CRISPR/Cas9 for Human Genome Engineering and Disease Research

Xin Xiong,⁠¹ Meng Chen,²,³,⁴,⁵ Wendell A. Lim,¹ Dehua Zhao,² and Lei S. Qi²,³,⁴

¹Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94158; email: xiongxin1009@gmail.com, wendell.lim@ucsf.edu
²Department of Bioengineering, Stanford University, Stanford, California 94305; email: dehua@stanford.edu, stanley.qi@stanford.edu
³Department of Chemical and Systems Biology, Stanford University, Stanford, California 94305
⁴ChEM-H, Stanford University, Stanford, California 94305
⁵Gladstone Institute of Cardiovascular Disease, San Francisco, California 94158; email: meng.chen@gladstone.ucsf.edu

Keywords
gene editing, gene regulation, genetic screening, human diseases, gene therapy

Abstract
The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated 9 (Cas9) system, a versatile RNA-guided DNA targeting platform, has been revolutionizing our ability to modify, manipulate, and visualize the human genome, which greatly advances both biological research and therapeutics development. Here, we review the current development of CRISPR/Cas9 technologies for gene editing, transcription regulation, genome imaging, and epigenetic modification. We discuss the broad application of this system to the study of functional genomics, especially genome-wide genetic screening, and to therapeutics development, including establishing disease models, correcting defective genetic mutations, and treating diseases.
1. INTRODUCTION: TOOLS FOR GENOME EDITING AND REGULATION

Over a meter of linear DNA encodes more than 20,000 protein-coding and noncoding genes in the nucleus of each human cell. A major goal of human genomic research has been to decode the functions of individual genes and identify the roles of key regulatory elements. Although accumulating data from comprehensive genetic studies began to reveal correlations between genetic variants and diseases decades ago, understanding the driving forces that cause certain disease phenotypes and correcting the mutations in order to cure them require modifying the genome. However, precisely modifying genetic information in the vast genome remained a major challenge. The development of potent genetic tools that allow one to robustly and flexibly edit or modulate a genome is key to gaining a more comprehensive understanding of genetic function and to creating more effective therapeutics.

In terms of genome engineering, the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated 9 (Cas9) system is a broadly used tool but not the first member in its class. Programmable protein-based genome engineering systems, including zinc-finger nucleases (ZFNs) (5, 72, 92) and transcription activator–like effector nucleases (TALENs) (15), have been developed and widely applied. These molecules allow precise targeting and cutting at a specific genomic locus to generate double-strand breaks (DSBs) and therefore allow precise genome editing. Studies with these two classes of nucleases have led to important scientific discoveries and therapeutics development. In fact, a ZFN-based treatment of HIV that disables the HIV coreceptor C-C chemokine receptor type 5 (CCR5) in human primary T cells is currently in clinical trials and has shown great promise (ClinicalTrials.gov IDs NCT01252641, NCT00842634, and NCT01044654) (82). However, the target DNA sequence recognition by these protein-based genome engineering systems is determined by protein sequences. Tedious protein engineering and optimization are therefore required for each target DNA sequence, and delivering many of these proteins into cells for simultaneous multiplexed genetic manipulation is challenging. Because of these difficulties, their use for large-scale genomic manipulation or genetic screens has been very limited.

The CRISPR/Cas9 system offers a simple RNA-guided mechanism for introducing precise mutations at a target locus. Bacteria and archaea encode different types of natural CRISPR/Cas systems that recognize and eliminate invading foreign DNA species (3, 32, 66). The system encodes a set of Cas protein genes and a set of CRISPR RNA (crRNA) genes (117). Utilizing a complex of protein and RNA, CRISPRs recognize foreign DNA based largely on RNA-DNA base pairing, which subsequently triggers cleavage of foreign DNA by the Cas proteins.

The CRISPR/Cas9 system offers a simple RNA-guided mechanism for introducing precise mutations at a target locus. Bacteria and archaea encode different types of natural CRISPR/Cas systems that recognize and eliminate invading foreign DNA species (3, 32, 66). The system encodes a set of Cas protein genes and a set of CRISPR RNA (crRNA) genes (117). Utilizing a complex of protein and RNA, CRISPRs recognize foreign DNA based largely on RNA-DNA base pairing, which subsequently triggers cleavage of foreign DNA by the Cas proteins.

The discovery of the type II CRISPR/Cas9 system has inspired the development of a new approach for RNA-mediated DNA targeting (28, 41). Several discoveries were integral to its use as a genome engineering tool. The type II CRISPR/Cas9 system from Streptococcus thermophilus was the first one demonstrated to specifically cleave double-stranded DNA via a process mediated by Cas9 (93). Later, a short DNA sequence adjacent to an RNA-binding site, termed the protospacer-adjacent motif (PAM), was identified as a crucial element that helps Cas9 discriminate self and nonself DNA (70, 74). A trans-activating crRNA (tracrRNA) specific to the type II CRISPR directs the processing and maturation of the crRNA (20). In 2012, Jinek et al. (41) demonstrated that the Cas9 protein from Streptococcus pyogenes can bind with a tracrRNA-crRNA RNA complex to generate DSBs in vitro at a specific DNA sequence targeted by the 5′-terminal 20 nucleotides (nt) of the crRNA via Watson–Crick base pairing. The same study also showed that directing Cas9 to bind and cut a specific DNA sequence did not require using an RNA complex; instead, using a designed, chimeric single guide RNA (sgRNA) was sufficient. These fundamental biological
discoveries paved the way for the use of CRISPR/Cas9 for genome engineering, including gene editing and gene expression regulation, epigenetic modification, and genome imaging, as detailed below (Figure 1).

2. TARGETED GENOME EDITING WITH CRISPR/CAS9

Cas9 is a highly programmable nuclease tool for modifying DNA sequences in the genomes of various organisms. Directed by sgRNAs with a 20-nt DNA binding sequence, Cas9-induced sequence-specific DNA DSBs have been used to introduce nonhomologous end joining (NHEJ)-mediated sequence-specific insertion or deletion (indel) mutations in human endogenous genomic loci, which often lead to loss of function of target genes (17, 42, 68). The CRISPR/Cas9 system is highly programmable and multiplexable. When introducing multiple sgRNAs, it can simultaneously edit several sites within a mammalian genome (17, 69) and can generate animals that carry mutations in multiple genes (40, 55, 111). Multiple DSBs simultaneously induced by Cas9 and multiple sgRNAs can promote large or small chromosomal rearrangements between these DSBs, including interchromosomal translocations and intrachromosomal inversions, and could therefore serve as a potential tool for the study of genomic rearrangements (14, 122). The rearrangements and inversions likely occur through the NHEJ pathway of DSB repair, as they involve the joining of mismatched ends (Figure 2a).

