Science
AAAS

# Supplementary Materials for

# Unconstrained Genome Targeting with near-PAMless Engineered CRISPR-Cas9 Variants

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

Correspondence to: bkleinstiver@mgh.harvard.edu

**This PDF file includes:**

**Materials and Methods:**

Plasmids and oligonucleotides

Brief descriptions of plasmids used in this study can be found in **Table S2**, and new plasmids have been deposited with Addgene (https://www.addgene.org/Benjamin_Kleinstiver/). Target site sequences for sgRNAs and oligonucleotide sequences are available in **Tables S1** and **S3**, respectively. The SpCas9 nuclease human expression plasmid was generated by subcloning the SpCas9 open reading frame from pX330 (Addgene plasmid 42230) into the NotI and AgeI sites of JDS246 (Addgene plasmid 43861). Nuclease constructs harboring a C-terminal BP(SV40)NLS-3xFLAG-P2A-EGFP sequence were utilized for all human cell experiments unless otherwise indicated. Cytidine base editor (CBE) constructs were generated by subcloning the open reading frame of BE4max (Addgene plasmid 112099) into the NotI and AgeI sites of pCAG-CFP (Addgene plasmid 11179). Adenine base editor (ABE) variants were generated by modifying ABEmax (Addgene plasmid 112101). All modifications to plasmids, including generation of point mutations, altered nuclear localization architectures, and the addition of P2A-EGFP were generated through standard molecular cloning and isothermal assembly. Human cell expression plasmids for U6 promoter-driven SpCas9 sgRNAs were generated by annealing and ligating duplexed oligonucleotides corresponding to spacer sequences into BsmBI-digested BPK1520(*8*). Plasmids for in vitro transcription of SpCas9 sgRNAs were generated by annealing and ligating oligonucleotides corresponding to spacer sequence duplexes into BsaI-digested MSP3485 for T7 promoter-driven transcription of sgRNAs.

Plasmid libraries with 8 nt randomized PAM sequences on the 3' end of the target sites were generated from two oligonucleotides encoding separate spacer sequences (**Table S3**), similar to as previously described(*39*). Briefly, Klenow(-exo) (New England Biolabs, NEB) was used to generate the bottom strand of the dsDNA sequence, and the product was digested with EcoRI prior to ligation into EcoRI and SphI digested p11-lacY-wtx1 (Addgene plasmid 69056). Ligated plasmids were transformed into electrocompetent XL1-Blue *E. coli*, recovered in 9 ml of super optimal broth with catabolite repression (SOC) at 37 °C for approximately 60 minutes, and then grown for 16 hours in 150 mL of Luria-Bertani (LB) medium with 100 µg/mL carbenicillin. The complexity of each library was estimated to be greater than $10^5$ unique PAMs based on the number of transformants. Plasmid libraries were linearized with PvuI (NEB) prior to use in the *in vitro* cleavage reactions.

Structural modeling of SpCas9

The crystal structures of WT SpCas9 (PDB:4UN3)(*20*), SpCas9-VQR (PDB:5B2R)(*25*), and SpCas9-VRER (PDB:5B2T)(*25*) were visualized using PyMOL version 2.3.3 (for commands see **10**).

Human cell culture

Human HEK 293T cells (ATCC) were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% heat-inactivated FBS (HI-FBS) and 1% penicillin/streptomycin. The supernatant media from cell cultures was analyzed monthly for the presence of mycoplasma using MycoAlert PLUS (Lonza).

Transfection of human cells

All experiments were performed with at least 3 independent biological replicates. For all human cell experiments, transfections were performed between 20 and 24 hours following seeding of $2x10^4$ HEK 293T cells per well in 96-well plates. For nuclease experiments, 29 ng of nuclease and 12.5 ng of sgRNA expression plasmids (unless otherwise indicated) were mixed with 0.3 µL of TransIT-X2 (Mirus) in a total volume of 15 µL Opti-MEM (Thermo Fisher Scientific), incubated for 15 minutes at room temperature, and added to HEK 293T cells. For CBE and ABE experiments, 70 ng of base-editor and 30 ng of sgRNA expression plasmids were mixed with 0.72 µL of TransIT-X2 in a total volume of 15 µL Opti-MEM, incubated for 15 minutes at room temperature, and added to HEK 293T cells. Nuclease and CBE experiments were halted after 72 hours, and ABE experiments after 120 hours. Genomic DNA was collected by discarding the media, resuspending the cells in 100 µL of quick lysis buffer (20 mM Hepes pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 5% glycerol, 25 mM DTT, 0.1% Triton X-100, and 60 ng/ul Proteinase K (NEB)), heating the lysate for 6 minutes at 65 ºC, heating at 98 ºC for 2 minutes, and then storing at -20 ºC.

For GUIDE-seq experiments, 29 ng of nuclease and 12.5 ng of sgRNA expression plasmids, 1 pmol of the GUIDE-seq double-stranded oligodeoxynucleotide (dsODN; oSQT685/686) tag(*31*), and 0.3 µL of TransIT-X2 (Mirus) were mixed in a total volume of 16 µL Opti-MEM, incubated for 15 minutes at room temperature, and added to HEK 293T cells. Genomic DNA was extracted ~72 hours post-transfection by discarding the media, resuspending the cells in 100 µL of overnight lysis buffer (100 mM Tris pH 7.5, 100 mM NaCl, 5 mM EDTA, 0.05% SDS, 5 µL PK (NEB), and 25 mM DTT), and incubating lysate at 55°C for 15-18 hours shaking at

approximately 200 rpm. Following incubation, genomic DNA was purified using a 0.7x ratio of paramagnetic beads prepared as previously described(*39*, *42*).

Assessment of nuclease and base editor activities in human cells

The efficiency of genome modification by nucleases, CBEs, and ABEs were determined by next-generation sequencing using a 2-step PCR-based Illumina library construction method. Briefly, genomic loci were amplified from approximately 100 ng of genomic DNA using Q5 High-fidelity DNA Polymerase (NEB) and the primers listed in **Table S3**. PCR products were purified using paramagnetic beads prepared as previously described(*39*, *42*). Approximately 20 ng of purified PCR product was used as template for a second PCR to add Illumina barcodes and adapter sequences using Q5 and the primers listed in **Table S3**. PCR products were purified prior to quantification via capillary electrophoresis (Qiagen QIAxcel), normalization, and pooling. Final libraries were quantified by qPCR (Illumina Library qPCR Quantification Kit, KAPA Biosystems) and sequenced on a MiSeq sequencer using a 300-cycle v2 kit (Illumina).

On-target genome editing activities were determined from sequencing data using CRISPResso2(*40*) in pooled mode with custom input parameters for nucleases: --min_reads_to_use_region 100; for CBEs: --min_reads_to_use_region 100 -w 20 --cleavage_offset -10 --base_editor_output; for ABEs: min_reads_to_use_region 100 -w 20 --cleavage_offset -10 --base_editor_output --conversion_nuc_from A --conversion_nuc_to G. For quantifying GUIDE-seq on-target dsODN tag integration, CRISPResso2 was run in non-pooled mode by supplying the target site spacer, the reference amplicon, and both the forward and reverse dsODN-containing sequences as 'HDR' alleles with custom parameters: -w 25 -g GUIDE --plot_window_size 50. Total dsODN capture was calculated by dividing the sum of the fwd-dsODN modified and unmodified reads + the rev-dsODN modified and unmodified reads by the total number of reads; cumulative on-target editing by was calculated by dividing the sum of the wild-type modified reads, the fwd-dsODN modified and unmodified reads, and the rev-dsODN modified and unmodified reads by the total number of reads for each replicate sample.

GUIDE-seq samples were prepared for sequencing as previously described(*31*) and sequenced on an Illumina NextSeq sequencer in manual mode for custom Index2 read length. Binary base call files were converted to fastq format using bcl2fastq v2.17.1.14. GUIDE-seq data was analyzed using *guideseq* v1.0.2 (https://github.com/aryeelab/guideseq) with custom input parameters: demultiplex_min_reads 500,000 for all

nucleases; max_mismatches 6 and an NGG PAM for WT SpCas9 samples; max_mismatches 7 and an NGN PAM for SpG samples; and max_mismatches 8 and an NNN PAM for SpRY samples; cell-type specific SNP correction was not performed.

Summaries of genome editing nuclease and GUIDE-seq on-target dsODN tag integration data are found in **Table S4**; summaries of CBE and ABE data are found in **Table S5**; and summaries of GUIDE-seq data are found in **Table S6**. The edit window for base editor constructs was defined as PAM-distal spacer positions 3-9 for CBEs and positions 5-7 for ABEs.

*in vitro* transcription of sgRNAs

SpCas9 sgRNAs were *in vitro* transcribed at 37 °C for 16 hours from roughly 1 µg of HindIII linearized sgRNA T7-transcription plasmid template (cloned into MSP3485) using the T7 RiboMAX Express Large Scale RNA Production Kit (Promega). The DNA template was degraded by the addition of 1 µL RQ1 DNase at 37 °C for 15 minutes. sgRNAs were purified with the MEGAclear Transcription Clean-Up Kit (ThermoFisher) and refolded by heating to 90 ºC for 5 minutes and then cooling to room temperature for 15 minutes.

Expression of SpCas9 and base editor proteins in human cells and normalization of lysates

To generate SpCas9 and variant proteins from human cell lysates, approximately 20-24 hours prior to transfection $1.5x10^5$ HEK 293T cells were seeded in 24-well plates. Transfections containing 500 ng of human codon optimized nuclease expression plasmid (with a -P2A-EGFP signal) and 1.5 µL TransIT-X2 were mixed in a total volume of 50 µL of Opti-MEM, incubated at room temperature for 15 minutes, and added to the cells. The lysate was harvested after 48 hours by discarding the media and resuspending the cells in 100 ul of gentle lysis buffer (1X SIGMAFAST Protease Inhibitor Cocktail, EDTA-Free (Millipore Sigma), 20 mM Hepes pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 5% glycerol, 1 mM DTT, and 0.1% Triton X-100). The amount of SpCas9 or base editor protein was approximated from the whole-cell lysate based on EGFP fluorescence. SpCas9 nuclease and base editor lysates were normalized to 150 and 300 nM Fluorescien (Sigma), respectively, based on a Fluorescein standard curve. Fluorescence was measured in 384-well plates on a DTX 880 Multimode Plate Reader (Beckman Coulter) with $\lambda_{ex}$ = 485 nm and $\lambda_{em}$= 535 nm.

High-throughput PAM Determination assay for nucleases

The high-throughput PAM determination assay (HT-PAMDA) was performed using linearized randomized PAM-containing plasmid substrates that were subject to *in vitro* cleavage reactions with SpCas9 and variant proteins. First, SpCas9 ribonucleoproteins (RNPs) were complexed by mixing 4.375 µL of normalized whole-cell lysate (150 nM Fluorescien) with 8.75 pmol of *in vitro* transcribed sgRNA and incubating for 5 minutes at 37 °C. Cleavage reactions were initiated by the addition of 43.75 fmol of randomized-PAM plasmid library and buffer to bring the total reaction volume to 17.5 µL with a final composition of 10 mM Hepes pH 7.5, 150 mM NaCl, and 5 mM $MgCl_2$. Reactions were performed at 37 °C and aliquots were terminated at timepoints of 1, 8, and 32 minutes by removing 5 µL aliquots from the reaction and mixing with 5 µL of stop buffer (50 mM EDTA and 2 mg/ml Proteinase K (NEB)), incubating at room temperature for 10-minutes, and heat inactivating at 98 °C for 5 minutes. For all variants characterized, time courses were completed on both libraries harboring distinct spacer sequences for n = 2; several variants were characterized with additional replicates to evaluate reproducibility of the assay (e.g. **Fig. S2C**), where for those variants the final data is an average of all replicates.

