ISSN 1607-6729, Doklady Biochemistry and Biophysics, 2018, Vol. 479, pp. 90–94. © Pleiades Publishing, Ltd., 2018.
Original Russian Text © A.V. Khromov, V.A. Gushchin, V.I. Timerbaev, N.O. Kalinina, M.E. Taliansky, V.V. Makarov, 2018, published in Doklady Akademii Nauk, 2018, Vol. 479, No. 3, pp. 343–347.

BIOCHEMISTRY, BIOPHYSICS, AND MOLECULAR BIOLOGY

Guide RNA Design for CRISPR/Cas9-Mediated Potato Genome Editing

A. V. Khromov[a, b], V. A. Gushchin[a, b], V. I. Timerbaev[a, c], N. O. Kalinina[a, b], M. E. Taliansky[a, d], and V. V. Makarov[a, b, *]

Presented by Academician L.P. Ovchinnikov July 11, 2017

Received September 27, 2017

Abstract—The activity of the pool of sgRNA molecules designed for different regions of potato coilin and phytoene desaturase genes was compared in vitro. Due to the presence of nucleotides unpaired with DNA, sgRNA is able not only to inhibit but also to stimulate the activity of the Cas9–sgRNA complex in vitro. Although the first six nucleotides located in the DNA substrate proximally to the PAM site at the 3' end are the binding sites for cas9, they had no significant effect on the activity of the Cas9–sgRNA complex.

DOI: 10.1134/S1607672918020084

Potato is one of the most important crops in the world. However, global climate changes, the emergence of new pathogens, and the need to improve the quality and storability of potato necessitate the design of new potato cultivars with desired properties. Genome editing methods open up entirely new possibilities for designing such varieties within a short time and allow researchers to manipulate with virtually any gene in various cells, tissues, and organisms. The most popular method is the simple and convenient gene editing system based on the RNA-guided DNA nucleases, which is known as CRISPR/Sas9 [1]. In nature, CRISPR (clustered regularly interspaced short palindromic repeats)—short palindromic repeats regularly arranged in groups that are separated by unique sequences (spacers)—were found in prokaryotic genomes and originated from foreign genetic elements (plasmids and bacteriophages). RNA molecules transcribed from the CRISPR loci, together with the associated nuclease Cas, complementarily bind to DNA of foreign elements with their subsequent cutting by the Cas protein [2]. A simplified version of this system was designed for genome editing. It includes two main components: endonuclease Cas9 and a single guide RNA (sgRNA), which binds a complementary region of the target DNA 20 nucleotides in length (called protospacer), thus providing a targeted delivery of Cas9 to the motif adjoining the protospacer (PAM site, usually NGG for Cas9), whose interaction with endonuclease leads to an effective cutting of the target DNA sequence.

This technology provides an efficient and accurate genome modification, causing target breaks in DNA, which activate cellular mechanisms of DNA repair by non-homologous end joining (NHEJ) or homology-directed recombination (HDR) mechanisms. The CRISPR/Cas9 system is widely used as a powerful platform for genome editing, affecting the transcriptome, epigenetic modulation, and visualizing the genome [3–5]. A key role in a successful development of applications based on this technology is the rational design of sgRNA. The design of potentially possible sgRNAs is usually performed using bioinformatics services [9]. However, in practice, not all of the predicted sgRNAs exhibit high activity. Thus, the selection of sgRNA to a unique sequence in the genome is very important and depends on not fully investigated factors. In addition, in the selection of sgRNA, it is necessary to exclude the possibility of a non-targeted editing of the genome (the so-called "off-target effect"), which may result from the interaction of sgRNA with other nucleotide sequences in the genome of a given species that have a certain homology with the target sequence.