Besides NHEJ, DSBs introduced by CRISPR can also trigger DNA repair through homology-directed repair (HDR). In the presence of a single-stranded oligonucleotide or double-stranded plasmid DNA donor template, HDR can mediate the precise replacement or insertion of DNA sequences from the template (Figure 2a). This allows precise gene modifications such as coding sequence replacements, including but not limited to targeted mutagenesis, gene correction, and
a Gene editing based on Cas9 nuclease activity

- Cas9 → DSB → NHEJ or HDR
  - Deletions
  - Insertions
  - Donor DNA
  - Precise insertion or modification

b Transcription repression mediated by dCas9

- RNA polymerase roadblock
  - dCas9
  - dCas9
  - Transcription repressor

- RNA polymerase

- dCas9

- Transcription activator

- Transcription activation

- i. VP64 transactivator fused to dCas9

- ii. SunTag system: recruit 24 copies of VP64 to dCas9

- iii. Tripartite transactivator

- iv. RNA scaffold recruits multiple copies of VP64 to CRISPR complex

- v. Synergistic activation mediator (SAM) system

- Single-chain fragment variable (scFv) antibody
  - 24 peptide epitopes

- dCas9

- MCP
  - MS2
  - VP64
  - Rta
  - HSF1

- dCas9

- VP64

- p65

- MCP

- MS2

- M52
insertion of genetic coding sequences such as fluorescence markers, protein tags, or recombination sites at human genomic loci (91, 126).

NHEJ naturally occurs more frequently than HDR in mammalian cells. Shifting from NHEJ to HDR can increase the efficiency of precise homologous recombination–based genome editing. Many efforts have been made toward achieving enhanced HDR. Lin et al. (56) found that, because the phase of the cell cycle during which the DNA repair happens largely determines the choice between NHEJ and HDR, the HDR rate increased from 9% to up to 33% of the total detected DSB repair events when delivering CRISPR components in the format of a Cas9-sgRNA complex into synchronized M-phase cells. In another study, Yu et al. (131) carried out a large-scale small-molecule screening and identified two molecules—L755507 (a β3-adrenergic receptor agonist) and brefeldin A (an inhibitor of intracellular protein transport from the endoplasmic reticulum to the Golgi apparatus)—that can enhance the efficiency of HDR-mediated gene insertion severalfold. Another strategy to enhance HDR is to inhibit the NHEJ pathway, because HDR and NHEJ are usually competing processes. Two separate studies have shown significant enhancement of precise HDR-mediated genome editing by antagonizing DNA ligase IV, a key enzyme in the NHEJ pathway, either by treating cells with the DNA ligase IV inhibitor Scr7 (71) or by knocking down the ligase with gene silencing (16). Cas9 can also be engineered into a nickase protein (by introducing point mutations to silence either the HNH or RuvC nuclease domain) in order to facilitate HDR with minimal mutagenic activity (90). The recent discovery of CRISPR from Prevotella and Francisella 1 (Cpf1), a class II CRISPR system that creates a staggered cut instead of a blunt-end cut, could potentially increase the frequency of HDR (133).

3. DEVELOPING CRISPR/DCAS9 TECHNOLOGY FOR GENE REGULATION

Beyond editing the genome (changing the genomic DNA sequence), technologies that allow one to switch gene expression on or off at the transcription level provide a powerful way to study gene function. Engineered DNA-binding proteins such as zinc-finger or transcription activator–like effector (TALE) proteins have been applied to activate or repress gene expression by fusing to transcription effector domains (5, 135). However, because these protein-based DNA recognition molecules are technically difficult to manufacture and deliver into cells, using them for genome-scale studies remains challenging. In general, modulating transcription can be done in two ways: downregulation (repression) and upregulation (activation).
For loss of function, a nuclease-deactivated form of Cas9 termed dCas9 was first repurposed as an RNA-guided platform that could efficiently repress gene expression. dCas9 can bind to the coding sequence of a gene or its promoter region to affect the activity of the RNA polymerase via complementary binding by an sgRNA. This binding is sufficient to repress transcription in microbial organisms (such as *Escherichia coli* and *Saccharomyces cerevisiae*) by blocking the elongating RNA polymerase (when binding to the coding sequence) or by interfering with the binding of the RNA polymerase to cognate promoter sequences (6, 87).

In mammalian cells, efficient transcription repression requires fusing dCas9 to a transcription repressor domain. Previous research has demonstrated that fusing dCas9 to the Krüppel-associated box (KRAB) domain could efficiently silence both reporter and endogenous gene expression in mammalian cells (30), a method referred to as CRISPR interference (CRISPRi) (Figure 2b). sgRNAs that target different regions of the gene locus showed different levels of repression efficiency, as demonstrated by a high-throughput experiment that targeted the genomic DNA of human K562 leukemia cells using 54,810 sgRNAs that tiled within a 10-kb sequence window around the transcription start sites of 49 genes (29). The optimal repression by dCas9-KRAB was achieved when targeting the 50–100-base-pair (bp) region downstream of the transcription start site. Furthermore, using sgRNAs with a statistical protospacer length of 18–21 nt leads to better repression, whereas the targeted DNA strand and sgRNA GC content are not crucial factors in CRISPRi efficiency (29).

RNA interference (RNAi) is the conventional approach for repressing gene expression on a large scale, but its off-target effects remain a major concern. In comparison, CRISPRi exhibits minimal off-target activity from properly designed sgRNAs (29, 30). Evidence suggests that CRISPRi is highly sensitive to mismatches between the target DNA and the base-pairing sgRNA, as even a single mismatch at the 3’ end near the PAM sequence dramatically decreases CRISPRi activity (29).

For gain of function, CRISPR/dCas9 has fused to transcription activators such as multiple copies of transactivator domain VP16. Catalytically inactive dCas9 could localize the activator domain to the promoter regions of target genes and activate their expression (12, 30, 65, 83). This technology is referred to as CRISPR activation (CRISPRa) (Figure 2c).

The CRISPRa system could be used to simultaneously activate many genes when multiple sgRNAs targeting these genes are expressed (12, 78). In mammalian cells, these sgRNAs could be expressed effectively from multiple constructs using an RNA polymerase III promoter such as U6. To express sgRNA from RNA polymerase II promoters, Nissim et al. (78) developed a Csy4 endoribonuclease system, which also allows expression of multiple sgRNAs from a single transcript.

The dCas9-VP16 fusions usually show modest gene activation activity. Several studies have shown that tiling a given promoter region with several sgRNAs can significantly increase the efficiency of gene activation compared with a single sgRNA (12, 65, 67, 83). However, using multiple sgRNAs is tedious, and genome-scale gain-of-function screens necessitate using a single sgRNA to efficiently switch on a gene. Therefore, improving the efficiency of endogenous gene activation and pushing the limits of activation potency with CRISPRa has been the focus for technological development.

Tremendous efforts have been made to increase activation efficiency either by recruiting more copies of the same activator or by recruiting different activators. One strategy to recruit multiple copies of an activator is to fuse a protein scaffold called SunTag to the dCas9 protein (106). In this system, a repeating peptide array containing up to 24 copies of antibody epitopes mediates high-affinity recruitment of a second fusion protein consisting of a single-chain fragment variable
(scFv) antibody and VP64. This system can induce robust transcription activation of endogenous genes with a single sgRNA and enable genome-wide gain-of-function screens (29) (Figure 2c).

Another strategy to enhance transcription activation efficiency is to fuse multiple activators to the dCas9 protein. When Chavez et al. (9) fused a tripartite activator system, consisting of a fusion of three transactivators called VP64-p65-Rta (VPR), to dCas9, the system activated endogenous gene expression much more efficiently than a VP64 fusion did. The enhanced synergistic effect of the tripartite activator suggests a potentially interesting transcription regulation mechanism and indicates that effective transcription activation may require coordination between many transcription factors. The synergistic activation mediator (SAM) system also exhibited this synergistic effect of recruiting multiple transcription activators to the CRISPRa complex (53). This system comprises a modified sgRNA with hairpin structures that recruits the fusion protein of RNA-binding protein and two activators, p65 activation domain (p65AD) and heat shock factor 1 (HSF1) (Figure 2c). Together with dCas9-VP64, the SAM system efficiently activates multiple endogenous genes, and it has been used for genome-wide gain-of-function screening (53).

