ISSN 0026-8933, Molecular Biology, 2017, Vol. 51, No. 4, pp. 514–525. © Pleiades Publishing, Inc., 2017.
Original Russian Text © A.V. Bannikov, A.V. Lavrov, 2017, published in Molekulyarnaya Biologiya, 2017, Vol. 51, No. 4, pp. 582–594.

REVIEWS

UDC 577.2.08

CRISPR/CAS9, the King of Genome Editing Tools

A. V. Bannikov[a, *] and A. V. Lavrov[a, b]

[a] *Medical Genetic Research Center, Moscow, 115478 Russia*
[b] *Pirogov Russian National Research Medical University, Moscow, 117997 Russia*
**e-mail: broyler3@mail.ru*

Received June 15, 2016; in final form, October 27, 2016

Abstract—The discovery of CRISPR/Cas9 brought a hope for having an efficient, reliable, and readily available tool for genome editing. CRISPR/Cas9 is certainly easy to use, while its efficiency and reliability remain the focus of studies. The review describes the general principles of the organization and function of Cas nucleases and a number of important issues to be considered while planning genome editing experiments with CRISPR/Cas9. The issues include evaluation of the efficiency and specificity for Cas9, sgRNA selection, Cas9 variants designed artificially, and use of homologous recombination and nonhomologous end joining in DNA editing.

Keywords: CRISPR/Cas9, genome editing, sgRNA
DOI: 10.1134/S0026893317040033

INTRODUCTION

Elements of the CRISPR/Cas system were first described in 1987 [1]. The role that the new repetitive sequences play in bacterial DNA remained unknown for a long period of time. In 2007, nucleases of the Cas family and CRISPR elements were shown to play a role in antiviral immunity of bacteria [2]. A similar system was found more recently in Archaea [3]. The studies provided a basis for a large-scale bioinformatics search for similar repeats in the genomes of various species [4]. CRISPR/Cas proved to be relatively widespread, being found in 48% of bacteria and 84% of archaea.

Nucleases are a main tool in gene therapy. The idea of gene therapy is simple. A double-strand break (DSB) is introduced in a necessary genome site and is then repaired via two main mechanisms, nonhomologous end joining (NHEJ) and homology-directed repair (HDR). In NHEJ, DNA ends are ligated so that a deletion or insertion of one or several nucleotides may arise. In HDR, a template is used to copy the DSB locus. The other, intact chromosome or exogenous DNA, such as a plasmid or an oligonucleotide, can serve as a template. Thus, it is possible to correct a pathogenic mutation or to introduce a new sequence in the genome by making a DSB in a certain site and repairing it with the use of exogenous DNA. This is the gist of gene therapy. Specific programmable nucleases have long been employed in genome editing, but Cas9 is superior to all competitors in one important parameter, the cost. A reprogramming of previously known nucleases requires a new amino acid sequence to be designed, and the process is rather laborious and often fails to meet with success. The potentials and advantages of various nucleases have been discussed [5]. To reprogram Cas9, it is only necessary to synthesize a nucleotide sequence (DNA or RNA, depending on the delivery method) of approximately 20 nt, which is possible to do fast, with a high accuracy, and at a low cost.

NATURAL CRISPR/Cas SYSTEMS

All CRISPR/Cas systems are classified into six types, I–VI. The two main components of a system are an operon, which codes for a Cas protein(s), and clustered regularly-interspaced short palindromic repeats (CRISPR), which provide a template for synthesizing CRISPR RNAs (crRNAs) to direct nuclease activity [4]. One or several Cas proteins perform the following main functions in various systems: the adaptor function to bind crRNA, the RNase function to ensure crRNA maturation, and the nuclease function to cleave the target DNA. All Cas proteins belong to ten families, Cas1–10.

The CRISPR locus consists of palindromic repeats and spacers. The spacers arise when the host genome incorporates virus or plasmid genome fragments, which are termed the protospacers until insertion. A spacer-containing crRNA is synthesized from the

Abbreviations: NGS, next-generation sequencing; NHEJ, nonhomologous end joining; HDR, homology-directed repair; DSB, double-strand break; CRISPR, clustered regularly-interspaced short palindromic repeats; PAM, protospacer adjacent motif; crRNA, CRISPR RNA; sgRNA, single guide RNA; tracrRNA. trans-activating crRNA; Cas, CRISPR-associated protein; spCas9, Streptococcus pyogenes Cas9.

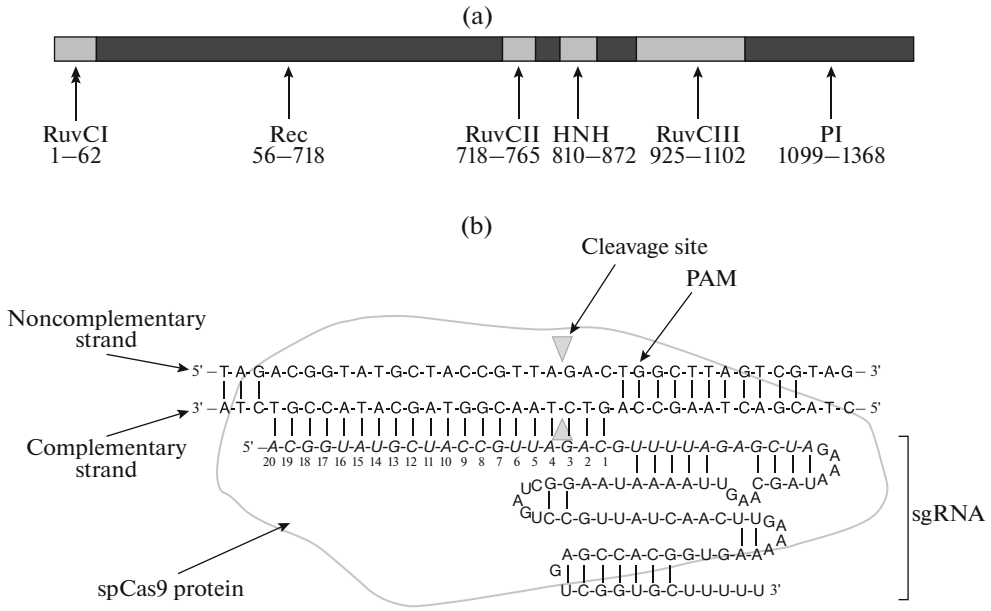

Fig. 1. (a) Domain structure of spCas9 and (b) a scheme of spCas9 binding with target DNA.

locus. Then the crRNA binds directly to the Cas proteins of type I and type III systems or through an accessory tracrRNA to the Cas proteins of type II systems. The specificity of interactions between the Cas–RNA complex and DNA is determined by hybridization of the crRNA spacer with the complementary target DNA. The Cas nuclease is thereby directed to the protospacer present in a virus genome. Depending on the Cas type, a protospacer adjacent motif (PAM) should flank the 5' or 3' end of the protospacer. The PAM is absolutely essential for Cas binding to DNA and usually consists of 3–5 bp. *Streptococcus pyogenes* Cas9 (spCas9) recognizes NGG as a PAM, while *Staphylococcus aureus* Cas9 (saCas9) recognizes NNGRRT (Fig. 1) [6]. PAMs are not fully specific. For instance, NAG is recognized as a PAM by spCas9, although less efficiently than NGG. The PAM is present in virus DNA, but is absent from the CRISPR array, and the bacterial genome is thereby protected from Cas-dependent cleavage. This mechanism is utilized in the type I and type II systems. In the type III systems, spacer DNA is flanked by special repeats, which prevent its cleavage.

