

SCIENTIFIC REPORTS



OPEN

Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence

Received: 12 May 2016
Accepted: 09 December 2016
Published: 18 January 2017

Ting Zheng^{1,*}, Yingzi Hou^{1,*}, Pingjing Zhang^{2,3}, Zhenxi Zhang¹, Ying Xu¹, Letian Zhang¹, Leilei Niu¹, Yi Yang¹, Da Liang¹, Fan Yi¹, Wei Peng^{2,3}, Wenjian Feng^{2,3}, Ying Yang⁴, Jianxin Chen^{2,3}, York Yuanyuan Zhu^{2,3}, Li-He Zhang¹ & Quan Du¹

Targeting specificity is an essential issue in the development of CRISPR-Cas technology. Using a luciferase activation assay, off-target cleavage activity of sgRNA was systematically investigated on single nucleotide-mismatched targets. In addition to confirming that PAM-proximal mismatches are less tolerated than PAM-distal mismatches, our study further identified a “core” sequence that is highly sensitive to target-mismatch. This sequence is of 4-nucleotide long, located at +4 to +7 position upstream of PAM, and positioned in a steric restriction region when assembled into Cas9 endonuclease. Our study also found that, single or multiple target mismatches at this region abolished off-target cleavage mediated by active sgRNAs, thus proposing a principle for gene-specific sgRNA design. Characterization of a mismatch sensitive “core” sequence not only enhances our understanding of how this elegant system functions, but also facilitates our efforts to improve targeting specificity of a sgRNA.

With the development of CRISPR-Cas9 system, researchers are able to induce specific double-stranded breaks in chromosomal DNA, whose repair leads to either local mutations or homologous recombination with donor DNA. In most CRISPR-Cas systems, Cas9 endonuclease from *Streptococcus pyogenes* is used, which prefers a proto-spacer adjacent motif (PAM) of NGG and a guide RNA of 20 nucleotides^{1–3}. In the process of the RNA-mediated DNA targeting, endonuclease Cas9 is guided to a specific target DNA by means of sequence complementarity between a single-stranded guide RNA (sgRNA) and target DNA. Under certain scenarios, sgRNA may guide Cas9 to recognize and cleave imperfectly matched DNAs, this leads to so called “off-target” cleavage. As a major concern in the development of CRISPR-Cas9 technology, off-target activity can cause genome instability and disruption of normal gene functions. Initial studies revealed that imperfectly matched DNAs with up to 6 mismatches were efficiently cleaved, causing evident off-targeting effects^{4,5}. Using a battery of endogenous loci-targeting sgRNA, recent studies reported that while some mismatched genome loci were cleaved as efficiently as that of perfectly matched target^{6,7}, many single-nucleotide mismatched targets however resist cleavage by an active sgRNA^{8,9}. Besides revealing widespread off-target effects, these studies also demonstrated that some mismatches are not tolerated by CRISPR-Cas system. Furthermore, conformation of the HNH nuclease domain has been suggested to contribute to DNA cleavage activity¹⁰.

More recently, several groups reported to solve the crystal structure of *Streptococcus pyogenes* Cas9, complexed with guide sgRNA and target DNA^{11–13}. In which, a bilobed architecture was revealed for Cas9 endonuclease, composing of a target recognition lobe and a nuclease lobe. The recognition lobe is essential to sgRNA and DNA binding, the nuclease lobe contains two nuclease domains and a PAM-interacting domain¹¹. In the initiation of RNA-guided DNA targeting, Cas9 interacts with the PAM-proximal region and the repeat: anti-repeat duplex of sgRNA, assembling a Cas9-sgRNA binary complex. The binary complex subsequently interacts with a target DNA with a sequence complementary to a 20-nt guide region of sgRNA, to form a Cas9-sgRNA-target DNA ternary

¹State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing 100191, China. ²Biomics Biotechnologies co. Ltd, Nantong 226016, Jiangsu Province, China. ³Small RNA Technology and Application Institute, Nantong University, Nantong 226016, Jiangsu Province, China. ⁴Department of Stomatology, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Q.D. (email: quan.du@pku.edu.cn)

