Topic Introduction

Optimization Strategies for the CRISPR–Cas9 Genome-Editing System

Charles E. Vejnar,[1,4] Miguel A. Moreno-Mateos,[1,4] Daniel Cifuentes,[1,3] Ariel A. Bazzini,[1] and Antonio J. Giraldez[1,2,5]

[1]*Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06510;* [2]*Yale Stem Cell Center, Yale University School of Medicine, New Haven, Connecticut 06520*

The CRISPR–Cas9 system uncovered in bacteria has emerged as a powerful genome-editing technology in eukaryotic cells. It consists of two components—a single guide RNA (sgRNA) that directs the Cas9 endonuclease to a complementary DNA target site. Efficient targeting of individual genes requires highly active sgRNAs. Recent efforts have made significant progress in understanding the sequence features that increase sgRNA activity. In this introduction, we highlight advancements in the field of CRISPR–Cas9 targeting and discuss our web tool CRISPRscan, which predicts the targeting activity of sgRNAs and improves the efficiency of the CRISPR–Cas9 system for in vivo genome engineering.

OPTIMIZING CRISPR–Cas9 TARGETING EFFICIENCY

The CRISPR–Cas9 system has facilitated rapid in vivo reverse genetics studies across multiple systems (Friedland et al. 2013; Hwang et al. 2013; Wang et al. 2013; Bassett and Liu 2014), but the optimal design of single guide RNAs (sgRNAs) is essential to maximize the efficiency of the system. Using a cell proliferation screen, Wang et al. (2014) first suggested that GC-rich sgRNAs improved targeting efficiency, whereas poly(U) stretches close to the protospacer-adjacent motif (PAM) sequence (Fig. 1) were associated with sgRNAs of lower efficiency. This effect was later attributed to premature termination during sgRNA transcription, given the resemblance of the poly(U) stretches to the RNA polymerase III termination sequence (Wu et al. 2014).

Using a loss-of-function screen targeting nine genes coding for cell-surface proteins, Doench et al. (2014) identified nucleotide biases affecting the activity of thousands of sgRNAs in mammalian cell lines. By analyzing the targeted sequence and the flanking nucleotides (Fig. 1), they observed a significant guanine enrichment 1 nt upstream of the PAM sequence. This strong bias was also observed in vivo (Gagnon et al. 2014; Farboud and Meyer 2015). Combining multiple CRISPR–Cas9 screens, Xu et al. (2015) proposed an improved model for the sgRNA design. They observed a large overlap of the nucleotide biases among these screens and proposed a model to predict sgRNA efficiency based on the consensus features. Recently, Chari et al. (2015) applied a high-throughput sequencing approach to measure sgRNA activity in a large-scale screen. In contrast with the Doench et al. (2014) approach, they used a shorter activity time of CRISPR–Cas9 (72 h vs. 2 wk) and a non-phenotype-based readout

[3]Present address: Department of Biochemistry, Boston University School of Medicine, Boston, Massachusetts 02118
[4]These authors contributed equally to this work.
[5]Correspondence: antonio.giraldez@yale.edu

Cite this introduction as *Cold Spring Harb Protoc*; doi:10.1101/pdb.top090894

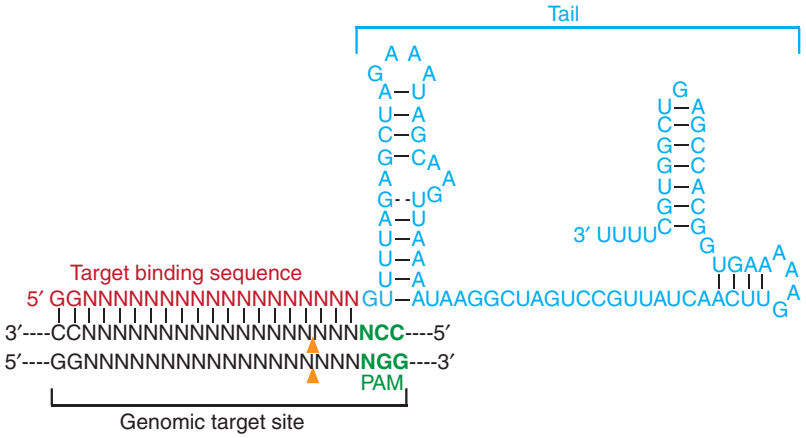

FIGURE 1. sgRNA–target site interaction. An sgRNA (target-binding sequence in red; tail in light blue) binds to its genomic target site (black), which is adjacent to the PAM sequence 5′-NGG (green). Cleavage is predicted to occur at the sites indicated by orange triangles. (Adapted by permission from Macmillan Publishers Ltd: *Nature Methods* [Moreno-Mateos et al. 2015], © 2015.)

(DNA sequencing vs. protein detection). Although the same G-rich bias upstream of the PAM sequence was reported, there was a weak correlation between the most efficient sgRNAs predicted by each study. This might underline the specificity of Cas9 activity in different systems and approaches, although some rules, such as the G-rich upstream of the PAM sequence, are widely applicable.