Next, approximately 3 ng of digested PAM library for each SpCas9 variant and reaction timepoint was PCR amplified using Q5 polymerase (NEB) and barcoded using unique combinations of the i5 and i7 primers listed in **Table S3**. PCR products were pooled for each time point, purified using paramagnetic beads, and prepared for sequencing using one of two library preparation methods. Pooled amplicons were prepared for sequencing using either (1) the KAPA HTP PCR-free Library Preparation Kit (KAPA BioSystems), or (2) a PCR-based method where pooled amplicons were treated with Exonuclease I, purified using paramagnetic beads, amplified using Q5 polymerase and the primers listed in **Table S3** with approximately 250 pg of pooled amplicons at template, and again purified using paramagnetic beads. Libraries constructed via either method were quantified using the Universal KAPA Illumina Library qPCR Quantification Kit (KAPA Biosystems) and sequenced on a NextSeq sequencer using either a 150-cycle (method 1) or 75-cycle (method 2) NextSeq 500/550 High Output v2.5 kits (Illumina). Identical cleavage reactions prepared and sequenced via either library preparation method did not exhibit substantial differences.

Sequencing reads were analyzed using custom Python scripts (*41*) to determine cleavage rates for all SpCas9 nucleases on each substrate with unique spacers and PAMs, similar to as previously described(*39*). Briefly, reads were assigned to specific SpCas9 variants based on based on custom pooling barcodes (see **Table S3**),

assigned timepoints based on the combination of i5 and i7 primer barcodes, assigned to a plasmid library based on the spacer sequence, and assigned to a 3 (NNNN) or 4 (NNNN) nt PAM based on the identities of the DNA bases adjacent to the spacer sequence. Counts for all PAMs were computed for every SpCas9 variant (**Table S7**), plasmid library, and timepoint, corrected for inter-sample differences in sequencing depth, converted to a fraction of the initial representation of that PAM in the original plasmid library (as determined by an untreated control), and then normalized to account for the increased fractional representation of uncut substrates over time due to depletion of cleaved substrates (by selecting the five PAMs with the highest average fractional representation across all time points to represent the profile of uncleavable substrates). The depletion of each PAM over time was then fit to an exponential decay model ($y(t) = Ae^{-kt}$, where y(t) is the normalized PAM count (**Table S7**), t is the time (seconds), k is the rate constant, and A is a constant), by nonlinear regression. Reported rates are the average across both spacer sequences and across technical replicates when performed. Nonlinear least squares curve fitting was utilized to model Cas9 nuclease and CBE activities, whereas linear least squares curve fitting was previously used for our Cas12a PAMDA assay(*39*).

CBE-HT-PAMDA

The cytosine base editor high-throughput PAM determination assay (CBE-HT-PAMDA) was performed using a linearized randomized PAM-containing plasmid library that was subjected to *in vitro* reactions with base editor variants. First, base editor proteins were complexed with sgRNAs by mixing 8.75 µL of normalized whole-cell lysate (300 nM Fluorescien) with 14 pmol of *in vitro* transcribed sgRNA and incubating for 5 minutes at 37 °C. Cleavage reactions were initiated by the addition of 43.75 fmol of randomized-PAM plasmid library and buffer to bring the total reaction volume to 17.5 µL with a final composition of 10 mM Hepes pH 7.5, 150 mM NaCl, and 5 mM $MgCl_2$. Reactions were performed at 37 °C and aliquots were terminated at timepoints of 4, 32, and 256 minutes by removing 5 µL aliquots from the reaction and mixing with 5 µL of stop buffer (50 mM EDTA and 2 mg/ml Proteinase K (NEB)), incubating at room temperature for 10-minutes, and heat inactivating at 98 °C for 5 minutes. Deamination and nicking events were converted to double strand breaks through the addition of 1 unit of USER enzyme (NEB) in 5 µL of 1x NEB buffer 4 to each reaction, bringing the total volume to 15 µL. After an hour incubation at 37 ºC, reactions were stopped by adding of 5 ul of 4 mg/mL Proteinase K in 1 mM Tris pH 8.0, incubating at room temperature for 10-minutes, and heat inactivating at 98 °C for 5 minutes. Reactions were carried

out on a single plasmid library for each base editor. Samples were subsequently processed as described above for HT-PAMDA for nucleases, with the exception that depletion rates are for a single spacer sequence for CBE-HT-PAMDA, rather than the average of two spacer sequences as in the nuclease analysis. The CBE-HT-PAMDA data is included in **Table S7**.

5

## Supplementary Text

**1.** Rationale for targeted engineering of SpCas9 PAM preference

Crystal structures of wild-type (**WT**) SpCas9 have elucidated the molecular mechanism of PAM recognition by SpCas9, which occurs via bidentate hydrogen bonds between R1333 and R1335 residues in the PAM-interacting (**PI**) domain and $dG_2$ and $dG_3$ of the NGG PAM, respectively(*20*) (**Fig. S1A**). We and others have previously shown that modifications to R1333 or R1335 abrogate nuclease activity(*8*, *20*), confirming their indispensable roles for SpCas9 PAM recognition and activity. These observations suggested that other residues must be altered in concert with changes to R1333 and R1335 to retain SpCas9 function while altering PAM specificity. Our previous work to evolve SpCas9 variants that target non-canonical PAMs utilized an unbiased mutagenic approach and identified other PAM-proximal amino acids whose substitution were critical to for altering activity (including D1135, S1136, G1218, E1219, and T1337)(*8*, *21*). The SpCas9-VQR (D1135V/R1335Q/T1337R), VRER (D1135V/G1218R/R1335E/T1337R), and VRQR (D1135V/G1218R/R1335Q/T1337R) variants achieve efficient targeting of non-canonical PAMs through the coordinated substitution of R1335 (to alter the 3$^{rd}$ PAM position preference) along with other amino acids mutations in these neighboring positions(*8*, *21*).

Structural studies of SpCas9-VQR and VRER revealed the mechanisms of non-canonical PAM recognition by these variants(*24*, *25*). The formation of new base-specific contacts through the R1335Q and R1335E mutations are essential, but not sufficient, for altering recognition of the 3$^{rd}$ position of the PAM. For example, structures of SpCas9-VQR bound to an NGAG PAM revealed bidentate hydrogen bonds between R1335Q and $dA_3$ of the non-target strand (**NTS**) (**Fig. S1B**). However, the shortened sidechain length of R1335Q necessitates a 1.5 Å displacement of the PAM DNA essential for this interaction (**Fig. S1C**), a molecular rearrangement supported by the D1135V substitution in the minor groove(*8*, *24*, *25*). Additionally, the T1337R substitution of SpCas9-VQR can form base-specific hydrogen bond contacts with $dG_4$ of the NTS, contributing to PAM DNA displacement (**Fig. S1D**). Crystal structures of SpCas9-VRER, an SpCas9 variant that we previously engineered to target NGCG PAM sequences, suggest that the G1218R mutation can form electrostatic interactions with the DNA phosphate group between the third and fourth nucleotides of the NTS(*8*, *24*, *25*) (**Fig. S1E**). We speculate that the G1218R substitution of SpCas9-VRQR might perform a similar function.

Importantly, for both SpCas9-VQR and SpCas9-VRER no single substitution altered PAM preference while maintaining potent activity, suggesting a strong interdependence and co-evolutionary relationship of the residues surrounding the PAM DNA bases for PAM recognition(*8*). Together, our previous engineering studies and subsequent structural work on SpCas9 PAM variants suggest three important considerations and mechanisms for engineering SpCas9 PAM preference: (1) amino acid substitutions that create novel base-specific contacts, (2) substitutions that displace the PAM DNA bases to accommodate novel base-specific contacts, and (3) the addition of non-specific contacts to stabilize PAM binding. Furthermore, the observation that individual substitutions did not generate functional variants with altered PAM preferences foretold the necessity of a higher-throughput method to analyze larger collections of variants bearing more complex combinations of substitutions.

**2.** Hypotheses for engineering an SpCas9 variant with a relaxed PAM preference

To engineer a more broadly targeting SpCas9 variant, we focused on modifying six PAM-proximal residues (D1135, S1136, G1218, E1219, R1335, and T1337). We utilized SpCas9-VRQR as a scaffold for our engineering approach to relax PAM preference since it already possessed a somewhat relaxed PAM preference of NGA>NGNG, it displayed improved activities relative to SpCas9-VQR, and because we could leverage the structural studies of SpCas9-VQR and VRER to infer potential mechanisms of PAM recognition(*8*, *21*, *24*, *25*).

Through structure-motivated substitutions of the six PAM-proximal amino acids, we determined how their alteration could lead to variants with relaxed PAM preferences. Because SpCas9-VRQR had demonstrated the ability to target NGNG PAMs (and thus possessed a relaxed tolerance in the 3$^{rd}$ position of the PAM), we elected to maintain the R1335Q substitution of SpCas9-VRQR while varying the other five positions. Since a D1135V substitution contributes to the displacement of the PAM DNA bases, we hypothesized that we could tune the displacement of the PAM bases with a combination of hydrophobic substitutions at D1135 and S1136 and that modulating this displacement could facilitate interactions within the major groove (**Fig. S1C**). We also sought to form non-specific contacts to the DNA by varying the identity of G1218, similar to those formed by G1218R in SpCas9-VRQR and -VRER (**Fig. S1E**). In SpCas9-VQR the E1219 side chain forms hydrogen bonds that stabilize the R1335Q-d$A_3$ interaction (**Figs. S1B** and **S1F**), suggesting that E1219 may prevent R1335Q from adopting alternate conformations capable of accommodating different bases in the 3$^{rd}$ position of the PAM. We hypothesized that a more electrostatically neutral E1219Q substitution might support alternative conformations of R1335Q that

could lead to a more relaxed tolerance within the 3rd PAM position (**Figs. S1F-S1I**). Finally, we hypothesized that positively charged arginine or lysine substitutions at position T1337 might form compensatory contacts to stabilize PAM DNA binding and contribute to the displacement of the DNA towards R1335Q.

**3.** Optimization and validation of HT-PAMDA

To facilitate a large-scale rational engineering approach to develop SpCas9 variants capable of targeting NGN PAM sequences, we required a high-throughput PAM determination assay (**HT-PAMDA**) that could rapidly and comprehensively profile the PAM preferences of dozens or even hundreds of SpCas9 variants. A scalable assay to fulfill these criteria would: (1) preclude protein expression and purification (as we and others have previously done for Cas12a variants(*39*, *43*)), (2) would optimally be performed *in vitro* with conditions approximating a human cell context, and (3) would not be performed in bacteria or bacterial lysates (as we had done previously for SpCas9 and SaCas9 variants(*8*, *44*)). To enable our studies, we developed the HT-PAMDA that relies on the expression of SpCas9 variants in human cells, a step that can be easily arrayed and thus performed in high-throughput (**Fig. S2A**). The variable expression of SpCas9 proteins across different transfections is measurably linked to the expression of a 2A-EGFP fluorescence, permitting the normalization of SpCas9 protein concentrations by using a defined amount of EGFP-expressing lysate based on a fluorescein standard curve. A constant amount of SpCas9 human cell lysate is then subject to a time-course *in vitro* cleavage reaction of two separate libraries harboring distinct spacer sequences and 8 nucleotide randomized PAM sequences (**Fig. S2A**). Targeted sequencing of the libraries at various time points allowed quantitation of the rate of depletion of each PAM from the population over time; the rate constant for each PAM therefore enables us to calculate comprehensive PAM preferences for each SpCas9 variant. In general, we found that the HT-PAMDA profiles for WT SpCas9 and SpCas9-VQR were highly reproducible across two different spacer sequences (**Fig. S2B**) and across technical replicates (**Fig. S2C**). Furthermore, the HT-PAMDA profiles of WT SpCas9, SpCas9-VQR, and SpCas9-VRER were consistent with their previously described PAM preferences(*7*, *8*) established using alternate methods (**Fig. S2D**). These results demonstrate that HT-PAMDA recapitulates known PAM preferences and can in principle be scaled to large numbers of SpCas9 variants.

