GENE EDITING

Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants

Russell T. Walton[1,2]*†, Kathleen A. Christie[1,2,3], Madelynn N. Whittaker[1,2], Benjamin P. Kleinstiver[1,2,3]‡

Manipulation of DNA by CRISPR-Cas enzymes requires the recognition of a protospacer-adjacent motif (PAM), limiting target site recognition to a subset of sequences. To remove this constraint, we engineered variants of *Streptococcus pyogenes* Cas9 (SpCas9) to eliminate the NGG PAM requirement. We developed a variant named SpG that is capable of targeting an expanded set of NGN PAMs, and we further optimized this enzyme to develop a near-PAMless SpCas9 variant named SpRY (NRN and to a lesser extent NYN PAMs). SpRY nuclease and base-editor variants can target almost all PAMs, exhibiting robust activities on a wide range of sites with NRN PAMs in human cells and lower but substantial activity on those with NYN PAMs. Using SpG and SpRY, we generated previously inaccessible disease-relevant genetic variants, supporting the utility of high-resolution targeting across genome editing applications.

The requirement for DNA-targeting CRISPR-Cas enzymes to recognize a short sequence motif adjacent to target sites in foreign DNA is a critical step for CRISPR systems to distinguish self from nonself (*1*, *2*). For genome editing applications, however, the necessity of protospacer-adjacent motif (*3*–*6*) (PAM) recognition by Cas9 and Cas12a proteins constrains targeting and affects editing efficiency and flexibility. The prototypical Cas9 from *Streptococcus pyogenes* (SpCas9) naturally recognizes target sites with NGG PAMs (where N is A, C, G, or T) (*5*, *7*, *8*), making it one of the most targetable CRISPR enzymes characterized to date. Although other naturally occurring Cas orthologs can, in principle, expand targeting by recognizing divergent noncanonical PAMs, most Cas9 and Cas12a enzymes (*9*–*12*) require extended motifs that limit their utility for genome editing. Thus, the PAM requirement prevents the accurate positioning of CRISPR target sites and is a major barrier for genome editing applications that command high-resolution target site positioning [e.g., targeting small genetic elements, base editing, generating efficient homology-directed repair–mediated alterations, performing tiling screens, etc. (*13*–*19*)].

One method to improve the targeting range of genome editing technologies is to purposefully engineer CRISPR enzymes that can target previously inaccessible PAMs. SpCas9 primarily recognizes its optimal NGG PAM by direct molecular readout of the guanine DNA bases via the amino acid side chains of R1333 and R1335 (*20*) (Fig. 1, A and B, and fig. S1A). Modification of either arginine alone ablates SpCas9 nuclease activity against sites with NGG, NAG, or NGA PAMs (*8*, *20*), necessitating the use of molecular evolution to alter PAM preference by mutation of other amino acids in the PAM-interacting (PI) domain. Several protein engineering strategies have been pursued toward expanding targeting with SpCas9, including using directed evolution or structure-guided engineering to develop variants with altered PAM profiles (e.g., SpCas9-VQR, VRQR, and VRER) (*8*, *21*) or relaxed PAM preferences (e.g., SpCas9-NG and xCas9) (*22*, *23*). Although these variants expand the potential targeting space of SpCas9, target sites encoding most noncanonical PAMs still remain inaccessible for genome editing.

In this study, we used structure-guided engineering to nearly completely relax the PAM requirement of SpCas9. This approach enabled the generation of a highly enzymatically active NGN PAM variant (named SpG), and subsequent optimization of SpG led to a variant (named SpRY) capable of editing nearly all PAMs. We demonstrate that SpG and SpRY improve editing resolution and offer previously unrealized genome editing capabilities. More broadly, the molecular strategy described herein should in principle be extensible to a wide diversity of Cas orthologs, paving a path toward the development of a suite of editing technologies no longer constrained by their inherent targeting limitations.

Structure-guided mutagenesis to relax SpCas9 PAM preference

In our attempt to eliminate the PAM requirement of SpCas9, we first developed a variant capable of recognizing a reduced NGN PAM relative to the canonical NGG sequence. Our previous efforts to alter SpCas9 PAM preference illuminated several PAM-proximal residues important for PAM recognition (*8*) (Fig. 1B), observations that are supported by structural studies (*20*, *24*, *25*) [fig. S1, A to E, and (*26*)]. Characterizations of SpCas9-VQR [harboring D1135V/R1335Q/T1337R substitutions (where D1135V indicates an Asp^{1135}→Val substitution, etc.)] and the derivative SpCas9-VRQR variant (that additionally encodes G1218R) suggest that these enzymes target an expanded number of noncanonical PAMs, including those with variable bases in the third position of the PAM (NGAN>NGNG, where > symbol indicates relative activity) (*8*, *21*). We therefore hypothesized that R1335Q-harboring variants with other PI domain substitutions could recognize an expanded number of PAMs (fig. S1, F to I). Thus, we used SpCas9-VRQR as a molecular scaffold to further relax SpCas9 PAM preference.

To more thoroughly investigate the impacts of amino acid substitutions in PI domain residues, we developed a high-throughput PAM determination assay (HT-PAMDA) to comprehensively profile the PAM preferences of a large number of SpCas9 variants [fig. S2, A to D, and (*26*)]. HT-PAMDA accurately replicated the PAM profiles of wild-type (WT) SpCas9 and other previously described variants (Fig. 1C and fig. S2D). To engineer an SpCas9 variant capable of more-relaxed targeting, we used HT-PAMDA to sequentially determine the contributions of substitutions at five PI-critical positions, D1135, S1136, G1218, E1219, and T1337, in the context of SpCas9-VRQR [fig. S2E and (*26*)]. We identified several variants bearing combinations of rational substitutions at these five important residues that exhibited more-balanced tolerances for any nucleotide at the third and fourth PAM positions (Fig. 1C and fig. S2E). One variant bearing D1135L/S1136W/G1218K/E1219Q/R1335Q/T1337R substitutions, henceforth referred to as SpG, exhibited the most-even targeting of NGA, NGC, NGG, and NGT PAMs (Fig. 1C and fig. S2E).