In this study, we analyzed the features of the functional activity of sgRNA after the break by nuclease Cas9 of different DNA regions in two potato genes: the gene encoding phytoene desaturase and the gene encoding coilin, the main protein of subnuclear Cajal

[a] *OOO Doka Gene Technologies, Rogachevo, Moscow oblast, 141880 Russia*
[b] *Moscow State University, Moscow, 119991 Russia*
[c] *Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, 117997 Russia*
[d] *The James Hutton Institute, Dundee, DD2 5DA UK*
* *e-mail: makarovvalentine@gmail.com*

bodies. Phytoene desaturase is involved in the carotenoid biosynthetic processes, the deficiency of which results in depigmentation of plants and, thus, may serve as a convenient model system for visualization and evaluation of efficiency of genome editing methods for plants, including potato [8]. Coilin deficiency in plants may enhance plant resistance to certain viral infections and abiotic (salt) stress [6, 7], which makes the coilin gene an attractive target for genome editing.

Since the key parameters for sgRNA design are the PAM site position, the complementarity of the spacer sequence of sgRNA to the protospacer, and the possibility of pairing with non-target DNA sites in the genome, we performed a comparative analysis of the activity of sgRNA molecules designed for producing gaps in different regions of phytoene desaturase and coilin potato genes. In addition, we investigated the effects of unpaired bases in sgRNA and the role of the nucleotide sequences located in the 3'-region proximally to the PAM site, which are involved in the interaction between DNA and Cas9, on the sgRNA activity in vitro.

To compare the efficiency of break induction in different DNA regions of coilin and phytoene desaturase genes by Cas9 to each of these genes using the bioinformatics web service (http://crispr.mit.edu/), we selected sequences (five for coilin and six for phytoene desaturase) that were complementary to the potential editing sites. Then, the selected sequences were synthesized as oligonucleotides, cloned into genetically engineered expression constructs containing the other required elements of sgRNA, accumulated by transcription in vitro, and purified using the Guide-it IVT RNA Clean-Up Kit (catalogue no. 632638, New England BioLabs, United States). Quality control of thus obtained sgRNAs was performed by analyzing their motility and the absence of fragmentation in agarose gel. The activity of the synthesized sgRNAs was determined in vitro according to the method available on the website of the manufacturer of the recombinant Cas9 (catalogue no. M0386, New England BioLabs) [10] using the genetically engineered constructs containing inserted fragments of coilin and phytoene desaturase genes of potato cultivar Chicago. The cloned gene fragments were sequenced to determine the exact sequence of the gene and its integrity as well as for the subsequent design of sgRNA on their basis. Genetically engineered constructs containing coilin and phytoene desaturase gene fragments, which were linearized for the *Mro*XI site, were used as a substrate for the complexes. To form the ribonucleoprotein complex, the recombinant Cas9 (30 nmol) and the synthetic sgRNA (30 nmol) were incubated at 25°C for 10 min. Then, the resulting complex was incubated with DNA substrates at 37°C for 15 min, and the reaction was stopped by the addition of proteinase K. The result of the cutting reaction was determined by electrophoresis in agarose gel. The linearized DNA substrate at the appropriate concentration without the complex of the endonuclease and the guide RNA was used as a negative control. It was found that five out of the six sgRNAs for the phytoene desaturase gene were active and could effectively guide Cas9 for cutting of the target DNA region in vitro. However, in the case of coilin, such activity was found for only one of the five sgRNAs. Figure 1b shows the results of fragments analysis for six sgRNAs for the phytoene desaturase gene and the only active sgRNA for the coilin gene. The size of the fragments obtained as a result of after cutting corresponded to the expected ones, indicating the specificity of cutting the target DNA. The specific nature of cutting was also confirmed by sequencing the resulting fragments according to Sanger (data not shown). It should be noted that the bioinformatics analysis did not reveal any consistent patterns explaining the presence or absence of activity of sgRNA depending on the sequence of the target region of the potato genome, which indicates the possibility of existence of other factors that control the activity of sgRNA. In particular, we found no correlations between the efficiency predicted using the web service (https://crispr.cos.uni-heidelberg.de/) and the actual activity of sgRNA. For example, the highest efficiency was predicted for sgRNA4 for the coilin gene; however, it did not show any activity in studies in vitro.