In biological events such as organism development, transcription is elaborately regulated in time, space, and dosage. Understanding, interrogating, or reprogramming these natural transcriptional events therefore requires precisely controlling the spatiotemporal pattern and dosage of transcriptional activity. Drug-inducible control of the CRISPR/Cas9 system has been achieved by expressing either the Cas9 protein (22, 31) or sgRNA (2) with drug-inducible promoters. Inducing the reassembly of the two fragments from split Cas9 with drug-inducible dimerization domains also allowed drug induction of CRISPR gene editing (118, 134). Recent efforts have generated photocaging Cas9, which allows optogenetic activation of the Cas9 protein with UV light (34). Light-inducible transcription control with CRISPRa or CRISPRi has also been achieved by recruiting the transcription activator or repressor to dCas9 with light-inducible dimerization domains (76, 77, 85). However, these systems are significantly less efficient than direct fusions of transcription factors to dCas9. In the future, a more robust and efficient system with inducible and reversible gene regulation capability should be established.

Using the CRISPR/Cas9 system as a building block, multiple studies have made efforts to build complex gene regulatory circuits. Liu et al. (59) constructed a promoter-based AND gate as a detector of bladder cancer cells by driving the two components of the CRISPR/Cas9 system—the nuclease Cas9 and sgRNA—with two cancer cell–specific promoters; this system allows the assembly of the CRISPR only when both promoters are activated. A more complex circuit with multilayer regulatory control was achieved by interconnecting cascaded dCas9-based transcription regulation events (48, 78). In these studies, the authors created sophisticated feedback-loop and multioutput circuits built by the combination of microRNA machinery, RNA-processing mechanism, and CRISPR-based transcription regulation (78).

One major advantage of the CRISPR/Cas9 system is its flexibility to modulate multiple genes at once. To enable simultaneous repression and activation in the same cells, Zalatan et al. (132) engineered scaffold RNA by appending orthogonal protein-interacting RNA hairpin structural modules to the 3’ ends of sgRNAs, which then recruit their cognate binding proteins fused with VP64. This enabled them to construct simultaneous ON/OFF gene regulatory switches using orthogonal RNA-binding proteins fused to either transcription activators or repressors, which were then recruited to the corresponding orthogonal RNA hairpin motifs integrated in the sgRNAs targeting distinct genes.

Epigenetic modification is an inheritable form of transcription regulation. The ability to control and modify epigenetic marks in a locus-specific way will enable the engineering of transcription regulation across multiple cell generations. Precisely modifying an epigenetic mark at a target locus may also lead to gene therapies based on epigenome editing. Although the concept is intriguing,
the development of CRISPR-based epigenetic engineering tools is still in its infancy. A recent study by Hilton et al. (35) reported a programmable CRISPR-based system created by fusing dCas9 to the catalytic core of the human acetyltransferase p300 core. The authors showed robust transcriptional activation of target genes when targeting proximal or distal enhancers. In another study, Kearns et al. (45) fused dCas9 to lysine demethylase 1 (LSD1), and the fusion protein could effectively repress enhancers of pluripotency regulation factors in mouse embryonic stem cells, including octamer-binding transcription factor 4 (Oct4) and T-box 3 (Tbx3), resulting in changes in colony morphology and increases of differentiation-associated markers. Targeted epigenetic engineering tools offer great advantages for gene regulation in terms of potency and potential inheritability over time, and therefore they can be useful for cell-based therapeutic applications.

4. APPLICATION OF CRISPR/CAS9 TO GENOME-WIDE SCREENING

4.1. CRISPR/Cas9 Nuclease Function–Based Loss-of-Function Screening

An important conventional approach to identifying unknown genes and understanding gene function is to utilize genetic screening in order to determine the genes responsible for certain phenotypic changes. Genetic screens based on random DNA mutagenesis have led to the discovery of many important pathways and basic biological mechanisms. However, this approach also has significant limitations: The resulting mutants are typically heterozygous, and the random mutations are unknown. The development of RNAi, which targets specific mRNA molecules for degradation, has revolutionized forward genetic screening in the past decade. RNAi-based screens allow large-scale targeted genetic screens and have generated valuable information about gene functions, such as gene targets that confer drug resistance or sensitivity (4). However, the application of RNAi to screens has been hindered because RNAi knockdown is usually inefficient and creates significant off-target effects.

As a highly programmable sequence-specific nuclease, CRISPR/Cas9 has been widely applied to high-throughput functional genomic studies. By varying the unique sgRNA sequence, one can use CRISPR/Cas9 to target any gene and efficiently introduce mutations or deletions in the targeted regions. Therefore, the CRISPR/Cas9 system combines the advantages of the permanently mutagenic nature of classical mutagens and the programmability of RNAi.

CRISPR-based knockout screens provide a method that is mechanistically distinct from RNAi for systematic perturbation of gene function. RNAi reduces protein expression by targeting RNA, whereas CRISPR knockout introduces loss-of-function mutations into the genomic DNA to permanently silence the target gene. Although some indel mutations are expected to maintain the open reading frame, complete loss-of-function knockout yields high screening sensitivity, which is important in cases where incomplete knockdown retains gene function. In addition, RNAi is limited to transcripts, whereas CRISPR can target elements across the entire genome, including promoters, enhancers, introns, and intergenic regions.

The CRISPR approach is particularly powerful in pooled genetic screens. The ease of designing and synthesizing DNA oligonucleotides that encode sgRNAs allows the generation of sgRNA libraries with a scale up to almost 100,000 and covering up to 100,000 genes. Creating a lentiviral sgRNA library usually requires four steps: (a) computationally designed oligonucleotide libraries containing the target-specific sequences are synthesized (by a commercial vendor), (b) these oligonucleotides are cloned as a pool to create a lentiviral vector library, (c) this vector library is used for pooled viral particle production, and (d) viral sgRNA libraries are used to transduce cells at a low multiplicity of infection to ensure that each cell can take at most one sgRNA viral particle (reviewed in 99). Multiple sgRNAs are usually included for each gene to...
statistically reduce off-target effects, because it is unlikely that two sgRNAs targeting the same gene have the same off-target effects (off-target effects are further discussed in Section 9, below). The true target rate could be enhanced when using a cutoff of a minimum of two different sgRNA hits per gene to select candidate genes for validation (52).

CRISPR-based screening has been successful in both gain- and loss-of-function genetic screens. Positive selection screens are usually designed to select the perturbations that lead to resistance to unfavorable growth conditions, such as toxins, drugs, or pathogens (62, 98, 114, 138). Therefore, cells with resistance will enrich in the environment, and the corresponding sgRNAs and their targeted gene could be identified by deep-sequencing analysis. Negative selection screens are designed to select the perturbations that cause cells to be less favorable during selection, therefore targeting genes that are necessary for survival under the chosen selective pressure. These genes could be identified by comparing the frequency of each sgRNA between a late time point and an early time point. With these screening paradigms, many studies have been carried out to identify genes that are essential for cell viability in cancer and pluripotent stem cells (98).