We then compared the activities of WT SpCas9 with those of SpG and nearly all intermediate variants in human embryonic kidney (HEK) 293T cells to corroborate our HT-PAMDA findings. Using an optimal nuclear localization signal (*27*) (fig. S3) and nonsaturating nuclease expression conditions in HEK 293T cells, we examined the human cell editing activities of this large collection of variants on four sites with NGA, NGC, NGG, and NGT PAMs [fig. S4, A and B, and (*26*)]. These experiments revealed high-activity editing on the four NGN PAM sites with SpG (Fig. 1D and fig. S4B), results that were consistent with the PAM preference of SpG characterized using HT-PAMDA (Fig. 1C and fig. S2F). We then sought to bolster the activity of SpG through the addition of nonspecific contacts mediated by L1111R and A1322R substitutions, mutations that are necessary for the NGN PAM tolerance of SpCas9-NG (*22*) (fig. S4, C and D). However, the L1111R and A1322R substitutions were detrimental to the human

[1]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA. [2]Department of Pathology, Massachusetts General Hospital, Boston, MA 02114, USA. [3]Department of Pathology, Harvard Medical School, Boston, MA 02115, USA.
*Present address: Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. †Present address: Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.
‡Corresponding author. Email: bkleinstiver@mgh.harvard.edu

Downloaded from http://science.sciencemag.org/ on February 21, 2021

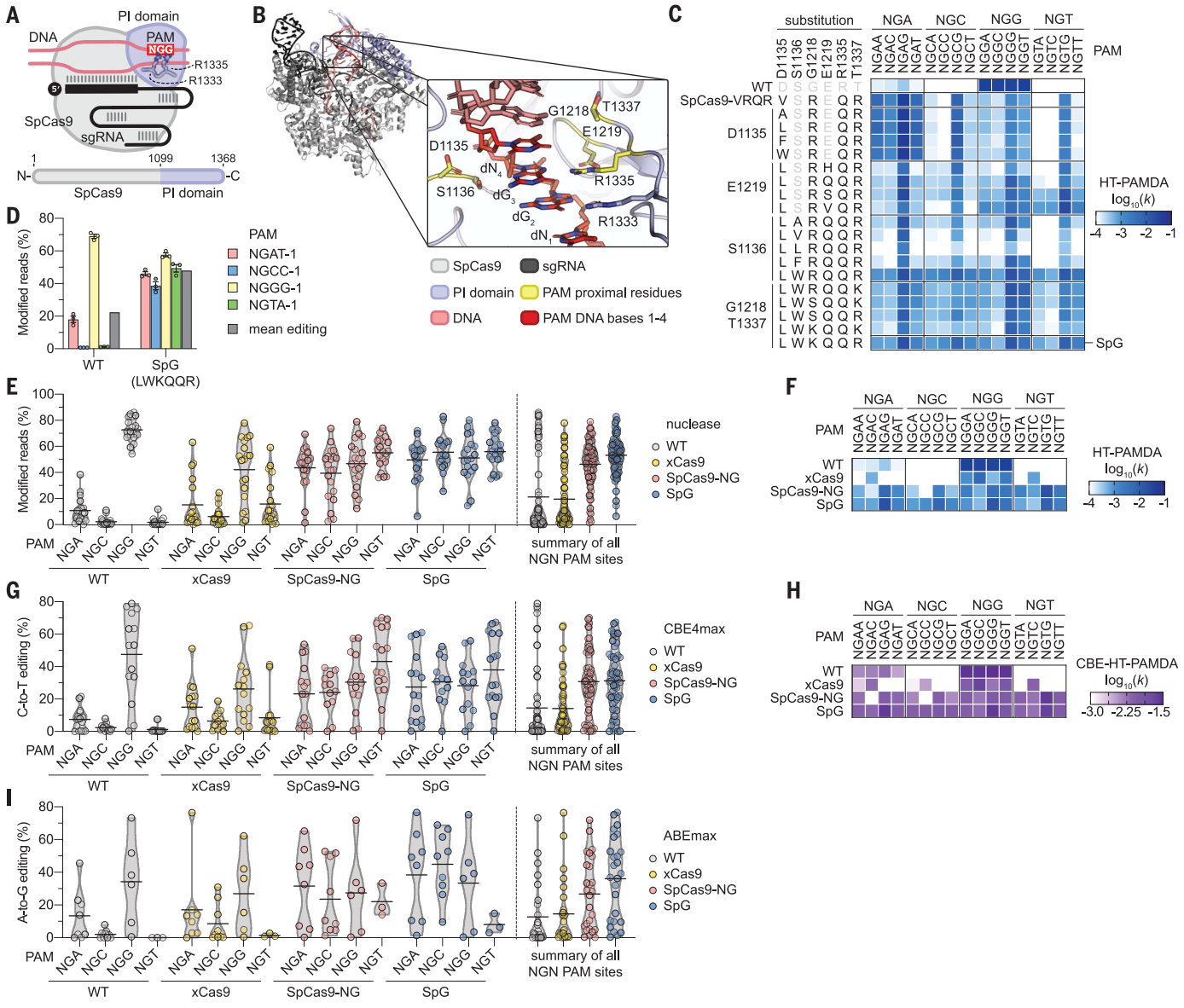


Fig. 1. Engineering and characterization of SpCas9 variants capable of targeting NGN PAMs. (**A**) Schematic of SpCas9, highlighting the PI domain along with R1333 and R1335 that make base-specific contacts to the guanines of the NGG PAM. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. sgRNA, single-guide RNA; N, N terminus; C, C terminus. (**B**) Rendering of a crystal structure of SpCas9 with amino acid side chains proximal to the second guanine of the NGG PAM shown in yellow. In the zoomed image, the nontarget strand (NTS) is hidden for clarity. Image was generated from Protein Data Bank (PDB) ID 4UN3 (*20*). (**C**) HT-PAMDA characterization of WT SpCas9 and engineered variants to illustrate their NGNN PAM preferences. The $\log_{10}$ rate constants (k) are the mean of at least two replicates against two distinct spacer sequences (see also fig. S2, A to C and E). (**D**) Modification of endogenous sites in human cells bearing canonical and noncanonical PAMs with WT SpCas9 and SpG. Editing assessed by targeted sequencing; mean, SEM, and individual data points shown for n = 3 biological replicates. (**E**) Mean nuclease activity plots for WT, xCas9 (*23*), SpCas9-NG (*22*), and SpG on 78 sites with NGN PAMs in human cells. The horizontal black lines represent the mean of 19 to 20 sites for each PAM class (see also fig. S5A), and the gray outlines are violin plots. (**F**) HT-PAMDA characterization of WT, xCas9, SpCas9-NG, and SpG to illustrate their NGNN PAM preferences. The $\log_{10}$ rate constants are the mean of at least two replicates against two distinct spacer sequences. (**G**) Mean C-to-T editing plots for WT, xCas9, SpCas9-NG, and SpG CBEs on 57 cytosines within the editing windows (positions three through nine) of 20 target sites harboring NGN PAMs in human cells. The horizontal black lines represent the mean of 12 to 16 cytosines for each PAM class (see also fig. S6A), and the gray outlines are violin plots. (**H**) CBE-HT-PAMDA data for WT, xCas9, SpCas9-NG, and SpG to illustrate their NGNN PAM preferences. The $\log_{10}$ rate constants are single replicates against one spacer sequence (see also fig. S6, C and D). (**I**) Mean A-to-G editing plots for WT, xCas9, SpCas9-NG, and SpG ABEs on 24 adenines within the editing windows (positions five through seven) of 21 target sites harboring NGN PAMs in human cells. The horizontal black lines represent the mean of three to nine adenines for each PAM class (see also fig. S7A), and the gray outlines are violin plots.

Downloaded from http://science.sciencemag.org/ on February 21, 2021

cell editing activities of SpG, albeit without alteration of PAM preference [figs. S4, E and F, and S2, F and G, and see (*26*)].

