided genome editing platform relative to meganucleases, ZFNs or TALENs. For example, expression of Cas9 and multiple gRNAs has been used to induce small and large deletions or inversions between the DSBs, to simultaneously introduce mutations in three genes in rat cells, five genes in mouse ES cells, and five genes in the somatic cells of a single zebrafish.

The simplicity of Cas9 targeting has also inspired the generation of large gRNA libraries using array-based oligonucleotide synthesis. These libraries can be engineered to encompass multiple gRNAs for almost every gene in a host organism, thereby facilitating forward genetic screens and selection. In contrast to short hairpin RNA libraries, which mediate only gene knockdown, these gRNA libraries have been used with Cas9 nuclease to generate libraries of cells with knockout mutations. Libraries consisting of between ~64,000 and ~87,000 distinct gRNAs have demonstrated this strategy for positive and negative forward genetic phenotype screens in human and mouse cells.

The CRISPR-Cas systems can be engineered by fusing Cas9 to effector domains, such as transcription activator libraries (TALs), zinc finger nucleases (ZFNs), and TALENs. These systems can be used to target specific loci in the genome and induce targeted DNA cleavage. The hybridization of the gRNA to the target DNA results in the cleavage of the DNA using the Cas9 nuclease. This can lead to the introduction of mutations, such as indels or insertions, at the targeted loci. This technology has been used to generate large libraries of CRISPR knockout mouse models, which can be used to study the function of individual genes in a mouse model system.

Figure 2 Overview of various Cas9-based applications. (a,b) gRNA-directed Cas9 nuclease can induce indel mutations (a) or specific sequence replacement or insertion (b). (c) Pairs of gRNA-directed Cas9 nuclease can stimulate large deletions or genomic rearrangements (e.g., inversions or translocations). (d-f) gRNA-directed dCas9 can be fused to activation domains (d) to mediate upregulation of specific endogenous genes, heterologous effector domains (e) to alter histone modifications or DNA methylation, or fluorescent proteins (f) to enable imaging of specific genomic loci. TSS, transcription start site.

Figure 3 Naturally occurring and engineered CRISPR-Cas systems. (a) Naturally occurring CRISPR systems incorporate foreign DNA sequences into CRISPR arrays, which then produce crRNAs bearing “protospacer” regions that are complementary to the foreign DNA site. crRNAs hybridize to tracrRNAs (also encoded by the CRISPR system) and this pair of RNAs can associate with the Cas9 nuclease. crRNA-tracrRNA:Cas9 complexes recognize and cleave foreign DNAs bearing the protospacer sequences. (b) The most widely used engineered CRISPR-Cas system utilizes a fusion between a crRNA and part of the tracrRNA sequence. This single gRNA complexes with Cas9 to mediate cleavage of target DNA sites that are complementary to the 5’ 20 nt of the gRNA and that lie next to a PAM sequence. (c) Example sequences of a crRNA-tracrRNA hybrid and a gRNA.
Cas9 variants that cut one strand rather than both strands of the target DNA site (known as ‘nickases’) have also been shown to be useful for genome editing. Introduction of a D10A or H840A mutation into the RuvC1- or HNH-like nuclease domains in Cas9 (Fig. 4a)\textsuperscript{41,42} results in the generation of nickases that cut either the complementary or noncomplementary DNA target strands, respectively, \textit{in vitro}\textsuperscript{7–9,29} (Fig. 4b.c). Consistent with previous studies with ZFN-derived nickases\textsuperscript{34–46} (J.K.J. and colleagues\textsuperscript{44}), previous studies with Cas9 nickases can, at some sites, induce HDR with reduced levels of concomitant NHEJ-mediated indels\textsuperscript{13,14}. However, although at some sites Cas9 nickases can induce HDR with efficiencies similar to those of the original Cas9 nuclease\textsuperscript{13,14}, these rates can be much lower at other sites\textsuperscript{47}. Notably, the frequencies of indel mutations introduced by nickases have also been high at certain sites\textsuperscript{13,47–49} (J.K.J. and colleagues\textsuperscript{48}). Notably, the frequencies of indel mutations introduced by nickases can induce HDR with efficiencies similar to those of the original Cas9 nuclease\textsuperscript{13,14,47,49}.