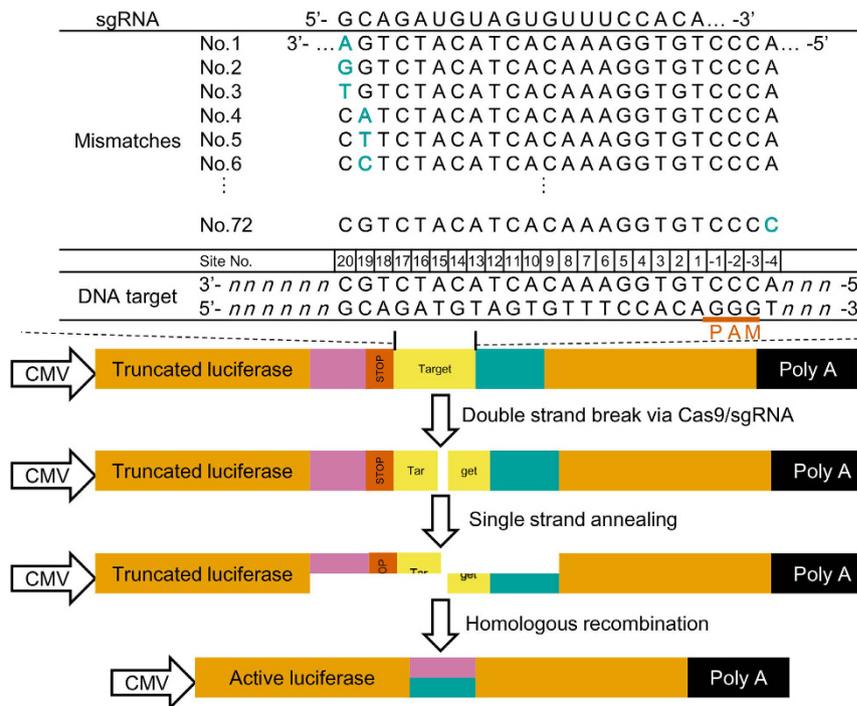


Figure 1. Schematic of the luciferase activation assay. sgRNA target sites carrying all possible single nucleotide mutations were tested for potential off-target effects.

complex. Upon the assembly of the ternary complex, the two nuclease domains approach target DNA and mediate a specific double-stranded cleavage¹³.

Despite these insights gained in the functional and structural analyses, a comprehensive specificity profile is still missing. Taking advantage of a sensitive luciferase activation assay, targeting specificity of CRISPR-Cas9 system was systematically investigated in the present study.

Results and Discussion

For genome engineering purposes, utility of CRISPR-Cas system relies on accurate selection of target DNA, in which process three components are engaged, including Cas9 protein, sgRNA and target DNA. Correct target selection derives from the base-pairing between a 20-nt sgRNA sequence and the DNA, as well as the presence of a PAM sequence. To mimic the situation in off-target cleavage, researchers mutate either the sgRNA or the target DNA, to generate mismatches in the base-pairing of sgRNA and target DNA. Considering the facts that sgRNA is assembled into Cas9 endonuclease prior to its interaction with target DNA, as well as the assembly of sgRNA causes Cas9 to take a dramatic conformational rearrangement¹², we prefer to mutate target DNA instead of sgRNA in the present study. Compared to mutating sgRNA, mutating target DNA will not affect the Cas9 assembly process of sgRNA. This, on the one hand, provides a better simulation for *in vivo* situation; and on the other hand, enables us to focus on the effects of target mismatch.

To precisely evaluate the effects of target mutation in off-target cleavage, a single-strand annealing recombination assay and a luciferase reporter have been used (Biomics Inc, China). Previous research has used this robust process to investigate the activity of zinc finger nucleases (ZFNs) and transcription activated-like effector nucleases (TALENs)¹⁴. The system consists of three plasmids, a pCas9 plasmid encoding a cas9 endonuclease, a psgRNA plasmid encoding sgRNA sequence, and a pTarget plasmid encoding an inactive form of *firefly* luciferase reporter gene. In its inactive form, the luciferase reporter is divided into two segments, separated by an in-frame stop codon and a sgRNA target site. To enable a recombination after Cas9 cleavage, both of the segments contain a 1000 bp homologous region in direct repeat orientation. When sgRNA/Cas9 complex mediates a double-strand cleavage between the two segments, the homologous regions will initiate an efficient homologous recombination, and lead to the formation of an active *firefly* luciferase reporter (Fig. 1). Using inactive luciferase as reference, activation of luciferase gene can be quantitatively measured by assessing the gain of luciferase signals in sgRNA-treated cells. This sensitive luciferase activation assay enables us to detect subtle changes of the cleavage activity.

To assess sgRNA specificity in human cells, cleavage specificity of six effective sgRNAs were tested with all possible single-nucleotide mutated target sites. Mutations were systematically made at every position across the 20-nt target site, as well as the PAM region and 1-nt downstream PAM (Fig. 1). For each sgRNA, 73 synthetic DNA fragments bearing original or mutated target site were synthesized and cloned into pTarget vector (Biomics Inc), generating a set of target plasmid carrying all single-nucleotide mutation. The resulting plasmids were then transfected into human embryonic kidney 293 A cells (HEK293), together with pCas9 plasmid, psgRNA plasmid