Type I and type III systems are similar in mechanism of action and composition. A complex of adaptor proteins is assembled around the crRNA. In type III systems, nuclease (Cas10) binds to the complex immediately. In type I systems, nuclease (Cas3) binds after DNA binding. The spatial structures of the CRISPR/Cas complexes are also very similar between type I and type III systems. The most remarkable feature of the structures is that the Cas7 adaptor protein interacts with the spacer so that several Cas7 molecules polymerize to produce a chain, which binds parallel to the crRNA spacer. Because the crRNA is substantially bent in complex with Cas7, its hybridization to DNA proceeds in a stepwise manner, by approximately 5 nt, as Cas7 loses its contact with the crRNA. This mechanism is thought to be more precise (less sensitive to individual mismatching nucleotides) than simultaneous hybridization [6].

Type II CRISPR/Cas systems substantially differ from type I and type III systems. A type II system includes only one protein, Cas9, which interacts with the crRNA and possesses two nuclease domains, RuvC and HNH (Fig. 1). Two well-separated lobes are distinguishable in the Cas9 structure. One harbors all nuclease domains with the exception of a part of HNH, and the other includes all binding domains. The crRNA binds between the two lobes [7]. The spacer interacts with the arginine-rich α helix that connects the two Cas9 lobes. The interaction is especially strong at the start of the spacer. This circumstance possibly explains why the so-called seed sequence, which includes 4 nt at the 3' end of the spacer, is of special importance for recognizing the DNA nucleotide sequence. A stronger interaction with these nucleotides provides for their more exact positioning in contact with DNA. Hybridization starts with these nucleotides and, probably, proceeds in a stepwise manner, like in the type I and type III systems. A guide RNA (gRNA) of a type II system consists of two separate RNA molecules, crRNA and tracrRNA, which do not bind to Cas9 until hybridizing to each other.

Type IV, type V, and type VI CRISPR/Cas systems were identified relatively recently and are less understood. A type IV system lacks CRISPR and is directed

by a DNA molecule [8]. A type V system utilizes RNA of 41–42 nt to direct nuclease activity; the RNA corresponds to Cas9 crRNA [9]. A type VI system similarly utilizes a short guide RNA, but cleaves RNA rather than DNA [10].

The functional principles of CRISPR immunity and the variants of systems found in various microorganisms have been reviewed in detail [11].

Only Cas9 is now broadly employed in genome editing. The other systems have not been studied as comprehensively, but have certain features that may allow them to equal Cas9 in the future.

MOLECULAR MECHANISMS OF THE Cas9 FUNCTION

Construction of a single guide RNA (sgRNA) was an important achievement in the field of genome editing [12]. The crRNA and tracrRNA components were artificially combined in a single RNA, which was termed the sgRNA and used to direct Cas9 without impairing its activity. The sgRNA consists of approximately 80 nt, including a 20-nt spacer. Thus, a substitution of only 20 nt makes it possible to target Cas9 to a new DNA site.

Standard promoters are commonly used to express Cas9 in eukaryotic cells: the cytomegalovirus (CMV) promoter for the protein and the U6 (RNA polymerase III) promoter for the sgRNA. The genes are transcribed in the nucleus, and the Cas9 mRNA is exported into the cytosol and translated. The sgRNA lacks export signals and consequently remains in the nucleus. Cas9 has a nuclear localization signal (NLS) to determine its transfer into the nucleus. It is known now that the NLS is not essential for the nuclear import of Cas9 [13, 14]. This circumstance is explained by a high positive charge of the nuclease. However, a Cas9 designed to allow its allosteric regulation with 5-hydroxytamoxifen was less active than the wild-type protein in the absence of 5-hydroxytamoxiphen [14]. Lack of the NLS further reduced the activity of allosterically regulated Cas9. It seems that Cas9 nuclear transport becomes limiting only when nuclease activity or concentration is low. Once in the nucleus, Cas9 binds with sgRNA to produce a ribonucleoprotein complex, which is capable of specifically cleaving DNA.

Cas9 finds its binding sites in DNA by colliding with DNA during diffusion. A sliding along DNA has not been observed for Cas9. This is not surprising given that Cas9 highly specifically interacts with DNA and primarily the PAM. However, Cas9 is capable of DNA binding in the absence of a sgRNA, but the binding is nonspecific. The dissociation constant is approximately 25 nM in this case, while decreasing to 0.5 nM in the presence of a sgRNA [15].

In the case of specific interactions, Cas9 first recognizes the PAM. Then Cas9 binds to DNA, the interaction with the PAM stimulates helicase activity, and the DNA helix starts melting. The PAM is most likely necessary for exertion or stimulation of Cas9 helicase activity. There are data that the PAM is not essential when the first two nucleotides of the protospacer always occur in a melted state [15]. The seed sequence plays an important role in this case by stabilizing the interaction of Cas9 with melting DNA.

The efficiency of Cas9–DNA binding depends on how accessible the target site is. The eukaryotic genome varies in the extent of compaction, which affects the accessibility of its sites. Cas9 is most likely capable of initiating chromatin decondensation in vivo [16]. In vitro experiments showed that even nucleosomes cause a six- to eightfold decrease in Cas9 activity [17]. Little is known on the Cas9 interaction with chromatin, but it is safe to say that chromatin condensation affects in fact the efficiency of genome editing. The conclusion is supported by data from a library-on-library approach, which is detailed below (see Selection of the sgRNA).

Ample data indicate that Cas9 extremely tightly binds to target DNA [18, 19]. The lifetime of the Cas9–DNA complex is several hours, and Cas9 does not dissociate from the complex even in the presence of 0.5 M NaCl [15]. Cas9 is sometimes characterized as a single-turnover enzyme. The assumption is supported by the following observation. It takes approximately 1 h for a mammalian cell to eliminate UV-induced damage and approximately 15 h to repair Cas-induced DSBs [20]. A possible explanation is that Cas9 remains associated with DNA for a long period of time after cleavage and thus protects the DNA ends from being recognized by the repair system.

Unfortunately, the specificity of Cas 9 binding to DNA is not absolute. The nuclease can bind not only to its target, but also to the off-target sites that only slightly differ from the target and possess the PAM. The probability for an off-target site to be cleaved depends primarily on the number, nature, and positions of nucleotide substitutions. Statistical data have been collected for the effect of these and other factors on the specificity of DNA binding and cleavage by the Cas9–sgRNA complex. The data are considered below together with sgRNA design.