The selection criteria of screening could be tailored to the specific purpose of the screening, such as the appearance or absence of certain cell surface markers (80) or the evolution of cancer cell metastasis indicated by cell migration (11). Some screens have utilized a sorted clonal Cas9 stably integrated cell line because individual clones harboring Cas9 may vary in how efficiently they generate sgRNA-mediated indels. The efficiency of CRISPR/Cas9 for genetic screening also depends on the targeting site in the gene structure. For example, targeting exons encoding functional protein domains generated a higher proportion of null mutations and increased the screening potency (100).

Genetic screens in primary cells can be quite challenging because of the difficulty of delivering exogenous DNA materials into many cell types and the general challenge of long-term culture of primary human cells. The generation of CRISPR/Cas9 knock-in mice (84) facilitates such screening because primary cells generated from Cas9 transgenic mice will have Cas9 stably integrated in the genome. Efforts have been made to develop convenient methods for generating CRISPR sgRNA library–based knockout mice for genetic screening (137). For example, a pooled CRISPR screening has taken advantage of this approach by generating bone marrow–derived dendritic cells from Cas9 mice and screening for regulatory factors of innate immune circuits responsible for the host response to pathogens (80). The pooled Cas9-sgRNA-integrated cell lines could also be introduced in vivo by transplanting cells into animal models and assessing physiological outputs such as cancer metastasis (11).

4.2. CRISPR/Cas9 Transcription Regulation–Based Screening

CRISPR/Cas9-based transcription regulation carried out by fusing nuclease-inactive dCas9 to various transcription regulation domains has enabled both genome-wide loss-of-function and gain-of-function screens. By using the dCas9-KRAB system to robustly repress gene expression as part of a CRISPRi-based system, Gilbert et al. (29) successfully demonstrated that this strategy can be applied to genome-wide genetic screening. This screen revealed genes and pathways that modulate cellular response to the AB toxin ricin and a chimeric cholera-diphtheria fusion toxin (CTx-DTA). Partial and reversible repression with CRISPRi is especially useful when studying essential genes, which cannot be done with nuclease Cas9 because this approach permanently disrupts the gene and causes cell death.

In the past, gain-of-function screens have been challenging. The use of small-scale cDNA overexpression libraries has been the primary discovery method to identify key gene factors for oncogenesis, development, and cell proliferation. However, because of the complexity of the
transcript isoform variance, designing a cDNA library that covers this information is challenging. In addition, large cDNA sequences are often difficult to clone into size-limited viral expression vectors, and synthesizing cDNA libraries on a large scale is expensive. Therefore, the utility of gain-of-function screens using cDNA libraries is limited.

Excitingly, the development of the CRISPRa system also enabled systematic genome-scale gain-of-function perturbations at endogenous loci. The dCas9-VP64 fusion protein can activate endogenous genes with multiple sgRNAs tiled along the promoter sequence. However, because large-scale screening requires using a single sgRNA for each gene, effective gain-of-function screens require more efficient CRISPRa systems. The SunTag and SAM systems significantly enhance transcription activation efficiency and allow genome-wide screening with CRISPRa. Gilbert et al. (29) used the SunTag system to perform gain-of-function screening and identified genes involved in cell resistance to ricin. A genome-wide gain-of-function screen using the SAM system provided insight into the epidermal growth factor receptor (EGFR) signaling pathway components that mediate resistance against the BRAF inhibitor PLX04720 (53). As these results demonstrate, robust activation from the endogenous gene loci with CRISPRa has provided an excellent platform for genome-wide gain-of-function screening studies.

5. APPLICATION OF CRISPR/CAS9 TO STUDIES OF GENOMIC STRUCTURE

The ability to visualize endogenous genomic loci in living cells to track their dynamics has provided an ideal research tool for studying genomic structure. The most popular method for imaging the genome sequence is probably fluorescence in situ hybridization (FISH), which uses fluorescence-tagged nucleic acid probes (RNA or DNA) based on sequence complementarity to label specific DNA sequences. Emerging genome engineering tools have also advanced our ability to visualize genomic sequences. For example, by fusing fluorescent proteins to TALEs, several groups have been able to image repetitive genomic elements (64, 73, 107). However, the difficulty of constructing many TALE proteins and delivering them into the cells has prevented this application from imaging nonrepetitive sequences. The CRISPR method, by contrast, is able to image both repetitive and nonrepetitive sequences. By fusing a green fluorescent protein (GFP) to the nuclease-inactive S. pyogenes Cas9, Chen et al. (10) were able to label the DNA sequence that is complementary to the sgRNA in mammalian cells. In this study, the authors also captured the dynamics of repetitive genomic loci (telomeres) throughout the cell cycle. When tiling multiple sgRNAs along the target locus, this platform can be used to image sequence-specific nonrepetitive genomic elements in living human retinal pigment epithelial cells (10). Another study reported labeling endogenous centromeres, pericentric regions, and telomeres in living mouse embryonic stem cells (1). The fluorescent signal for CRISPR imaging could be enhanced by using the dCas9-SunTag system, which allows recruitment of multiple copies of GFP to each dCas9 protein (106).

Another development is multicolor CRISPR imaging. Ma et al. (63) paired orthologous dCas9 proteins from three bacterial species with three different fluorescent proteins. Each pair was targeted to distinct genomic loci in living human cells, guided by their cognate sgRNAs. This strategy allowed the authors to study the intranuclear distance between loci on different chromosomes as well as two loci on the same chromosome with a resolution of approximately 2 Mb. DNA compaction of a chromosome region could potentially be inferred by comparing the measurable fluorescent distance of intrachromosomal loci with their linear distance on the chromosome’s physical map.

The CRISPR/Cas9 system could also be adapted to study DNA-binding proteins, which are important components of genomic structure. Fusion of the dCas9 protein with affinity protein tags
and immunoprecipitation of the tagged dCas9 protein enable the pull-down of proteins that are bound to a specific genomic region targeted by guide RNA(s), and these DNA-binding proteins could then be characterized by proteomic studies (27). This CRISPR-based immunoprecipitation method thus has the potential to be developed into a tool for characterizing protein-DNA interactions at specific genomic loci. Given the potential for off-target dCas9 binding events (discussed further below), future applications of this tool for studying genomic structure will require proper controls and validations.

### 6. APPLICATION OF CRISPR/CAS9 TO STUDIES OF HUMAN DISEASES

Molecular genetics plays a key role in exploring the molecular mechanisms of diseases. Genetically modified animal models are crucial tools for understanding gene functions and pathogenesis in human diseases. For the creation of transgenic mouse models, genome modifications are achieved primarily through homologous recombination in mouse embryonic stem cells, followed by microinjection of these cells into blastocysts for germline transmission—a process that is very time consuming because of its inefficiency in triggering genomic modifications. For mammalian species other than mice, it is difficult to culture embryonic stem cells in vitro to generate chimeric animals. Compared with traditional approaches, the CRISPR/Cas9 system offers an easier and more efficient technology for multiplexed genome editing in generating animal disease models.