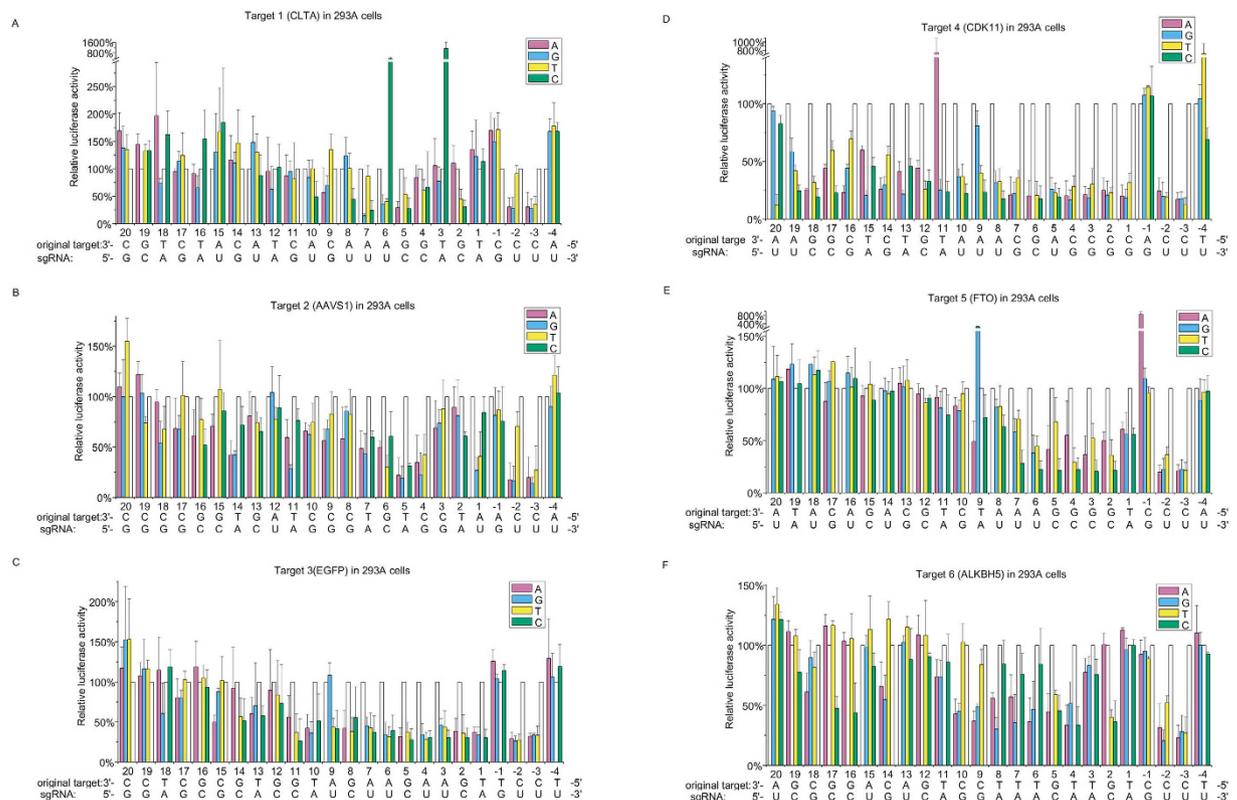


Figure 2. The mutation discrimination profile across sgRNA target site. Relative luciferase activities of sgRNA were plotted in terms of position and then the identity of the mutated nucleotide in sgRNA target sites. Target positions were numbered in the 5' to 3' direction of the DNA sequence, starting from the nucleotide immediately downstream of PAM complementary sequence. The positions of PAM region were accordingly numbered as -1, -2 and -3. Horizontal axis, sgRNA target site; vertical axis, relative luciferase activity displayed in percentage. White columns, normalized luciferase activity of original target. Scrambled target site was used as negative control. (A) Relative luciferase activities of sgRNA-1. Relative luciferase activity of negative control is 10.9%. (B) Relative luciferase activities of sgRNA-2. Relative luciferase activity of negative control is 22.4%. (C) Relative luciferase activities of sgRNA-3. Relative luciferase activity of negative control is 21.9%. (D) Relative luciferase activities of sgRNA-4. Relative luciferase activity of negative control is 10.4%. (E) Relative luciferase activities of sgRNA-5. Relative luciferase activity of negative control is 8.9%. (F) Relative luciferase activities of sgRNA-6. Relative luciferase activity of negative control is 14.6%. The data were average values of three biological replicates and every assay was conducted in triplicates.

and a reference *Renilla* luciferase plasmid. By means of dual-luciferase assay, the activity of *firefly* luciferase was quantified, and was used to evaluate the target cleavage efficacy of a sgRNA.

Initial experiments revealed that, all the six sgRNAs were able to mediate efficient cleavage on their perfectly matched targets, leading to an increase of luciferase signal up to 10-fold. For the sake of simplicity, cleavage of the perfectly matched target was used as a reference and set as 100% in the study. Relative to the perfectly matched reference target, background luciferase activities were measured with scrambled target sites. For sgRNA 1–6, this led to background luciferase activity of 10.9%, 22.4%, 21.9%, 10.4%, 8.9% and 14.6%. With a reference of matched targets, relative luciferase activity of mutated targets was normalized and aligned in terms of the position and the identity of the mismatched target nucleotide (Fig. 2).

PAM recognition is a critical step in DNA targeting and cleavage¹². In type II CRISPR system, Cas9 endonuclease was shown to bind to target site by recognizing a dominant PAM sequence of NGG¹⁵. In our study, analysis of PAM showed that while mutations at position -1 were generally tolerated, mutations at position -3 were not tolerated at all, leading to greatly compromised cleavage efficacies. For sgRNA-1, sgRNA-2, sgRNA-5 and sgRNA-6, when nucleotide C was mutated to T at position -2 in complementary strand and generated a PAM sequence of NAG in non-complementary strand, the cleavage efficacy was decreased relative to that of the NGG PAM, they are however much higher than that of the NTG or NCG PAM. This demonstrated that, at least for some sgRNAs, both NGG and NAG PAM sequence can be recognized and cleaved by Cas9. While for sgRNA-3 and sgRNA-4, NAG PAM could not be recognized. Together with published study¹⁶, these findings on one hand raise more concerns over off-target effects in CRISPR/Cas9-mediated genome engineering, on the other hand greatly expand potential targets for genome editing, especially when Cas9 is modified using structural information and improve its specificity¹⁷.