METHODS TO EVALUATE THE CRISPR/Cas9 SPECIFICITY

As already mentioned, the specificity of CRISPR/Cas9 is not absolute, which is unallowable for certain important applications, such as gene therapy. Hence, one of the primary tasks is to improve the specificity of the system. The issue gave impetus to developing techniques for DSB mapping at the whole-genome scale. DSB mapping techniques are indispensable for evaluating the specificity of a nuclease because they provide an objective estimate for its off-target activity. Several methods were developed for

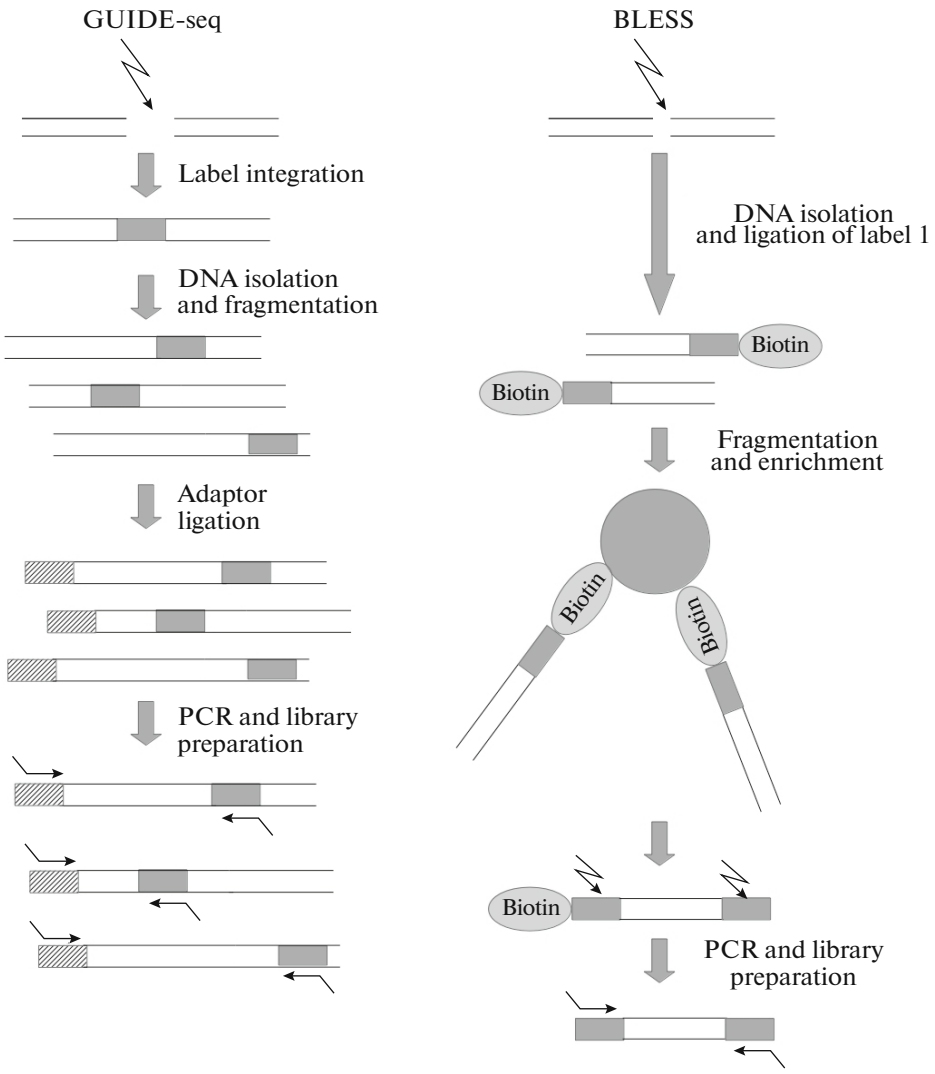

Fig. 2. Flow charts of GUIDE-seq and BLESS.

DSB mapping, each being employed in only few studies. Two methods, GUIDE-seq and BLESS, seem to be the most promising.

Genome-wide Unbiased Identification of DSBs Enabled by Sequencing (GUIDE-seq)

To perform GUIDE-seq (Fig. 2) [21], cells are grown in the presence of phosphorothioate-protected short double-stranded oligodeoxyribonucleotides (dsODNs), which are incorporated in sites of Cas9-induced DSBs during NHEJ repair. The label (dsODN) may harbor a restriction site so that its incorporation is possible to assay by a restriction fragment length polymorphism analysis. The dsODNs are used for sequencing and DSB mapping. To obtain a DNA library for next-generation sequencing (NGS), genomic DNA is isolated and digested into fragments of 500 bp on average. An adaptor is then ligated to the fragments at one end. Amplification is carried out with primers directed to the adaptor and the label, and the resulting library consequently includes only DSB-flanking regions. The sequence prevalence in the library depends on the label integration efficiency. An analysis of the GUIDE-seq results [22] showed that the coefficient of correlation between the number of reads and the efficiency of label integration in all sites of five genes was only 0.43. When sites of only one gene, *FANCF*, were analyzed, the coefficient of correlation was far higher, approximately 0.9. It seems that other factors act together with the label integration efficiency to affect the prevalence of a sequence in the library. PCR is one of the most likely factors. The dependence of PCR efficiency on the GC content has long been discussed in the context of library preparation for NGS.

It is clear that GUIDE-seq is a sensitive method because label can be integrated at any time while DSB exists. However, the number of reads in GUIDE-seq

results does not directly report how likely a DSB is to arise in a given site. GUIDE-seq results obtained for one site (with one sgRNA) can be used to quantitatively compare the activities of different nucleases, for instance, different Cas9 mutants. However, a direct quantitative comparison of two different sites is impossible.

Direct In Situ Break Labeling, Enrichment on Streptavidin, and Next-Generation Sequencing (BLESS)

In BLESS (Fig. 2), DSBs are labeled after chromatin extraction [23]. A biotinylated label makes it possible to enrich the preparation in DSB-containing sequences. Then DNA is digested, and the fragments are ligated with a second label. Subsequent PCR is carried out with primers directed to the two labels.

BLESS is far less sensitive than GUIDE-seq because only the DSBs that exist at the time of DNA isolation are possible to map. Rare DSBs will most likely be repaired and will thereby escape detection in the BLESS results. Currently, the sensitivities of the two methods can be compared for two sgRNAs. GUIDE-seq detects more off-target sites than BLESS with either sgRNA: 8 vs. 4 with one and 32 vs. 10 with the other [22, 24].

Other Methods

High-throughput genome-wide translocation sequencing (HTGTS) [25] and its variant linear amplification-mediated HTGTS (LAM-HTGTS) are other high-throughput methods to evaluate DSBs. First, the so-called constant restriction site, which is always cleaved in vivo, is introduced in the cell genome. There is a chance that DSB repair in any genomic site ligates the DSB ends with the ends of the constant DSB, resulting in a translocation. Given that the sequence is known for the constant DSB locus, any DSB arising in the genome can be mapped via NGS. Because translocations are approximately 200–1000 times less frequent than deletions and insertions resulting from NHEJ, a large amount of DNA is necessary for LAM-HTGTS. Moreover, HTGTS was initially designed to analyze the DSB formation in the immunoglobulin genes with a constant DSB site inserted in their vicinity, which is of importance because intrachromosomal translocations are more likely to arise than interchromosomal ones. Because of their low sensitivity, translocation-based methods to estimate nuclease activity are replaced by their competitors, such as GUIDE-seq.

The T7 and Surveyor nuclease activity assays also deserve attention [26]. The methods are based on using the nucleases that cleave DNA in mismatch sites. Cells transfected with a plasmid expressing Cas9 and a corresponding sgRNA are used to isolate DNA. The editing region is amplified in conventional PCR. Some of the copies have mutations as a result of NHEJ, while the others are wild type. When the PCR product is rehybridized, hybridization of mutant strands with wild-type strands is approximately as likely as with their counterparts. The resulting duplexes contain mismatches in mutation sites and, therefore, are cleaved upon digestion with the T7 or Surveyor nuclease. The fragments are then examined. The maximum number of cleaved duplexes has been obtained at a 50% efficiency of editing. The T7 and Surveyor assays are relatively inexpensive and have consequently found broad application. Yet their potential is limited in that only known sites can be tested for DSBs and that the lower sensitivity threshold is several percent, thus being inferior to NGS.