SpG activities as a nuclease, CBE, and ABE

Given the broad compatibility of SpG with NGN PAMs as determined by HT-PAMDA but our limited testing on only 16 target sites in HEK 293T cells, we more thoroughly compared its nuclease activity in human cells against WT SpCas9, xCas9(3.7), and SpCas9-NG. We assessed the editing activities of SpG and these three nucleases on 78 sites bearing NGNN PAM sequences that encompassed an approximately even distribution of nucleotide identities in the third and fourth positions of the PAM (fig. S5A and table S1). Our assessment recapitulated the PAM preference of WT SpCas9 (*7*, *8*), with a mean editing activity of 72.8% on sites with NGG PAMs and a reduced 4.7% mean editing across the remaining NGH sites (where H is A, C, or T) (Fig. 1E). SpG exhibited the highest mean editing activities across all NGN PAM sites, averaging 51.2% on sites with NGG PAMs and 53.7% on sites with NGH PAMs (Fig. 1E). Of the two variants reported to recognize sites with NGN PAMs, xCas9 displayed more-modest editing (42.2% on NGG and 12.5% across NGH), while SpCas9-NG editing activities were even across NGN PAM sites (46.9% on NGG and 46.0% across NGH) but lower compared with those of SpG (Fig. 1E).

To better understand the PAM requirements of each of the NGN PAM variants, we used HT-PAMDA to profile SpG, xCas9, and SpCas9-NG (Fig. 1F and fig. S2F). These experiments demonstrated that SpG exhibited the most even and robust targeting of all the NGN PAMs (targeting efficacy ranked as follows: SpG > SpCas9-NG > xCas9) (Fig. 1F), consistent with the results of our 78-site human cell experiment (Fig. 1E). Closer inspection of our HT-PAMDA and human cell data did not reveal substantial first PAM, fourth PAM, or first spacer position preferences for WT SpCas9, SpG, or SpCas9-NG [fig. S5, B to D, and (*26*)]. This analysis also attributed the decreased activities observed with xCas9 to a preference for a C in the fourth position of the PAM, which likely makes targeting of sites with NGND PAMs (where D is A, G, or T) less efficient [fig. S5E; for a more complete analysis of xCas9 PAM preference, see (*26*)]. For the four nucleases, HT-PAMDA values for each NGNN PAM class correlated with mean editing activities in HEK 293T cells (fig. S5F). Together, these results indicate that SpG is an efficient and broadly targeting nuclease across sites bearing NGN PAMs.

Given the ubiquitous use of base editor (BE) technologies to mediate single nucleotide substitutions in various organisms (*17*, *18*, *28*), we investigated whether the improved activities of SpG could enhance BE activities across sites with NGN PAMs. We compared C-to-T editing with WT SpCas9, xCas9, SpCas9-NG, and SpG BE4max cytosine base editor (*29*) (CBE) constructs across 22 endogenous sites in human cells bearing NGNN PAMs (fig. S6A). We observed that, whereas WT- and xCas9-CBE exhibited mean C-to-T editing efficiencies >15% only on sites with NGG PAMs, both SpG- and SpCas9-NG-CBE were capable of mean C-to-T editing >23% across NGN sites (Fig. 1G) and displayed typical CBE substrate preferences (fig. S6B). To ensure that the CBE versions of the PAM variants harbored the same PAM profiles as the nucleases, we developed a modified CBE high-throughput PAM determination assay (CBE-HT-PAMDA) [fig. S6C and (*26*)]. The PAM compatibilities of the CBEs were largely consistent with the preferences of the nucleases (compare Fig. 1, H and F, respectively, and see fig. S6, D and E).

Beyond C-to-T editing, adenine base editor (ABE) constructs have also been developed that mediate A-to-G edits (*18*). Thus, we also compared the A-to-G editing potencies of WT SpCas9, xCas9, SpCas9-NG, and SpG in the ABEmax architecture (*29*) across 21 endogenous sites harboring NGNN PAMs (fig. S7A). Similar to our observations for CBEs, WT- and xCas9-ABE could only efficiently perform A-to-G

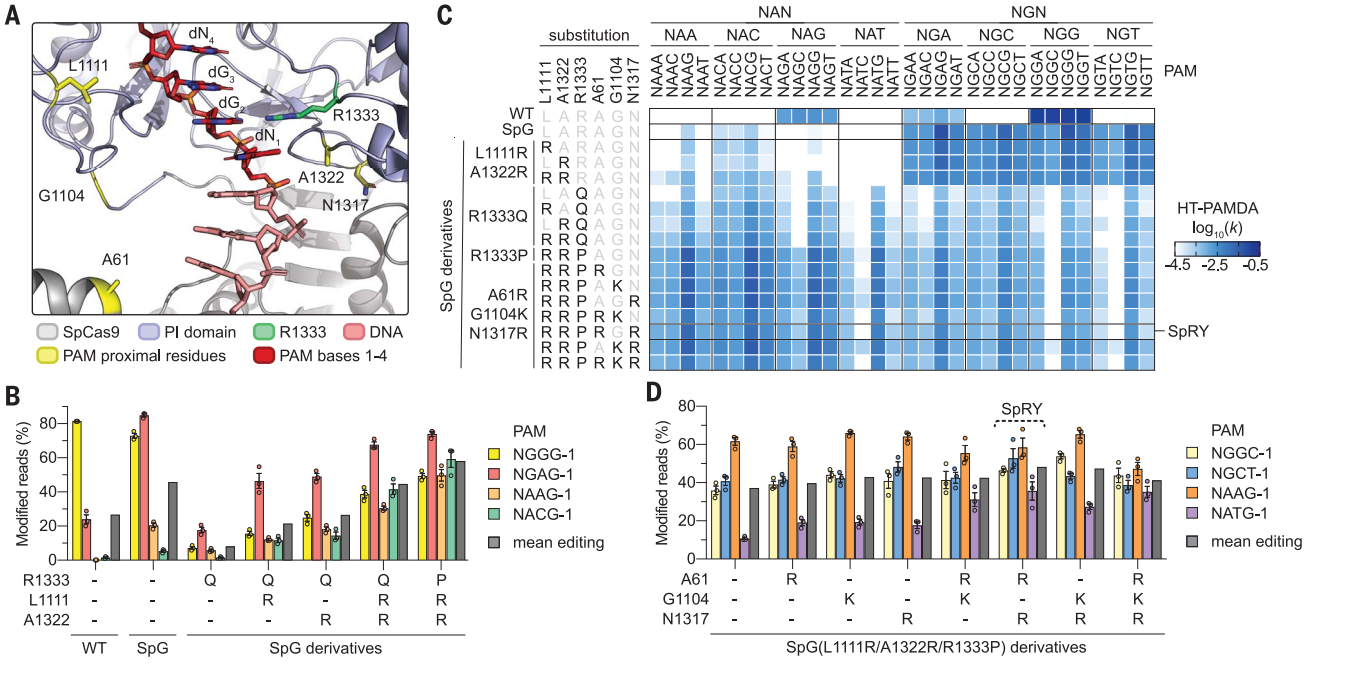


Fig. 2. Engineering and characterization of SpCas9 variants capable of targeting NRN PAMs. (**A**) Crystal structure of SpCas9 to illustrate amino acid side chains of R1333 and selected PAM-proximal residues. The NTS is hidden for clarity. Image generated from PDB ID 4UN3 (*20*). (**B**) Modification of endogenous sites in human cells bearing different NRN PAMs with WT SpCas9, SpG, and SpG derivatives. Editing assessed by targeted sequencing; mean, SEM, and individual data points shown for $n = 3$. (**C**) HT-PAMDA characterizations of WT SpCas9, SpG, and SpG derivatives to illustrate their NRNN PAM preferences. The $\log_{10}$ rate constants are the mean of at least two replicates against two distinct spacer sequences (see also fig. S2, A to C). (**D**) Modification of endogenous sites in human cells bearing different NRN PAMs by SpG(L1111R/A1322R/R1333P) and derivatives bearing additional substitutions. See fig. S8C for all variants tested. Editing assessed by targeted sequencing; mean, SEM, and individual data points shown for $n = 3$.