Similarly to traditional methods of generating genetically modified mouse models, CRISPR/Cas9 has been used to manipulate genes in the germline or zygote stage, but it enables a faster, more efficient, and multiplexable process. Wang et al. (111) demonstrated that CRISPR/Cas9-mediated genome editing allows rapid, efficient, and simultaneous knockout of several genes in mouse embryonic stem cells. They showed that Cas9-encoding mRNA and sgRNAs can be directly injected into the fertilized eggs of mice, efficiently producing mice carrying biallelic mutations in one or more genes. In addition to NHEJ-mediated gene knockout, the same group demonstrated that CRISPR/Cas9 technology can be used for precise HDR-mediated genome editing by coinjecting Cas9 mRNA, sgRNAs, and single-stranded DNA oligonucleotides into a one-cell embryo and successfully targeted several genes with specific modifications (126). This one-step procedure allowed them to generate mice carrying a reporter gene, a conditional allele, or a tag in endogenous genes with a specific modification of interest in a one-cell zygote. The study provides a proof of principle for using the CRISPR/Cas9 system to carry out multiplexed editing of animal embryonic stem cells or zygotes in order to create mouse disease models. Platt et al. (84) later generated a Cre-dependent Rosa26 Cas9 knock-in mouse that can be used in conjunction with specific sgRNAs, providing a promising mouse model for studying in vivo gene functions in biological processes and diseases.

Although CRISPR/Cas9 has been harnessed to manipulate genes in animals at the germline stage, it remains challenging to efficiently deliver Cas9 (in DNA, RNA, or protein format) in vivo owing to its large size and other factors. Nevertheless, there have been reports of successful in vivo delivery of Cas9-sgRNA into postnatal mice. Using hydrodynamic injection to deliver a plasmid encoding Cas9 and sgRNAs that target the tumor suppressor genes phosphatase and tensin homolog (Pten) and p53 to the liver, Xue et al. (124) successfully generated a liver cancer mouse model that phenocopies the reported effects of gene deletion using the traditional Cre-LoxP technology. Swiech et al. (104) injected dual-system adeno-associated viruses (AAVs) expressing Cas9 and an sgRNA targeting the methyl-CpG binding protein 2 (MeCP2) gene into the hippocampal dentate gyrus of adult male mice. After the crucial gene for contextual learning was deleted, behavioral tests on these mice revealed impaired contextual memory ability (summarized in Table 1). Carroll et al. (7) generated cardiac-specific Cas9 transgenic mice and used AAV9 to deliver an sgRNA to
Table 1 CRISPR/Cas9-based disease mouse models

<table>
<thead>
<tr>
<th>Disease mouse model</th>
<th>Targeted genes</th>
<th>Delivery method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver cancer</td>
<td>Pten, p53</td>
<td>Hydrodynamic injection to deliver a CRISPR plasmid DNA expressing Cas9 and sgRNAs</td>
<td>124</td>
</tr>
<tr>
<td>Contextual memory</td>
<td>Mecp2</td>
<td>Stereotactical injection of a mixture (1:1 ratio) of AAV expressing Cas9 and sgRNAs into the hippocampal dentate gyrus</td>
<td>104</td>
</tr>
<tr>
<td>Bronchial alveolar adenoma</td>
<td>Kras, p53, Lkb1</td>
<td>Intratracheal delivery of AAV expressing Kras-, p53-, and Lkb1-targeting sgRNAs into a Cre-dependent Rosa26 Cas9 knock-in mouse</td>
<td>84</td>
</tr>
<tr>
<td>Intestinal hyperplasia</td>
<td>Apc</td>
<td>Doxycycline-induced gene deletion in 4–5-week-old inducible CRISPR (both Cas9 and sgRNA) knock-in mice</td>
<td>22</td>
</tr>
<tr>
<td>Cardiomyopathy</td>
<td>Myh6</td>
<td>Intraperitoneal injection of postnatal cardiac-Cas9 transgenic mice with AAV9 encoding sgRNA against Myh6</td>
<td>7</td>
</tr>
<tr>
<td>Adrenal hypoplasia congenita and hypogonadotropic hypogonadism</td>
<td>DAX1</td>
<td>Microinjection of Cas9 mRNA and sgRNA into a one-cell monkey embryo</td>
<td>44</td>
</tr>
<tr>
<td>Hepatocellular carcinoma and intrahepatic cholangiocarcinoma</td>
<td>Somatic multiplex mutagenesis</td>
<td>Hydrodynamic tail vein injection of SB transposase and CRISPR-SB sgRNA vectors</td>
<td>116</td>
</tr>
</tbody>
</table>

Abbreviations: AAV, adeno-associated virus; Apc, adenomatous polyposis coli; DAX1, dosage-sensitive sex reversal, adrenal hypoplasia critical region on the X chromosome, gene 1; Kras, Kirsten rat sarcoma viral oncogene homolog; Lkb1, liver kinase B1; Mecp2, methyl-CpG binding protein 2; Myh6, myosin heavy chain 6; Pten, phosphatase and tensin homolog; SB, Sleeping Beauty; sgRNA, single guide RNA.

target the myosin heavy chain 6 (Myh6) locus exclusively in cardiomyocytes, and observed that the resulting mice displayed severe cardiomyopathy and loss of cardiac function. Weber et al. (116) used CRISPR/Cas9-based targeted somatic mutagenesis to mutate large gene sets and induce hepatocellular carcinoma and intrahepatic cholangiocarcinoma in mice. This provided a high-throughput analysis of gene function and functional annotation of cancer genomes in mice.

The CRISPR/Cas9 system also offers advantages when generating transgenic models of organisms other than mice. For example, several groups were able to efficiently modify endogenous genes in zebrafish by microinjecting zebrafish-codon-optimized Cas9 mRNA and sgRNAs (RNA form) into one-cell embryos (8, 39, 40). CRISPR/Cas9 has also been demonstrated to be a rapid and powerful tool in larger animal species, such as rat (55), sheep (18), goat (115), rabbit (125), pig (113), and monkey (79). In monkeys, coinjection of Cas9 mRNA and sgRNAs enabled precise and simultaneous disruption of two genes in one step with no detected off-target effects (79), providing a reliable and efficient platform to generate genetically modified monkey disease models. Kang et al. (44) generated a monkey model of X-linked adrenal hypoplasia congenita and hypogonadotropic hypogonadism using CRISPR/Cas9-mediated mutations in the dosage-sensitive sex reversal, adrenal hypoplasia critical region on the X chromosome, gene 1 (DAX1) locus. Because the larger mammalian models (especially primates) are more genetically and physiologically similar to human beings than smaller models are, they are meaningful for modeling human diseases and developing therapeutic strategies. Using CRISPR/Cas9 and a single sgRNA, Yang et al. (127) recently disrupted 62 repeated copies of the porcine endogenous retrovirus gene in porcine kidney epithelial cell line PK15, and the engineered cells showed over a 1,000-fold reduction in the
transmission of this retrovirus to human cells. This work demonstrates the promise of the clinical application of CRISPR/Cas9 technology for porcine-to-human organ xenotransplantation in the face of current challenges such as organ shortages.

7. POTENTIAL APPLICATIONS OF CRISPR/CAS9 FOR CURING DISEASES

While the CRISPR/Cas9 system has been widely used to modify targeted endogenous alleles in organisms for disease modeling, therapeutic applications based on the technology have also been taking off. Among its major uses are correcting defective genes in genetic diseases and eliminating viruses in the human genome to treat infectious diseases.

For example, Wu et al. (120, 121) microinjected the Cas9 mRNA, an sgRNA, and an oligonucleotide containing a correction sequence of crystallin gamma C (Crygc) into the zygotes of a cataract mouse disease model containing a single-base mutation in exon 3 of the gene. This corrected the mutant Crygc gene and rescued the cataract symptoms, and additionally corrected the mutation in spermatogonial stem cells harvested from the mouse. Fertilization using corrected spermatids gave rise to normal offspring with almost 100% efficiency. This study provides proof of concept for curing genetic disease in offspring through genetic correction in zygotes or spermatogonial stem cells.