### Table 1 Published examples of cell types and organisms modified by Cas9

<table>
<thead>
<tr>
<th>Cell type or organism</th>
<th>Cas9 form</th>
<th>Cell type</th>
<th>Reference numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human cells</td>
<td>Cas9 nuclease</td>
<td>HEK293FT, HEK293T, HEK293, K562, iPSC, HUES9, HUES1, BJ-RIPS, HeLa, Jurkat, U2OS</td>
<td>9,13–16,47, 49–51,54,59, 84,85</td>
</tr>
<tr>
<td></td>
<td>Cas9 nickase</td>
<td>HEK293FT, HEK293T</td>
<td>13,14,47,49</td>
</tr>
<tr>
<td></td>
<td>dCas9 (gene regulation)</td>
<td>HEK293FT, HEK293T</td>
<td>70–72,74,82</td>
</tr>
<tr>
<td></td>
<td>dCas9 (imaging)</td>
<td>HEK293T, UMUC3, HeLa</td>
<td>81</td>
</tr>
<tr>
<td>Mouse or mouse cells</td>
<td>Cas9 nuclease</td>
<td>Embryos</td>
<td>14,24–26</td>
</tr>
<tr>
<td></td>
<td>dCas9 (gene regulation)</td>
<td>Embryos</td>
<td>47</td>
</tr>
<tr>
<td>Rat</td>
<td>Cas9 nuclease</td>
<td>Embryos</td>
<td>26,36</td>
</tr>
<tr>
<td>Rabbit</td>
<td>Cas9 nuclease</td>
<td>Embryos</td>
<td>27</td>
</tr>
<tr>
<td>Frog</td>
<td>Cas9 nuclease</td>
<td>Embryos</td>
<td>28</td>
</tr>
<tr>
<td>Zebrafish</td>
<td>Cas9 nuclease</td>
<td>Embryos</td>
<td>17,33,37,60,85</td>
</tr>
<tr>
<td>Fruit fly</td>
<td>Cas9 nuclease</td>
<td>Embryos</td>
<td>29,30,61</td>
</tr>
<tr>
<td>Silkworm</td>
<td>Cas9 nuclease</td>
<td>Embryos</td>
<td>31</td>
</tr>
<tr>
<td>Roundworm</td>
<td>Cas9 nuclease</td>
<td>Adult gonads</td>
<td>32,62–67</td>
</tr>
<tr>
<td>Rice</td>
<td>Cas9 nuclease</td>
<td>Protoplasts, callus cells</td>
<td>21</td>
</tr>
<tr>
<td>Wheat</td>
<td>Cas9 nuclease</td>
<td>Protoplasts</td>
<td>21</td>
</tr>
<tr>
<td>Sorghum</td>
<td>Cas9 nuclease</td>
<td>Embryos</td>
<td>23</td>
</tr>
<tr>
<td>Tobacco</td>
<td>Cas9 nuclease</td>
<td>Protoplasts, leaf tissue</td>
<td>19,20,23</td>
</tr>
<tr>
<td>Thale cress</td>
<td>Cas9 nuclease</td>
<td>Protoplasts, seedlings</td>
<td>19,23</td>
</tr>
<tr>
<td>Yeast</td>
<td>Cas9 nuclease</td>
<td>Saccharomyces cerevisiae</td>
<td>18</td>
</tr>
<tr>
<td>Bacteria</td>
<td>Cas9 nuclease</td>
<td>Streptococcus pneumoniae, E. coli</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>dCas9 (gene regulation)</td>
<td>E. coli</td>
<td>69,70</td>
</tr>
</tbody>
</table>

HEK, human embryonic kidney; iPSCs, induced pluripotent stem cells; UMUC3, human bladder cancer.

These studies showed that mismatches are generally better tolerated at the 5’ end of the 20-nt targeting region of the gRNA than at the 3’ end; this result is consistent with previous experiments performed \textit{in vitro} and in bacterial cells, which suggested that the 8–12 bp at the 3’ end of the targeting sequence (also known as the ‘seed’ sequence) are crucial for target recognition\textsuperscript{7,8,14,52,53}. However, the effects of single and double mismatches are not always predictable based on their location within the gRNA targeting region; some mismatches in the 5’ end can have dramatic effects, whereas some in the 3’ end do not greatly affect Cas9 activity\textsuperscript{50}. In addition, not all nucleotide substitutions at a given position necessarily have equivalent effects on activity\textsuperscript{51}.