**4.** Engineering of SpG for targeting sites with NGN PAMs

To relax the PAM preference of SpCas9, we generated a series of variants bearing structure-motivated substitutions in residues D1135, S1136, G1218, E1219, R1335, and T1337 using SpCas9-VRQR as a scaffold. Based on our hypotheses outlined above, we sequentially tested hydrophobic substitutions at D1135, substitutions bearing different charges at E1219, and hydrophobic substitutions at S1136. The PAM preferences for variants bearing these substitutions were determined by HT-PAMDA, revealing differential contributions to PAM recognition by substitutions at D1135, S1136, and E1219 (**Figs. 1C** and **S2E**). Examination of the activities of this collection of variants against sites bearing NGAT, NGCC, NGGG, and NGTA PAMs in human cells recapitulated our HT-PAMDA observations and demonstrated that the concurrent substitution of several amino acids enables high activity targeting of non-canonical PAMs (**Fig. S4A**). The D1135L/S1136W/G1218R/E1219Q/R1335Q/T1337R variant improved human cell editing relative to SpCas9-VRQR by 2- to 53-fold on the sites with NGCC, NGGG, and NGTA PAMs, and maintained 87% of the activity on the NGAT site (**Fig. S4A**). While this variant displayed robust activity, our method of sequential mutagenesis to this point did not allow us to examine the more complete interdependence of the amino acid substitutions. We proceeded to assess all combinations of wild-type residues and the substitutions S1136W, E1219S/V/H/Q, G1218R/K/S, and T1337R/K. The forty-eight variants bearing combinations of the wild-type and substituted residues were profiled by HT-PAMDA (**Figs. 1C** and **S2E**) and assessed for on-target editing using the same four sites in human cells with NGAT, NGCC, NGGG, and NGTA PAMs (**Fig. S4B**). The combined HT-PAMDA and human cell analysis revealed that the variant encoding D1135L/S1136W/G1218K/E1219Q/R1335Q/T1337R substitutions (henceforth named **SpG**) displayed the most even PAM tolerance and the highest average activity across the four sites in human cells.

**5.** Energetic supplementation via non-specific contacts

To further improve the on-target activity of SpG, we wondered whether SpG could tolerate other substitutions intended to form non-specific DNA contacts and thus improve the overall interaction energy of SpG with the PAM. A similar strategy was previously described for SpCas9-NG, which harbors L1111R and A1322R substitutions hypothesized to form DNA backbone contacts to compensate for the loss of base-specific interactions to the 3$^{rd}$ position of the PAM caused by the R1335V substitution(*22*). To investigate this hypothesis, we first determined whether the L1111R and A1322R substitutions are essential for the activities of SpCas9-NG. We compared the on-target editing of SpCas9-NG to the R1111L, R1322A, and R1111L/R1322A derivative variants

that lack the supplementary energetic contacts across 16 sites harboring NGNN PAMs in human cells (**Figs. S4C** and **S4D**). We found that the inclusion of both arginine substitutions is necessary for the activities observed with SpCas9-NG.

We then determined whether the same substitutions could improve the editing efficiencies of SpG by generating derivative variants harboring L1111R, A1322R, or both substitutions. When we assessed the activities of these variants across the same 16 sites harboring NGNN PAMs in human cells, we surprisingly observed a reduction in the on-target activities for 14 of 16 sites with most variants (**Figs. S4E** and **S4F**). These results suggest that the substitutions in SpG that result in expanded PAM recognition do not require energetic supplementation, or possibly that the L1111R and A1322R substitutions are not compatible with SpG residues. HT-PAMDA experiments to determine the PAM profiles of these variants showed that the presence or absence of L1111R and A1322R in either variant did not appear to alter their PAM preferences (**Figs. S2F** and **S2G**), supporting the hypothesis that the major roles of L1111R and A1322R substitutions are energetic rather than PAM preference altering.

**6.** Sequence preferences of WT, xCas9, SpCas9-NG, and SpG

We utilized our HT-PAMDA and human cell datasets (**Figs. S2** and **S5** respectively) to more thoroughly characterize the sequence preferences of WT SpCas9, xCas9(3.7)(*23*), SpCas9-NG(*22*), and SpG. We observed correlations between average human cell editing efficiencies on NGNN PAMs and the $\log_{10}$ PAMDA rate constants, verifying that the preferences revealed in HT-PAMDA were replicated in human cells (**Fig. S5F**). Our characterization of WT SpCas9 PAM preference is consistent with prior reports(*5*, *7*, *8*), revealing that sites encoding NGG PAMs were most efficiently targeted followed by a minor preference for sites with NAG>NGA PAMs (**Fig. S2D**). For WT SpCas9, xCas9, SpCas9-NG, and SpG, we observed minimal evidence of a 1$^{st}$ PAM position preference (**Figs. S2D**, **S2F**, and **S5B**) and minor influence of the 1$^{st}$ position of the spacer (**Fig. S5D**).

Based on our HT-PAMDA and human cell editing data (**Figs. S2F** and **S5A**, respectively), we examined the 3$^{rd}$ PAM position preferences for each NG variant. We found that xCas9 preferred NG<u>G</u>>>NG<u>A</u>=NG<u>T</u>>NG<u>C</u>, and that SpCas9-NG and SpG exhibited more even tolerances of NG<u>N</u> PAMs (see **Figs. 1E** and **S2F**). When considering the 4$^{th}$ PAM position, xCas9 preferred NGN<u>C</u>>NGN<u>D</u> where once again SpCas9-NG and SpG were capable of more even targeting of NGN<u>N</u> PAMs (see **Figs. S2F** and **S5C**). Our human cell-based cytosine and adenine base editor (CBE and ABE, respectively) PAM characterizations of xCas9, SpCas9-NG, and SpG were

consistent with our nuclease observations, with differences likely attributable to the fact that we performed fewer CBE and ABE experiments (22 and 21 sites, respectively) compared to our nuclease datasets (78 sites; **Figs. 1E**, **1G** and **1I**).

**7.** PAM profile for xCas9

While the initial description of xCas9 reported targeting capabilities including NGN PAMs in human cells(*23*), our data suggests the targeting range of xCas9 to be more narrow. Across 78 sites in human cells, xCas9 averaged lower modification rates than WT SpCas9 and did not surpass 20% mean modification of sites encoding NGA, NGC, or NGT PAMs (**Fig. 1E**). HT-PAMDA also revealed that xCas9 can only appreciably target NGAC, NGTC, and NGGN PAMs (**Fig. S2F**). Interestingly, xCas9 also exhibited reduced activity on NGG PAMs compared to WT SpCas9, SpCas9-NG, and SpG, despite the retention of both critical R1333 and R1335 residues for NGG PAM recognition (**Figs. 1E** and **S2F**). Of the substitutions in xCas9, only E1219V is in close proximity to the PAM; this substitution might disrupt the E1219-R1335-$dG_3$ hydrogen-bonding network that stabilizes NGG PAM recognition, consistent with prior structural analysis(*45*). We characterized the PAM preference of the SpCas9(E1219V) single mutant and found that the PAM preference of xCas9 was attributable to this single mutation (**Fig. S2G**). Analysis of our human cell data corroborated HT-PAMDA findings that xCas9 has a preference for cytosine in the 4$^{th}$ position of the PAM, both in the context of NGG and NGH PAMs that nearly completely explained non-canonical PAM targeting with xCas9 (**Figs. S5C** and **S5E**). A preference for cytosine in the 4$^{th}$ position of the PAM is not observed with WT SpCas9, SpCas9-NG, and SpG (**Figs. S5C** and **S5E**). Collectively, our human cell (**Figs. S5C** and **S5E**), HT-PAMDA (**Figs. S2F** and **S2G**), and CBE-HT-PAMDA (**Fig. S6D**; see below) characterizations of xCas9 are more consistent with a PAM preference of NGGC>NGGD≈NGWC.

**8.** Optimization and validation of CBE-HT-PAMDA

The PAM requirements of base editor protein fusions have generally been assumed to be consistent with the PAM requirements of CRISPR nucleases, yet it remains possible that they exhibit distinctive preferences. To determine whether or not SpCas9 nucleases and base editors (BEs) exhibit consistent PAM profiles, we adapted the HT-PAMDA assay to function in the absence of SpCas9-mediated DNA cleavage. The PAM profiles generated by HT-PAMDA are dependent on the depletion of library members over time due to plasmid cleavage, yet base editors

do not intentionally cleave DNA (rather, DNA binding events are followed by nicking and deamination events). To directly address this question, we adapted HT-PAMDA to develop a cytosine base editor high-throughput PAM determination assay (**CBE-PAMDA-HT**; **Fig. S6C**). CBE-HT-PAMDA is similar to HT-PAMDA, but instead of double-strand DNA cleavage, the assay relies on SpCas9-based nicking and deamination of a cytosine by the tethered rAPOBEC1 domain. The combination of a target strand nick and deamination of the non-target strand is later converted to a double strand break via USER enzyme to remove the uracil base and cleave the non-target strand backbone, depleting CBE-targetable PAM-containing substrates from the library (**Fig. S6C**). With CBE-PAMDA-HT, we observed comparable CBE-based PAM profiles for WT-SpCas9, xCas9, SpCas9-NG, and SpG (**Fig. S6D**) relative to the nuclease-based PAM profiles for the same proteins (**Figs. S2D** and **S2F**). We also observed agreement the between HT-PAMDA and CBE-HT-PAMDA $\log_{10}$ rates for the PAMs of the same four variants (**Fig. S6E**). Thus, nuclease and CBE versions of different SpCas9 variants exhibit comparable PAM profiles.

**9.** Characterization of sequence preferences of SpRY

We explored if the range of activities displayed by SpRY across sites bearing NR and NY PAMs could be explained by aspects of PAM preference or sequence context. Our HT-PAMDA characterization of SpRY revealed a NR>NY PAM preference, and suggested a number of preferences at other positions (**Fig. S8D**). When comparing the HT-PAMDA $\log_{10}$ rate constants to the mean human cell editing efficiencies on NNN PAMs, we found that while HT-PAMDA accurately characterized the PAM preference of wild-type SpCas9 in human cells, the assay was less effective for SpRY (**Fig. S9F**). We expect that this discrepancy is due to the vastly different PAM compatibilities of WT SpCas9 and SpRY, leading to almost an order of magnitude difference in the number of targetable library members in the HT-PAMDA assay for each nuclease. Because HT-PAMDA infers PAM preference based on the rate of depletion of targetable PAMs by sequencing library members harboring non-compatible PAMs, there is a stoichiometric difference between WT SpCas9 and SpRY when considering targetable and non-targetable library members.

We also analyzed our human cell modification data to determine whether SpRY displayed PAM or PAM-proximal sequence preferences. In addition to the NR>NY preference in the 2$^{nd}$ position of the PAM, we observed varying degrees of sequence tolerance in 1$^{st}$, 3$^{rd}$, and 4$^{th}$ positions of the PAM as well as the 1$^{st}$ position of the spacer (**Figs. S9G-S9J**). Given the moderate number of sites that we examined (32 endogenous sites bearing NANN

PAMs, 32 with NGNN PAMs, and 31 with NYNN PAMs; **Figs. S9A**, **S9B**, and **S9E**, respectively) additional experiments may be required to more completely assess the *in cellulo* or *in vivo* PAM requirements of SpRY.