Cas9 ENGINEERING

The primary objective of Cas9 engineering is to reduce off-target activity of Cas9. Gene-engineering Cas9 variants are possible to classify into two groups. One includes mutant proteins obtained by introducing one or more amino acid substitutions. The other includes split Cas9 proteins. That is, Cas9 is expressed as two peptides, which are then joined to form one nuclease.

Mutant Proteins

Several Cas9 mutants have been obtained and characterized to date in order to reduce off-target activity [22, 24].

Helicase activity of Cas9 has been modified for the purpose [24]. When DNA is melted, both the sgRNA and the helicase domain of Cas9 can stabilize the single-stranded DNA state. A decrease in helicase activity increases the relative contribution of the sgRNA to stabilizing ssDNA, thus presumably reducing the Cas9 sensitivity to incomplete complementarity between the sgRNA and target DNA. A total of 32 amino acid residues were mutated in various combinations, and a higher specificity was observed for two variants, K810A/K1003A/R1060A and K848A/K1003A/R1060A [24]. BLESS was used to estimate the Cas9 activity in the study.

Elimination of nonspecific contacts should weaken the interaction with off-target sites to a far greater extent than with target sites. A study has been performed to reduce the nonspecific electrostatic interactions of Cas9 with DNA [22]. Four Cas9 variants were constructed and examined: HF1 (N497A/R661A/Q695A/Q926A), HF2 (D1135E/N497A/R661A/Q695A/Q926A), HF3 (L169A/N497A/R661A/Q695A/Q926A), and HF4 (Y450A/N497A/R661A/Q695A/Q926A) [22]. A crystal structure analysis of Cas9 made it possible to choose the amino acid residues that were then mutated to obtain HF1. Each of the other variants had one additional mutation characterized earlier. D1135 most likely interacts with the PAM [27]; a decrease in

affinity for the PAM reduces off-target activity and the seed sequence comes to play a greater role in binding to DNA. L169 is involved in Van der Waals interactions with DNA [10] and presumably allows the enzyme to distinguish DNA and RNA. Y450 is displaced by 120° upon DNA binding [27] and is most likely involved in hydrophobic interactions. Cas9 activity was assayed by GUIDE-seq.

The potential of Cas9 mutagenesis is certainly far from being exhausted. A combined effect is necessary to check for the mutations obtained in the above two studies. Second, specificity estimates obtained for a new mutant nuclease depend on the methods employed, which have not been firmly established as of yet. The above studies provide an illustrative example, one employing BLESS and the other, GUIDE-seq. Either method clearly demonstrates that a mutant nuclease works better than the wild-type enzyme, but GUIDE-seq better reflects the absolute nuclease specificity owing to its higher sensitivity. GUIDE-seq results are more reliable to use when choosing the nuclease for gene therapy, although certain problems are possible in this case as well.

Cas9 is possible to convert to a nickase by inactivating one of its nuclease domains [28]. Only the use of two nickases, one with inactivated HNH and the other with inactivated RuvC, may produce a DSB. Off-target activity is lower with nickases because a simultaneous binding of two proteins is required for a DSB to arise. However, the nickase system is more difficult to design and is consequently of limited use.

Cas9 variants that utilize other PAMs may also be useful, allowing more freedom in choosing the target site. For instance, directed evolution was used to obtain three Cas9 variants: D1135V/R1335Q/T1337R for the NGAN PAM; D1135E/R1335Q/T1337R for the NGAG PAM; and D1135V/G1218R/R1335E/T1337R for the NGCG PAM [29]. Such enzymes may provide the only solution when DSBs are necessary to introduce in sites inaccessible to other Cas9 variants.

Split Cas9

Wright et al. [30] have split Cas9 into the nuclease and binding lobes, which were expressed separately and joined via noncovalent interactions. In vitro experiments showed that split Cas9 was approximately 10 times slower than the wild-type enzyme, while its maximum activity was much the same. In vivo, split Cas9 had a far lower efficiency (0.6 and 2% vs. 22 and 34% in two experiments), which was explained by a potential inequality in expression of the two lobes and relatively high expression of the sgRNA, which might titrate the lobes apart from each other when in an excess. The hypothesis is supported by the fact that an increase in sgRNA concentration decreased the Cas9 efficiency.

Ligand-dependent dimerization was used to improve the association of the two lobes of split Cas9. For this purpose, one Cas9 lobe was fused with FKBP (FK506 binding protein 12) and the other, with the FKBP-binding domain [31]. Dimerization of the two domains occurs in the presence of rapamycin. Thus, an inducible split Cas9 enzyme was obtained. However, activity of the noninduced system was still high, approximately one-third of induced activity. The C lobe of Cas9 was then supplemented with two NLSs; and the N lobe, with a nuclear export signal (NES). The lobes are transferred into different cell compartments after translation. When the Cas9 lobes dimerize in the cytosol, the two NLSs ensure the enzyme transfer into the nucleus in spite of the presence of the NES. Noninduced activity was not observed for the construct. Activity of dimerized Cas9 was approximately half that of the wild-type enzyme (95 and 43% on the target site), and off-target activity was absent. It is of interest that rapamycin-dependent activation was stable; i.e., a single 2-h exposure had the same effect as continuous activation.

The two Cas9 lobes were joined via inteins, which are short amino acid sequences that possess autocatalytic peptidase activity [32]. An intein consists of two, C- and N-terminal, parts, which may belong to one (*cis*-inteins) or different (*trans*-inteins) protein chains. *Cis*-inteins do not need any cofactor to excise themselves from their host proteins. The halves of a *trans*-intein first covalently bind with each other and then catalyze self-excision like *cis*-inteins.

The Cas9 enzyme assembled with the use of inteins had the same efficiency as the wild-type enzyme [33]. The intein-containing Cas9 was delivered into the cell with an adeno-associated virus (AAV). AAVs are currently considered to be the safest vectors, but their packaging capacity is limited to approximately 4.7 kb. The full-size spCas9 system is slightly greater than the limit. This circumstance complicates its use and, in particular, restricts the choice of regulatory sequences. The intein-containing Cas9 solves the packaging problem because two virus particles can be used to deliver the system.

New Cas9 Enzymes

Delivery of Cas9 into the cell is an important problem. The Cas9 size is inconvenient for delivering Cas9 with an efficient and safe AAV, and shorter variants of the enzyme are consequently sought. The most promising variant is *S. aureus* Cas9 (saCas9). The saCas9 cDNA is only 3252 bp, and saCas9 functions with a higher precision than spCas9 according to the available data. Based on off-target activity evaluated by BLESS, saCas9 is more precise than wild-type spCas9 [34] and less precise than highly specific mutant spCas9(1.1) [24]. GUIDE-seq was used to evaluate the specificity of editing at a target site in *VEGFA* for wild-type spCas9, mutant spCas9, and saCas9 [22, 35].

The results showed eight off-target sites for saCas9 and more than 30 for spCas9; i.e., saCas9 has a higher specificity.