A reciprocal, and perhaps more relevant, approach for studying specificity is to assess the activities of Cas9 at potential off-target genomic DNA target sites, (i.e., sites that have a few nucleotide differences compared to the intended target). A number of studies have examined potential off-target sites that differ at one to six positions from the on-target site in human cells\textsuperscript{9,47,48,50,51,54}. Collectively, these reports have found cases of off-target mutations at sites that differ by as many as five positions within the protospacer region\textsuperscript{50} and/or that have an alternative PAM sequence of the form NAG\textsuperscript{51}. Interestingly, indel mutation frequencies at these off-target sites can be high enough (>2–5%) to detect using the relatively insensitive T7 endonuclease I (T7E1) mutation mismatch assay and sometimes are comparable to the on-target site mutation frequency\textsuperscript{48,50,54}. In addition, more sensitive deep sequencing assays have identified lower frequency off-target mutations\textsuperscript{48,51,55}. It is important to note that not all of these directed studies examined only a subset of the much larger number of potential off-target sites in the genome. For example, any given 20-nt protospacer will typically have hundreds to thousands of potential off-target sites that differ at four or five positions, respectively, in 6 x 10\textsuperscript{8} bases of random DNA. In addition, although it has been suggested that higher GC content at the RNA:DNA hybridization interface might potentially help to stabilize binding of the RGN to DNA, high rates of mutagenesis have been observed for off-target sites with as little as 30% matched GC content\textsuperscript{9,50}.

A somewhat more comprehensive strategy for examining Cas9 specificities is to identify off-target sites from a partially degenerate library of variants that is based on the intended on-target sequence. One recent report identified sites from such libraries based on their abilities to be bound by a catalytically inactive form of Cas9 fused to a transcriptional activation domain (see Fig. 4 and further discussion below)\textsuperscript{49}. This study found sites that were mismatched by as many as three (and possibly more) positions relative to the on-target site\textsuperscript{49}. These results are similar to those of another study, which used \textit{in vitro} selection for Cas9 nuclease cleavage activity to identify potential off-target sites from a partially degenerate library of target site variants. Some of the off-target sites identified by these \textit{in vitro} selections (with up to four mismatches) were also shown to be mutated in human cells\textsuperscript{49}.

A recent study using whole-exome sequencing did not find evidence of Cas9-induced, off-target mutations in three modified human K562 cell line clones\textsuperscript{56}. Although the authors acknowledge that the high false-negative rate associated with exome sequencing analysis limits interpretation of these data, these results do suggest that with careful target selection, it may be possible to isolate Cas9-edited cells with otherwise intact exomes. Additional examples with deeper sequencing coverage and whole genome (rather than whole exome) sequencing will be needed to determine how readily cells that do not have off-target mutations can be isolated. The ability to do so would encourage broader research application of Cas9 technology. However, it is worth noting that deep sequencing the genomes of individual cells...
clones is expected to be neither sensitive nor effective for defining the full genome-wide spectrum of Cas9 off-target sites because each clone would likely only carry mutations at a small proportion of, if any, possible off-target sites.

Overall, the various published studies strongly suggest that off-target sites of RNA-guided Cas9 nucleases can be variable in frequency and challenging to predict. For any given target site, it is not currently possible to predict how many mismatches can be tolerated, nor do we fully understand why some sites are cleaved whereas other are not. We also do not know how genomic and/or epigenomic context might affect the frequency of cleavage. Although some initial evidence suggests that DNA methylation does not inhibit Cas9-based genome editing51, it seems plausible and likely that chromatin structure could play a role in off-target site accessibility. A more comprehensive understanding of Cas9 off-target effects will have to await the development of unbiased, global measures of Cas9 specificity in cells.

Methods for reducing off-target effects of Cas9 nucleases

Even with an incomplete understanding of RNA-guided Cas9 nuclease specificity, researchers have begun to explore various approaches to reduce off-target mutagenic effects. One potential strategy is to test the effects of reducing the concentrations of gRNA and Cas9 expressed in human cells. Results with this approach have been mixed; one group observed proportionately larger decreases in rates of off-target relative to on-target mutagenesis for two gRNAs51, whereas our group observed nearly proportionate decreases at both off-target and on-target mutagenesis for two gRNAs50. The use of modified gRNA architectures with truncated 3′ ends (within the tracrRNA-derived sequence) or with two extra guanine nucleotides appended to the 5′ end (just before the complementarity region) also yielded better on-target to off-target ratios but generally with considerably lower absolute efficiencies of on-target genome editing50.56.