**10.** Commands for crystal structure visualization in PyMOL

```
cmd.set("bg_rgb", 'white')
cmd.set("ambient", '0.21000')
cmd.set("direct", '0.40000')
cmd.set("reflect", '0.43000')
cmd.set("power", '2.00000')
cmd.set("spec_reflect", '-0.01000')
cmd.set("line_width", '3.00000')
cmd.set("cache_display", 'off')
cmd.set("shininess", '30.00000')
cmd.set("cartoon_sampling", '7')
cmd.set("cartoon_loop_radius", '0.15000')
cmd.set("cartoon_oval_length", '1.00000')
cmd.set("auto_color_next", '1')
cmd.set("max_threads", '4')
cmd.set("specular_intensity", '0.30000')
cmd.set("button_mode_name", '3-Button Viewing')
cmd.set("seq_view", 'on')
cmd.set("cartoon_ring_mode", '3')
```

**Figures S1-S10:**

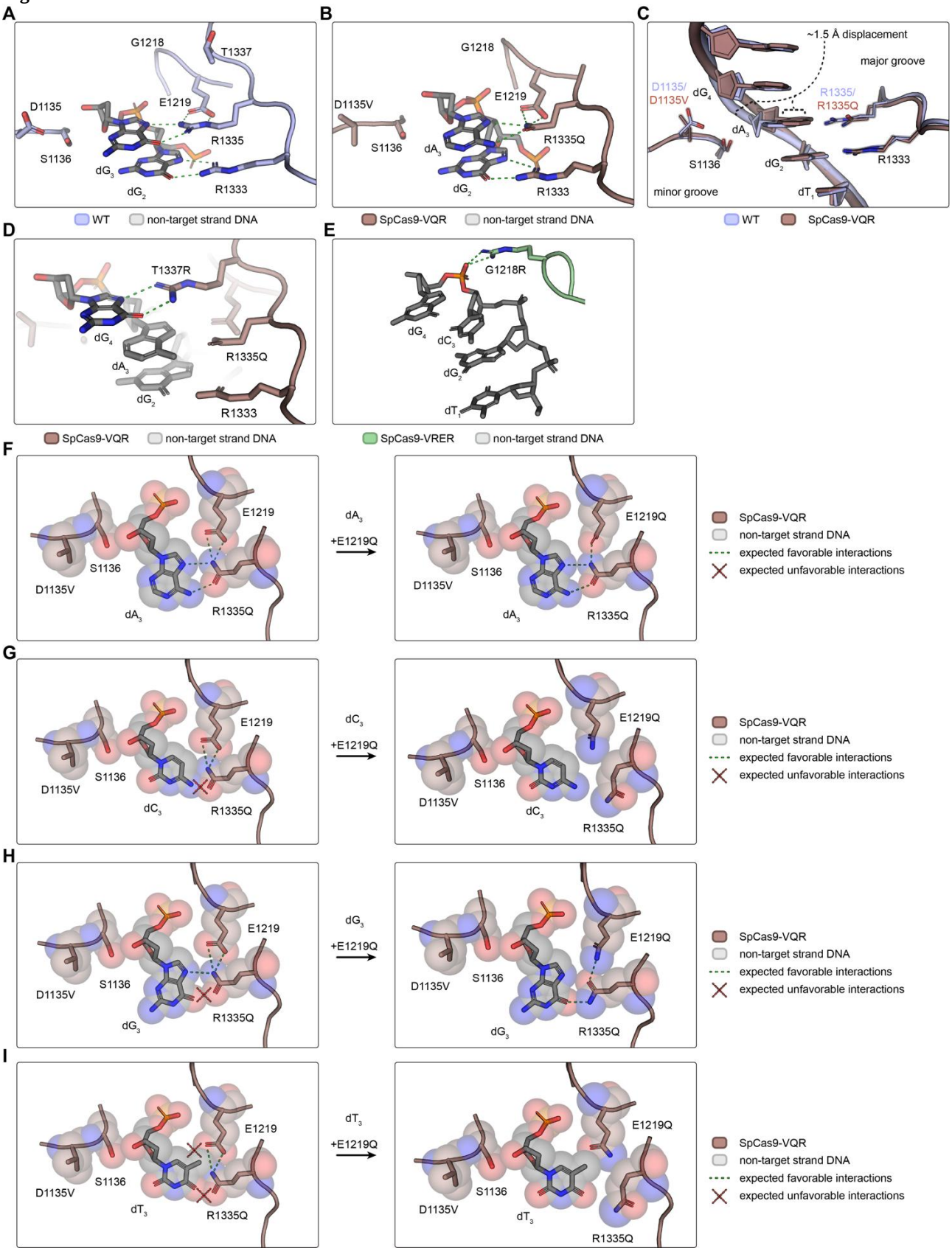

**Fig. S1. Structural models of PAM recognition by SpCas9 and engineered variants.**

(**A**, **B**) Molecular mechanisms of NGG PAM recognition by WT SpCas9 (**panel A**) and NGA PAM recognition by SpCas9-VQR(*8*) (**panel B**) with key residues shown. (**C**) Overlay of the non-target strand PAM DNA from WT SpCas9 (blue) and SpCas9-VQR (maroon) crystal structures with key residues labeled. The 3$^{rd}$ base of the PAM in the SpCas9-VQR structure is displaced towards R1333Q on the major groove side of the duplex relative to the wild-type structure. (**D**) In the context of SpCas9-VQR, the T1337R substitution forms a base-specific contact with $dG_4$ of an NGAG PAM. (**E**) In the context of SpCas9-VRER(*8*), the G1218R substitution interacts with the PAM DNA phosphate backbone. (**F-I**) Structural predictions for how the addition of E1219Q to SpCas9-VQR might enable recognition of NGN PAM sequences, modeled on NGA (**panel F**), NGC (**panel G**), NGG (**panel H**), and NGT (**panel I**) PAMs. For all panels, structures were visualized in PyMOL (v 2.3.3) using PDB IDs 4UN3(*20*) (WT SpCas9), 5B2R(*25*) (SpCas9-VQR), and 5B2T(*25*) (SpCas9-VRER); certain protein and nucleic acid residues were omitted for clarity.

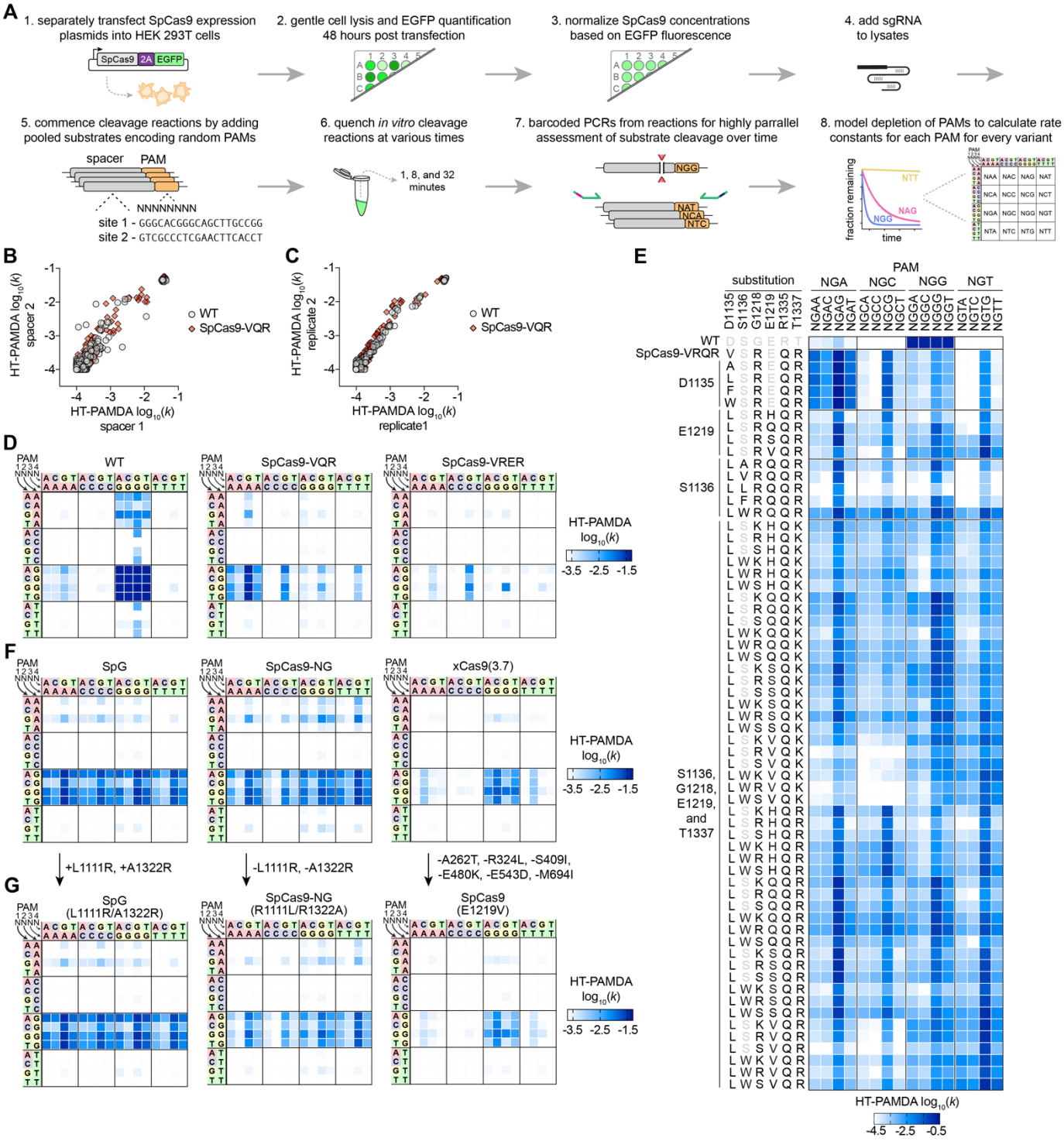


**Fig. S2. Optimization and application of a high-throughput PAM determination assay (HT-PAMDA).**

(**A**) Schematic of the HT-PAMDA workflow. SpCas9 proteins are expressed in human cells and harvested by gentle lysis, with SpCas9 concentrations normalized by EGFP fluorescence. Two libraries harboring randomized PAMs with separate spacer sequences are subjected to timecourse *in vitro* cleavage reactions using SpCas9 lysate

complexed with sgRNAs. PAM depletion over time is monitored by deep sequencing and modeled to generate rate constants for each PAM. (**B**) Correlation of HT-PAMDA $\log_{10}$ rates ($k$) for NNNN PAMs across two randomized PAM libraries with distinct spacer sequences (wild-type SpCas9: $r(256) = 0.9574$, $p < 0.0001$; SpCas9-VQR: $r(256) = 0.9521$, $p < 0.0001$). (**C**) Correlation of HT-PAMDA rates for NNNN PAMs across two technical replicates, where each technical replicate is the average of experiments on the two libraries harboring distinct spacer sequences (wild-type SpCas9: $r(256) = 0.9955$, $p < 0.0001$; SpCas9-VQR: $r(256) = 0.9819$, $p < 0.0001$). In **panels B** and **C,** HT-PAMDA $\log_{10}(k)$ were set to a minimum value of -4; reported correlations are Pearson correlation coefficients with statistical significance evaluated with a two-tailed t-test. (**D**) HT-PAMDA NNNN profiles of the well-characterized WT SpCas9, SpCas9-VQR, and SpCas9-VRER nucleases(*5*, *7*, *8*). (**E**) HT-PAMDA NGNN profiles of WT SpCas9 and engineered variants; some variants are shown twice for completeness. (**F**) HT-PAMDA NNNN PAM profiles of SpG, SpCas9-NG, and xCas9(3.7). (**G**) HT-PAMDA NNNN PAM profiles of SpG with L1111R and A1322R substitutions, SpCas9-NG without the requisite L1111R and A1322R substitutions, and xCas9(3.7) without the A262T, R324L, S409I, E480K, E543D, and M694I substitutions. For **panels D-G**, HT-PAMDA $\log_{10}(k)$ are the mean of at least two replicates against two distinct spacer sequences.