Downloaded from http://science.sciencemag.org/ on February 21, 2021

edits on target sites with NGG PAMs (Fig. 1I). However, both SpG- and SpCas9-NG-ABE efficiently edited target sites with NGNN PAMs, with SpG-ABE exhibiting the most-robust activity across all NGNN sites (Fig. 1I) with typical ABE substrate preferences (fig. S7B).

Collectively, these results demonstrate that SpCas9 PAM preference can be relaxed to a single NGN nucleotide motif by designing a more tolerant PI domain. The SpG variant developed through this strategy exhibits robust nuclease, CBE, and ABE activities across NGN PAMs.

Engineering SpCas9 variants capable of targeting NRN PAMs

Notwithstanding the efficient modification of sites with NGN PAMs using SpG, many genomic regions remain inaccessible to genome editing. Because we observed efficient modification of sites bearing NGN PAMs with SpG, we speculated that SpG could be used as a molecular scaffold upon which to further relax PAM specificity. To alter recognition of the second position of the PAM, we hypothesized that substitution of R1333 to glutamine might enable access to sites harboring NAN PAMs by forming a base-specific contact with the adenine base in the second position of the PAM (*8*, *20*, *24*) (Fig. 2A). However, our initial tests of SpG(R1333Q) nearly abolished activity in human cells against four sites bearing NRN PAMs (where R is A or G) (Fig. 2B), revealing that the R1333Q alone was insufficient to enable highly active targeting of NAN PAMs [consistent with observations for WT SpCas9 (*8*, *20*)]. Notably, contrary to our previous finding that L1111R and A1322R substitutions negatively affected SpG activity (fig. S4, E and F), we observed that these nonspecific DNA contacts were able to rescue some activity of SpG(R1333Q) across the four sites bearing NRN PAMs in human cells (Fig. 2B). Concurrent HT-PAMDA experiments to analyze the same variants corroborated a general relaxation of PAM specificity against NRN PAMs but with a lower overall activity (Fig. 2C).

Next, to determine whether the R1333Q substitution was the most permissive for recognition of an expanded number of PAMs, we used HT-PAMDA to investigate whether variants harboring other amino acid substitutions at residue 1333 might be more amenable to highly active and broad targeting of NRN PAMs. Systematic evaluation of SpG(L1111R/A1322R) variants harboring all 20 possible amino acids at residue 1333 revealed that the range of substitutions at this position cause different second PAM position preferences and overall levels of activity (fig. S8A). Unexpectedly, variants bearing R1333 substitutions to alanine, cysteine, or proline conferred the most-efficient collective targeting of NRN PAMs. Experiments in HEK 293T cells against the same four sites harboring NRN PAMs demonstrated that one SpG(L1111R/A1322R) variant that also harbored the R1333P substitution exhibited greater activity on NRN PAMs compared with the precursor R1333Q-containing variants (Fig. 2B). HT-PAMDA experiments confirmed these observations (Fig. 2C).

Given that the addition of L1111R and A1322R to SpG-R1333Q improved on-target activity, we sought to determine whether additional analogous substitutions could further enhance editing of sites with NRN PAMs. We used crystal structures of SpCas9 (*20*, *24*, *25*) to identify other positions in the PI domain where amino acid substitutions to positively charged residues might be expected to increase activity by forming novel nonspecific DNA contacts (Fig. 2A). We used HT-PAMDA to determine the single or combinatorial effects of three such substitutions—A61R, G1104K, and N1317R—in the context of SpG(L1111R/A1322R) variants also bearing R1333A, R1333C, or R1333P substitutions (Fig. 2C and fig. S8B). This analysis revealed that different combinations of the three nonspecific substitutions were well tolerated by nearly all variants and that the NRN PAM preferences of variants harboring R1333A, R1333C, or R1333P substitutions were similar, as demonstrated by HT-PAMDA.

To determine which variant exhibited the highest activities in human cells, we tested this large series of variants against four additional sites bearing NRN PAMs (fig. S8C). We observed that the SpG(L1111R/A1322R) variant harboring the R1333P substitution and a combination of A61R/N1317R offered the greatest mean editing against NRN PAMs (Fig. 2D). Use of HT-PAMDA to examine the sequential effects of the substitutions encoded by this variant demonstrated a stepwise progression from NGN to NRN PAM preference and also revealed the unexpected finding that this variant may target some NYN PAMs (where Y is C or T) (fig. S8D). Together, our human cell and HT-PAMDA data suggest that the SpG (L1111R/A1322R) derivative containing A61R, N1317R, and R1333P substitutions (henceforth referred to as SpRY, for SpCas9 variant capable of targeting NRN>NYN PAMs) enables targeting of sites with NRN PAMs.

SpRY activities as a nuclease, CBE, and ABE in human cells

Having established the potential of SpRY to widely expand sequence targeting, we more thoroughly assessed its nuclease activities in HEK 293T cells. We compared the on-target editing of WT SpCas9 and SpRY across 64 sites, 32 harboring NANN PAMs and 32 harboring NGNN PAMs (fig. S9, A and B, respectively). We observed that, consistent with prior reports (*7*, *8*), WT SpCas9 preferred NGG>NAG>NGA PAMs with negligible targeting of the remaining NRN PAMs (Fig. 3A). In comparison, SpRY was more effective than WT at targeting sites encoding NRN PAMs, except for those harboring canonical NGG PAMs (Fig. 3A). Across the 32 sites with NGN PAMs, SpRY often exhibited comparable activities to SpG, although SpG remained the most effective NGN PAM variant (figs. S9, B and C). Overall, SpRY was capable of efficiently editing most sites with NRN PAMs, where the range of activities could not necessarily be explained by PAM preference alone. These results demonstrate the ability to effectively target a range of sites with NAN PAMs using a Cas9 variant.