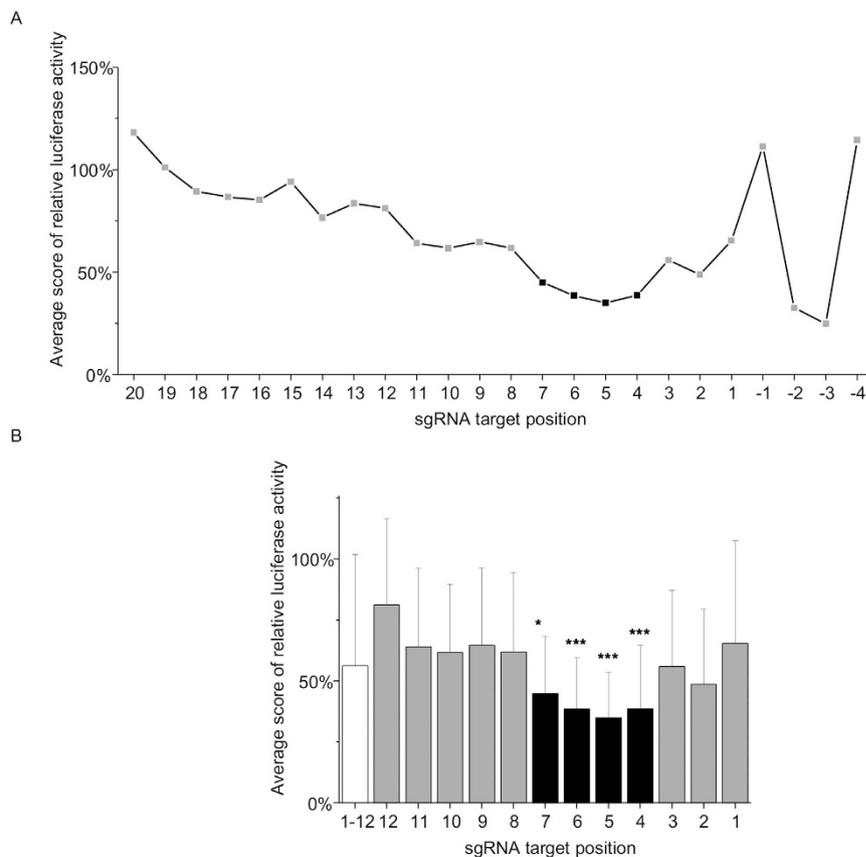


Figure 3. The tolerance profile of sgRNA target mutation. (A) A tolerance profile is obtained by averaging the relative luciferase activity of all sgRNA target sequences. **(B)** Target specificity of ‘seed sequence’. A comprehensive tolerance of target mutation is calculated by averaging the relative luciferase activity of all target positions. In relative to this reference, target specificity of seed sequence is obtained by means of Mann-Whitney test. * $p < 0.05$; *** $p < 0.001$.

In consistence with previous reports^{7,9}, our study also showed that the entire 20-base-pair target site can contribute to target specificity to some extent, depending on the sequence of sgRNA. While the cleavage efficacy varied with the position and the identity of the mismatched target nucleotide, positional effects were the most profound. Along the length of sgRNA, while modest changes in cleavage efficacy were revealed for the PAM-distal mismatches, profound effects were shown for some of the PAM-proximal mismatches. These results indicated that PAM-proximal mismatches are less tolerated than PAM-distal mismatches, and confirmed the existence of a “seed” region as previously proposed^{18,19}.

Actually, the seed sequence of the CRISPR-Cas system has been known for quite some time²⁰. Cas9 is guided by sgRNAs containing an uninterrupted 12 nt seed sequence at the 3′ end of the spacer segment⁵. Recently, structural analyses revealed that when complexed with Cas9, the entire 20-nt guide RNA segment is engaged in an A-form helical interaction with a target DNA strand¹¹. Proper orientation of this structure in Cas9 endonuclease will correctly positions target DNA for cleavage by the nuclease domains. Therefore, maintaining an intact A-form conformation is critical for target recognition and cleavage. This might explain the fact that the entire 20-base-pair target site can contribute to target specificity, although with varying significance.

In our study, while the cleavage efficacy varied with the position and the nucleotide identity in the seed region depending on the sgRNA sequence, the most profound compromising effects were surprisingly observed within a short “core” sequence, ranging from positions +4 to +7 (Fig. 3). Target cleavage was abolished by most of the single-nucleotide mismatches at this “core” sequence, rendering Cas9 activity highly sensitive to the mismatches. This also suggests a strict requirement of an intact A-form architecture for this region, which might attribute to its spatial restriction within Cas9. In structural analyses¹³, the seed region was shown to thread through a narrow nucleic acid-binding channel formed between the two Cas9 lobes. While the terminal nucleotides +1, +2, and +8 to +10 are exposed to bulk solvent, the internal nucleotides +3 to +7 are shielded from solvent by helical protein domains. Being located in such a spatial restricted region may render the “core” sequence subject to steric hindrance. This may further limit its conformation flexibility, and lead to a high sensitivity to target mismatch. Taken together with published studies¹¹, our data characterized a mismatch-sensitive “core” sequence in sgRNA, which might be a critical determinant of targeting specificity.

