Engineered CRISPR-Cas9 nuclease with expanded targeting space

Citation: Nishimasu, Hiroshi et al. "Engineered CRISPR-Cas9 nuclease with expanded targeting space." Science 361, 6408 (August 2018): 1259-1262 © The Authors

As Published: http://dx.doi.org/10.1126/science.aas9129

Publisher: American Association for the Advancement of Science (AAAS)

Persistent URL: https://hdl.handle.net/1721.1/126396

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2019 March 21.

Published in final edited form as:

Science. 2018 September 21; 361(6408): 1259–1262. doi:10.1126/science.aas9129.

Engineered CRISPR-Cas9 nuclease with expanded targeting space

Hiroshi Nishimasu[1,*], **Xi Shi**[2,3], **Soh Ishiguro**[4,5,6], **Linyi Gao**[2,7], **Seiichi Hirano**[1], **Sae Okazaki**[1], **Taichi Noda**[8], **Omar O. Abudayyeh**[2,3,9], **Jonathan S. Gootenberg**[2,3,9], **Hideto Mori**[4,5,6], **Seiya Oura**[8,10], **Benjamin Holmes**[2,3], **Mamoru Tanaka**[4], **Motoaki Seki**[4], **Hisato Hirano**[1], **Hiroyuki Aburatani**[4], **Ryuichiro Ishitani**[1], **Masahito Ikawa**[8,10,11], **Nozomu Yachie**[1,4,5,6], **Feng Zhang**[2,3,7,9], and **Osamu Nureki**[1,*]

[1]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

[2]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.

[3]McGovern Institute for Brain Research, Cambridge, MA 02139, USA.

[4]Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan.

[5]Institute for Advanced Biosciences, Keio University, 14-1 Baba-cho, Tsuruoka, Yamagata 997-0035, Japan.

[6]Graduate School of Media and Governance, Keio University, 5322 Endo, Fujisawa, Kanagawa 252-0882, Japan.

[7]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

[8]Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan.

[9]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

[10]Graduate School of Pharmaceutical Sciences, Osaka University, 1-6 Yamadaoka, Suita, Osaka 565-0871, Japan.

[*] Corresponding author. nisimasu@bs.s.u-tokyo.ac.jp (H.N.); nureki@bs.s.u-tokyo.ac.jp (O.N.).

Author contributions: H.N. designed the variants, performed in vitro cleavage experiments, and determined the crystal structure; X.S., L.G., O.O.A., J.S.G., B.H., and F.Z. performed PAM screens, indel detection, and GUIDE-seq analyses; S.I., H.M., M.T., M.S., H.A., and N.Y. performed base-editing analyses; T.N., S.Ou., and M.I. performed indel detection; S.H. performed crystallization; H.N., S.H., S.Ok., and H.H. prepared the proteins; R.I. assisted with the structural determination; H.N. wrote the manuscript with help from all authors; and H.N. and O.N. supervised the research.

Competing interests: H.N. is a scientific adviser for EdiGENE. F.Z. is a co-founder and scientific adviser for Editas Medicine, Pairwise Plants, Beam Therapeutics, and Arbor Biotechnologies. F.Z. serves as a director for Beam Therapeutics and Arbor Biotechnologies. O.N. is a co-founder, board member, and scientific adviser for EdiGENE. H.N., H.H., and O.N. have filed a patent application related to this work.

Data and materials availability: The atomic coordinates of SpCas9-NG have been deposited in the Protein Data Bank with PDB code 6AI6. All data are available in the manuscript or the supplementary material.

[11]Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.

Abstract

The RNA-guided endonuclease Cas9 cleaves its target DNA and is a powerful genome-editing tool. However, the widely used *Streptococcus pyogenes* Cas9 enzyme (SpCas9) requires an NGG protospacer adjacent motif (PAM) for target recognition, thereby restricting the targetable genomic loci. Here, we report a rationally engineered SpCas9 variant (SpCas9-NG) that can recognize relaxed NG PAMs. The crystal structure revealed that the loss of the base-specific interaction with the third nucleobase is compensated by newly introduced non–base-specific interactions, thereby enabling the NG PAM recognition. We showed that SpCas9-NG induces indels at endogenous target sites bearing NG PAMs in human cells. Furthermore, we found that the fusion of SpCas9-NG and the activation-induced cytidine deaminase (AID) mediates the C-to-T conversion at target sites with NG PAMs in human cells.