In a mouse disease model of Duchenne muscular dystrophy (DMD), precise correction of the dystrophin gene (Dmd) mutation by coinjection of the Cas9 mRNA, an sgRNA, and an oligonucleotide produced genetically mosaic animals containing 2–100% correction of the Dmd gene. Interestingly, the degree of muscle phenotypic rescue in mosaic mice exceeded the efficiency of gene correction, which likely reflected an advantage of the corrected cells and their contribution to regenerating muscle (61). Using the DMD model, three recent reports developed and applied CRISPR/Cas9 to correct genomic mutations leading to DMD in vivo (60, 75, 105), which further established the potential of CRISPR/Cas9 to treat DMD and other genetic diseases. Similarly, in a study that used a mouse model of the human disease hereditary tyrosinemia with a fumarylacetate hydrolase (Fah) gene mutation in hepatocytes, hydrodynamic injection of the components of the CRISPR/Cas9 system resulted in initial expression of the wild-type Fah protein in approximately 1 out of 250 liver cells (130). Expansion of Fah-positive hepatocytes further rescued the body weight loss phenotype. Using an improved delivery method of CRISPR/Cas9, Yin et al. (129) combined lipid nanoparticle-mediated delivery of Cas9 mRNA, AAVs encoding an sgRNA, and a repair template to successfully repair a Fah-splicing mutation. Another work demonstrated a partial cure of a liver genetic disease using AAV-mediated CRISPR/Cas9 correction of a urea cycle disorder enzyme, ornithine transcarbamylase (OTC) (128).

Taken together, these studies indicate that CRISPR/Cas9-mediated genome editing provides a potential gene therapy scheme for precisely correcting human genetic diseases. This technology opens up exciting possibilities for future treatment of postnatal somatic diseases.

Beyond correcting disease mutations in mouse models, the CRISPR/Cas9 system has been utilized in correcting genetic diseases in primary human patient cells (summarized in Table 2). Schwank et al. (95) isolated intestinal stem cells from cystic fibrosis patients with a homozygous deletion of phenylalanine at position 508 (F508) in the cystic fibrosis transmembrane conductor regulator gene (CFTR), corrected the F508 deletion mutation using CRISPR/Cas9-mediated HDR, and confirmed the functionality of the corrected allele in the expanded organoid system. This could be a potential therapeutic strategy for treating intestinal diseases by transplanting the in vitro expanded and corrected organoids into the patients. In another study, Xie et al. (123) used the CRISPR/Cas9 system to correct disease-causing mutations in the human hemoglobin beta gene
<table>
<thead>
<tr>
<th>Disease</th>
<th>Targeted gene/DNA</th>
<th>Correction method</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genetic diseases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cataract</td>
<td>Crygc</td>
<td>Injection of Cas9 mRNA, sgRNA, and ssODN as a template for HDR-mediated gene repair into zygotes</td>
<td>120</td>
</tr>
<tr>
<td>Duchenne muscular dystrophy</td>
<td>Dmd</td>
<td>Injection of Cas9 mRNA, sgRNA, and ssODN as a template for HDR-mediated gene repair into zygotes</td>
<td>61, 60, 75, 105</td>
</tr>
<tr>
<td>Hereditary tyrosinemia</td>
<td>FAH</td>
<td>Lipid nanoparticle–mediated delivery of Cas9 mRNA with AAVs encoding an sgRNA and a repair template</td>
<td>129</td>
</tr>
<tr>
<td>Hereditary tyrosinemia type I</td>
<td>FAH</td>
<td>Hydrodynamic tail vein injection of plasmids expressing Cas9, sgRNA, and ssDNA donor</td>
<td>130</td>
</tr>
<tr>
<td>Cystic fibrosis</td>
<td>CFTR</td>
<td>Cotransfection of a plasmid expressing Cas9 and sgRNA together with a donor plasmid encoding wild-type CFTR sequences</td>
<td>95</td>
</tr>
<tr>
<td>β-Thalassemia</td>
<td>HBB</td>
<td>Homologous recombination mediated by a footprint-free piggyBac system</td>
<td>123</td>
</tr>
<tr>
<td>Urea cycle disorder</td>
<td>OTC</td>
<td>One AAV expressing Cas9 and one AAV expressing a guide RNA and the donor DNA</td>
<td>128</td>
</tr>
<tr>
<td><strong>Infectious diseases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBV</td>
<td>HBV cccDNA</td>
<td>Plasmid transfection or lentiviral transduction for in vitro assays</td>
<td>21, 46, 57, 88, 96</td>
</tr>
<tr>
<td>HIV-1</td>
<td>HIV-1 LTR</td>
<td>Hydrodynamic injection of plasmids encoding Cas9 and sgRNAs for in vivo assays</td>
<td>23, 37</td>
</tr>
<tr>
<td>EBV</td>
<td>Latent EBV in Burkitt’s lymphoma cell line</td>
<td>Nucleofection of plasmid encoding Cas9 and sgRNA</td>
<td>112</td>
</tr>
<tr>
<td>HPV</td>
<td>HPV oncogenes E6 and E7 in cancer cell line</td>
<td>Transfection of plasmid encoding Cas9 and sgRNA</td>
<td>47, 136</td>
</tr>
</tbody>
</table>

Abbreviations: AAV, adeno-associated virus; cccDNA, covalently closed circular DNA; CFTR, cystic fibrosis transmembrane conductor regulator; Crygc, crystallin gamma C; Dmd, dystrophin; EBV, Epstein-Barr virus; FAH, fumarylacetoacetate hydrolase; HBB, hemoglobin beta; HBV, hepatitis B virus; HDR, homology-directed repair; HIV, human immunodeficiency virus; HPV, human papillomavirus; LTR, long terminal repeat; OTC, ornithine transcarbamylase; sgRNA, single guide RNA; ssDNA, single-stranded DNA; ssODN, single-stranded oligodeoxynucleotide.

(HBB) from β-thalassemia patients. The authors precisely modified the patient-derived induced pluripotent stem cells to correct the mutations in the HBB gene. The corrected patient-derived induced pluripotent stem cells can potentially be used to restore normal function for potential therapeutic transplantation. For T cell engineering, Schumann et al. (94) reported that delivery of CRISPR/Cas9 ribonucleoprotein, a complex assembled with the Cas9 protein and an sgRNA in vitro, allowed efficient gene knockout and knock-in in primary T cells. Su et al. (103) also reported gene knockout of programmed death 1 (PD-1) by electroportation of plasmids encoding Cas9 and sgRNA in primary T cells derived from cancer patients. Together, these studies demonstrate great
promise for the use of CRISPR/Cas9 to facilitate existing therapies, such as the use of chimeric antigen receptors and T cells to treat cancers, infectious diseases, primary immune deficiencies, and autoimmune diseases.

Programmable nucleases, including ZFNs, TALENs, and CRISPR/Cas9, could also be used as potential treatments for bacterial and viral infections. Compared with ZFNs and TALENs, the CRISPR/Cas9 approach is more robust and efficient in targeting invading pathogenic microbes. Several groups have used CRISPR/Cas9 to successfully target hepatitis B virus genomic DNA (21, 46, 57, 88, 96). In these studies, expression of Cas9 and designed sgRNAs targeting the hepatitis B virus genomic DNA significantly decreased the viral protein levels. Notably, CRISPR/Cas9 can target viral covalently closed circular DNA in replicating cells, chronically infected hepatocytes, and mouse models, implying that it could potentially be used to treat acute and chronic hepatitis B virus infection.