In further work, we used the active sgRNA for the coilin gene, because it did not ensure a complete cutting of the template and, therefore, the efficiency of DNA cutting could be increased by optimizing the sgRNA sequence. For this purpose, degenerate nucleotide substitutions were inserted into the sequence of the sgRNA site interacting with the protospacer. During the oligonucleotide synthesis, nucleotides were successively replaced with random ones in a pairwise manner. The synthesis of sgRNA to assess the effects of the degenerate nucleotide substitutions on its activity was performed as described above. The activity of the synthesized sgRNAs when cutting the coilin gene fragment in vitro was assessed by real-time PCR using a pairs of primer adjacent to the cutting site. In the case when the template was cut using the Cas9–sgRNA complex, the efficiency of amplification of respective fragment was reduced and PCR was blocked. We analyzed the activity of ten different sgRNAs containing degenerate nucleotide substitutions (Fig. 2a). Some degenerate substitutions (nucleotides 3–4 (d2), 5–6 (d3), 9–10 (d5), 15–16 (d8), 17–18 (d9), and 19–20 (d10)) led to a decrease in activity, whereas the substitution of nucleotides 7–8 (d4) caused a complete inhibition of the activity of the respective sgRNA. Substitutions 1–2 (d1) and 11–12 (d6) did not affect the activity of sgRNA, whereas substitution 13–14 (d7) significantly increased the efficiency of cutting. Thus, a complete complementarity between sgRNA and the target DNA region is not a factor that determines the efficiency of induction of breaks by DNA nuclease. Moreover, the absence of