Thus, saCas9 may have other advantages in addition to a shorter length. However, published data are not ample enough to state this for certain.

CHOICE OF sgRNA

The sgRNA nucleotide sequence is known to affect both the efficiency of target site cleavage and off-target activity. Statistical data were collected to predict the sgRNA efficiency from the nucleotide sequence and to characterize the effects of the number, nature, and positions of mismatches on the sgRNA efficiency, thus predicting off-target activity.

In all studies, the sgRNA efficiency is inferred from the efficiency of knocking out the sgRNA target gene. A knockout arises when DSB is repaired via NHEJ, which is often associated with a loss of several nucleotides and far less frequently with an insertion. Frame shifting by 1, 2, or 3 bp is equally possible; i.e., a frameshift mutation leading to a gene knockout occurs in approximately two-thirds of all cases [36]. In such experiments, cells were treated with a DNA library coding for several thousand or several tens of thousands of sgRNAs, which are targeted to the genes whose knockouts are easy to detect. The set may include surface receptor genes, which are possible to test for expression by flow cytometry, or vital genes, whose knockout efficiency is directly proportional to the cell viability. Cells are treated with a DNA library so that only one sgRNA is expressed in a cell on average. For instance, to identify the genes involved in mismatch repair (MMR), tests were performed with 73300 sgRNAs targeting 7330 genes, at 10 sgRNAs per gene [37]. Cells transfected using the CRISPR/Cas9 system were cultured in the presence of 6-thioguanine, which was used in a concentration that results in cell death because repair proteins recognize 6-thioguanine and arrest the cell cycle. The sgRNAs that target the MMR genes distort the cell-cycle control, and cells continue proliferating in the presence of 6-thioguanine. A NGS-based analysis of the surviving cells identified the sgRNAs and, therefore, the genes whose knockout turns off the MMR pathway. The efficiency was compared for different sgRNAs targeting one gene, and specific sgRNA motifs were thereby associated with sgRNA efficiency. The data are used to study the nucleotide sequence dependence of sgRNA efficiency.

The sgRNA efficiency always depends on several factors, including the sgRNA stability within the cell, affinity for Cas9, and nucleotide sequence. The roles of these factors were analyzed in detail and separated in several studies. In one of them, the Cas9 mRNA and sgRNA were injected in zebrafish embryos [36]. Injections of sgRNA alone were used to estimate its intracellular stability, and injections of both sgRNA and the Cas9 mRNA served to estimate sgRNA affinity for the nuclease (with due regard to the stability). It was observed that the higher the G content, the more stable is the sgRNA. The dependence is explained by the formation of quadruplex structures, which was verified experimentally. In contrast, sgRNAs with lower G contents were found to have higher affinity for Cas9. In spite of this, the nucleotide composition of efficient sgRNAs was similar to that of stable sgRNAs, i.e., there is an antagonistic relationship between stability and affinity, a high and low G content, but stability is a more important factor. Similar results were reported from other studies [38]. All active sgRNAs had a GC content higher than 50%. It should be noted that the method employed in delivering the CRISPR/Cas9 components is likely to affect the results. For instance, the sgRNA stability is far more important when sgRNA is injected in the cell than in the case of its continuous expression. This results in a dominance of the corresponding characteristics of efficient sgRNAs.

The accessibility of the target site, that is, the extent of its DNA compaction, is another important factor. The so-called library-on-library method was used to eliminate the chromatin organization factor [39]. The idea of the method is quite simple. Only sgRNA library is used in a conventional screening, while target sites are in the genome. The library-on-library method additionally utilizes a library of target sites. Cells are transfected not only with plasmids carrying the components of the Cas9 system, but also with plasmids carrying target sites. The two libraries are delivered into cells with lentivirus vectors, which are integrated in highly accessible chromatin regions. The effect of epigenetic factors on sgRNA activity is excluded by using an integrated target site to evaluate the activity and is possible to estimate by comparing the activities between the integrated and genomic sites. The available data indicate that epigenetic factors reduce Cas9 activity; the finding is supported by the correlation between sgRNA activity and DNase I sensitivity. However, an association between DNA chromatization and sgRNA activity was not uniformly observed in all studies [38].

Another difficulty arises because 1-, 2-, and 3-bp frame shifts are not always similarly likely [40]. A bias towards one or another repair outcome was observed depending on the DSB position in the genome [36, 41]. For instance, if a site is repaired to produce mostly a 3-bp deletion, the gene knockout efficiency and, therefore, the apparent sgRNA efficiency will be low. It remains unknown how strongly this circumstance affects sgRNA activity measurements.

The sgRNA Designer Program

Special programs were developed to select sgRNAs with regard to various parameters that affect both target and off-target sgRNA activities. The programs

available for designing sgRNA have been reviewed in [42]. As an example, we consider in detail the data collection for the sgRNA Designer program (Broad Institute, United States; http://www.broadinstitute.org/rnai/public/analysis-tools/sgrna-design), which is one of the most common tools used for the purpose.

The first study to collect experimental data for developing sgRNA Designer analyzed target activity as dependent on the nucleotide sequence for 1841 sgRNAs directed to surface receptor genes [43]. Target gene expression was assayed by flow cytometry. The data were used to create the so-called rule set no. 1 for sgRNA design. The rules were updated to include additional factors more recently [44]. A Gini index was developed to indicate the importance of a particular factor for sgRNA activity: the nature of nucleotides and dinucleotides depending on their positions (58%), the total number of nucleotides and dinucleotides (16%), the sgRNA position in a protein-coding gene (13%), and melting temperature (11%). To evaluate off-target activity, the effect of mismatches on sgRNA activity was studied and summarized in the cutting frequency determination (CFD) index. Statistical data for CFDs were collected using 65 sgRNA that targeted *CD33* with activities verified experimentally. A total of 9914 sgRNAs were obtained by introducing all possible modifications, including single nucleotide substitutions, deletions, and insertions. The effect of a substitution in a given position was determined as an average decrease in activity across all sgRNAs carrying the given substitution. For instance, to estimate the effect of the G>T substitution in position 5, averaging was performed across all sgRNAs with the substitution regardless of all other substitutions, if any. To evaluate the efficiency of CDF, cross-validation was performed with 89 sgRNAs by calculating the coefficient of correlation between the CDF of a particular sgRNA and its activity measured experimentally. The coefficient of correlation was 0.572 for sgRNAs with one substitution, 0.512 for sgRNAs with two substitutions, and 0.164 with three or more substitutions.

Other studies confirmed the results obtained with sgRNA Designer. Three sgRNA sets were examined, including 2073 sgRNAs against 58 human ribosomal genes, 1667 sgRNAs against 163 human nonribosomal genes, and 1443 sgRNAs against mouse nonribosomal genes [45]. With all three sets, a good agreement was obtained only for the seed sequence region, where G is preferable to occur in positions 1 and 2, C is preferable to occur in position 3, and T is undesirable in the four first positions of the spacer.

The above three studies [43–45] provide the most ample sets of statistical data on sgRNA activity. The results are used in many online sgRNA design tools, such as CHOPCHOP [46] and WU-Crispr [47].

It is impossible to define stringent rules for sgRNA design on the basis of the available data. The existing programs are probably suitable for designing a large library, while choosing from a few sgRNAs is a more common problem. Experimental testing is the best solution in such cases.