Combined with the observation of modest levels of NYN targeting with SpRY in HT-PAMDA (fig. S8D), structural analysis of the R1333P substitution in SpRY led us to speculate that R1333P-containing variants might also enable targeting of any base in the second PAM position (including thus far unexamined NYN PAMs) (fig. S9D). We examined the activities of WT SpCas9 and SpRY across 31 sites with NYN PAMs (15 NCNN and 16 NTNN sites) (fig. S9E). SpRY was able to edit 13 of 31 sites (42%) with NYN PAMs to levels of modification above 20%, compared with zero sites with WT SpCas9 (Fig. 3B). Although the mean editing activities on sites with NYN PAMs were about half of what we observed on sites with NRN PAMs, the activities were greater than the essentially negligible editing with WT SpCas9 (Fig. 3B). The PAM preference of SpRY as determined by HT-PAMDA was generally consistent with the mean nuclease editing levels for each PAM class [fig. S9F and (*26*)], although additional experiments are needed to more thoroughly characterize the substrate requirements of SpRY (fig. S9, G to J). Collectively, these results demonstrate the ability to target sites with NRN PAMs and some NYN PAMs using an SpCas9 variant.

Because SpRY enables nuclease targeting of many sites with NNN PAMs in human cells, we examined its compatibility with base editors, which are dependent upon the availability of PAMs to appropriately position the CBE or ABE edit windows (*28*). Assessment of SpRY-CBE across 14 sites bearing NRN PAMs revealed mean C-to-T editing of 38.0% across all substrate cytosines (fig. S10A), with SpRY-CBE achieving >20% modification of at least one cytosine per site for all but one site (Fig. 3C). Comparatively, WT-CBE most efficiently modified sites bearing NGG PAMs and was also capable of modifying sites bearing NAG and NGA PAMs, albeit at lower efficiency than SpRY (Fig. 3C and fig. S10A). We also assessed the activities of SpRY-CBE on five high-activity NYN PAM sites from our nuclease datasets (see fig. S9E). For these preselected high-activity sites, we observed robust levels of editing compared with negligible editing with WT-CBE (Fig. 3D and fig. S10B). Similar to our observations for SpG-CBE, SpRY-CBE exhibited typical CBE substrate preferences (fig. S10C).

Downloaded from http://science.sciencemag.org/ on February 21, 2021

We then examined the A-to-G editing activities of SpRY-ABE across 13 sites with NRN PAMs and five high-activity sites with NYN PAMs (NYN sites selected from fig. S9E). For the NRN PAM sites, we observed mean A-to-G editing activities of 34.7% with SpRY-ABE on substrate adenines (fig. S10D) and achieved >20% modification on at least one adenine for 10 of 13 sites (Fig. 3E). With WT-ABE, the most efficient editing was observed on the NGG PAM site, and minor editing was detected on three sites with noncanonical PAMs (Fig. 3E). Across the five preselected high-activity sites harboring NYN PAMs, A-to-G editing activities were generally more modest with SpRY, but