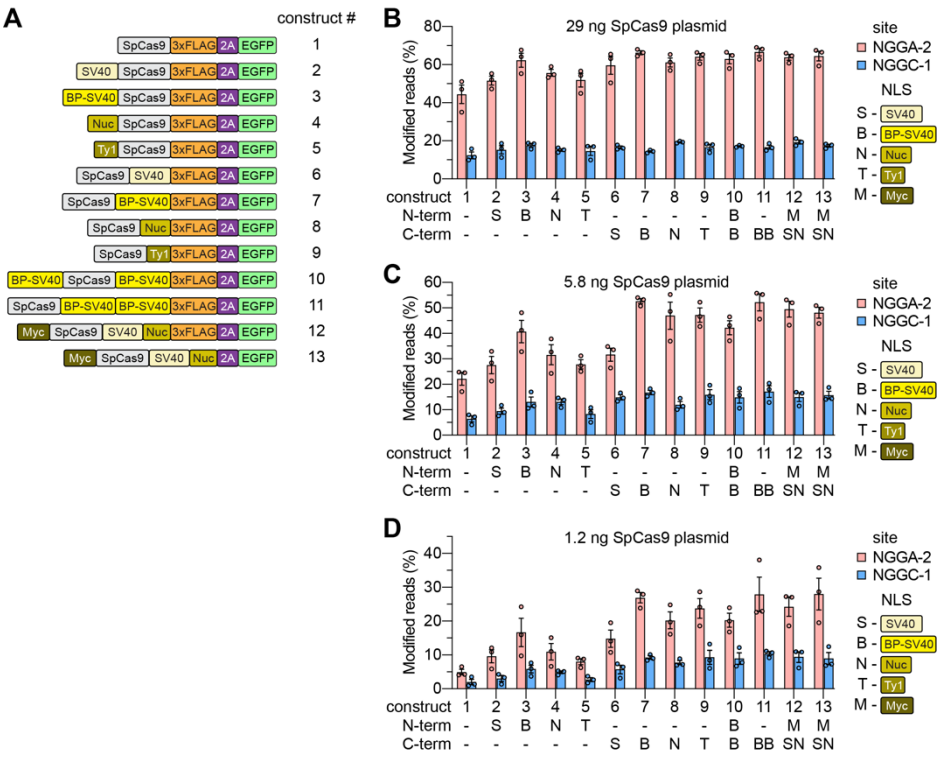


**Fig. S3. Nuclear localization signal (NLS) architecture assessment for SpCas9.**

(**A**) Schematics of the thirteen NLS architectures tested with wild-type SpCas9. Constructs 7 and 13 are similar to as previously described(*27*, *46*). (**B**-**D**) Modification of two sites in HEK 293T cells bearing NGG PAMs using WT SpCas9 with the NLS architectures described in (**panel A**), using 29 ng (**panel B**), 5.8 ng (**panel C**), or 1.2 ng (**panel D**) of SpCas9-expression plasmid. Percent modified reads in **panels B**-**D** assessed by targeted sequencing; mean, s.e.m., and individual data points shown for n = 3. Constructs harboring a C-terminal BP(SV40)NLS-3xFLAG-P2A-EGFP sequence (construct #7) were utilized for all human cell experiments unless otherwise indicated.

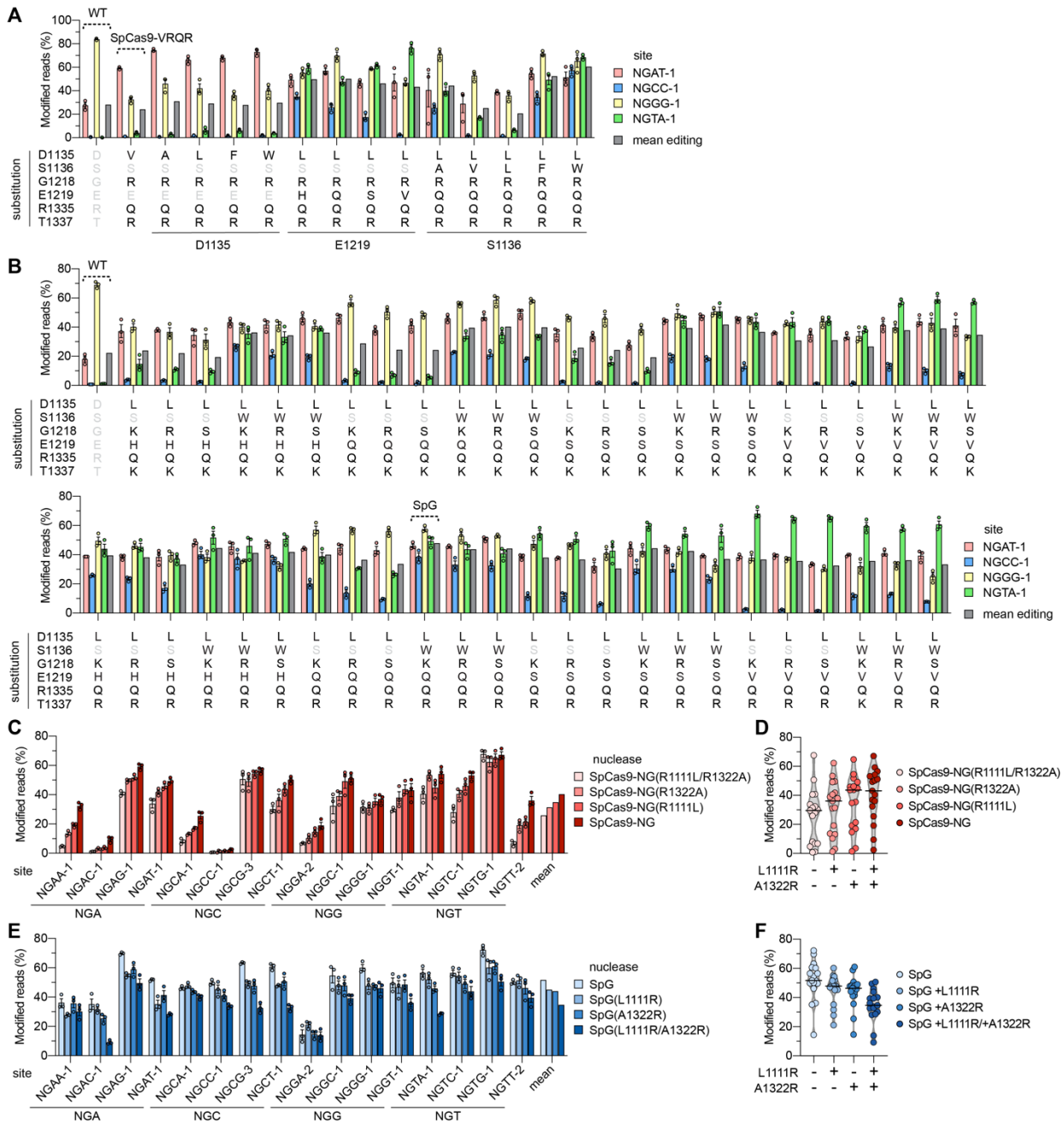


**Fig. S4**. **Engineering SpG for efficient targeting of NGN PAMs.**

(**A**) Modification of endogenous sites in HEK 293T cells bearing NGAT, NGCC, NGGG, and NGTA PAMs by WT SpCas9, SpCas9-VRQR(*21*), and derivative variants of SpCas9-VRQR. (**B**) Modification of endogenous sites in HEK 293T cells bearing NGAT, NGCC, NGGG, and NGTA PAMs by WT SpCas9 and variants bearing substitutions at positions D1135, S1136, G1218, E1219, R1335, and T1337. (**C**-**F**) Modification of 16 sites in HEK293T cells bearing NGNN PAMs by SpCas9-NG(*22*) and derivatives lacking one or both of the requisite

L1111R and A1322R substitutions (bar plots of data in **panel C**; summary of data in **panel D**), and by SpG and derivatives harboring one or both of the L1111R and A1322R substitutions (bar plots of data in **panel E**; summary of data in **panel F**). Percent modified reads in **panels A-F** assessed by targeted sequencing; mean, s.e.m., and individual data points shown for n = 3.

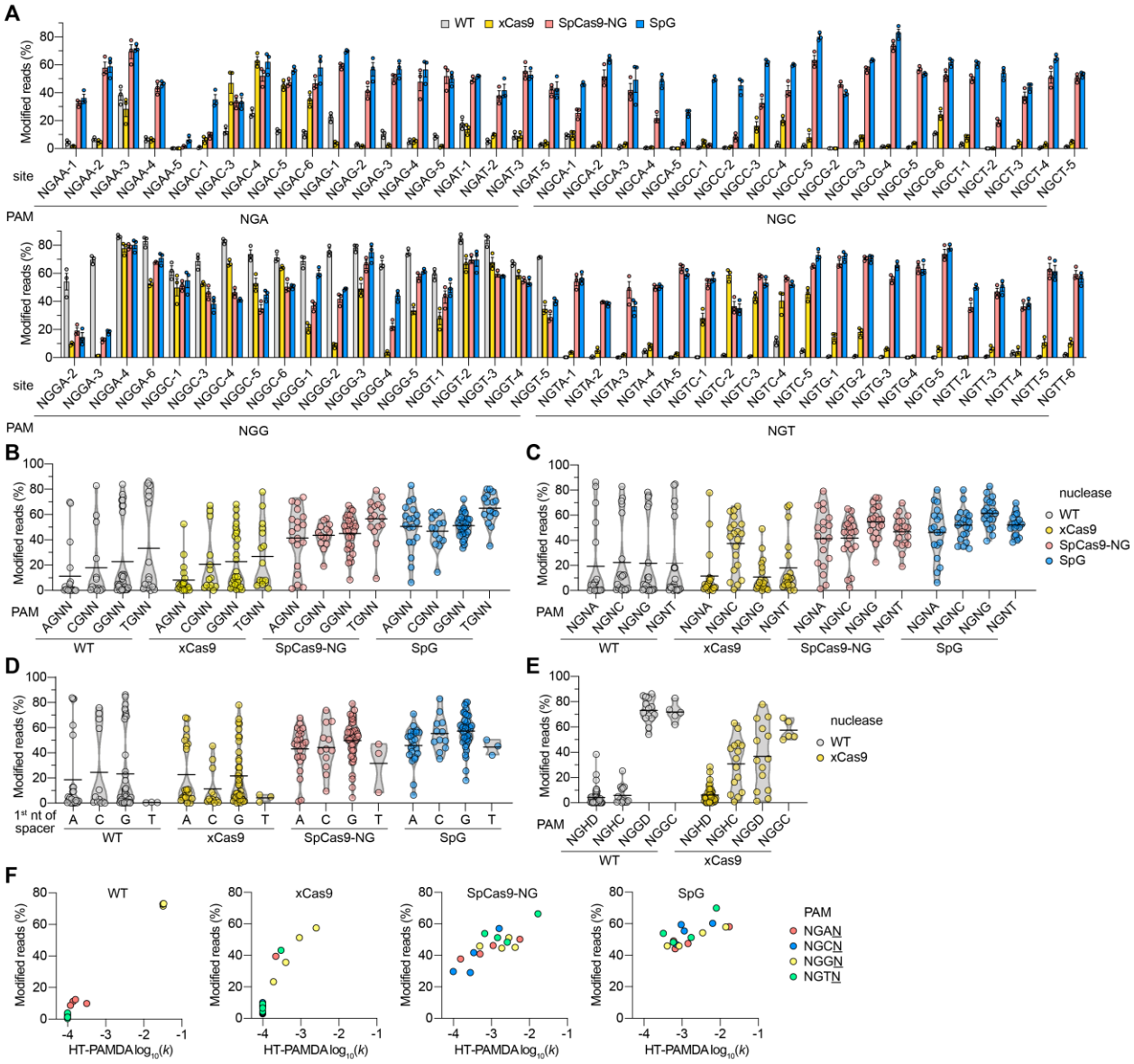


**Fig. S5. Characterization of SpCas9 nucleases capable of targeting NGN PAMs.**

(**A**) Modification of 78 endogenous sites in HEK 293T cells bearing NGNN PAMs by WT SpCas9, xCas9, SpCas9-NG, and SpG. Percent modification assessed by targeted sequencing; mean, s.e.m., and individual data points shown for n = 3. (**B-D**) Summaries of the results in **panel A**, but grouped by the base identity of the 1st spacer proximal nt of the PAM (NGNN; **panel B**), the 4th nt of the PAM (NGNN; **panel C**), or the 1st PAM proximal nt of the spacer (**panel D**). (**E**) Summary of the results in **panel A** for WT and xCas9, but separated by both canonical NGG PAM versus non-canonical NGH PAMs (where H = A, C, or T) and the corresponding 4th nt base of the PAM for C versus D (where D = A, G, or T). In **panels B-E**, mean modification of each site from **panel A** is shown; mean modification across each subset of PAMs is shown as a black line), and the grey outline is a violin plot. (**F**)

Correlation between HT-PAMDA $\log_{10}(k)$ (see **Figs. S2E** and **S2F**) and mean human cell modification from **panel A** for each NGNN PAM (WT SpCas9: $r(16) = 0.9959$, $p < 0.0001$; xCas9: $r(16) = 0.9335$, $p < 0.0001$; SpCas9-NG: $r(16) = 0.8038$, $p = 0.0002$; SpG: $r(16) = 0.6897$, $p = 0.0031$). HT-PAMDA $\log_{10}(k)$ were set to a minimum value of -4.