The CRISPR RNA-guided endonuclease Cas9 cleaves double-stranded DNA targets complementary to the RNA guide (1) (fig. S1) and has been harnessed for genome editing in eukaryotic cells (2). However, the widely used Cas9 from *S. pyogenes* (SpCas9) strictly recognizes an NGG sequence (where N is any nucleobase) as the protospacer adjacent motif (PAM) (3), thereby restricting the targetable genomic loci. Structure-guided directed evolution approaches to address this limitation yielded several SpCas9 variants with altered PAM specificities, such as the SpCas9 VQR and VRER variants, which recognize the NGA and NGCG PAMs, respectively (4). In addition, Cas9 and Cas12a (also known as Cpf1) enzymes with distinct PAM specificities, such as *Staphylococcus aureus* Cas9 (SaCas9) (5), *Acidaminococcus* sp. Cas12a (AsCas12a), and *Lachnospiraceae bacterium* Cas12a (LbCas12a) (6), have extended the targeting range in CRISPR-Cas-mediated genome editing.

To expand the targeting range of CRISPR-Cas9, we sought to engineer a SpCas9 variant with relaxed preferences for the third nucleobase of the PAM. Previous studies revealed that the second and third nucleobases in the NGG PAM are recognized by Arg^{1333} and Arg^{1335} of SpCas9, respectively (7) (fig. S2). We thus hypothesized that the PAM constraint can be reduced by eliminating the base-specific interaction between Arg^{1335} and the third G, and compensating for the loss of this base-specific interaction by introducing non-base-specific interactions with the PAM duplex. We first measured the in vitro cleavage activities of purified wild-type SpCas9 and the R1335A ($Arg^{1335} \rightarrow$ Ala) mutant toward a target plasmid with the TGG PAM and confirmed that, whereas wild-type SpCas9 efficiently cleaves the TGG target, R1335A has almost no activity (fig. S3, A to C).

We next examined whether the R1335A activity is restored by the substitution of residues surrounding the PAM duplex, and found that the replacements of Leu^{1111}, Gly^{1218}, Ala^{1322}, and Thr^{1337} with Arg partially restored the activity of the R1335A mutant (fig. S3, A to C). Furthermore, the R1335A/L1111R/G1218R/A1322R/T1337R variant (referred to as ARRRR) efficiently cleaved the TGG target (fig. S3, A to C). However, the cleavage kinetics of ARRRR was slower than that of wildtype SpCas9 (fig. S3, D and E). In the previously

reported VQR (D1135V/R1335Q/T1337R) and VRER (D1135V/G1218R/R1335E/T1337R) variants, the D1135V mutation provides interactions with the sugar-phosphate backbone of the PAM duplex (8,9). In addition, molecular modeling suggested that the E1219F mutation forms hydrophobic interactions with the ribose moiety of the second G, and that the R1335V mutation stabilizes Arg^{1333} and Phe^{1219} (E1219F) (fig. S3F). Indeed, the addition of the D1135V, E1219F, and R1335V mutations enhanced the cleavage activity (fig. S3, D and E). We designated the R1335V/L1111R/D1135V/G1218R/E1219F/A1322R/T1337R variant as VRVRFRR.

We next measured the cleavage activities of VRVRFRR toward the target plasmid with TGN PAMs. Relative to wild-type SpCas9, VRVRFRR slowly but more efficiently cleaved the TGA, TGT, and TGC targets (Fig. 1, A to C, and fig. S4). Although VRVRFRR is less active than wild-type SpCas9 and SaCas9, its cleavage activity was comparable to those of AsCas12a and LbCas12a (5, 6) (fig. S5). Using an in vitro PAM identification assay (10), we confirmed that whereas wildtype SpCas9 is specific to NGG PAMs, VRVRFRR preferentially recognizes NG PAMs (Fig. 1D and fig. S6). Although the PAM identification assay revealed that VRVRFRR slightly recognizes NAN PAMs (Fig. 1D), in vitro cleavage experiments demonstrated that VRVRFRR is less active toward TAN PAMs than toward TGN PAMs (fig. S7). Thus, we concluded that VRVRFRR recognizes a relaxed PAM, and we refer to this variant as SpCas9-NG, as it has increased activity on NGH (H = A, T, or C) PAMs, albeit with reduced relative activity on NGC.