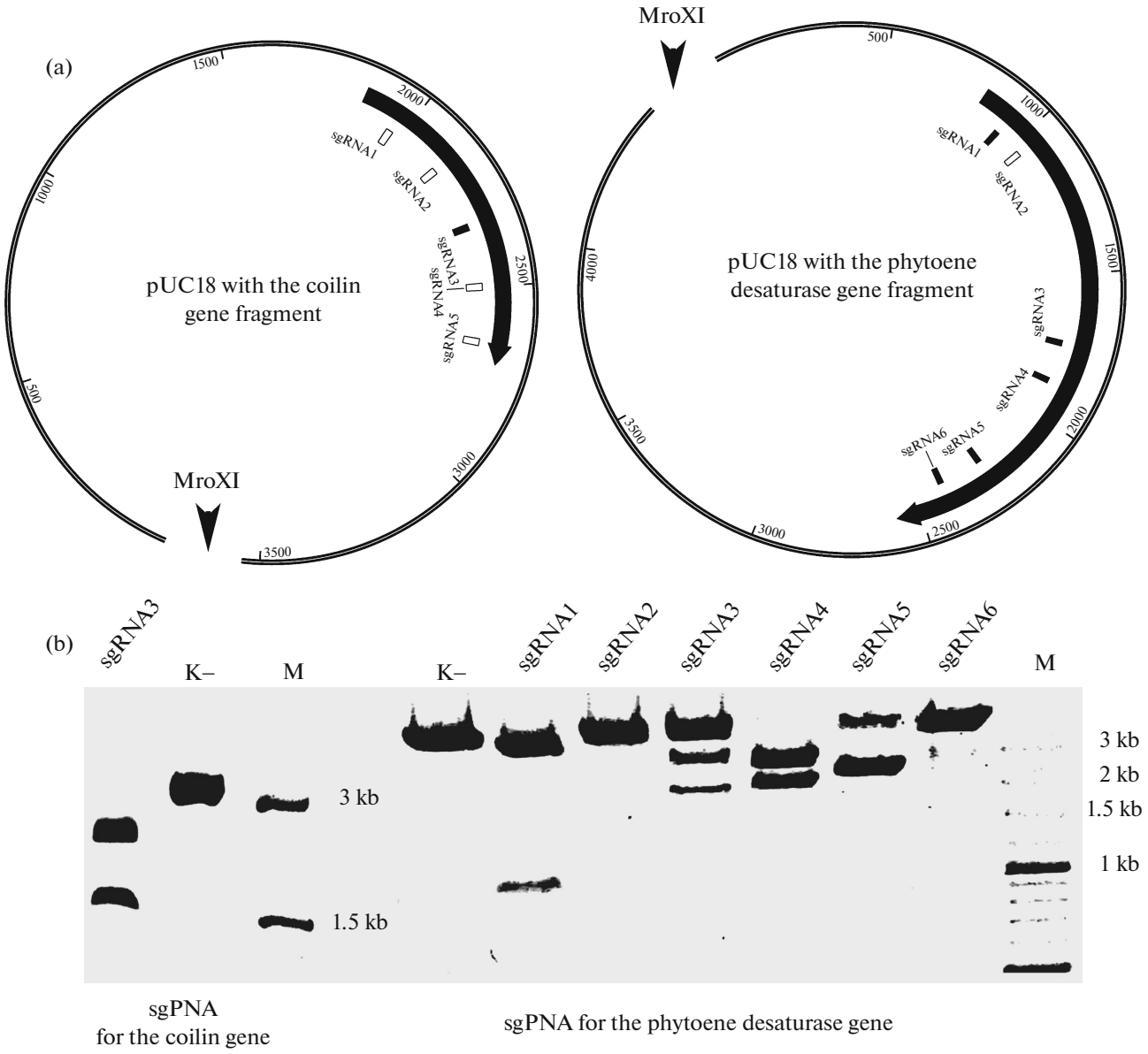

Fig. 1. Activity of the Cas9–sgRNA complex in vitro in cutting the target DNA substrate—coilin and phytoene desaturase genes of *S. tuberosum*. (a) Scheme of location of sgRNA on the target fragments of coilin and phytoene desaturase genes in the genetically engineered constructs. Black rectangles mark the active sgRNAs. The arrows indicate the linearization sites. (b) Visualization of the cutting reaction using fragment analysis in agarose gel. The linearized DNA substrate at the appropriate concentration without the complex of the endonuclease and the guide RNA was used as a negative control. The comparison of panels (a) and (b) shows that the sizes of the fragments obtained after cutting corresponded to the expected ones. In the case of the coilin gene fragment, data for only the active sgRNA are shows.

complete pairing between sgRNA and DNA can lead to an increase in the cutting efficiency.

It is known [11] that, in addition to the interaction of the Cas9–sgRNA complex with the target DNA region, there are also direct interactions between the Cas9 protein and six nucleotides in the DNA substrate adjoining the PAM site and not involved in the interaction with sgRNA. In view of this, we assumed that this sequence of the DNA substrate may be one of the factors influencing the activity of the Cas9–sgRNA complex. To experimentally verify this assumption, we obtained four mutant variants of the potential editing region of the coilin gene, in which six nucleotides located in the 3'-region proximally to the PAM site were replaced with six cytosines (mutant 6C), six guanines (mutant 6G), six adenines (mutant 6A), and six thymines (mutant 6T), respectively. The analysis of the efficiency of cutting the target DNAs, performed by real-time PCR, revealed no significant differences between the cutting efficiency of the target DNAs both