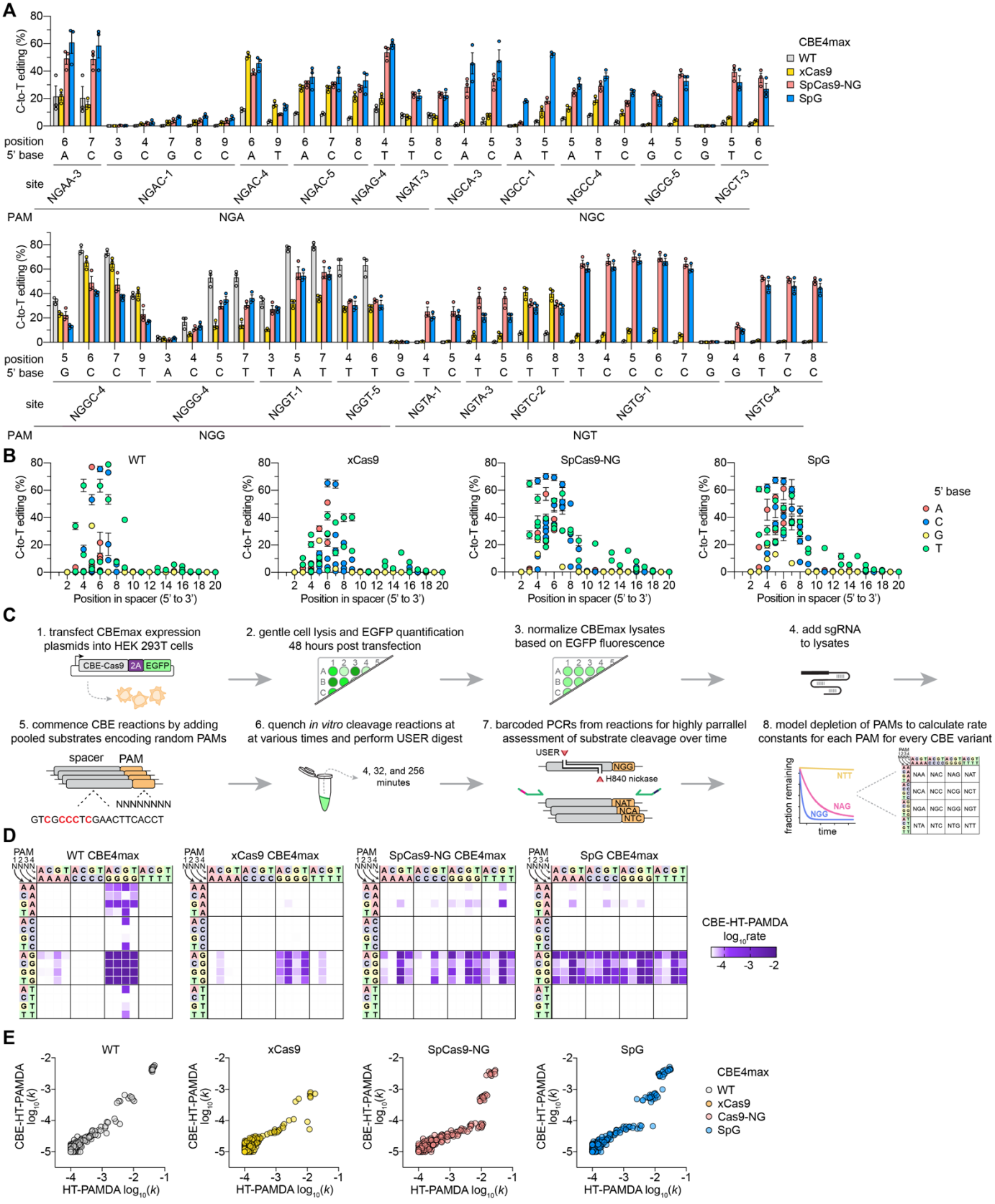
A
C-to-T editing (%)
CBE4max
WT
xCas9
SpCas9-NG
SpG
position
5' base
site
PAM
NGA
NGC
NGAA-3
NGAC-1
NGAC-4
NGAC-5
NGAG-4
NGAT-3
NGCA-3
NGCC-1
NGCC-4
NGCG-5
NGCT-3
NGGC-4
NGGG-4
NGGT-1
NGGT-5
NGTA-1
NGTA-3
NGTC-2
NGTG-1
NGTG-4
NGG
NGT
B
WT
xCas9
SpCas9-NG
SpG
Position in spacer (5' to 3')
5' base
C
1. transfect CBEmax expression plasmids into HEK 293T cells
CBE-Cas9
2A
EGFP
2. gentle cell lysis and EGFP quantification 48 hours post transfection
3. normalize CBEmax lysates based on EGFP fluorescence
4. add sgRNA to lysates
5. commence CBE reactions by adding pooled substrates encoding random PAMs
spacer
PAM
NNNNNNNN
GTCGCCCTCGAACTTCACCT
6. quench in vitro cleavage reactions at various times and perform USER digest
4, 32, and 256 minutes
7. barcoded PCRs from reactions for highly parrallel assessment of substrate cleavage over time
USER
NGG
H840 nickase
NAT
NCA
NTC
8. model depletion of PAMs to calculate rate constants for each PAM for every CBE variant
fraction remaining
time
NTT
NAG
NGG
D
WT CBE4max
xCas9 CBE4max
SpCas9-NG CBE4max
SpG CBE4max
CBE-HT-PAMDA log10rate
E
CBE-HT-PAMDA log10(k)
HT-PAMDA log10(k)
CBE4max
WT
xCas9
Cas9-NG
SpG

**Fig. S6**. **Characterization of C-to-T base editors capable of targeting NGN PAMs.**

(**A**) C-to-T base editing of 57 cytosines across 20 endogenous sites in HEK 293T cells bearing NGN PAMs with WT SpCas9, xCas9, SpCas9-NG, and SpG CBE4max contructs. C-to-T editing within the edit window (positions 3 through 9) assessed by targeted sequencing; mean, s.e.m., and individual data points shown for $n = 3$. (**B**) Scatterplots of C-to-T editing of all cytosines in the spacer compared to position in the spacer for sites from **panel A**, to define the edit window and the impact of the identity of the preceeding base 5' of each cytosine. Mean and s.e.m. shown for n = 3. (**C**) Schemcatic of the cytosine base editor (CBE) HT-PAMDA (CBE-HT-PAMDA) workflow. CBE4max variants are expressed in human cells and harvested by gentle lysis, with CBE4max concentrations normalized by EGFP fluorescence. A library harboring randomized PAMs is subjected to timecourse *in vitro* reactions using CBE4max lysate complexed with sgRNAs. Following termination of each reaction, USER enzyme is added to convert C-to-U deamination events to double-strand breaks when they co-occur with SpCas9-HNH domain mediated DNA nicks. PAM depletion over time is monitored by deep sequencing and modeled to generate rate constants for each PAM. (**D**) CBE-HT-PAMDA NNNN profiles for WT SpCas9, xCas9, SpCas9-NG, and SpG CBE4max constructs. (**E**) Correlation of HT-PAMDA $\log_{10}$ rate constants ($k$) with CBE-HT PAMDA $\log_{10}$ rate constants (WT CBE4max: $r(256) = 0.9845$, $p < 0.0001$; xCas9 CBE4max: $r(256) = 0.9248$, $p < 0.0001$; SpCas9-NG CBE4max: $r(256) = 0.9134$, $p < 0.0001$; SpG CBE4max: $r(256) = 0.9722$, $p < 0.0001$). Reported correlations are Pearson correlation coefficients with statistical significance evaluated with a two-tailed t-test. HT-PAMDA $\log_{10}(k)$ and CBE-HT-PAMDA $\log_{10}(k)$ were set to minimum values of -4 and -5, respectively.

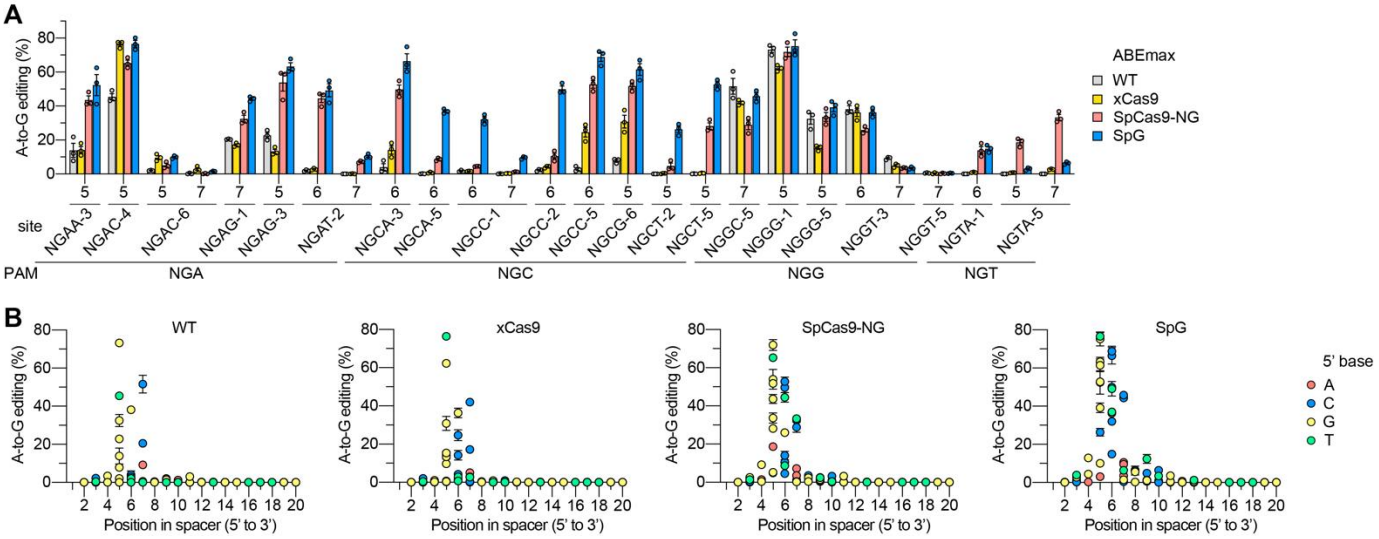


**Fig. S7. Characterization of A-to-G base editors capable of targeting NGN PAMs.** (**A**) A-to-G base editing of 26 adenines across 21 endogenous sites in HEK 293T cells bearing NGN PAMs with WT SpCas9, xCas9, SpCas9-NG, and SpG ABEmax constructs. A-to-G editing within the edit window (positions 5 through 7) assessed by targeted sequencing; mean, s.e.m. and individual data points shown for $n = 3$. (**B**) Scatterplots of A-to-G editing of all adenines in the spacer compared to position in the spacer for sites from **panel A**, to define the edit window and the impact of identity of the preceding base 5' of each adenine. Mean and s.e.m. shown for n = 3.