We compared SpCas9-NG with the xCas9 enzyme (a SpCas9 variant with the A262T/R324L/S409I/E480K/E543D/M694I/E1219V mutations), which was engineered via directed evolution and recognizes NG PAMs (11) (fig. S8). We measured their in vitro cleavage activities toward the target plasmid with the TGN PAMs. Under our assay conditions (50 nM Cas9), xCas9 showed almost no activity toward the TGA, TGT, and TGC targets, and it cleaved the TGG target with lower efficiency than did SpCas9-NG (Fig. 1, A to C and E, and fig. S9A). At a higher concentration (200 nM Cas9), xCas9 cleaved the TGA, TGT, and TGC targets (fig. S9, D and G); nonetheless, the cleavage kinetics of xCas9 was slower than that of SpCas9-NG (fig. S9, C, D, F, and G), and the cleavage activity of xCas9 toward the TGA, TGT, and TGC targets was comparable to or lower than that of wild-type SpCas9 toward the TGA target (fig. S9, B, D, E, and G). These results demonstrate that SpCas9-NG outperforms xCas9 in recognizing NGH PAMs in vitro.

To clarify the NG PAM recognition mechanism, we determined the crystal structure of SpCas9- NG, in complex with a single guide RNA and its target DNA containing the TGG PAM, at 2.7 Å resolution (fig. S10A and table S1). The second G in the PAM (dG2*) forms bidentate hydrogen bonds with Arg^{1333} (Fig. 2A), which is stabilized by Val^{1335} (R1335V). The third G (dG3*) forms a hydrogen bond with Arg^{1337} (T1337R) (Fig. 2A), consistent with the preference of SpCas9-NG for the third G (Fig. 1C). Arg^{1111} (L1111R), Val^{1135} (D1135V), and Phe^{1219} (E1219F) interact with the sugar-phosphate backbone of the PAM duplex, whereas Arg^{1218} (G1218R) does not directly interact with the DNA backbone (Fig. 2B). A structural comparison with the SpCas9 R-loop complex (12) suggested that Arg^{1322} (A1322R) interacts with the nontarget DNA strand (fig. S10B).

To assess the activity of SpCas9-NG in mammalian cells, we next measured the formation of indels (insertions or deletions) induced by wildtype SpCas9, SpCas9-NG, and xCas9 at 69 endogenous target sites with NGN PAMs in human embryonic kidney (HEK) 293FT cells (table S2). As expected, wild-type SpCas9 induced indels predominantly at the NGG sites, with some recognition of NGA sites (Fig. 3, A and B). Wild-type SpCas9 achieved a >20% indel rate at NGG (17 of 17 sites) and NGA sites (3 of 19 sites), but not NGT (0 of 18 sites) or NGC sites (0 of 15 sites) (Fig. 3, A and B). In contrast, SpCas9-NG edited NGA, NGT, and NGG sites, with lower activity at NGC (Fig. 3, A and B), consistent with the in vitro cleavage preference for the third D (D = A, T, or G) PAM nucleotide (Fig. 1C). SpCas9-NG achieved a >20% indel rate at 13 NGG, 13 NGA, 15 NGT, and 5 NGC sites (Fig. 3, A and B). Relative to SpCas9-NG, xCas9 had lower editing efficiency at all of the NGH sites (Fig. 3, A and B), consistent with our in vitro cleavage data (Fig. 1E). xCas9 induced a >20% indel rate at 15 NGG sites, five NGA sites, and four NGT sites, but none at NGC sites (Fig. 3, A and B).