NHEJ or HDR?

Two main mechanisms, NHEJ and HDR, are known to repair Cas9-induced DSBs in the cell. NHEJ is often accompanied by deletions and insertions, which may cause a frameshift and, consequently, a gene knockout when occurring in the coding region. Zhang and colleagues [48] developed protocols to construct modified cell lines by using CRISPR/Cas. HDR utilizes a template, which is copied in the course of repair. Plasmid DNA with proper homology regions (approximately 1000 bp at each DSB flank) or a shorter ssDNA of no more than 120 nt is possible to use as a template in gene therapy. Thus, HDR makes it possible to introduce an insert and to modify the genome sequence. HDR is basically more difficult to employ than NHEJ. HDR of DSBs is far less likely in vivo. The problem of enhancing HDR is still poorly understood, and the efficiency of genome editing with HDR is usually estimated at few percent.

Various methods are used to examine NHEJ and HDR in estimating the Cas9 efficiency. The most common one is a fluorescence analysis with HDR and/or NHEJ restoring the function of a fluorescent protein, such as the traffic light reporter (TLR), which includes two proteins, mCherry and GFP. The mCherry gene carries a frameshift mutation and is consequently reparable by NHEJ. Thus, the content (%) of mCherry+ cells is approximately one-third of all cells where NHEJ has been employed in repair. The GFP gene carries a deletion and is therefore reparable only by HDR.

A digital PCR-based method was proposed for evaluating the NHEJ and HDR efficiencies [49]. Two probes are used in the method. One serves to detect NHEJ event; this probe carries fluorophore 1, covers the DSB region, and is fully complementary to it. The other probe serves to detect HDR events; the probe is directed to the repair region, is complementary to the repaired sequence, and carries fluorophore 2. A third probe anneals outside of the editing region and is used to score the sites, carrying fluorophore 1. The NHEJ probe does not bind to sequences with deletions or insertions; i.e., a decrease in fluorescence from fluorophore 1 is a measure of NHEJ, and an increase in fluorescence from fluorophore 2 is a measure of HDR.

Transfected cell selection also deserves attention. Cells are usually transfected simultaneously with Cas9 and a selective marker (antibiotic resistance gene), and only Cas9-expressing cells are used in further experiments [50].

Because a low efficiency of HDR prevents construction of stably transfected cell lines, NHEJ was proposed for insertion of large genes [51]. A linearized

plasmid with blunt ends serves as a donor in this case. The plasmid is inserted into the DSB via NHEJ. The efficiency of HDR-mediated *GFP* insertion ranged from 1.5 to 6% in eight cell lines. NHEJ increased the efficiency to 15–20%. The efficiency of inserting larger fragments of 12 and 34 kb was 7.49 and 1.18%, respectively.

Two approaches have been used to date to improve the HDR efficiency. One is silencing the genes for NHEJ proteins, and the other is synchronizing Cas9 expression with the cell cycle. The Ku70 protein is thought to play a main role in initiating NHEJ by binding to the DNA ends and preventing their binding with factors necessary for HDR. The genes for Ku70 and DNA ligase IV (*LIG4*), which joins the ends in NHEJ, were silenced using RNA interference, the low-molecular-weight inhibitor SCR7 [50], and protein inhibitors. Simultaneous inhibition of Ku70 and LIG4 caused a substantial, five- to sevenfold, increase in HDR, from 5 to 25–36%. The SCR7 effect was not confirmed [52]. However, SCR7 used together with injection of Cas components into rat zygotes increased the HDR efficiency approximately twice [53].

The other approach takes advantage of the fact that the repair mode depends on the cell-cycle phase. HDR occurs in S and G2, while NHEJ activity remains the same throughout the cell cycle. If Cas9 expression is restricted to S and G2, HDR may have a greater chance to occur as compared with NHEJ. Many cell cycle-dependent proteins possess degradation (ubiquitination) signals, which determine their hydrolysis once the cell enters a certain phase of the cell cycle. Geminin is one of the best-studied proteins of the set. The first 110 amino acid residues of geminin contain a signal for ubiquitination in G1. Fusion of Cas9 with the fragment should cause its degradation in G1. In fact, the chimeric Cas9-Gemi protein with the geminin fragment increased the HDR rate by a factor of 1.28–1.72 depending on the transfected plasmid amount. This is probably explained by the fact that Cas9-Gemi is expressed to a lower level, and the amount of the plasmid carrying its gene may therefore act as a limiting factor [52].

The HDR efficiency is possibly to modulate by changing the Cas9 level. Both target and off-target activities of Cas9 were shown to linearly depend on the amount of the plasmid transfected into the cell [54]. Thus, the specificity can be increased by decreasing the total Cas9 level, at the sacrifice of target activity.

It is important to note, first, that the repair mechanism (NHER/HDR ratio and the lengths of deletions and insertions) depends on the locus and cell type. Target sites are possible to characterize with an average length of deletions and insertions and an average HDR efficiency. The parameters differ even when sites are only a few nucleotides apart. Second, the Cas9 efficiency depends on the cell type; for instance, the HDR efficiency is the lowest in pluripotent stem cells [55, 56]. Third, there is no correlation between NHEJ and HDR [49, 51], and this circumstance should be taken into account when estimating Cas9 activity.

SELECTION OF DONOR DNA

The HDR efficiency depends on the nature of donor DNA, which may be a plasmid or an oligonucleotide. Single-stranded oligodeoxyribonucleotides (ssODNs) are known to be the most efficient. Studies of the mechanism of action for spCas9 showed that spCas9 dissociates symmetrically from the two DNA strands, but that the nontarget strand (containing the NGG) remains relatively motile and capable of "breathing" while spCas9 is on DNA [19]. The finding gave ground for assuming that ssODNs complementary to the nontarget strand may increase the HDR efficiency. The assumption was confirmed experimentally; i.e., proper ssODNs increased the HDR rate by a factor of up to 2.6.

The ssODN length is another important parameter and should not exceed 100–120 nt because longer ssODNs are cytotoxic [56]. The intracellular stability of ssODN also plays a certain role. In some cases, phosphorothioate-protected ssODNs (PS-ssODNs) provide for a higher editing efficiency as compared with nonprotected ones (PO-ssODNs) [57]. For instance, a dramatic difference in efficiency was observed between PS-ssODNs and PO-ssODNs (62 and 9.5%, respectively) in editing of the mouse zygotic genome. Yet no difference in efficiency was observed for the two ssODN types when another gene was edited in the rat zygote. The discrepancy was attributed to the insertion length, which was 40 nt in the mouse genome and 1 nt in the rat genome. An analysis of the ssODNs tested makes it possible to conclude that a shortening of the PAM-adjacent arm in a ssODN exerts a greater effect on the ssODN efficiency and that the total ssODN length is more important when the ssODN arms are rather long (at least 20 nt). The conclusion is supported by the data that the highest efficiency is characteristic of the ssODNs that have arms of 91 and 36 nt (PAM-adjacent arm) as measured from the cleavage site [19].

Use of dsODNs only slightly reduces the repair efficiency, but is accompanied by additional mutations in the majority of cases [57].

EDITING OF NONCODING SEQUENCES

A special problem is to edit noncoding sequences, which can be classified as transcribed (microRNAs (miRNAs) and long noncoding RNAs (lncRNAs)) and nontranscribed (regulatory). It is of interest to achieve a knockout of RNA genes in the former case and parts of the regulatory sequences in the latter in order to investigate their functional significance.