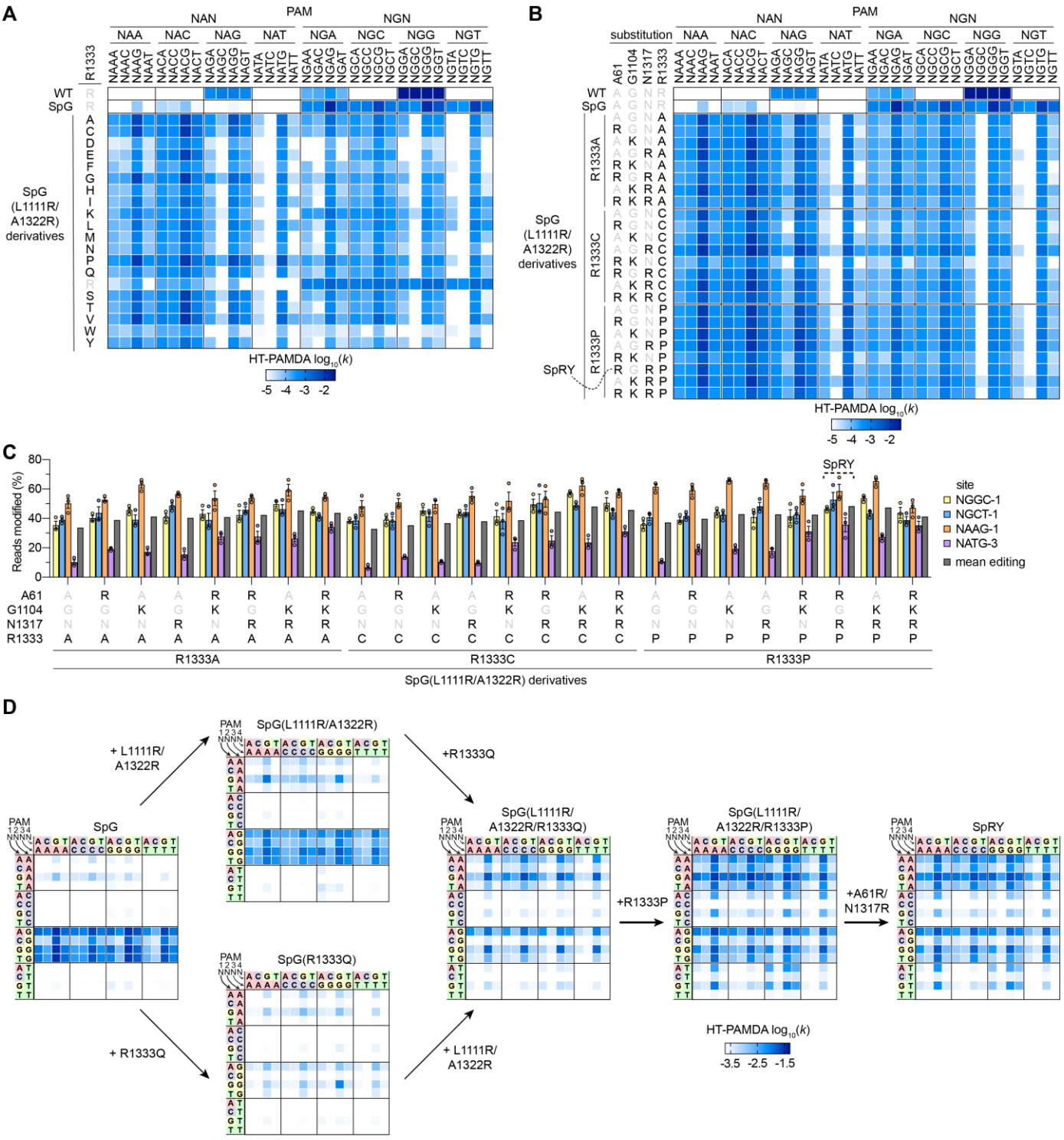


**Fig. S8. Engineering SpRY for efficient targeting of NR PAM sequences.**

(**A**, **B**) HT-PAMDA NRNN nt PAM profiles of SpG(L1111R/A1322R) derivatives bearing all possible amino acid substitutions at R1333 (**panel A**), and of SpG(L1111R/A1322R) derivatives bearing substitutions at A61, G1104, and N1317 in the context of R1333A/C/P (**panel B**). (**C**) Modification of four endogenous sites in HEK 293T cells

bearing NRN PAMs by SpG(L1111R/A1322R) derivatives. Percent modification assessed by targeted sequencing; mean, s.e.m., and individual data points shown for n = 3. (**D**) HT-PAMDA NNNN PAM profiles of SpG, SpRY, and selected intermediate variants. For **panels A**, **B**, and **D**, HT-PAMDA $\log_{10}(k)$ are the mean of at least two replicates against two distinct spacer sequences.

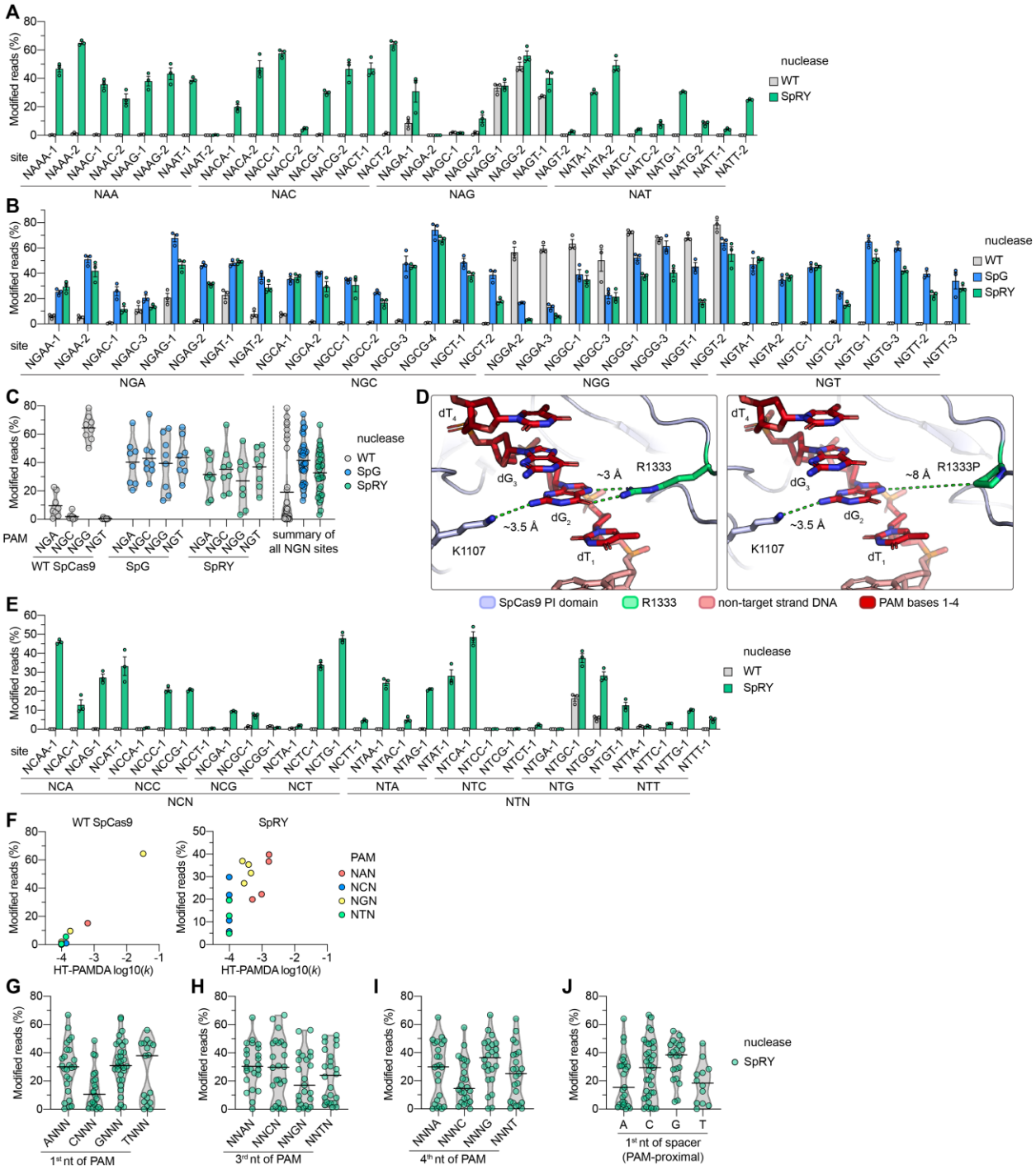


**Fig. S9. Characterization of SpRY nuclease activities.**

(**A**, **B**) Modification of 32 endogenous sites in HEK 293T cells bearing NANN PAMs with WT SpCas9 and SpRY (**panel A**), and 32 sites bearing NGNN PAMs with WT SpCas9, SpG, and SpRY (**panel B**). Percent modification assessed by targeted sequencing; mean, s.e.m., and individual data points shown for n = 3. (**C**) Summary of the results in **panel B**, grouped by 3rd nt of the PAM for WT, SpG, and SpRY. (**D**) Molecular mechanisms for

recognition of the 2nd base of the PAM for WT SpCas9 (**left panel**) and the R1333P variant (**right panel**) with key residues shown. Structures were visualized in PyMOL (v 2.3.3) using PDB ID 4UN3 (WT SpCas9)(*20*); certain protein and nucleic acid residues were omitted for clarity. (**E**) Modification of 31 endogenous sites bearing NYNN PAMs by WT SpCas9 and SpRY. Percent modification assessed by targeted sequencing; mean, s.e.m., and individual data points shown for n = 3. (**F**) Correlation between HT-PAMDA $\log_{10}(k)$ for WT and SpRY (see **Figs. S2E** and **S8D**, respectively) and mean human cell modification from **panels A**, **B**, and **E** for each NNN PAM (WT SpCas9: $r(16) = 0.9934$, $p < 0.0001$; SpRY-Cas9: $r(16) = 0.6774$, $p = 0.0039$;). HT-PAMDA $\log_{10}(k)$ were set to a minimum value of -4; reported correlations are Pearson correlation coefficients with statistical significance evaluated with a two-tailed t-test.. (**G-J**) Summaries of the results in **panels A**, **B**, and **E** grouped by 1st nt of the PAM (**panel G**), 3rd nt of the PAM (**panel H**), 4th nt of the PAM (**panel I**), and 1st PAM proximal nt of the spacer (**panel J**) for WT and SpRY. The black line represents the mean and the grey outline is a violin plot.

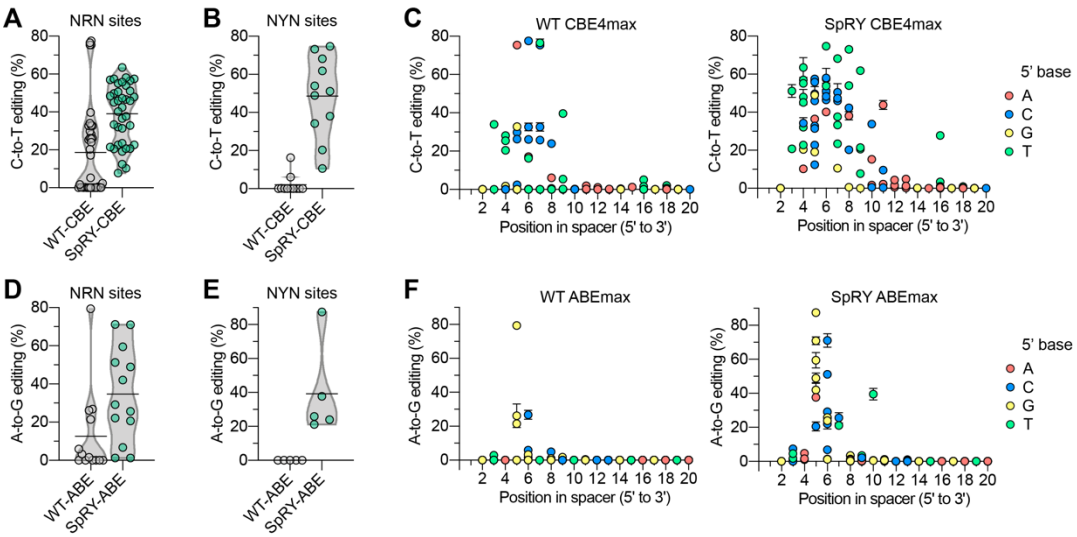


**Fig. S10. Characterization of SpRY base editor activities.**

(**A, B**) Summaries of the C-to-T editing activities of WT- and SpRY-CBE from **Figs. 3C** and **3D**, respectively. Each data point represents the mean of the 3 replicates; the black line signifies the mean), and the grey outline is a violin plot. (**C**) Scatterplots of C-to-T editing of all cytosines in the spacer compared to position in the spacer for sites from **Figs. 3C** and **3D**, to define the edit window and the impact of the identity of the preceding base 5' of each cytosine for WT- and SpRY-CBE, respectively. Mean and s.e.m. shown for n = 3. (**D**, **E**) Summaries of the A-to-G editing activities of WT- and SpRY-ABE from **Figs. 3E** and **3F**, respectively. Each data point represents the mean of the 3 replicates; the black line signifies the mean), and the grey outline is a violin plot. (**F**) Scatterplots of A-to-G editing of all adenines in the spacer compared to position in the spacer for sites from **Figs. 3E** and **3F**, to define the edit window and the impact of the identity of the preceding base 5' of each adenine for WT- and SpRY-ABE, respectively. Mean and s.e.m. shown for n = 3.