Using GUIDE-seq (genome-wide, unbiased identification of double-stranded breaks enabled by sequencing) (13), we examined the specificity of wild-type SpCas9 and SpCas9-NG in human cells at two previously characterized target sites *(EMX1* and *VEGFA).* We found that wild-type SpCas9 and SpCas9-NG have comparable numbers of off-target sites for the EMX1 target, and that the K848A/K1003A/R1060A mutations in the high-fidelity SpCas9 variant [eSpCas9(1.1)] (14) substantially reduced the off-target cleavage by SpCas9-NG (Fig. 3C and fig. S11). These results show that SpCas9-NG has cleavage specificity comparable to that of wild-type SpCas9, and that its specificity can be enhanced by the high-fidelity mutations. As expected, SpCas9-NG had a different off-target profile with off-target sites harboring NGH PAMs (fig. S11), further confirming the relaxed PAM recognition.

The nuclease-inactive version of SpCas9 can be applied to numerous technologies such as base editing (15,16). We examined whether the SpCas9-NG D10A nickase fused to the activation-induced cytidine deaminase (nSpCas9-NG-AID, referred to as Target-AID-NG) mediates C-to-T conversion at 32 endogenous target sites with NG PAMs in human cells (table S3). nSpCas9-AID (Target- AID) efficiently induced C-to-T conversion at the NGG sites, with lower activity at NGA sites and no activity at NGT or NGC sites (Fig. 4A and figs. S12A and S13). In contrast, Target- AID-NG showed base-editing activity toward all PAMs assessed, albeit with lower efficiency at NGC sites (Fig. 4A and figs. S12A and S13), consistent with the indel data (Fig. 3A). In addition, Target-AID-NG showed higher base-editing efficiencies than did the xCas9 D10A nickase fused to the APOBEC1 cytidine deaminase (xCas9-BE4) at most of the tested poly-C–containing target sites (Fig. 4B and figs. S12B and S13). These results demonstrate that the catalytically inactive version of SpCas9-NG can serve as a useful RNA-guided DNA targeting platform.

In this study, we engineered an SpCas9-NG variant with an increased targeting range (fig. S14). Nonetheless, the cleavage activity of SpCas9-NG is lower than that of wild-type SpCas9 at NGG sites, and SpCas9-NG is less active at NGC sites relative to NGD sites. Thus, it will be important to improve the activity of SpCas9-NG by further molecular engineering. Overall, the rationally designed SpCas9-NG and its high-fidelity variants can serve as useful genome-editing tools with increased versatility across genomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank A. Kurabayashi for assistance with sample preparation, T. Yamano for Cas12a preparation, B. P. Kleinstiver and J. K. Joung for assistance with GUIDE-seq, and the beamline scientists at BL41XU at SPring-8 for assistance with data collection.

Funding: Supported by JST (JPMJPR13L8) and JSPS (26291010 and 15H01463) (H.N.); JSPS (16J06287) (S.I.); JSPS (17H01394), AMED (JP18gm5010001), Takeda Science Foundation, NICHD (P01HD087157 and R01HD088412), and the Bill & Melinda Gates Foundation (OPP1160866) (M.I.); NEDO (Genome Editing Program), AMED (17g6110007h0002), JST (10814) and JSPS (18H02428) (N.Y.); NIMH (5DP1-MH100706 and 1R01-MH110049), NIDDK (5R01DK097768-03), and the New York Stem Cell, Simons, Paul G. Allen Family, and Vallee foundations (F.Z.); AMED (16am0301002h0003) and CSTI, SIP (O.N.). F.Z. is a New York Stem Cell Foundation Robertson Investigator.