Although current anti-human immunodeficiency virus (HIV) therapies can inhibit HIV-1 replication, the viruses that have integrated within the host genome in a latent state can still potentially reactivate at any time. The CRISPR/Cas9 system may be useful for eliminating latent HIV-1 by targeting its genomic DNA. Several groups have reported that using CRISPR/Cas9 to target long terminal repeats eradicated the HIV-1 genome integrated in the host chromosome and effectively immunized the targeted cells against HIV-1 reactivation with high specificity and efficiency (23, 37).

CRISPR/Cas9 also provides a therapeutic strategy to cure other infectious diseases. Wang & Quake (112) reported that patient-derived cells with latent Epstein-Barr virus infection showed dramatic proliferation arrest and a concomitant decrease of viral titers after they applied CRISPR/Cas9 targeting to the viral genome. Furthermore, targeting human papillomavirus E6 or E7 genes in cervical carcinoma cells resulted in cell cycle arrest and senescence. These data provide preliminary evidence that CRISPR/Cas9-mediated genomic editing could offer effective treatment for virus-induced cancers (38, 47, 136) (summarized in Table 2).

Although CRISPR/Cas9-mediated genome editing provides dramatic advantages over conventional approaches and is moving rapidly toward treatments, there are several concerns with its broad application. First, although its targeting specificity seems much higher than that of RNAi, there are reported off-target mutagenesis effects of the system (see Section 8). Second, in vivo delivery of Cas9 to cells is limited owing to the large size of Cas9 (an average Cas9 coding sequence is on the order of 3–5 kb). For example, although AAV vectors are commonly used for in vivo gene delivery because they have low immunogenicity, remain episomal rather than integrating into the genome, and have a variety of serotypes allowing for infection of certain tissues, their maximal packaging capability is only about 4.5 kb, leaving limited space for additional regulatory regions (such as promoters) with the S. pyogenes Cas9 (approximately 4.1 kb, with 1,368 amino acids). Employing smaller Cas9 orthologs (such as Staphylococcus aureus Cas9, which is approximately 3.2 kb, with 1,053 amino acids) or engineering a minimal Cas9 could facilitate in vivo delivery for therapeutic purposes. Finally, diminishing adverse immune responses to the bacterial Cas9 protein might be necessary. Little work has been performed to characterize the immunogenicity of the Cas9 protein in humans. Developing CRISPR/Cas9-based therapeutics will require devoting more effort to addressing the potential problems caused by the immunogenicity of the Cas9 protein before its application in clinical trials. For example, approaches to minimize the adverse immune response—including humanizing the relevant peptide fragments and optimizing parameters for drug delivery, such as dosage and drug formulation—could help reduce potential immunogenicity effects.
8. CHALLENGES AND FUTURE DIRECTIONS OF APPLYING CRISPR/CAS9 TO THERAPIES WITH REDUCED OFF-TARGET EFFECTS

CRISPR/Cas9 recognizes its genomic target by Watson-Crick base pairing between the sgRNA and the target DNA. Therefore, the tolerance of mismatches of sgRNA is a key factor determining the specificity of CRISPR/Cas9. Several groups have created sgRNA variants containing different numbers of nucleotide mismatches (up to four) in the complementary region and tested the Cas9-mediated cleavage activity with these sgRNA variants at reporter genes or endogenous genes (25, 36). The results showed that the mismatches at the 3′ end of the unique 20-nt target sequence of the sgRNA are generally less tolerated than mismatches at the 5′ end, which may be explained by biochemical studies suggesting that the sequence at the 3′ end of the targeting sequence is crucial for target recognition, and therefore is regarded as the seed sequence (97).

The off-target activities of Cas9 can also be characterized by directly assessing the potential off-target genomic DNA sites defined by the sequences that have a few (one to six) nucleotide differences compared with the intended target sequence. A given 20-nt target sequence might have hundreds to thousands of such potential off-targets in random DNA within the human genome. Off-target binding can happen at a locus with as many as five mutations within the sgRNA (25) or with an alternative PAM sequence (36). Moreover, Cas9 can cleave off-target sites with extra or missing nucleotides that form a DNA or RNA bulge (58). Pattanayak et al. (81) used high-throughput sequencing to assess off-target effects with preselection libraries containing more than $10^{12}$ individual potential off-target sites for specific target sequences, and found that there was a trade-off between cleavage efficiency (on-target binding) and specificity (off-target binding). They also found that a shorter, less active sgRNA was more specific than a longer, more active sgRNA and that a higher concentration of the Cas9-sgRNA complex showed more off-target sites.

An unbiased approach has been used to test the genome-wide off-target effects of the CRISPR/Cas9 DNA-binding event. To uncouple DNA binding from cleavage, these studies used nuclease-inactive dCas9. There are two major approaches to profiling the binding events of Cas9: identifying binding events by gene activation mediated by dCas9–transcription activators (67, 86) and using chromatin immunoprecipitation of the dCas9-sgRNA complex followed by high-throughput sequencing (ChIP-Seq) analysis to identify the bound DNA sequences (54, 119). These studies revealed that, depending on the design of the sgRNA, significant off-target dCas9 binding could occur, with some showing thousands of off-target binding sites.

However, binding and cleavage are not necessarily coupled. For example, in vitro experiments have confirmed that off-target binding sites with mismatches distal from the cleavage site (for S. pyogenes Cas9, this cleavage site is 3 bp from the PAM with the binding region) could be tightly bound but not cleaved (102). In fact, genome-wide detection of DSBs on the DNA provides a more direct way to assess the specificity of Cas9-mediated DNA cleavage, and several methods have been developed for this purpose (Figure 3). In one method—called genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-Seq)—the Cas9-sgRNA-induced DSBs are tagged in the genomes of living cells by integrating a blunt, double-stranded oligodeoxynucleotide during the end-joining process following a DSB. The double-stranded oligodeoxynucleotide integration sites are then amplified and deep sequenced (109). Another method—called high-throughput, genome-wide translocation sequencing (HTGTS)—allows the detection of DSBs based on translocation to other endogenous or ectopic DSBs. This allows detection of junctions mediated by genome-wide DSBs using the target DSB as bait to catch the prey sequences transacted to the target DSB (24). A third method—called breaks labeling, enrichments on streptavidin, and next-generation sequencing (BLESS)—labels DSBs in fixed cells using biotinylated oligonucleotides,
Approaches for unbiased genome-wide measurement of double-strand breaks (DSBs) and off-target effects. Next-generation sequencing has greatly facilitated unbiased detection of DSBs in the genome. However, depending on the experimental needs, the upstream DSB labeling and capture and sample preparation for library construction can be very different. Four approaches for capturing DSBs in the genome are shown here: genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-Seq); high-throughput, genome-wide translocation sequencing (HTGTS); breaks labeling, enrichments on streptavidin, and next-generation sequencing (BLESS); and digested genome sequencing (Digenome-Seq). The steps in light-brown boxes are events that occur in the live cells; those in light-blue boxes are cell-free events after DNA extraction. Additional abbreviations: dsODN, double-stranded oligodeoxynucleotide; gDNA, genomic DNA; PCR, polymerase chain reaction; sgRNA, single guide RNA.

allowing the enrichment of the DSB-containing sequences followed by deep sequencing (19, 89). Finally, another method, digested genome sequencing (Digenome-Seq), uses cell-free genomic DNA for in vitro Cas9-mediated digestion followed by whole-genome sequencing to profile genome-wide Cas9 off-target effects (49). All published studies have suggested that the off-target effect of Cas9-sgRNA could vary in frequency depending on the sgRNA design and target sequence. Notably, accurately predicting the off-target cleavage sites remains a challenge.