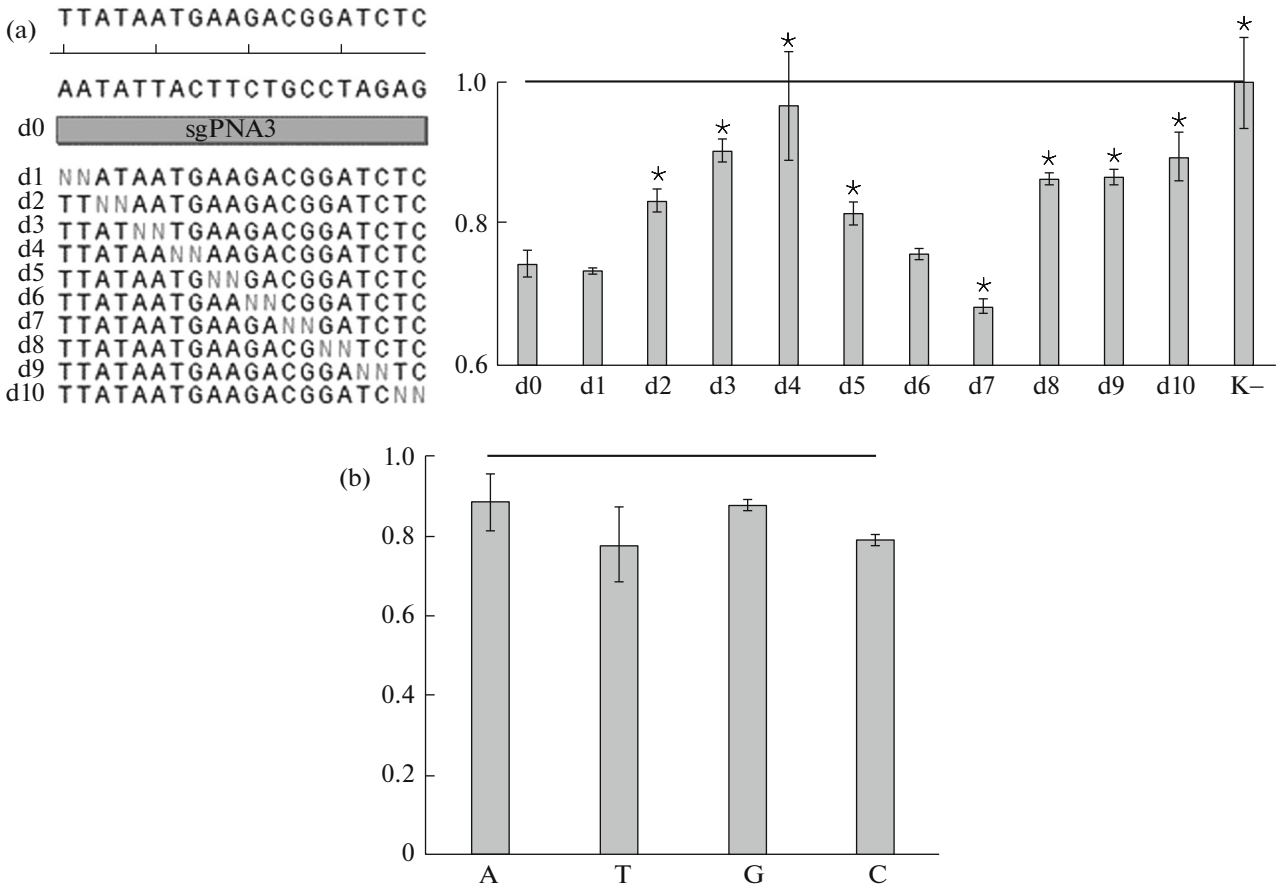

Fig. 2. Effect of the degenerate nucleotide substitutions in sgRNA and the target DNA region located in the 3'-region adjacent to the PAM site on the activity Cas9-sgRNA in vitro activity. (a) Effect of the degenerate nucleotide substitutions on the activity of sgRNA. The scheme of the arrangement of the guide RNAs containing the degenerate nucleotide substitutions is shown on the left. The cutting of the target DNA substrate using Cas9 and sgRNA containing the degenerate nucleotide substitutions is shown on the right. Quantification was performed using the real-time PCR. (b) Effect of the sequence of the target DNA region located in the 3'-region adjacent to the PAM site on the efficiency of functioning of the Cas9–sgRNA complex. The activity of the Cas9–sgRNA complex with respect to the DNA substrates containing nucleotides 6A, 6T, 6G, or 6C downstream of the PAM site is shown. Quantification was performed using the real-time PCR. The real-time PCR data were normalized to the initial template concentration (original DNA substrate, the horizontal line) and were represented as a ratio of the amount of the DNA that remained "uncut" after the reaction to the total amount of DNA added to the reaction mixture. Data are represented as $M \pm m$, $n = 5$, * $p < 0.05$ compared to the positive control (original guide RNA).

for the wild type and for all mutant variants (Fig. 2b). These data allowed us to assume that, although the six nucleotides located 3'-proximally to PAM are involved in the direct interaction with Cas9, their composition and primary structure do not play a significant role in determining the specificity of this interaction.

To conclude, it should be emphasized that the data obtained by us during the testing of sgRNA activity in vitro did not reveal their structural characteristics that determine the ability or inability of nuclease Cas9 associated with sgRNA to induce breaks in the potato genomic DNA. Using in vitro tests, we selected and optimized active sgRNAs for editing coilin and phytoene desaturase potato genes. At the next stage of our research, we will assess the efficiency of genome editing using these sgRNA in vivo.