REFERENCES AND NOTES

1. Jinek M et al., Science 337, 816–821 (2012). [PubMed: 22745249]
2. Cong L et al., Science 339, 819–823 (2013). [PubMed: 23287718]
3. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA, Nature 507, 62–67 (2014). [PubMed: 24476820]
4. Kleinstiver BP et al., Nature 523, 481–485 (2015). [PubMed: 26098369]
5. Ran FA et al., Nature 520, 186–191 (2015). [PubMed: 25830891]
6. Zetsche B et al., Cell 163, 759–771 (2015). [PubMed: 26422227]
7. Anders C, Niewoehner O, Duerst A, Jinek M, Nature 513, 569–573 (2014). [PubMed: 25079318]
8. Anders C, Bargsten K, Jinek M, Mol. Cell 61, 895–902 (2016). [PubMed: 26990992]
9. Hirano S, Nishimasu H, Ishitani R, Nureki O, Mol. Cell 61, 886–894 (2016). [PubMed: 26990991]
10. Gao L et al., Nat. Biotechnol. 35, 789–792 (2017). [PubMed: 28581492]
11. Hu JH et al., Nature 556, 57–63 (2018). [PubMed: 29512652]
12. Jiang F et al., Science 351, 867–871 (2016). [PubMed: 26841432]
13. Tsai SQ et al., Nat. Biotechnol. 33, 187–197 (2015). [PubMed: 25513782]
14. Slaymaker IM et al., Science 351, 84–88 (2016). [PubMed: 26628643]
15. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR, Nature 533, 420–424 (2016). [PubMed: 27096365]
16. Nishida K et al., Science 353, aaf8729 (2016). [PubMed: 27492474]

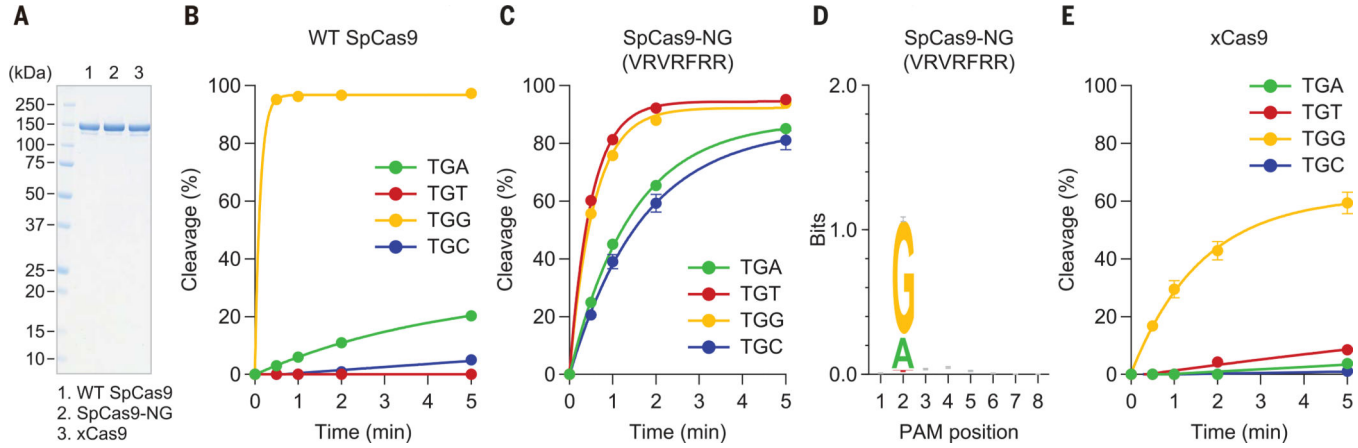

Fig. 1. In vitro cleavage activity.

(**A**) SDS-polyacrylamide gel electrophoresis analysis of wild-type SpCas9, SpCas9-NG, and xCas9. (**B**, **C**, and **E**) In vitro DNA cleavage activities of wild-type SpCas9 (B), SpCas9-NG (C), and xCas9 (E) toward the TGN PAM targets. Data are means ± SD (n = 3). (**D**) PAM preference of SpCas9-NG.