A knockout in microRNA is difficult to obtain because microRNAs are small in size and, consequently, suitable sites are few. For instance, three sgRNAs were targeted to RNA in order to achieve a miR-21 knockout, but miR-21 knockout cell lines were not obtained without selection [58]. Another attempt to obtain a knockout was performed with triploid cells and HDR, and a cassette with the puromycin resistance and GFP genes was inserted preliminarily in the gene of interest. Selection yielded five clones of cells with a complete miR-21 knockout. It was found that the cassette was integrated in one chromosome and that the other two miR-21 copies were inactivated by NHEJ.

A lncRNA knockout is difficult to obtain for opposite reasons, because lncRNAs are often comparable with proteins in molecular weight, but deletions or insertions usually fail to activate their genes because there is no reading frame. The problem of lncRNA knockouts was solved by deleting a large gene fragment with the use of two sgRNAs [58].

Simple Cas9-mediated editing allows a new level for the functional analysis of regulatory noncoding sequences. A main difficulty here is that the sequences are very long. A functional screening for an intronic enhancer was described for *BCL11A* [59]. For this purpose, 212, 174, and 147 sgRNAs were designed to target, respectively, three DNase-hypersensitive sites, which were 1284, 1264, and 1370 bp in length and contained a sufficient number of NGG sites. Only few sgRNAs proved to appreciably affect the enhancer activity. The accuracy and completeness of the analysis are difficult to evaluate.

CONCLUSIONS

Two main areas of application are possible to identify for CRISPR/Cas9, basic research and medicine (gene therapy). CRISPR/Cas9 has many research applications, from gene knockouts to antibody engineering and chromatin mapping. However, the greatest hope is associated with using Cas9 as a reliable tool for gene therapy. There are two obstacles to this application, off-target activity and a low HDR efficiency. Substantial achievements have been made to overcome the former problem. New Cas9 mutants have far lower off-target activities. The other problem is of a basic nature and may take a long while to solve. The question of whether Cas9 will provide the gold standard of gene therapy is still open.

REFERENCES

1. Ishino Y., Shinagawa H., Makino K., Amemura M., Nakata A. 1987. Nucleotide sequence of the *iop* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433.
2. Barrangou R., Fremaux C., Deveau H., Richards M., Boyaval P., Moineau S., Romero D.A., Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* **315**, 1709–1712.
3. Lander E.S. 2016. The Heroes of CRISPR. *Cell.* **164**, 18–28.
4. Mojica F.J., Díez-Villaseñor C., García-Martínez J., Soria E. 2005. Intervening sequences of regularly interspaced prokatyotic repeats derive from foreign generic elements. *J. Mol. Evol.* **60**, 174–182.
5. Vasil'eva E.A., Melino D., Barlev N.A. 2015. Application of the CRISPR/CAS genome editing system to pluripotent stem cells. *Tsitologiya.* **57**, 19–30.
6. Tsui T.K., Li H. 2015. Structure principles of CRISPR-Cas surveillance and effector complexes. *Annu. Rev. Biophys.* **44**, 229–255.
7. Nishimasu H., Ran F.A., Hsu P.D., Konermann S., Shehata S.I., Dohmae N., Ishitani R., Zhang F., Nureki O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell.* **156**, 935–949.
8. Koonin E.V., Krupovic M. 2015. Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat. Rev. Genet.* **16**, 184–192.
9. Fagerlund R.D., Staals R.H., Fineran P.C. 2015. The Cpf1 CRISPR-Cas protein expands genome-editing tools. *Genome Biol.* **16**, 251.
10. Shmakov S., Abudayyeh O.O., Makarova K.S., Wolf Y.I., Gootenberg J.S., Semenova E., Minakhin L., Joung J., Konermann S., Severinov K., Zhang F., Koonin E.V. 2015. Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol. Cell.* **60**, 385–397.
11. Pougach K.S., Lopatina A.V., Severinov K.V. 2012. CRISPR adaptive immunity systems of prokaryotes. *Mol. Biol.* (Moscow). **46** (2), 172–182.
12. Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* **337**, 816–821.
13. Kouranova E., Forbes K., Zhao G., Warren J., Bartels A., Wu Y., Cui X. 2016. CRISPRs for optimal targeting: Delivery of CRISPR components as DNA, RNA and protein into cultured cells and single-cell embryos. *Hum. Gene Ther.* **27**, 464–475.
14. Oakes B.L., Nadler D.C., Flamholz A., Fellmann C., Staahl B.T., Doudna J.A., Savage D.F. 2016. Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat. Biotechnol.* **34**, 646–651.
15. Sternberg S.H., Redding S., Jinek M., Greene E.C., Doudna J.A. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature.* **507**, 62–67.
16. Barkal A.A., Srinivasan S., Hashimoto T., Gifford D.K., Sherwood R. 2016. Cas9 functionally opens chromatin. *PLoS ONE.* **11** (3), e0152683. https://doi.org/10.1371/journal.pone.0152683.
17. Hinz J.M., Laughery M.F., Wyrick J.J. 2015. Nucleosomes inhibit Cas9 endonuclease activity in vitro. *Biochemistry.* **54**, 7063–7066.
18. Knight S.C., Xie L., Deng W., Guglielmi B., Witkowsky L.B., Bosanac L., Zhang E.T., El Beheiry M., Masson J.B., Dahan M., Liu Z., Doudna J.A., Tjian R. 2015. Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science.* **350**, 823–826.

19. Richardson C.D., Ray G.J., DeWitt M.A., Curie G.L., Corn J.E. 2016. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344.

20. Kim S., Kim D., Cho S.W., Kim J., Kim J.S. 2014. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012–1019.

21. Tsai S.Q., Zheng Z., Nguyen N.T., Liebers M., Topkar V.V., Thapar V., Wyvekens N., Khayter C., Iafrate A.J., Le L.P., Aryee M.J., Joung J.K. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197.

22. Kleinstiver B.P., Pattanayak V., Prew M.S., Tsai S.Q., Nguyen N.T., Zheng Z., Joung J.K. 2016. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature.* **529**, 490–495.

23. Hu J., Meyers R.M., Dong J., Panchakshari R.A., Alt F.W., Frock R.L. 2016. Detecting DNA double-stranded breaks in mammalian genomes by linear amplification–mediated high-throughput genome-wide translocation sequencing. *Nat. Protoc.* **11**, 853–871.

24. Slaymaker I.M., Gao L., Zetsche B., Scott D.A., Yan W.X., Zhang F. 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science.* **351**, 84–88.

25. Hu J., Meyers R.M., Dong J., Panchakshari R.A., Alt F.W., Frock R.L. 2016. Detecting DNA double-stranded breaks in mammalian genomes by linear amplification–mediated high-throughput genome-wide translocation sequencing. *Nat. Protoc.* **11**, 853–871.

26. Vouillot L., Thélie A., Pollet N. 2015. Comparison of T7E1 and Surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *Genes, Genomes, Genetics.* **5**, 407–415.

27. Jiang F., Zhou K., Ma L., Gressel S., Doudna J.A. 2015. A Cas9–guide RNA complex reorganized for target DNA recognition. *Science.* **348**, 1477–1481.