By testing 1280 sgRNAs in vivo using zebrafish embryos as a model system, we recapitulated the biases of the CRISPR–Cas9 system described above and uncovered specific features of efficient sgRNAs. We integrated these features into the CRISPRscan model, which we also validated in *Xenopus* (Moreno-Mateos et al. 2015). In addition, because we directly provided in vitro–transcribed sgRNAs, we were able to identify features associated with sgRNA stability that correlate with stronger activities. Recent studies of chemically modified sgRNAs have shown that more stable sgRNAs are more active in primary human cells (Hendel et al. 2015). These results highlight that the stability of the sgRNA molecule influences sgRNA activity when the sgRNA is exogenously provided rather than endogenously transcribed. Alternatively, differential stability influencing sgRNA activity may be controlled through delivery of preassembled Cas9–sgRNA ribonucleoprotein complexes (Gagnon et al. 2014; Kim et al. 2014).

EXPANDING THE TARGETING REPERTOIRE

Apart from the predicted targeting efficiency, precise mutagenesis is limited by the frequency of the specific PAM in the targeted genome and the sequence constraints to produce the sgRNA. The adapted CRISPR–Cas9 system used most extensively today for gene targeting is based on the Type II CRISPR–Cas9 system from *Streptococcus pyogenes* (Cong et al. 2013; Jinek et al. 2013; Mali et al. 2013). The Type II endonuclease Cas9 from these bacteria recognizes a PAM sequence next to the target consisting of 5′-NGG (Fig. 1). In addition, sgRNA sequences are further limited by their transcription requirements: sgRNAs require either (i) a G at the 5′ end of the molecule when using ex vivo RNA polymerase III–based systems or (ii) GA/GG when produced in vitro with SP6 or T7/T3 promoters. These restrictions limit the number of potential targets in the genome to 5′-G[N_{20}]GG in the case of RNA polymerase III–based systems and to 5′-G[G/A][N_{19}]GG for in vitro–transcribed sgRNAs.

To circumvent this constraint, various approaches have been recently developed. First, Cas9 orthologs from other bacteria (*Neisseria meningitidis*, *Streptococcus thermophilus*, and *Staphylococcus aureus*) have also been shown to target eukaryotic genomes (Esvelt et al. 2013; Hou et al. 2013; Ran et al. 2015). However, these orthologous Cas9s have PAM sequences longer or similar to those of

Cite this introduction as *Cold Spring Harb Protoc*; doi:10.1101/pdb.top090894

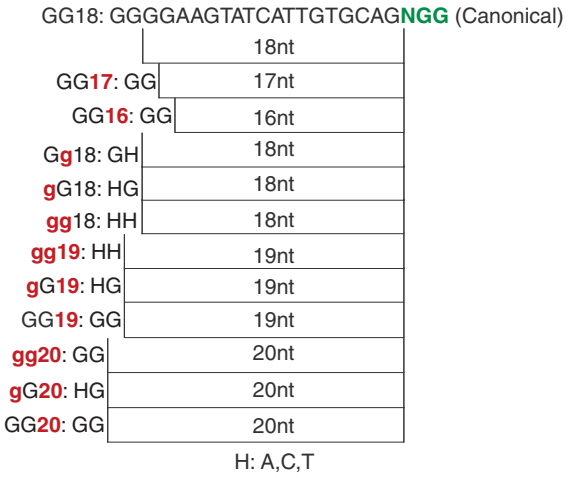

FIGURE 2. The 11 classes of alternative sgRNA targets analyzed in Moreno-Mateos et al. (2015). The PAM sequence and the alternative features are highlighted in green and in red, respectively. Mismatches between the sgRNAs and the targets are indicated by lowercase letters. (Adapted by permission from Macmillan Publishers Ltd: *Nature Methods* [Moreno-Mateos et al. 2015], © 2015.)

S. pyogenes (e.g., *S. aureus* PAM: 5′-NNGRRT), which do not dramatically increase the number of targets in a genome. To overcome this limitation, Kleinstiver et al. (2015) engineered the *S. pyogenes* Cas9 to recognize different PAM sequences (5′-NGA and 5′-NGCG), doubling the number of the targets in the human genome. More recently, a new endonuclease named Cpf1 was characterized (Zetsche et al. 2015), providing a significant increase in the number of genomic targets due to a drastically different PAM sequence (5′-TTTN).

As a complementary approach, we performed a large-scale analysis in which we sought efficient sgRNAs that target sequences other than the canonical G[G/A][N_{19}]GG (Moreno-Mateos et al. 2015). We analyzed the activity of 11 alternative targeting formulations in zebrafish embryos, varying the lengths of the sgRNAs, and introducing mismatches to the first two nucleotides of the target site (Fig. 2). We found that sgRNAs truncated by 1 or 2 nt or containing one mismatch in the first two positions of the sgRNA binding sequence were efficient alternatives to canonical sgRNAs, increasing the number of targets in the zebrafish genome by eightfold (Moreno-Mateos et al. 2015). Notably, the activities of truncated sgRNAs are similar to those of canonical sgRNAs in ex vivo systems as well (Fu et al. 2014), supporting the use of shorter sgRNAs for genome editing in vivo.