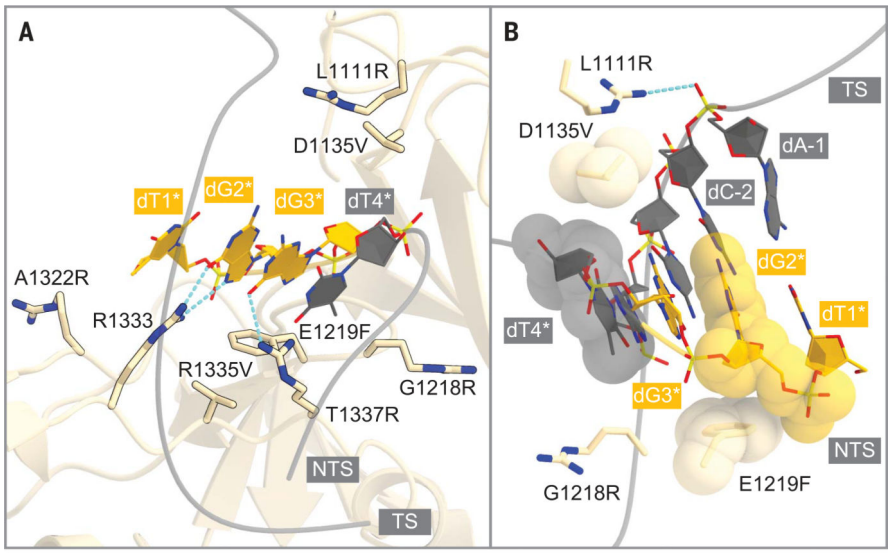

Fig. 2. Crystal structure of SpCas9-NG.

(**A**) Recognition of the PAM duplex. Arg1333 and the substituted residues are shown as stick models. (**B**) Non-base-specific interactions between the PAM duplex and the substituted residues. TS, target strand; NTS, nontarget strand.