Another proposed approach for improving specificity involves the use of ‘paired nickases’ in which adjacent off-set nick sites are generated at the target site using two gRNAs and Cas9 nickases47,49,56 (Fig. 4d), a strategy analogous to one originally performed with pairs of engineered zinc finger nickases46. In contrast to single Cas9 nickases (which can at some sites more favorably induce HDR events relative to NHEJ indels), paired Cas9 nickases targeted to sites on opposite DNA strands separated by 4 to 100 bp can efficiently introduce both indel mutations and HDR events with a single-stranded DNA oligonucleotide donor template in mammalian cells47,49,56. It has been proposed by some that the concerted action of paired nickases create a DSB that is then repaired by NHEJ or HDR47,55. Importantly, paired nickases can reduce Cas9-induced off-target effects of gRNAs in human cells; the addition of a second gRNA and substitution of Cas9 nickase for Cas9 nuclease can lead to lower levels of unwanted mutations at previously known off-target sites of the original gRNA47. However, an as-yet unanswered question is whether the second gRNA can itself induce its own range of Cas9 nickase–mediated off-target mutations in the genome. Multiple studies have shown that single monomeric Cas9 nickases can function on their own to induce indel mutations at certain genomic loci13,47–49, perhaps because an individual nick might be converted to a DSB when a replication fork passes through the locus57,58. Thus, an important improvement needed for the paired nickase system will be to make the activities of the two nickase monomers strictly co-dependent on each other for genome editing activity—that is, so that these nickase monomers are only active for genome editing when bound to DNA in close proximity to the other.

Our group has recently shown that off-target effects can be substantially reduced simply by using gRNAs that have been shortened...
at the 5′ end of their complementarity regions. These truncated gRNAs (which we refer to as ‘tru-gRNAs’) have 17 or 18 nucleotides of complementarity; they generally function as efficiently as full-length gRNAs in directing on-target Cas9 activity but show decreased mutagenic effects at off-target sites and enhanced sensitivity to single or double mismatches at the gRNA:DNA interface. This strategy avoids the technical challenges associated with expressing multiple gRNAs in a single cell for the paired nickase approach and should be straightforward to implement. tru-gRNAs could also be used in conjunction with other strategies for improving Cas9 specificity (e.g., we showed that tru-gRNAs further improve the specificity of paired nickases as well as the specificities of dCas9 fusion proteins for non-nuclease applications (described below)).

Practical considerations for implementing CRISPR-Cas technology

Owing to rapid progress in the field, potential users face a variety of choices about how to implement CRISPR-Cas technology. Here we discuss some of the parameters to consider when applying the methodology.

Choice of gRNA platform. It is important to note that the efficiency of Cas9 activity for any given locus can be influenced by the architecture of gRNA(s) used. As described above, most recent studies have used a single gRNA that is a fusion of a programmable crRNA and part of the tracrRNA, but earlier studies also used a ‘dual gRNA’ configuration in which the crRNA and tracrRNA are expressed separately. In general, studies using single gRNAs have consistently reported substantially higher editing rates than those using dual gRNAs.

These findings suggest that the single gRNA system may be more active than the double gRNA system, presumably because two components can assemble more efficiently than three components.

In addition, single gRNAs harboring variable lengths of tracrRNA sequence at their 3′ ends have been used by different groups (Supplementary Table 1). Systematic comparisons have generally demonstrated that longer single gRNAs (containing more of the 3′ portion of the tracrRNA sequence) yield higher editing rates than shorter ones. The most commonly used single gRNA design to date is ~100 nt in length (Supplementary Table 1). The tru-gRNAs described above are shortened versions of this ~100-nt single gRNA.