28. Ran F.A., Hsu P.D., Lin C.Y., Gootenberg J.S., Konermann S., Trevino A.E., Scott D.A., Inoue A., Matoba S., Zhang Y., Zhang F. 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell.* **154**, 1380–1389.

29. Kleinstiver B.P., Prew M.S., Tsai S.Q., Topkar V.V., Nguyen N.T., Zheng Z., Gonzales A.P., Li Z., Peterson R.T., Yeh J.R., Aryee M.J., Joung J.K. 2015. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature.* **523**, 481–485.

30. Wright A.V., Sternberg S.H., Taylor D.W., Staahl B.T., Bardales J.A., Kornfeld J.E., Doudna J.A. 2015. Rational design of a split-Cas9 enzyme complex. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 2984–2989.

31. Zetsche B., Volz S.E., Zhang F. 2015. A split Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33**, 139–142.

32. Elleuche S., Pöggeler S. 2010. Inteins, valuable genetic elements in molecular biology and biotechnology. *Appl. Microbiol. Biotechnol.* **87**, 479–489.

33. Truong D.J., Kühner K., Kühn R., Werfel S., Engelhardt S., Wurst W., Ortiz O. 2015. Development of an intein-mediated split–Cas9 system for gene therapy. *Nucleic Acids Res.* **43**, 6450–6458.

34. Ran F.A., Cong L., Yan W.X., Scott D.A, Gootenberg J.S., Kriz A.J., Zetsche B., Shalem O., Wu X., Makarova K.S., Koonin E.V., Sharp P.A., Zhang F. 2015. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature.* **520**, 186–191.

35. Friedland A.E., Baral R., Singhal P., Loveluck K., Shen S., Sanchez M., Marco E., Gotta G.M., Maeder M.L., Kennedy E.M., Kornepati A.V., Sousa A., Collins M.A., Jayaram H., Cullen B.R., Bumcrot D. 2015. Characterization of *Staphylococcus aureus* Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome Biol.* **16**, 257.

36. Moreno-Mateos M.A., Vejnar C.E., Beaudoin J.D., Fernandez J.P., Mis E.K., Khokha M.K., Giraldez A.J. 2015. CRISPRscan: Designing highly efficient sgRNAs for CRISPR-Cas9 targeting *in vivo. Nat. Methods.* **12**, 982–988.

37. Wang T., Wei J.J., Sabatini D.M., Lander E.S. 2014. Genetic screens in human cells using the CRISPR/Cas9 system. *Science.* **343**, 80–84.

38. Gagnon J.A., Valen E., Thyme S.B., Huang P., Akhmetova L., Pauli A., Montague T.G., Zimmerman S., Richter C., Schier A.F. 2014. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS ONE.* **9**, e98186.

39. Chari R., Mali P., Moosburner M., Church G.M. 2015. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat. Methods.* **12**, 823–826.

40. Koike-Yusa H., Li Y., Tan E.P., Velasco-Herrera Mdel C., Yusa K. 2014. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273.

41. Wang X., Raghavan A., Chen T., Qiao L., Zhang Y., Ding Q., Musunuru K. 2016. CRISPR-Cas9 targeting of PCSK9 in human hepatocytes in vivo. *Arterioscler. Thromb. Vasc. Biol.* **36**, 783–786.

42. Periwal V. 2016. A comprehensive overview of computational resources to aid in precision genome editing with engineered nucleases. *Brief Bioinform.* doi 10.1093/bib/bbw052

43. Doench J.G., Hartenian E., Graham D.B., Tothova Z., Hegde M., Smith I., Sullender M., Ebert B.L., Xavier R.J., Root D.E. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32,** 1262–1267.

44. Doench J.G., Fusi N., Sullender M., Hegde M., Vaimberg E.W., Donovan K.F., Smith I., Tothova Z., Wilen C., Orchard R., Virgin H.W., Listgarten J., Root D.E. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191.

45. Xu H., Xiao T., Chen C.H., Li W., Meyer C.A., Wu Q., Wu D., Cong L., Zhang F., Liu J.S., Brown M., Liu X.S. 2015. Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* **25**, 1147–1157.

46. Montague T.G., Cruz J.M., Gagnon J.A., Church G.M., Valen E. 2014. CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42**, W401–W407.

47. Wong N., Liu W., Wang X. 2015. WU-CRISPR: Characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol*. **16,** 218.

48. Ran F.A., Hsu P.D., Wright J., Agarwala V., Scott D.A., Zhang F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc*. **8**, 2281–2308.

49. Miyaoka Y., Berman J.R., Cooper S.B., Mayerl S.J., Chan A.H., Zhang B., Karlin-Neumann G.A., Conklin B.R. 2016. Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci. Rep*. **6**, 235–249.

50. Chu V.T., Weber T., Wefers B., Wurst W., Sander S., Rajewsky K., Kühn R. 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol*. **33,** 543–548.

51. He X., Tan C., Wang F., Wang Y., Zhou R., Cui D., You W., Zhao H., Ren J., Feng B. 2016. Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res*. **44**, e85.

52. Gutschner T., Haemmerle M., Genovese G., Draetta G.F., Chin L. 2016. Post-translational regulation of Cas9 during G1 enhances homology-directed repair. *Cell*. **14**, 1555–1566.

53. Ma Y., Chen W., Zhang X., Yu L., Dong W., Pan S., Gao S., Huang X., Zhang L. 2016. Increasing the efficiency of CRISPR/Cas9-mediated precise genome editing in rats by inhibiting NHEJ and using Cas9 protein. *RNA Biol*. **13**, 605–612.

54. Davis K.M., Pattanayak V., Thompson D.B., Zuris J.A., Liu D.R. 2015. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol*. **11**, 316–318.

55. Mali P., Yang L., Esvelt K.M., Aach J., Guell M., DiCarlo J.E., Norville J.E., Church G.M. 2013. RNA-guided human genome engineering via Cas9. *Science*. **339**, 823–826.

56. Yang L., Guell M., Byrne S., Yang J.L., De Los Angeles A., Mali P., Aach J., Kim-Kiselak C., Briggs A.W., Rios X., Huang P.Y., Daley G., Church G. 2013. Optimization of scarless human stem cell genome editing. *Nucleic Acids Res*. **41**, 9049–9061.

57. Renaud J.B., Boix C., Charpentier M., De Cian A., Cochennec J., Duvernois-Berthet E., Perrouault L., Tesson L., Edouard J., Thinard R., Cherifi Y., Menoret S., Fontanière S., de Crozé N., Fraichard A., et al. 2016. Improved genome editing efficiency and flexibility using modified oligonucleotides with TALEN and CRISPR-Cas9 nucleases. *Cell Rep*. **14,** 2263–2272.

58. Ho T.T., Zhou N., Huang J., Koirala P., Xu M., Fung R., Wu F., Mo Y.Y. 2015. Targeting non-coding RNAs with the CRISPR/Cas9 system in human cell lines. *Nucleic Acids Res*. **43**, e17.

59. Canver M.C., Smith E.C., Sher F., Pinello L., Sanjana N.E., Shalem O., Chen D.D., Schupp P.G., Vinjamur D.S., Garcia S.P., Luc S., Kurita R., Nakamura Y., Fujiwara Y., Maeda T., et al. 2015. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*. **527**, 192–197.

Translated by T. Tkacheva