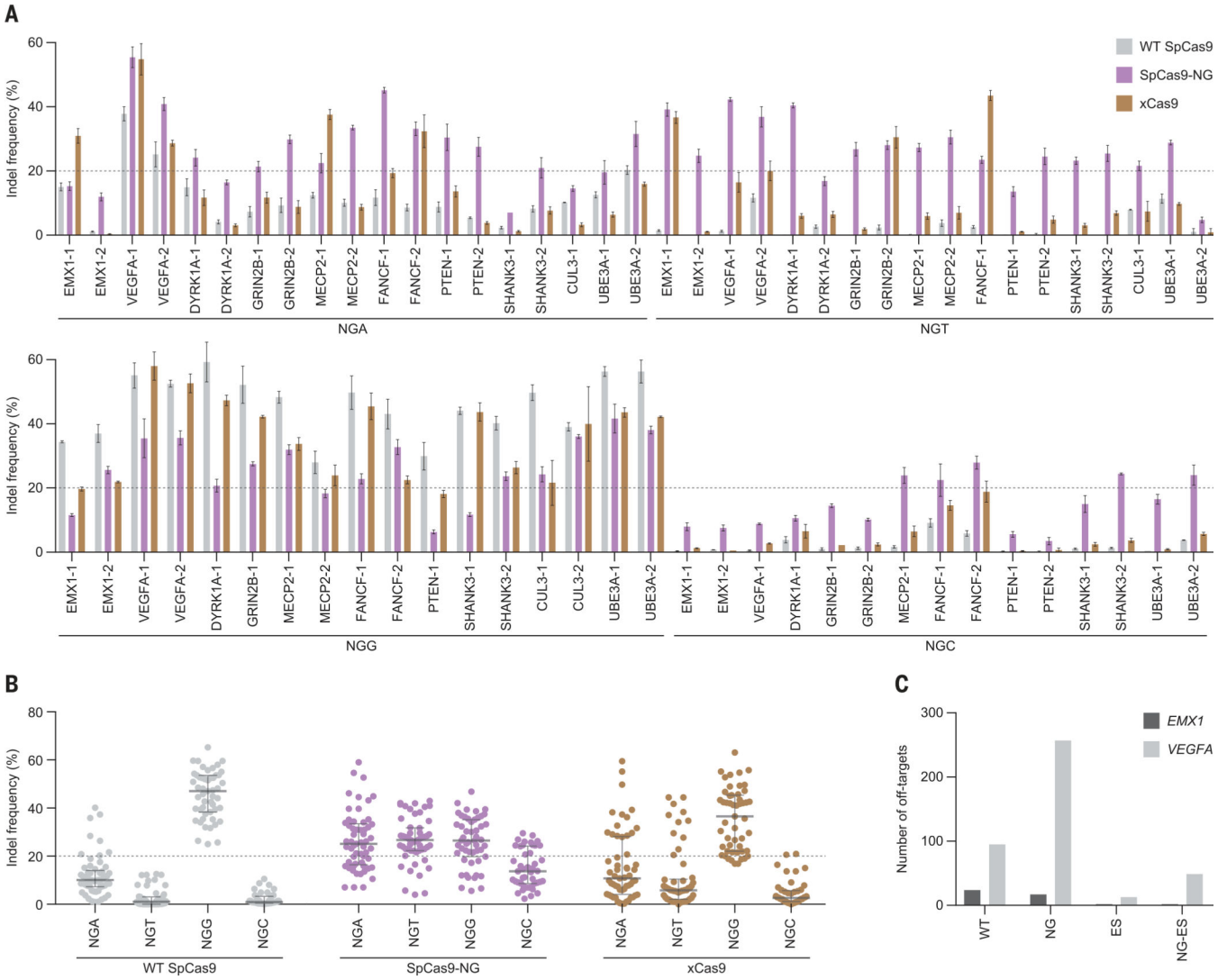

Fig. 3. Gene editing in human cells.

(**A**) Indel formation efficiencies of wild-type SpCas9, SpCas9-NG, and xCas9 at the 69 endogenous target sites in HEK293FT cells. Data are means ± SD ($n = 3$). (**B**) Summary of the editing efficiencies of SpCas9, SpCas9-NG, and xCas9. Medians and first and third quartiles are shown. In (A) and (B), 20% indel frequency is indicated by dashed lines. (**C**) Specificities of wild-type SpCas9, SpCas9-NG, and the enhanced-specificity versions of SpCas9 (ES) and SpCas9-NG (NG-ES). The off-target cleavages were evaluated by GUIDE-seq.

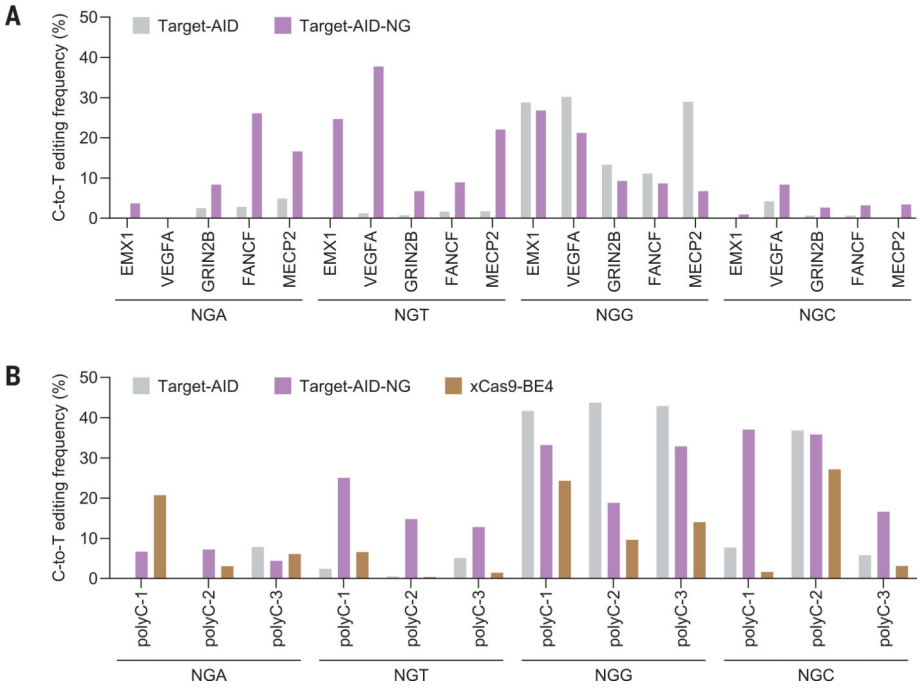

Fig. 4. Base editing in human cells.

(**A** and **B**) C-to-T conversion efficiencies at the 20 endogenous target sites (Target-AID and Target-AID-NG) (A) and at the 12 poly-C-containing target sites (Target-AID, Target-AID-NG, and xCas9-BE4) (B) in HEK293Tcells. The experiments were performed at least twice, and similar results were obtained.