In contrast to the unpredictable chromatin accessibility and the limited endogenous editing efficiency of CRISPR/Cas9 system, reporter gene assays used in the study are very sensitive and reproducible, enabling us to detect the minor functional difference of different target positions.

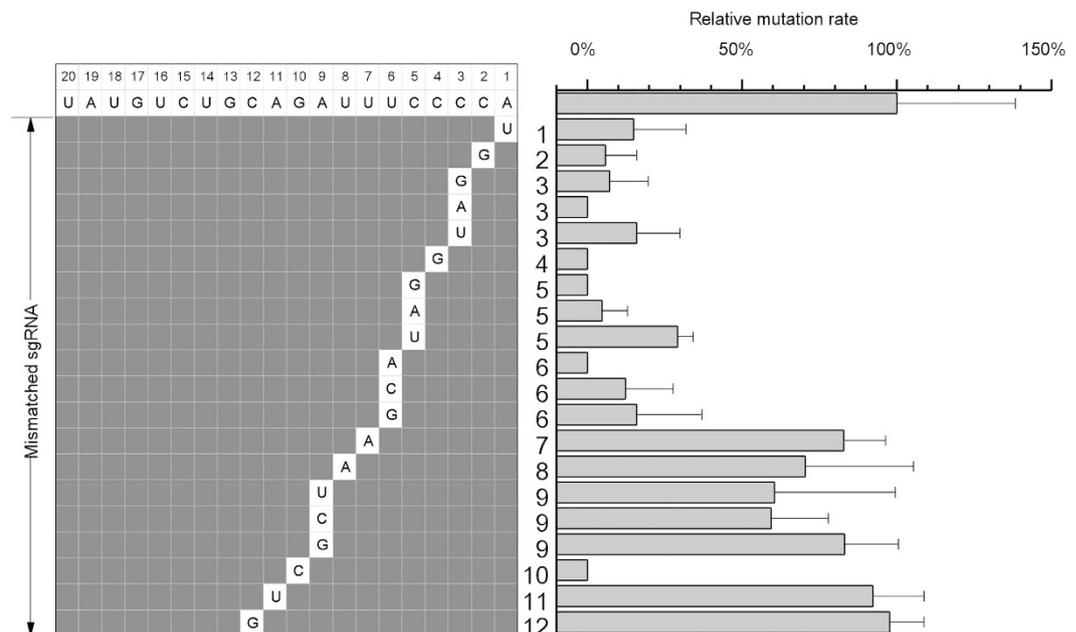


Figure 4. Activity of mismatched sgRNA-5 on endogenous target. Relative mutation rate of target DNA was measured by T7E1 assay. The identity and the position of the mismatched nucleotide are indicated.

To confirm the sensitivity of the core sequence, assays were conducted to target endogenous FTO. Within the seed region of sgRNA-5, position +5, +6 (core sequence) and +3, +9 (seed region but not core sequence) were randomly chosen to generate all possible single nucleotide mismatches, while single mismatch was generated for the other position. In agreement with our previous findings, T7E1 assays showed that mismatches at position +4, +5 and +6 almost abolished the cleavage activity of CRISP-cas9 system (Fig. 4).

Besides the positional effects, differential cleavage efficacies were revealed for different target mismatches, even at the same position. As for position +7 of sgRNA-1, while rU:dT mismatch was tolerated, rU:dC and rU:dG mismatches exhibited greatly compromised cleavage efficacies (Fig. 2A).

Considering a strong correlation between genome editing efficiency and sgRNA GC content of the six PAM-proximal nucleotides^{21,22}, we examined if sgRNA specificity is also related to the GC content of target sequence. In contrast to sgRNA-1, sgRNA-2, sgRNA-5 and sgRNA-6, almost all of the seed mismatches were intolerated for sgRNA-3 and sgRNA-4, abolishing the cleavage activity on mismatched target sites. That is to say, specificity of sgRNA-3 and sgRNA-4 are much higher than that of others. As to GC content of the PAM-proximal 6-nt, no obvious correlation was revealed.

To our surprising, assays of some target mutations demonstrated significantly enhanced cleavage activity. When the matched nucleotides at target position +3 and position +6 of sgRNA-1 were mutated to cytosine, cleavage efficacies were increased by a 11.4 and 3.6-fold, respectively. Even though both of the mismatched targets have cytosine nucleotide at the mismatched position, this effect was however not observed at the same positions of other sgRNAs. Meanwhile, rA:dA at site No. 11 of target 4 and rA:dG at site No. 9 of target 5 also reveal enhanced cleavage activity. Besides our observation, this phenomenon was also reported by other groups^{7,8}. However, further investigation on this issue revealed no obvious rule or potential mechanisms, indicating that additional efforts are needed to clarify this observation.