Targeting range and choice of gRNA target sites. The choice of promoter used to express gRNAs can limit the options for potential target DNA sites. For example, the RNA polymerase III–dependent U6 promoter or the T7 promoter require a G or GG, respectively, at the 5′ end of the sequence of the RNA that is to be transcribed (top panels of Fig. 5a,b). As a result, standard full-length or tru-gRNAs expressed from these promoters are limited to targeting sites that match the forms GNN16-18 or GN18-18; such sites are expected to occur every 1 in 32 bp or 1 in 128 bp, respectively, in random DNA sequence. Paired nickase strategies require the identification of two sites on opposite strands of DNA with appropriate spacing in between (as described above). One option to reduce these targeting range restrictions is to choose sites without regard to the identities of the first or first two bases at the 5′ end (that is, making gRNAs that are mismatched at these positions). Another potential strategy to bypass these restrictions is to append the required G or GG to the 5′ end of the gRNA, thereby encoding gRNA transcripts that are 1 or 2 bp longer (bottom panels of Fig. 5a,b). Both of these strategies have been used successfully to produce active gRNAs but with variable efficiencies in genome-editing activities with Cas9 nuclease.

Larger-scale studies are needed to clarify the effects of using either mismatched or extended gRNAs on the efficiencies and specificities of RGN-mediated cleavage. Several groups have provided web-based software that facilitates the identification of potential CRISPR RGN target sites in user-defined sequences (e.g., the ZifIT Targeter software and the CRISPR Design Tool).

Delivery of CRISPR-Cas components. RGNs have been delivered to a broad range of cell types and organisms using a variety of delivery methods. In cultured mammalian cells, researchers have used electroporation, nucleofection and Lipofectamine-mediated transfection of nonreplicating plasmid DNA to transiently express Cas9 and gRNAs. Lentiviral vectors have also been used to constitutively express Cas9 and/or gRNAs in cultured human cells and mouse cells. RNAs and/or plasmid DNA transcribed in vitro have been injected directly into the embryos of zebrafish, fruit flies, mice, and rats. Plasmid DNA and RNA have also been injected into the gonads of adult roundworms and mice, and in one study purified Cas9 protein complexed with gRNA was injected into roundworm gonads. In addition to animal models and cell lines, Cas9 has been used successfully in multiple plant species including wheat, rice, sorghum, tobacco and thale cress using a range of standard delivery methods including PEG-mediated transformation of protoplasts, Agrobacterium-mediated transfer in embryos and leaf tissue, and/or bombardment of callus cells with plasmid DNA. For most RGN applications, transient expression of gRNAs and Cas9 is typically sufficient to induce...
efficient genome editing. Although constitutive expression of RGN components might potentially lead to higher on-target editing efficiencies, extended persistence of these components in the cell might also lead to increased frequencies of off-target mutations, a phenomenon that has been previously reported with ZFNs68.

**Experimental strategies to control for RGN off-target effects.** The existence of CRISPR RGN-induced off-target effects and our current inability to comprehensively identify these alterations on a genome-wide scale mean that investigators need to account for the potentially confounding effects of these undesired mutations. Several strategies can be used to rule out off-target mutations as a potential alternative explanation for any phenotypes observed. For example, complementation with reintroduction of a wild-type gene can be used to confirm the effects of knockout mutations. In addition, similar to the strategy of targeting a gene with multiple RNA interference hairpins, one could easily create mutations in the same gene using gRNAs targeted to different sites. Presumably, each gRNA will be expected to have a different range of off-target effects and therefore if the same phenotype is observed with each of these different gRNAs it would seem unlikely that undesired mutations are the cause. The ease with which multiple gRNAs can be rapidly designed and constructed makes it simple and feasible to implement this type of strategy with the Cas9 system. The high efficacy of the Cas9 nuclease for inducing mutations makes it an attractive choice for creating mutant cell lines and whole organisms in spite of the need to account for off-target effects.

**Applications of CRISPR-Cas beyond genome editing.** Beyond enabling facile and efficient targeted genome editing, the CRISPR-Cas system has the potential to be used to regulate endogenous gene expression or to label specific chromosomal loci in living cells or organisms. Catalytically inactive or “dead” Cas9 (dCas9)—a variant bearing both the D10A and H840A mutations that does not cleave DNA—can be recruited by gRNAs to specific target DNA sequences. The high specificity of the Cas9 nuclease for inducing mutations makes it an attractive choice for creating mutant cell lines and whole organisms in spite of the need to account for off-target effects.