In addition, using degenerate nucleotide substitutions in which any of the four nucleotides can be found in a certain position with equal probability, we showed that the unpaired nucleotides in the spacer region of sgRNA are able not only to decrease but also to increase the efficiency of the Cas9–sgRNA complex. This fact increases the possibility of the nonspecific delivery of Cas9 to the partially complementary regions of the genome, increasing the risk of its nontargeted modification and, therefore, should be taken into account in the preliminary bioinformatic search for the target sites for genome editing.

Rational selection of sgRNA for a unique sequence in the genome provides a high editing specificity. Today, there are a number of principles of rational sgRNA design [12]. It is known that the most important region for the manifestation of the specific activity

of sgRNA is the so-called "seed" region in the guide RNA (10–12 nucleotides proximal to the PAM site) [12]. It should be noted that the sgRNAs with a low G/C content are usually much less active [12, 13]. However, these observations are valid only for large samples and cannot predict whether a certain sgRNA will be active. In addition, the majority of published data indicate the role of individual nucleotides rather than their combinations, which determined the pairwise nucleotide substitutions in sgRNA in our work.

Thus, we showed that the presence of nucleotides unpaired with DNA in sgRNA can not only inhibit but also stimulate the activity of the Cas9–sgRNA complex in vitro. In addition, it was shown that the first six nucleotides located in the 3'-region of the DNA substrate proximally to the PAM site, despite the fact that they are binding sites for Cas9, have no significant effect on the activity of the Cas9–sgRNA complex. The results obtained in this study are important for predicting the specific sites for targeted genome editing and preliminary testing sgRNAs in vitro.

ACKNOWLEDGMENTS

The work was supported by the Russian Science Foundation (project no. 16-16-04019).

REFERENCES

1. Mali, P., Yang, L., Esvelt Aach, J., Guell, M., Dicarlo, J.E., et al, *Science*, 2013, vol. 339, pp. 823–826.
2. Al-Attar, S., Westra, E.R., van der Oost, J., and Brouns, S.J., *Biol. Chem*., 2011, vol. 392, pp. 277–289.
3. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., et al., *Science*, 2013, vol. 339, pp. 819–823.
4. Taylor, J. and Woodcock, S., *J. Biomol. Screen*., 2015, vol. 20, pp. 1040–1051.
5. Nakamura, K., Fujii, M., Tsuboi, M., Tanihata, J., Teramoto, N., Takeuchi, S., et al., *Sci. Rep*., 2014, vol. 4, p. 5635.
6. Love, A.J., Yu, C., Petukhova, N.V., Kalinina, N.O., Chen, J., and Taliansky, M.E., *RNA Biol*., 2016, vol. 11, pp. 1–12.
7. Shaw, J., Love, A.J., Makarova, S.S., Kalinina, N.O., Harrison, B.D., and Taliansky, M.E., *Nucleus*, 2014, vol. 5, pp. 85–94.
8. Tian, S., Jiang, L., Gao, Q., Zhang, J., Zong, M., Zhang, H., Ren, Y., Guo, S., Gong, G., Liu, F., and Xu, Y., *Plant Cell Rept*., 2017, vol. 36, pp. 399–406.
9. http://crispr.mit.edu/
10. www.neb.com/protocols/2014/05/01/in-vitro-digestion-of-dna-with-cas9-nuclease-s-pyogenes-m0386
11. Hirano, H., Gootenberg, J.S., Horii, T., Abudayyeh, O.O., Kimura, M., Hsu, P.D., Nakane, T., Ishitani, R., Hatada, I., Zhang, F., Nishimasu, H., and Nureki, O., *Cell*, 2016, vol. 164, pp. 950–961.
12. Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E., *Nat. Biotechnol*., 2014, vol. 32, pp. 1262–1267.
13. 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., and Root, D.E., *Nat. Biotechnol*., 2016, vol. 34, pp. 184–191.

Translated by M. Batrukova