In addition to target mismatch, cell type may also play a role in off-target cleavage. To explore this issue, targeting specificity of sgRNA-1 was comparatively examined in HEK293 and Hela cells (Fig. 5). Compared to HEK293 cells, significantly enhanced off-target discrimination was revealed in Hela cells, in both the seed and the non-seed-region. For example at position +16, while off-target cleavage efficacy of rA:dG mismatched target is 65.2% in HEK293 cells; in Hela cells, the cleavage efficacy is only 9.9%. Despite this difference, a similar specificity profile was however revealed for both cells. rA:dC mismatched target is most tolerated, followed by rA:dA mismatched target and rA:dG mismatched target. Among the 24 target positions, similar specificity profiles were observed in 16 of them, suggesting a general targeting specificity profile for a given sgRNA. Interestingly, while the cleavage-enhancing mismatches were also observed at position +3 and +6, NAG PAM was not found active in Hela cells.

Having characterized a mismatch-sensitive “core” sequence, we speculated that this property may be used in gene-specific sgRNA design. As some of the mismatches are still tolerated in this region, we hypothesized that it is possible to further decrease or abolish off-target cleavage by introducing multiple mismatches. To this end, two or more mismatches were systematically made in the target sites of sgRNA-2. In total, 57 target plasmids were generated, and tested for potential off-target effects (Fig. 6). The results showed that all the targets bearing two or more mismatches were unable to be cleaved by sgRNA-2. These results indicate that multiple mismatches in the

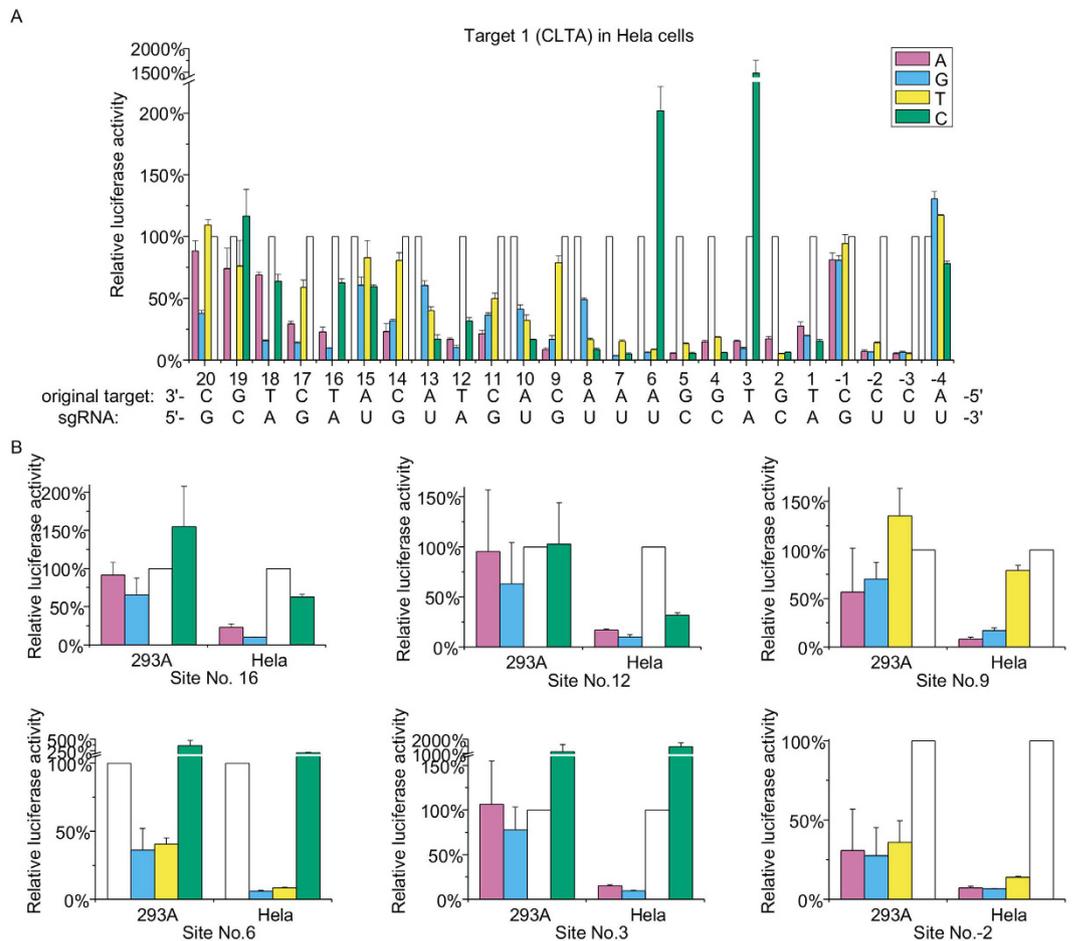


Figure 5. Activity of sgRNA-1 in HEK293 and HeLa cells. (A) Activity of sgRNA-1 in HeLa cells. **(B)** Comparative mismatch discrimination pattern in HEK293 and HeLa cells.

“core” sequence can significantly increase target specificity of a sgRNA and suggest that, it is possible to increase target specificity by choose sgRNA bearing multiple mismatches to potential off-target sites.