**Future directions**

Progress in the development of Cas9-based technologies over the past 18 months has been stunning, but many interesting questions and applications remain to be addressed and explored. First, methods for expanding the targeting range of RNA-guided Cas9 will be important for inducing precise HDR or NHEJ events as well as for implementing multiplex strategies, including paired nickases. As noted above, the targeting range for Cas9, paired Cas9 nickases and dCas9 fusions is restricted mainly by the need for a PAM sequence matching the form NGG. Alternative PAM sequences of the form NAG or NNNG can be exploited, as has been noted27,35,51, but more experiments are needed to ascertain how robustly these sequences are recognized and cleaved. Other gRNA-Cas9 platforms with different PAM sequences isolated from Streptococcus thermophilus, Neisseria meningitidis and Treponema denticola have also been characterized10,11,14 and identification of more of these systems from other species83 could further enhance the targeting range of the platform.

Second, the field urgently needs to develop unbiased strategies to globally assess the off-target effects of Cas9 nucleases or paired nickases in any genome of interest. Such methods will be crucial for evaluating how effectively improvements described to date enhance the specificity of the platform. In addition, although tru-gRNAs and paired nickases can reduce off-target effects, it is likely that further improvements will be needed, especially for therapeutic applications. Ideally, new strategies could be combined with existing approaches.
Examples of such improvements might involve using protein engineering to modify Cas9 and/or modifying the nucleotides used by the gRNA to mediate recognition of the target DNA site. Alternatively, the construction of inducible forms of Cas9 and/or gRNAs might provide a means to regulate the active concentration of these reagents in the cell and thereby improve the ratio of on- and off-target effects.

Third, methods for efficient delivery and expression of CRISPR-Cas system components will undoubtedly need to be optimized for each particular cell-type or organism to be modified. For example, some cell types or tissues might be refractory to transfection and/or infection by standard viral vectors. A related challenge will be to develop methods that enable expression of either the gRNAs or the Cas9 nuclelease that is specific to a tissue, cell type or developmental stage. Strategies that ensure efficient expression of large numbers of different gRNAs simultaneously from one vector would also allow more extensive use of the multiplex capability of CRISPR-Cas systems. Collectively, these advances will be important for research use and therapeutic applications.

Lastly, strategies for shifting the balance away from NHEJ-mediated indel mutations and toward HDR-driven alterations remain a priority for development. Although high rates of HDR can be achieved with the CRISPR RGNs and single-stranded DNA oligonucleotides, competing mutagenic NHEJ also occurs simultaneously. This limitation is particularly problematic when using HDR to induce point mutation changes (as opposed to insertions) in the protospacer part of the target site; alleles that have been successfully altered in this way can still be efficiently re-cut and then mutated by NHEJ, thereby reducing the yield of correctly edited sequences. One of the drawbacks to developing an approach to improve the HDR:NHEJ ratio is that inhibition of NHEJ is likely to be poorly tolerated by most cells, given its central role in normal DNA repair. For therapeutic applications seeking to exploit HDR, reduction or elimination of competing NHEJ will be crucially important.

The simplicity, high efficiency and broad applicability of the RNA-guided Cas9 system have positioned this technology to transform biological and biomedical research. The ease with which researchers can now make changes in the sequence or expression of any gene means reverse genetics can be performed in virtually any organism or cell type of interest. In addition, the construction of large libraries of gRNAs for altering or regulating genes of interest will enable facile, comprehensive forward genetic screens. All of these systems can also be multiplexed by expressing multiple gRNAs in a single cell, thereby further extending the complexity of forward and reverse genetic experiments that can be done. Although the off-target effects of Cas9 remain to be defined on a genome-wide scale, much progress has already been made toward improving specificity, and further advances will undoubtedly come rapidly, given the intensity of research efforts in this area. All of these recent advances—and those to come—in developing and optimizing Cas9-based systems for genome and epigenome editing should propel the technology toward therapeutic applications, opening the door to treating a wide variety of human diseases.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

J.K.J. is grateful for support from the US National Institutes of Health (NIH) (grants D1P GM105378 and R01 GM088040), the Defense Advanced Research Projects Agency (grant W911NF-11-2-0056) and The Jim and Ann Orr Massachusetts General Hospital Research Scholar Award. This material is based upon work supported fully or in part by the US Army Research Laboratory and the US Army Research Office under grant number W911NF-11-2-0056. The authors apologize to colleagues whose studies were not cited due to length and reference constraints.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests; details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.