Researchers have explored various approaches to improve the specificity of CRISPR/Cas9. The choice of proper target sequence is a key factor that helps improve the specificity. Predictive algorithms have been developed to facilitate this process by computationally searching target sequences that are distinct from any other sequence and thus include fewer off-target sites in the genome.

Precisely controlling the amount of Cas9 and sgRNA in cells helps improve the specificity. A few studies have shown that reducing the concentration of Cas9 and sgRNA in cells could
reduce off-target effects (25, 36). Furthermore, delivering the Cas9-sgRNA ribonucleoprotein complex resulted in fewer off-target effects compared with delivering the plasmids, likely because the ribonucleoprotein complex introduced on-target cleavage immediately after delivery and then was rapidly degraded by endogenous proteases (50). Modifying the sgRNA sequence also improved the specificity. For example, an sgRNA with a truncated base-pairing sequence (17 nt instead of 20 nt) enhanced the targeting specificity (26) because truncated sgRNAs have reduced binding affinity with the target DNA and thus are more sensitive to mismatches.

The other approach is to take advantage of the Cas9 nickase that contains mutations in one of the two nuclease domains, HNH or RuvC, which cleave the DNA strand complementary and noncomplementary (respectively) to the sgRNA (17, 28, 41). A pair of Cas9 nickases could generate two single-strand breaks adjacent to each other on opposite DNA strands when guided by two properly designed sgRNAs (13, 67, 90). The paired nickases show higher specificity in editing because the generation of DSBs requires two independent binding events, whereas the nuclelease Cas9 requires only one binding event. A similar strategy is to fuse dCas9 to the dimerizing FokI nuclease. The dCas9-FokI fusion is an RNA-guided nuclease that cleaves DNA only when a pair of FokI domains are in proximity and form a dimer. Studies have shown efficient cleavage when two target sites are spaced approximately 13–25 bp apart (33, 108). Moreover, because the FokI nuclease activity requires dimerization, this strategy also reduced the number of unwanted mutations compared with the Cas9 nickase strategy (33, 108). This is similar to the use of ZFNs and TALENs: Fusing a zinc-finger protein or a TALE protein to the dimerizing FokI enhances specificity in genome engineering (reviewed in 43, 110). However, these approaches improve CRISPR specificity at the cost of reduced efficiency.

Recently, two studies have reported improved specificity with a rationally engineered CRISPR/Cas9 system. Guided by the crystal structure of S. pyogenes Cas9, Slaymaker et al. (101) created systematic single or combined alanine substitutions in the positively charged residues that are predicted to be involved in stabilizing the nontarget strand of the target DNA, and have identified Cas9 variants that decrease off-target indel formation while preserving on-target activity. Using a similar strategy, Kleinstiver et al. (51) carried out an alanine scan at the four residues (N497, R661, Q695, and Q926) in S. pyogenes Cas9 that are predicted, based on the crystal structure, to make direct hydrogen bonds to the phosphate backbone of the target DNA strand. They found that the quadruple alanine substitution variant (spCas9-HF) retains high on-target activity while having minimum off-target activity. Application and further optimization of these high-fidelity Cas9 variants will increase the reliability of CRISPR/Cas9 as both a research tool and a therapeutic approach.

We are only beginning to investigate how best to apply CRISPR/Cas9 technologies to understand human genomics and develop new therapeutic methods. In the past decade, the ability to decode the human genome using high-throughput sequencing has increased significantly. CRISPR/Cas9 has the potential to do the same for large-scale genome engineering, for example, by systematically correcting disease-relevant mutations in primary human cells with high fidelity, specificity, and efficiency. The combination of genome engineering with other methods, such as single-cell DNA/RNA sequencing, epigenomic profiling, and proteomics, will create another horizon for understanding the complex biology in cells and tissues, transforming genomic research and disease treatment.

**DISCLOSURE STATEMENT**

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.
ACKNOWLEDGMENTS

X.X. acknowledges support from the Helen Hay Whitney Foundation. M.C. acknowledges support from Gladstone Institute BioFulcrum. D.Z. acknowledges support from startup funds in the L.S.Q. laboratory. L.S.Q. acknowledges support from the National Institutes of Health Office of the Director (OD). This work was supported by National Institutes of Health grant DP5 OD017887 and U01 EB021240 (to L.S.Q.).

LITERATURE CITED


Contents

The Great Adventure of an American Human Geneticist
    Arno G. Motulsky as told to Mary-Claire King .............................................. 1

CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation
    Matthias Merkenschlager and Elphège P. Nora ............................................. 17

Evolution of Gene Regulation in Humans
    Steven K. Reilly and James P. Noonan ......................................................... 45

The Properties of Long Noncoding RNAs That Regulate Chromatin
    Michael Rutenberg-Schoenberg, Alec N. Sexton, and Matthew D. Simon .................. 69

Advancements in Next-Generation Sequencing
    Shawn E. Levy and Richard M. Myers .......................................................... 95

Association Tests for Rare Variants
    Dan L. Nicolae .................................................................................................. 117

CRISPR/Cas9 for Human Genome Engineering and Disease Research
    Xin Xiong, Meng Chen, Wendell A. Lim, Dehua Zhao, and Lei S. Qi .......................... 131

Cystic Fibrosis and Its Management Through Established and Emerging Therapies
    David R. Spielberg and John P. Clancy ............................................................ 155

Developmental Origins of Common Disease: Epigenetic Contributions to Obesity
    Maya Kappil, Robert O. Wright, and Alison P. Sanders ...................................... 177

Genomic Analysis of the Emergence, Evolution, and Spread of Human Respiratory RNA Viruses
    Tommy T.-Y. Lam, Huachen Zbu, Yi Guan, and Edward C. Holmes ....................... 193

Germline Stem Cell Competition, Mutation Hot Spots, Genetic Disorders, and Older Fathers
    Norman Arnheim and Peter Calabrese ......................................................... 219
Old Dogs, New Tricks: Monogenic Autoinflammatory Disease Unleashed
Monique Stoffels and Daniel L. Kastner .................................................. 245
Recent Advances in Defining the Genetic Basis of Rheumatoid Arthritis
Chikashi Terao, Soumya Raychaudhuri, and Peter K. Gregersen .......... 273
Defining the Clinical Value of a Genomic Diagnosis in the Era of Next-Generation Sequencing
Natasha T. Strande and Jonathan S. Berg .............................................. 303
Host Factors in Ebola Infection
Angela L. Rasmussen .............................................................................. 333
Phenome-Wide Association Studies as a Tool to Advance Precision Medicine
Joshua C. Denny, Lisa Bastarache, and Dan M. Roden ....................... 353
Broad Consent for Genomic Research and Biobanking: Perspectives from Low- and Middle-Income Countries
Paulina Tindana and Jantina de Vries ...................................................... 375
Lessons from HeLa Cells: The Ethics and Policy of Biospecimens
Laura M. Beskow .................................................................................. 395

Errata
An online log of corrections to Annual Review of Genomics and Human Genetics articles may be found at http://www.annualreviews.org/errata/genom