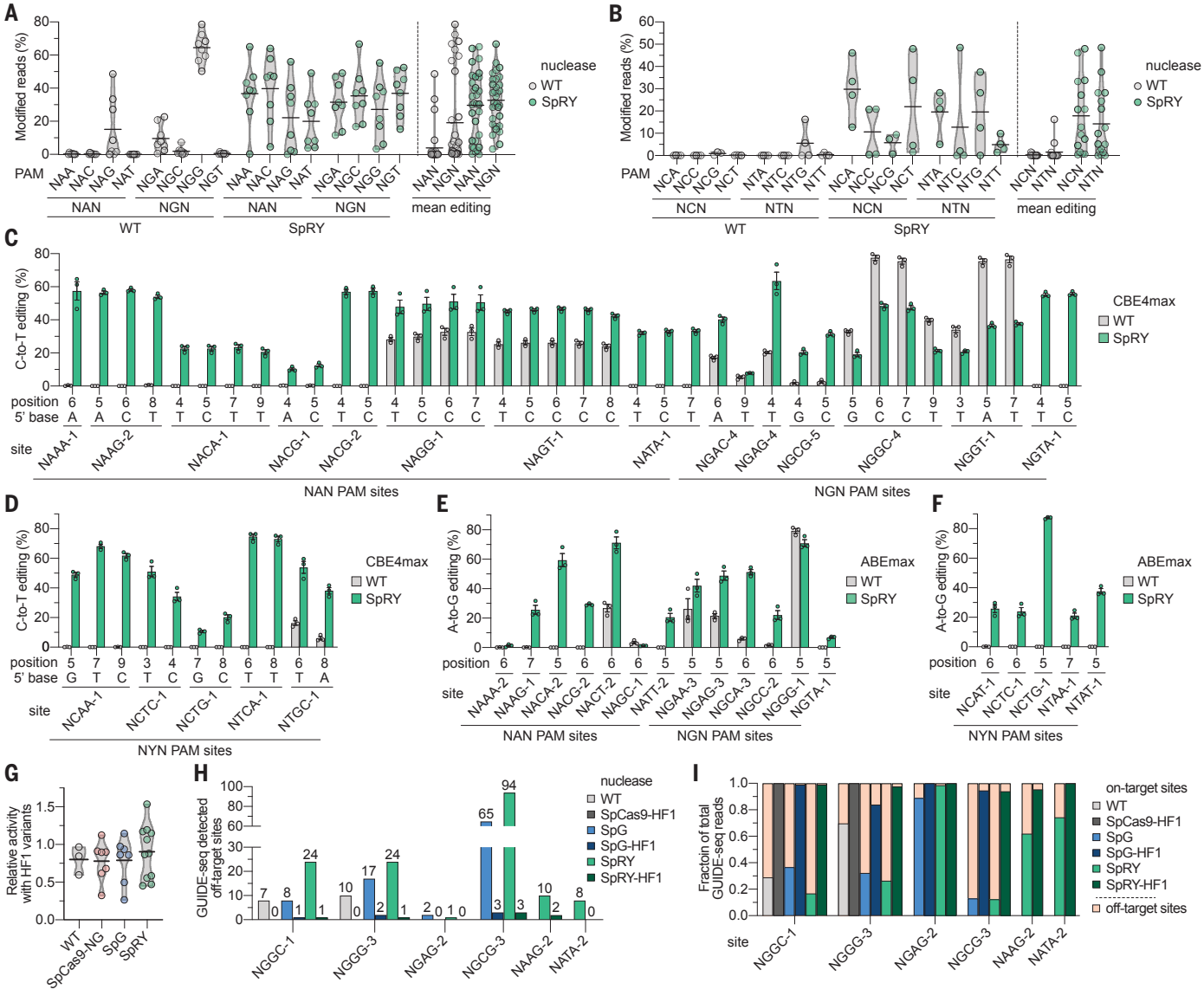


Fig. 3. Comparison of WT SpCas9 and SpRY nuclease and base editor activities across NNN PAM sites in human cells. (**A** and **B**) Mean nuclease activity plots for WT SpCas9 and SpRY on 64 sites with NRN PAMs (A) and 31 sites with NYN PAMs (B) in human cells. The horizontal black lines represent the mean of eight or three to four sites [(A) and (B), respectively] for each PAM of the indicated class (see also fig. S9, A to C), and the gray outlines are violin plots. (**C** and **D**) C-to-T base editing of endogenous sites in human cells bearing NRN and NYN PAMs [(C) and (D), respectively] with WT SpCas9 and SpRY-CBE4max constructs. Editing of cytosines in the edit window (positions three through nine) assessed by targeted sequencing; the five NYN PAM target sites were selected from high-activity sites in (B); mean, SEM, and individual data points shown for $n = 3$. (**E** and **F**) A-to-G base editing of endogenous sites in human cells bearing NRN and NYN PAMs [(E) and (F), respectively] with WT SpCas9 and SpRY-ABEmax constructs. Editing of adenines in the edit window assessed by targeted sequencing; the five NYN PAM target sites were selected from high-activity sites in (B); mean, SEM, and individual data points shown for $n = 3$. For base editing data in (C) to (F), see also table S5. (**G**) Relative nuclease activity plots for SpCas9-HF1, SpCas9-NG-HF1, SpG-HF1, and SpRY-HF1 compared with their parental variants across 3 to 10 endogenous sites in HEK 293T cells. Mean modification from sites in fig. S11A shown as dots; horizontal black lines represent the mean of those sites, and the gray outlines are violin plots. The HF1 variants additionally encode N497A, R661A, Q695A, and Q926A substitutions (*21*). (**H**) Histogram of the number of GUIDE-seq–detected off-target sites for SpCas9 variants across sites with NGG, NGN, and NAN PAMs (see fig. S12, A to C, respectively). (**I**) Fraction of GUIDE-seq reads attributed to the on- and off-target sites for WT SpCas9, SpG, SpRY, and their respective HF1 variants across two to six targets (see also fig. S12, A to C, and table S6).

Downloaded from http://science.sciencemag.org/ on February 21, 2021

we did observe editing of one adenine to near 90%; no editing with WT ABEmax was observed on any of the five sites (Fig. 3F). Overall, SpRY-ABE exhibited greater mean A-to-G editing compared with WT-ABE across sites containing NRN and NYN PAMs with typical ABE substrate requirements (fig. S10, D to F).

An important consideration for genome editing is the ability to mitigate potential off-target effects. To reduce off-target editing observed with WT SpCas9, we and others previously engineered high-fidelity (HF) variants of SpCas9 with improved genome-wide specificities (*21*, *30*). Because the relaxed PAM tolerances of SpG and SpRY can, in principle, lead to recognition of additional off-target sites, we first tested whether our variants were compatible with the fidelity-enhancing substitutions of SpCas9-HF1 (*21*). Across several target sites bearing different PAMs, we observed that WT, SpCas9-NG, SpG, SpRY, and their HF1 derivatives exhibited comparable levels of on-target modification (Fig. 3G and fig. S11A).

We then performed genome-wide, unbiased identification of double-strand breaks enabled by sequencing (GUIDE-seq) experiments (*31*) to analyze the genome-wide specificity profiles of these variants. In transfections containing the GUIDE-seq double-stranded oligodeoxynucleotide tag, we also observed similar levels of on-target editing between WT, SpG, SpRY, and their HF1 derivative variants (fig. S11, B to F). Analysis of GUIDE-seq experiments revealed that SpG and SpRY exhibited a somewhat increased propensity for off-target editing compared with WT SpCas9, albeit at similar absolute levels previously reported for WT SpCas9 (*21*, *31*) (Fig. 3H). Nearly all novel off-targets for SpG and SpRY were attributable to the expanded PAM recognition by these variants (fig. S12, A to C). Importantly, the HF1 variants were able to eliminate nearly all off-target editing events (Fig. 3H) and substantially increased the fraction of total editing events observed at the on-target sites (Fig. 3I). These results demonstrate that SpG- and SpRY-HF1 offer improved fidelity for applications that necessitate high specificity.

Expanded targeting of SpG and SpRY enables the generation of protective genetic variants

The necessity of DNA-targeting CRISPR enzymes to recognize a PAM fundamentally limits its precision targeting, a constraint that is exacerbated when using base editors whose activities are restricted to short "editing windows" (*17*, *18*, *28*) (Fig. 4A). In a proof-of-concept application, we leveraged the expanded targeting of SpG and the near-PAMless qualities of SpRY to generate biologically relevant substitutions implicated to protect individuals against coronary heart disease, type 2 diabetes, osteoporosis, chronic pain, and other pathologies (*32*–*38*). We first selected two genetic variants (*SOST* W124X and *MSTN* IVS1) with nearby NGG PAMs that might appropriately position the CBE edit window (Fig. 4B). We assessed the activities of WT-, SpG-, and SpRY-CBE constructs against target sites with a variety of PAMs and, as expected, observed the most-robust editing with WT-CBE on sites with NGG PAMs. However, for *MSTN* IVS1, we also observed potentially deleterious bystander editing of a nearby cytosine with WT-CBE; this collateral edit was avoided or reduced when using SpG- and SpRY-CBEs targeted to nearby sites bearing NGC and NAT PAMs (Fig. 4B).

Next, we systematically evaluated the capability to generate eight additional protective variants that lack nearby canonical NGG PAMs (Fig. 4C). We examined target sites harboring NRN PAMs using WT-, SpG-, and SpRY-CBEs