Taking together, our data indicate that target specificity of a sgRNA may be determined by a variety of mechanisms, depending on its interactions with target DNA, neighbouring protein domains, and solvent molecules. This is of particularly notable for a “core” sequence within the PAM-proximal region. Being located in a conformation restricted region makes this sequence highly sensitive to target mutation, a property might be exploited in gene-specific sgRNA design. To maximize target specificity of a sgRNA, we suggest that potential ‘off-target’ genomic sites should be examined by considering the following guidelines: (1) potential ‘off-target’ sites should not be followed by a NGG or NAG PAM sequences; (2) global sequence similarity between sgRNA and potential ‘off-target’ sites should be minimized; (3) single or multiple target mismatch in the “core” sequence is preferred.

Materials and Methods

Oligonucleotides and plasmids. All DNA oligonucleotides were synthesized by Invitrogen (Beijing, China). pCas9, psgRNA and pTarget plasmid were provided by Biomics Biotechnologies (China, Nantong).

Single-nucleotide mutated target fragment was cloned into linearized pTarget plasmid. The resulting fusion pTarget plasmid was sequenced for verification.

sgRNA activity assay. Human embryonic kidney (HEK293) cells were grown in Dulbecco’s modified Eagle’s medium (Corning) supplemented with 10% fetal bovine serum (Life Technologies), 100 units/ml penicillin and 100 µg/ml streptomycin (Life Technologies, Gibco). The cells were seeded into 48-well plates at a density of $\sim 2.5 \times 10^4$ cells/well one day before transfection. pCas plasmid (200 ng/well), psgRNA plasmid (200 ng/well), and pTarget plasmid (30 ng/well) carrying a target site of tested sgRNA was transfected into HEK293 cells at approximately 70% confluence, together with pRL-TK reference vector (5 ng/well). Activity of both luciferases was determined by a fluorometer (Synergy HT, BioTek, USA) before the *firefly* luciferase activity was normalized to *renilla* luciferase for each well. The cleavage efficacy of sgRNA was calculated by comparison with a sample without sgRNA treatment. All experiments were performed in triplicate and repeated three times.