CONCLUDING REMARKS

The CRISPR–Cas9 system has revolutionized gene targeting and genome engineering. However, using it at its full potential requires optimizations and instructions on how to apply it. In this introduction, we have reviewed two improvements that allow researchers to select the most active and convenient sgRNAs: optimization of sgRNA targeting efficiency and expansion of the potential targets in the genome. Both optimizations have been integrated into our protocol for in vivo genome targeting and can be found in CRISPRscan; see Protocol: **Optimized CRISPR–Cas9 System for Genome Editing in Zebrafish** (Vejnar et al. 2016).

ACKNOWLEDGMENTS

We thank Elizabeth Fleming and Hiba Codore for technical help; all the members of the Giraldez laboratory for intellectual and technical support; and Elizabeth Fleming, Cassandra Kontur, Timothy Johnstone, and Miler Lee for manuscript editing. The Swiss National Science Foundation (grant P2GEP3_148600 to C.E.V.), Programa de Movilidad en Áreas de Investigación priorizadas por la Consejería de Igualdad, Salud y Políticas Sociales de la Junta de Andalucía (M.A.M.-M.), the Eunice

Kennedy Shriver National Institute of Child Health and Human Development-National Institutes of Health (NIH) grant K99HD071968 (D.C.), and the NIH grants R21 HD073768 (A.J.G.), R01 HD073768 (A.J.G.), and R01 GM102251 (A.J.G.) supported our work.

REFERENCES

Bassett AR, Liu JL. 2014. CRISPR/Cas9 and genome editing in *Drosophila*. *J Genet Genomics* **41:** 7–19.

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

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339:** 819–823.

Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* **32:** 1262–1267.

Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* **10:** 1116–1121.

Farboud B, Meyer BJ. 2015. Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. *Genetics* **199:** 959–971.

Friedland AE, Tzur YB, Esvelt KM, Colaiacovo MP, Church GM, Calarco JA. 2013. Heritable genome editing in *C. elegans* via a CRISPR–Cas9 system. *Nat Methods* **10:** 741–743.

Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. 2014. Improving CRISPR–Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32:** 279–284.

Gagnon JA, Valen E, Thyme SB, Huang P, Ahkmetova L, Pauli A, Montague TG, Zimmerman S, Richter C, Schier AF. 2014. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One* **9:** e98186.

Hendel A, Bak RO, Clark JT, Kennedy AB, Ryan DE, Roy S, Steinfeld I, Lunstad BD, Kaiser RJ, Wilkens AB, et al. 2015. Chemically modified guide RNAs enhance CRISPR–Cas genome editing in human primary cells. *Nat Biotechnol* **33:** 985–989.

Hou Z, Zhang Y, Propson NE, Howden SE, Chu LF, Sontheimer EJ, Thomson JA. 2013. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc Natl Acad Sci* **110:** 15644–15649.

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK. 2013. Efficient genome editing in zebrafish using a CRISPR–Cas system. *Nat Biotechnol* **31:** 227–229.

Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. 2013. RNA-programmed genome editing in human cells. *eLife* **2:** e00471.

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

Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, Gonzales AP, Li Z, Peterson RT, Yeh JR, et al. 2015. Engineered CRISPR–Cas9 nucleases with altered PAM specificities. *Nature* **523:** 481–485.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. *Science* **339:** 823–826.

Moreno-Mateos MA, Vejnar CE, Beaudoin JD, Fernandez JP, Mis EK, Khokha MK, Giraldez AJ. 2015. CRISPRscan: Designing highly efficient sgRNAs for CRISPR–Cas9 targeting in vivo. *Nat Methods* **12:** 982–988.

Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, et al. 2015. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520:** 186–191.

Vejnar CE, Moreno-Mateos MA, Cifuentes D, Bazzini AA, Giraldez AJ. 2016. Optimized CRISPR–Cas9 system for genome editing in zebrafish. *Cold Spring Harb Protoc* doi: 10.1101/pdb.prot086850.

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R. 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153:** 910–918.

Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic screens in human cells using the CRISPR–Cas9 system. *Science* **343:** 80–84.

Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, Cheng AW, Trevino AE, Konermann S, Chen S, et al. 2014. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* **32:** 670–676.

Xu H, Xiao T, Chen CH, Li W, Meyer CA, Wu Q, Wu D, Cong L, Zhang F, Liu JS, et al. 2015. Sequence determinants of improved CRISPR sgRNA design. *Genome Res* **25:** 1147–1157.

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, et al. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system. *Cell* **163:** 759–771.

Cite this introduction as *Cold Spring Harb Protoc;* doi:10.1101/pdb.top090894

Optimization Strategies for the CRISPR–Cas9 Genome-Editing System

Charles E. Vejnar, Miguel A. Moreno-Mateos, Daniel Cifuentes, Ariel A. Bazzini and Antonio J. Giraldez

Cold Spring Harb Protoc; doi: 10.1101/pdb.top090894