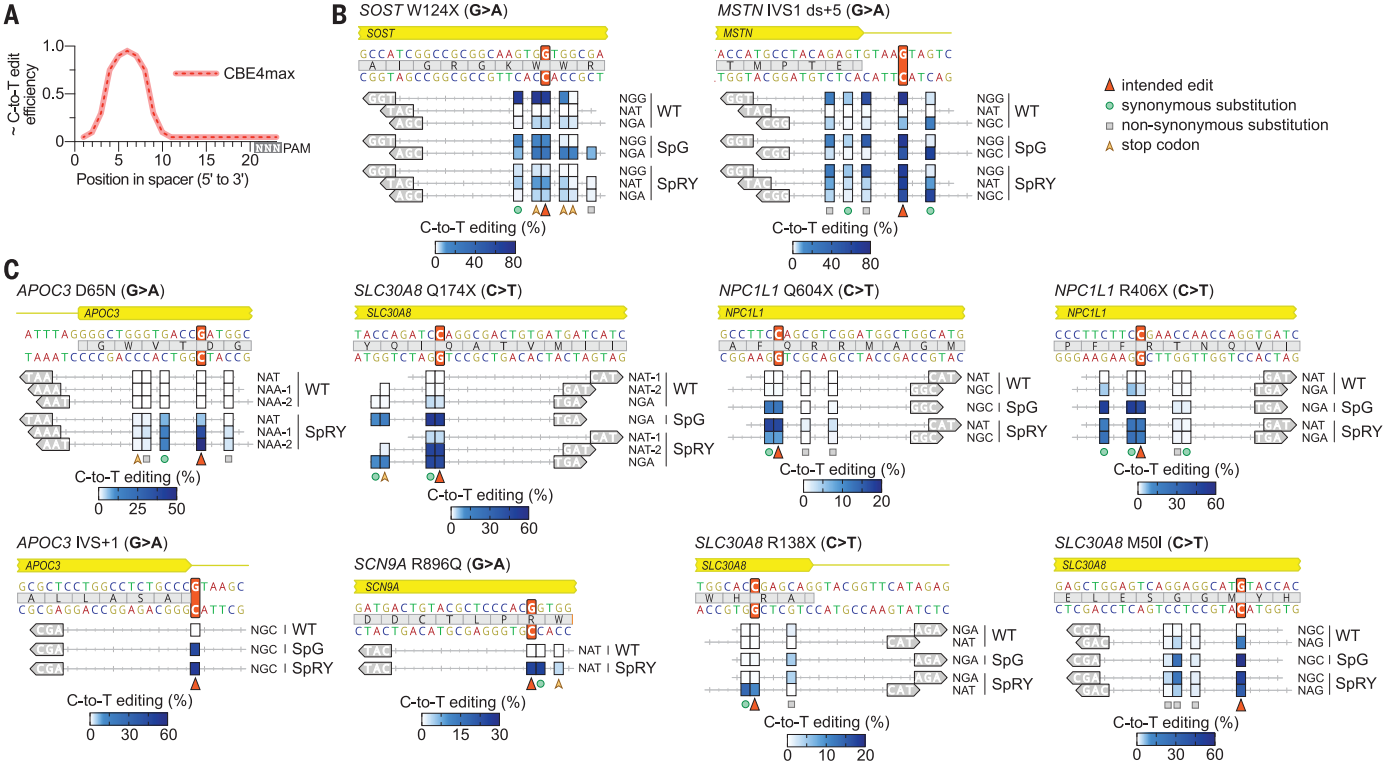


Fig. 4. Expanded capabilities of C-to-T base editors with SpG and SpRY to generate protective genetic variants. (**A**) Illustration of the relative C-to-T edit activity of CBE4max constructs depending on the position of the cytosine in the spacer [as reviewed in (*28*)]. (**B** and **C**) Comparison of the C-to-T editing activities of WT SpCas9, SpG, and SpRY CBE4max constructs across 22 target sites covering 10 previously described protective genetic variants accessible or inaccessible with target sites harboring NGG PAMs [(B) and (C), respectively]. The intended edit is highlighted with a dark-orange arrow. C-to-T editing for cytosines within the spacer that are edited above 1% by any variant is plotted for all appropriate variant and guide combinations. SpG was tested only on sites harboring NGN PAMs. Editing of cytosines assessed by targeted sequencing with mean C-to-T editing is shown for $n = 3$. The bystander synonymous, nonsynonymous, and stop codon C-to-T edits are indicated. The PAMs for each target site are shown in the gray arrow annotation. For raw data, see table S5.

Downloaded from http://science.sciencemag.org/ on February 21, 2021

that would position the intended C-to-T edit within the high-activity CBE edit window (Fig. 4A). With WT-CBE, we were unable to efficiently generate the intended edit for seven of eight substitutions, owing to the absence of nearby canonical PAMs, which reinforces the need for variants with expanded targeting capabilities. With SpG- and SpRY-CBEs, we screened additional target sites for each substitution and observed that we could efficiently introduce the intended C-to-T edit across all eight targets (Fig. 4C). For most genetic variants, we were able to avoid deleterious bystander editing with SpG and SpRY by selecting from multiple targetable sites that produce silent or tolerable collateral edits instead of those that cause nonsynonymous C-to-T changes. These results demonstrate the utility of SpG and SpRY for higher-resolution targeting of previously uneditable genomic sites and the capability to examine additional target sites to avoid detrimental bystander edits.

Outlook

While the PAM requirement of CRISPR systems enables bacteria to distinguish self from nonself, for genome editing applications, the necessity of PAM recognition constrains CRISPR-Cas systems for use across genomic loci that lack or only sparsely encode PAMs. The SpG and SpRY variants circumvent this limitation by relaxing or almost entirely removing the dependence of SpCas9 on a requisite PAM, extending targeting to sites with NGN and NAN PAMs and to many sites with NCN or NTN PAMs, albeit at a reduced relative efficiency. These variants should enable unconstrained targeting for a variety of applications that require precise DNA breaks, nicks, deamination, or binding events. Beyond the experiments presented here in HEK 293T cells, the utility of SpG and SpRY across different applications and delivery contexts requires additional investigation (e.g., for tiling putative regulatory elements in various cell types).

In principle, the strategy we used to reduce or eliminate the PAM requirement should be applicable to other native or engineered CRISPR-Cas9 and CRISPR-Cas12a orthologs for which structural information is available. Although future studies are needed to more precisely elucidate the molecular roles of the amino acid substitutions in SpG and SpRY, we speculate that SpRY achieves its expanded targeting range through the removal of the canonical base-specific interactions, displacement of the PAM DNA to facilitate interactions in the major groove of the PAM, and energetic compensation by the addition of novel nonspecific protein–DNA contacts. More practically, regarding which enzyme to use for experiments requiring on-target activity, we suggest WT SpCas9 for sites harboring NGG PAMs, SpG for NGH PAMs, and SpRY for targets encoding the remaining NHN PAMs (with NAN being preferable to NCN or NTN).

The potential for undesirable off-target effects requires methods to mitigate them. As observed when developing engineered CRISPR-Cas12a and CRISPR-Cas9 enzymes with expanded PAM tolerances, relaxation of the PAM can reduce specificity (*22*, *39*). However, both enAsCas12a and SpCas9-NG were compatible with substitutions that enhance genome-wide specificity. Similarly, SpG and SpRY are compatible with SpCas9-HF1 substitutions, which eliminate nearly all detectable off-target effects as determined by GUIDE-seq, enabling applications that require higher fidelity.

By developing SpCas9 variants capable of high-resolution editing, we demonstrate that protein engineering can eliminate biological constraints that limit applications of CRISPR-Cas enzymes. With SpRY supporting the editing of many sites containing NRN>NYN PAMs, nearly all majority of the genome is now targetable.