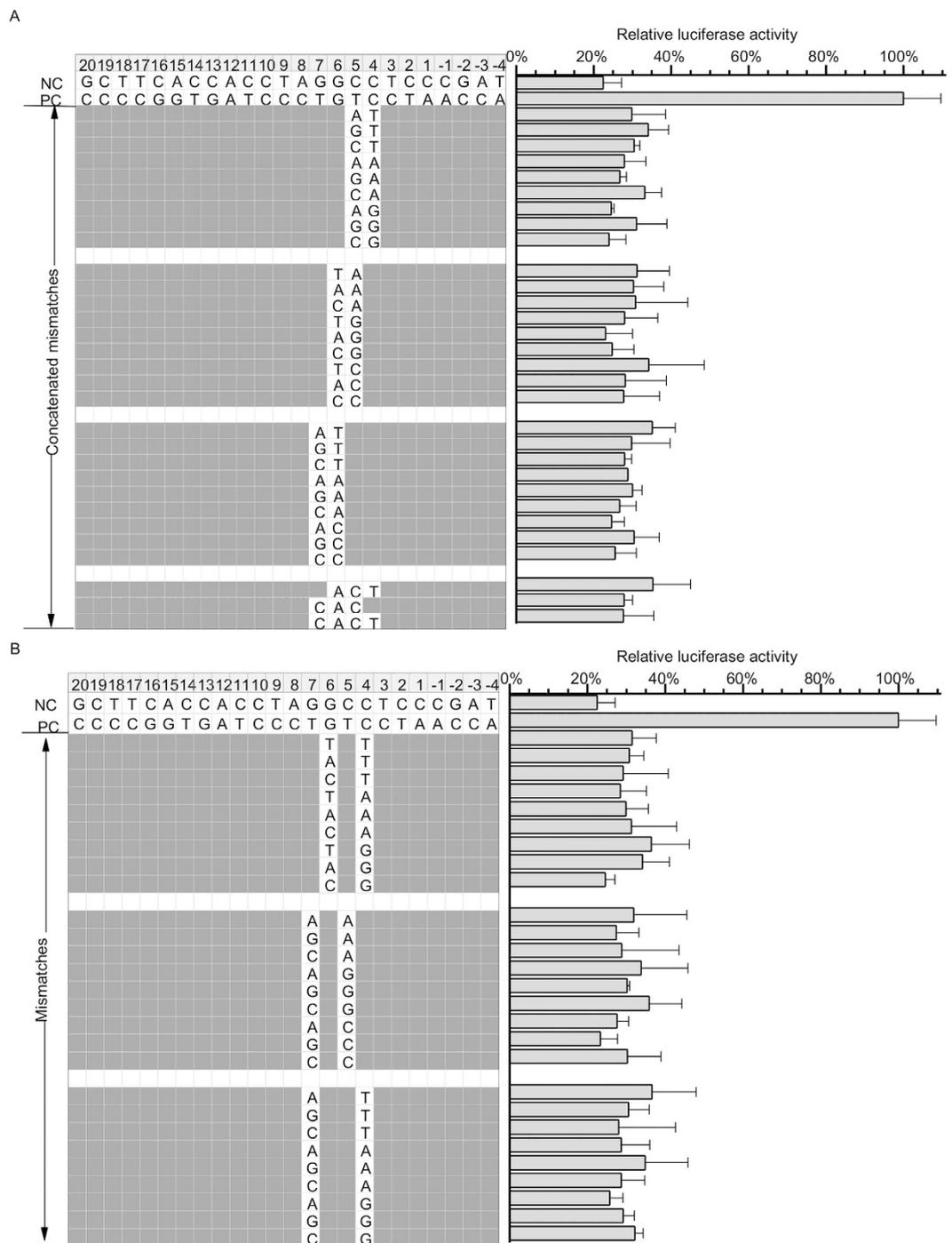


Figure 6. Activity of target sites with multiple mismatches. Relative luciferase activities were measured with 57 mismatched target sites bearing two or more mismatches. Mismatched target nucleotides are indicated.

T7E1 assay. Human embryonic kidney cells stably expressing Cas9 (HEK293 + Cas9) were grown in Dulbecco's modified Eagle's medium (Corning) supplemented with 10% fetal bovine serum (Life Technologies), 100 units/ml penicillin and 100 μ g/ml streptomycin (Life Technologies, Gibco). The cells were seeded into 24-well plates at a density of $\sim 4 \times 10^4$ cells/well one day before transfection. psgRNA plasmid (500 ng/well) was transfected into HEK293 + Cas9 cells at approximately 50% confluence following the manual of Lipofectamin 2000 (Life Technologies). Cells were collected 72 h after transfection and genomic DNA was extracted (Tiangen TIANamp Genomic DNA Kit, China). T7E1 assay was conducted following the manual of V-Solid T7 endonuclease I Kit (Beijing, China). All experiments were performed in triplicate and repeated three times.

References

1. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338, doi: 10.1038/nature10886 (2012).
2. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170, doi: 10.1126/science.1179555 (2010).
3. Terns, M. P. & Terns, R. M. CRISPR-based adaptive immune systems. *Current opinion in microbiology* **14**, 321–327, doi: 10.1016/j.mib.2011.03.005 (2011).
4. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823, doi: 10.1126/science.1231143 (2013).
5. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **31**, 233–239, doi: 10.1038/nbt.2508 (2013).
6. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822–826, doi: 10.1038/nbt.2623 (2013).
7. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827–832, doi: 10.1038/nbt.2647 (2013).
8. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**, 833–838, doi: 10.1038/nbt.2675 (2013).
9. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31**, 839–843, doi: 10.1038/nbt.2673 (2013).
10. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110–113, doi: 10.1038/nature15544 <http://www.nature.com/nature/journal/v527/n7576/abs/nature15544.html#supplementary-information> (2015).
11. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949, doi: 10.1016/j.cell.2014.02.001 (2014).
12. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573, doi: 10.1038/nature13579 (2014).
13. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481, doi: 10.1126/science.aab1452 (2015).
14. Cradick, T. J., Antico, C. J. & Bao, G. High-throughput cellular screening of engineered nuclease activity using the single-strand annealing assay and luciferase reporter. *Methods Mol Biol* **1114**, 339–352, doi: 10.1007/978-1-62703-761-7_22 (2014).
15. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67, doi: 10.1038/nature13011 (2014).
16. Zhang, Y. *et al.* Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci Rep* **4**, 5405, doi: 10.1038/srep05405 (2014).
17. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485, doi: 10.1038/nature14592 (2015).
18. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* **32**, 677–683, doi: 10.1038/nbt.2916 (2014).
19. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* **32**, 670–676, doi: 10.1038/nbt.2889 (2014).
20. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10098–10103, doi: 10.1073/pnas.1104144108/-/DCSupplemental (2011).
21. Ren, X. *et al.* Enhanced Specificity and Efficiency of the CRISPR/Cas9 System with Optimized sgRNA Parameters in *Drosophila*. *Cell reports* **9**, 1151–1162, doi: <http://dx.doi.org/10.1016/j.celrep.2014.09.044> (2014).
22. Gagnon, J. A. *et al.* Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One* **9**, e98186, doi: 10.1371/journal.pone.0098186 (2014).

Acknowledgements

This work was supported by the National High-tech R&D Program of China (2014AA021103), the Beijing Natural Science Foundation (5132015), and the Beijing Municipal Administration of Hospitals Incubating Program (PX2016065).

Author Contributions

T.Z., Y.H. and Q.D. conceived the project and edited the paper. T.Z. and Y.H. performed sgRNA activity assay and prepared figures. P.Z. designed and chose targets. Z.Z. analyzed statistics data. T.Z., Y.H., Y.X., L.Z., L.N., Yi.Y., D.L., F.Y. performed clone of pTarget plasmids. W.P. and W.F. designed and constructed the sgRNA plasmids. Ying.Y., J.C., Y.Y.Z. and L.-H.Z. made helpful suggestions and commented on the project and manuscript. Q.D. supervised the project. All authors reviewed the manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zheng, T. *et al.* Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence. *Sci. Rep.* **7**, 40638; doi: 10.1038/srep40638 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017