REFERENCES AND NOTES

1. L. A. Marraffini, E. J. Sontheimer, *Nature* **463**, 568–571 (2010).
2. R. Heler *et al*., *Nature* **519**, 199–202 (2015).
3. H. Deveau *et al*., *J. Bacteriol.* **190**, 1390–1400 (2008).
4. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, *Microbiology* **155**, 733–740 (2009).
5. M. Jinek *et al*., *Science* **337**, 816–821 (2012).
6. S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, J. A. Doudna, *Nature* **507**, 62–67 (2014).
7. W. Jiang, D. Bikard, D. Cox, F. Zhang, L. A. Marraffini, *Nat. Biotechnol.* **31**, 233–239 (2013).
8. B. P. Kleinstiver *et al*., *Nature* **523**, 481–485 (2015).
9. K. M. Esvelt *et al*., *Nat. Methods* **10**, 1116–1121 (2013).
10. I. Fonfara *et al*., *Nucleic Acids Res.* **42**, 2577–2590 (2014).
11. T. Karvelis *et al*., *Genome Biol.* **16**, 253 (2015).
12. B. Zetsche *et al*., *Cell* **163**, 759–771 (2015).
13. B. Elliott, C. Richardson, J. Winderbaum, J. A. Nickoloff, M. Jasin, *Mol. Cell. Biol.* **18**, 93–101 (1998).
14. G. M. Findlay, E. A. Boyle, R. J. Hause, J. C. Klein, J. Shendure, *Nature* **513**, 120–123 (2014).
15. M. C. Canver *et al*., *Nature* **527**, 192–197 (2015).
16. J. Shi *et al*., *Nat. Biotechnol.* **33**, 661–667 (2015).
17. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, *Nature* **533**, 420–424 (2016).
18. N. M. Gaudelli *et al*., *Nature* **551**, 464–471 (2017).
19. Y. Kan, B. Ruis, T. Takasugi, E. A. Hendrickson, *Genome Res.* **27**, 1099–1111 (2017).
20. C. Anders, O. Niewoehner, A. Duerst, M. Jinek, *Nature* **513**, 569–573 (2014).
21. B. P. Kleinstiver *et al*., *Nature* **529**, 490–495 (2016).
22. H. Nishimasu *et al*., *Science* **361**, 1259–1262 (2018).
23. J. H. Hu *et al*., *Nature* **556**, 57–63 (2018).
24. C. Anders, K. Bargsten, M. Jinek, *Mol. Cell* **61**, 895–902 (2016).
25. S. Hirano, H. Nishimasu, R. Ishitani, O. Nureki, *Mol. Cell* **61**, 886–894 (2016).
26. See supplementary materials.
27. K. Suzuki *et al*., *Nature* **540**, 144–149 (2016).
28. H. A. Rees, D. R. Liu, *Nat. Rev. Genet.* **19**, 770–788 (2018).
29. L. W. Koblan *et al*., *Nat. Biotechnol.* **36**, 843–846 (2018).
30. I. M. Slaymaker *et al*., *Science* **351**, 84–88 (2016).
31. S. Q. Tsai *et al*., *Nat. Biotechnol.* **33**, 187–197 (2015).
32. W. Balemans *et al*., *Hum. Mol. Genet.* **10**, 537–543 (2001).
33. M. Schuelke *et al*., *N. Engl. J. Med.* **350**, 2682–2688 (2004).
34. J. J. Cox *et al*., *Hum. Mutat.* **31**, E1670–E1686 (2010).
35. J. Flannick *et al*., *Nat. Genet.* **46**, 357–363 (2014).
36. TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute, *N. Engl. J. Med.* **371**, 22–31 (2014).
37. A. R. Harper, S. Nayee, E. J. Topol, *Nat. Rev. Genet.* **16**, 689–701 (2015).
38. Myocardial Infarction Genetics Consortium Investigators, *N. Engl. J. Med.* **371**, 2072–2082 (2014).
39. B. P. Kleinstiver *et al*., *Nat. Biotechnol.* **37**, 276–282 (2019).
40. K. Clement *et al*., *Nat. Biotechnol.* **37**, 224–226 (2019).
41. R. T. Walton, B. P. Kleinstiver, Scripts for analyzing High-Throughput PAM Determination Assay (HT-PAMDA) experimental data for CRISPR enzymes, Version 1, Zenodo (2020); http://doi.org/10.5281/zenodo.3710516.

ACKNOWLEDGMENTS

We thank S. Tsai and J. Lopez for advice on GUIDE-seq data analysis; J. Hsu for assistance with the PAMDA scripts; J. Grünewald for advice on NextSeq sequencing, Z. Hebert and the Molecular Biology Core Facilities (MBCF) at Dana-Farber Cancer Institute (DFCI) for support with NextSeq sequencing; K. Clement for guidance with CRISPResso2; S. Johnstone and B. Bernstein for assistance with and access to their Covaris E220; D. Gao for bioinformatic support; M. Prew and M. Welch for providing guide RNA plasmids; S. Katherisan for advice on protective variants; and J. K. Joung for suggestions on the manuscript. **Funding:** Research in the Kleinstiver lab was supported by NIH R00-CA218870, NIH P01-HL142494, a Career Development Award from the American Society of Gene & Cell Therapy, and the Margaret Q. Landenberger Research Foundation. **Author contributions:** R.T.W. and B.P.K. conceived of the study. R.T.W., K.A.C., and M.N.W. performed experiments. All authors analyzed data and interpreted results. R.T.W. and B.P.K. wrote the manuscript with input from all authors. **Competing interests:** R.T.W. and B.P.K. are inventors on various patents and patent applications that describe gene editing and epigenetic editing technologies. B.P.K. is a consultant for Avectas Inc. and an adviser for Acrigen Biosciences. **Data and materials availability:** Plasmids generated in this study are available through Addgene (www.addgene.org/Benjamin_Kleinstiver/; see also table S2). All data are available in the manuscript, in the supplementary materials, or in the following databases: High-throughput sequencing datasets from nuclease, CBE, ABE, HT-PAMDA, HT-PAMDA-CBE, and GUIDE-seq experiments are available through the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) under BioProject ID PRJNA605711 (www.ncbi.nlm.nih.gov/bioproject/605711). The commands for visualizing crystal structures in PyMOL are available in (*26*). CRISPResso2 (*40*) was used to analyze nuclease and base editor data (with nonstandard parameters detailed in the methods section of the supplementary materials). Custom Python scripts used to analyze HT-PAMDA and HT-PAMDA-CBE data are available through Zenodo (*41*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6488/290/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Tables S1 to S7
References (*42*–*46*)

View/request a protocol for this paper from *Bio-protocol*.

25 January 2020; accepted 16 March 2020
Published online 26 March 2020
10.1126/science.aba8853

Downloaded from http://science.sciencemag.org/ on February 21, 2021

Science

Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants

Russell T. Walton, Kathleen A. Christie, Madelynn N. Whittaker and Benjamin P. Kleinstiver

Science **368** (6488), 290-296.
DOI: 10.1126/science.aba8853originally published online March 26, 2020

A PAMless base editor

CRISPR-Cas DNA base editing typically requires a specific motif for targeting known as a protospacer-adjacent motif (PAM). This requirement limits the sequences within a genome that can be targeted. Walton *et al.* engineered specific variants of the *Streptococcus pyogenes* Cas9 enzyme named SpG and SpRY that could recognize and edit a wider array of PAMs. Using SpRY, the authors were able to correct previously uneditable mutations associated with human disease. Although off-target effects were observed for these engineered Cas enzymes at levels similar to those of the wild-type enzyme, depending on the context, these engineered enzymes widen the potential applications of precision genome editing.

Science, this issue p. 290

ARTICLE TOOLS http://science.sciencemag.org/content/368/6488/290

SUPPLEMENTARY MATERIALS http://science.sciencemag.org/content/suppl/2020/03/25/science.aba8853.DC1

RELATED CONTENT
http://stm.sciencemag.org/content/scitransmed/12/536/eaay0271.full
http://stm.sciencemag.org/content/scitransmed/11/503/eaaw3768.full
http://stm.sciencemag.org/content/scitransmed/10/449/eaao3240.full
http://stm.sciencemag.org/content/scitransmed/9/372/eaah3480.full

REFERENCES This article cites 46 articles, 7 of which you can access for free
http://science.sciencemag.org/content/368/6488/290#BIBL

PERMISSIONS http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works

Downloaded from http://science.sciencemag.org/ on February 21, 2021