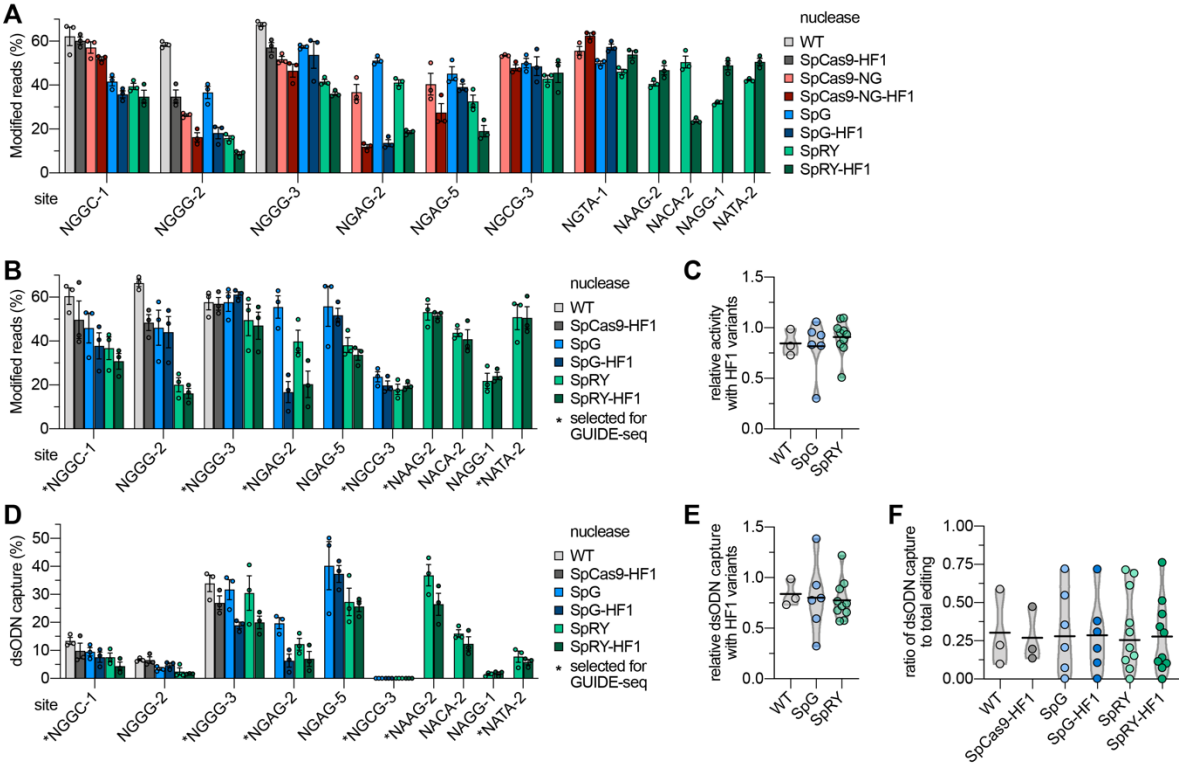


**Fig. S11. Evaluation of on-target editing with WT, SpG, SpRY, and their respective HF1 variants.**

**(A, B)** Modification of endogenous sites in HEK 293T cells by WT SpCas9, SpCas9-NG, SpG, SpRY, and their respective high-fidelity (HF1)(*21*) derivatives from transfections without (**panel A**) or containing (**panel B**) the GUIDE-seq double-stranded oligodeoxynucleotide (dsODN) tag(*31*). Percent modification assessed by targeted sequencing; mean, s.e.m., and individual data points shown for n = 3. (**C**) Summary of the mean relative nuclease activity of the HF1 derivatives of WT, SpG, and SpRY (from **panel B**). The black line represents the mean of 3-10 sites for each pair of variants), and the grey outline is a violin plot.. (**D**) GUIDE-seq dsODN capture at on-target sites by WT SpCas9, SpG, SpRY, and their respective HF1 variants. Percent modification assessed by targeted sequencing; mean, s.e.m., and individual data points shown for n = 3. (**E**) Summary of the mean relative dsODN capture of the HF1 derivatives of WT, SpG, and SpRY (from **panel D**). The black line represents the mean of 3-10 sites for each pair of variants), and the grey outline is a violin plot.. (**F**) Ratio of GUIDE-seq dsODN tag capture (**panel B**) to overall mutagenesis (**panel D**) for WT SpCas9, SpG, SpRY, and their respective HF1 variants. The black line represents the mean of 3-10 sites for each pair of variants), and the grey outline is a violin plot; HF1 variants encode N497A, R661A, Q695A, and Q926A substitutions.

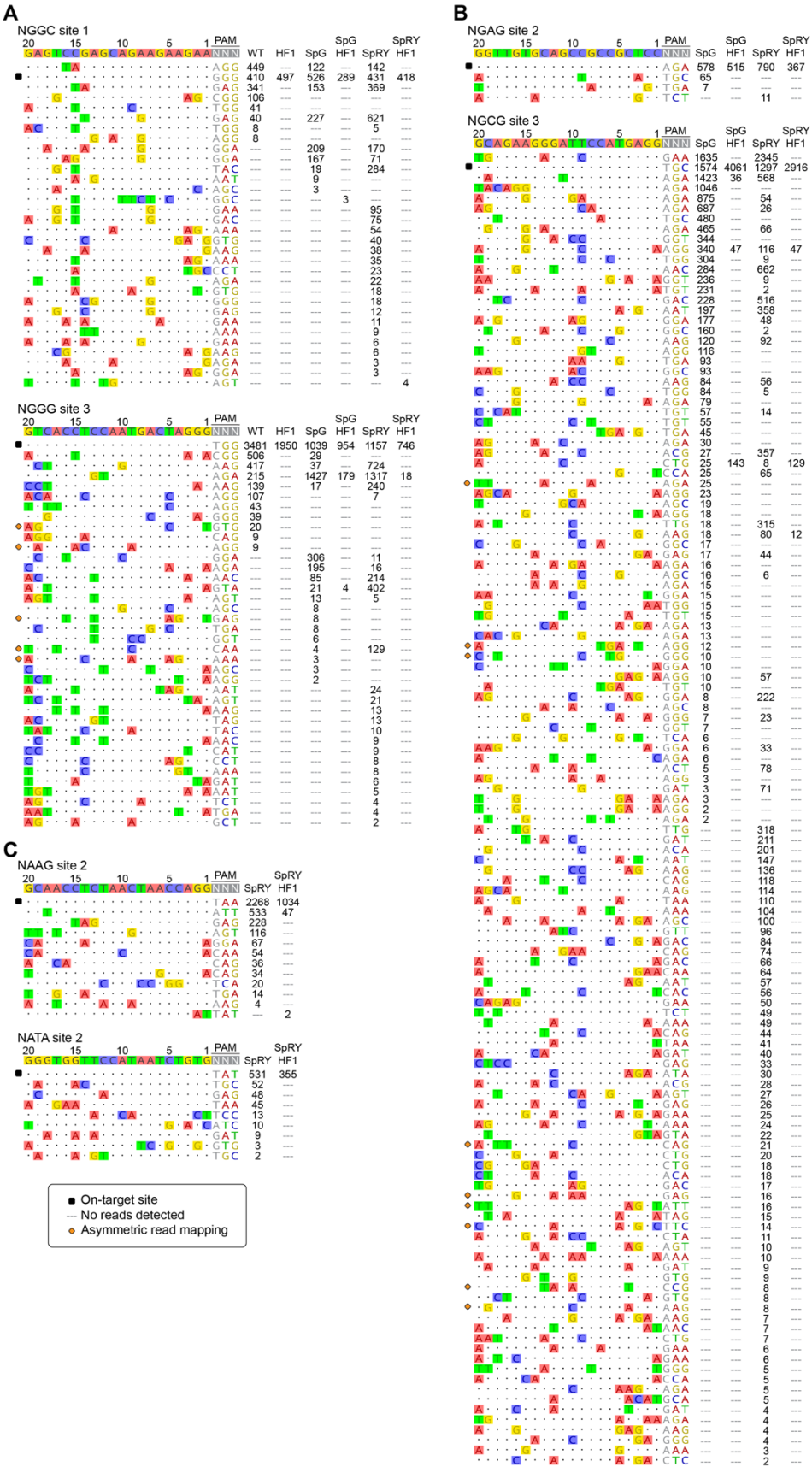
A
NGGC site 1
NGGG site 3
B
NGAG site 2
NGCG site 3
C
NAAG site 2
NATA site 2
On-target site
No reads detected
Asymmetric read mapping

**Fig. S12. Genome-wide specificity assessments of relaxed PAM variants.**

(**A**-**C**) GUIDE-seq genome-wide specificity profiles of WT, SpG, SpRY and their respective high-fidelity (HF1) (*21*) derivatives with gRNAs targets to sites harboring NGG (**panel A**), NGN (**panel B**), or NRN (**panel C**) PAMs. Mismatched positions in the spacers of the off-target sites are highlighted in color; GUIDE-seq read counts from consolidated unique molecular events for each variant are shown to the right of the sequence plots; orange diamonds indicate sites that are only supported by asymmetric GUIDE-seq reads; nucleotides within the PAMs of the on- and off-target sites are shown for clarity. HF1 variants encode N497A, R661A, Q695A, and Q926A substitutions.

**Tables S1-S7:**

**Table S1** *(attached separately).*

List of target sites.

**Table S2** *(attached separately).*

List of plasmids.

**Table S3** *(attached separately).*

List of oligonucleotides.

**Table S4** *(attached separately).*

Nuclease data summary.

**Table S5** *(attached separately).*

Base editor data summary.

**Table S6** *(attached separately).*

GUIDE-seq data summary.

**Table S7** *(attached separately).*

PAMDA data summary.

**References** *(#42-46 only in Supplementary Material)*

1. L. A. Marraffini, E. J. Sontheimer, Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*. **463**, 568–71 (2010).

2. R. Heler, P. Samai, J. W. Modell, C. Weiner, G. W. Goldberg, D. Bikard, L. A. Marraffini, Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature*. **519**, 199–202 (2015).

3. H. Deveau, R. Barrangou, J. E. Garneau, J. Labonte, C. Fremaux, P. Boyaval, D. A. Romero, P. Horvath, S. Moineau, Phage Response to CRISPR-Encoded Resistance in Streptococcus thermophilus. *J Bacteriol*. **190**, 1390–1400 (2007).

4. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiol Read Engl*. **155**, 733–40 (2009).

5. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. **337**, 816–821 (2012).

6. S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, J. A. Doudna, DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*. **507**, 62–7 (2014).

7. W. Jiang, D. Bikard, D. Cox, F. Zhang, L. A. Marraffini, RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*. **31**, 233–9 (2013).

8. B. P. Kleinstiver, M. S. Prew, S. Q. Tsai, V. V. Topkar, N. T. Nguyen, Z. Zheng, A. P. W. Gonzales, Z. Li, R. T. Peterson, J.-R. J. Yeh, M. J. Aryee, J. K. Joung, Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*. **523**, 481–5 (2015).

9. K. M. Esvelt, P. Mali, J. L. Braff, M. Moosburner, S. J. Yaung, G. M. Church, Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods*. **10**, 1116–21 (2013).

10. I. Fonfara, A. L. Rhun, K. Chylinski, K. S. Makarova, A.-L. Lécrivain, J. Bzdrenga, E. V. Koonin, E. Charpentier, Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res*. **42**, 2577–90 (2013).

11. T. Karvelis, G. Gasiunas, J. Young, G. Bigelyte, A. Silanskas, M. Cigan, V. Siksnys, Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol*. **16**, 253 (2015).

12. B. Zetsche, J. S. Gootenberg, O. O. Abudayyeh, I. M. Slaymaker, K. S. Makarova, P. Essletzbichler, S. E. Volz, J. Joung, J. van der Oost, A. Regev, E. V. Koonin, F. Zhang, Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*. **163**, 759–71 (2015).

13. B. Elliott, C. Richardson, J. Winderbaum, J. A. Nickoloff, M. Jasin, Gene Conversion Tracts from Double-Strand Break Repair in Mammalian Cells. *Mol Cell Biol*. **18**, 93–101 (1998).

14. G. M. Findlay, E. A. Boyle, R. J. Hause, J. C. Klein, J. Shendure, Saturation editing of genomic regions by multiplex homology-directed repair. *Nature*. **513**, 120–3 (2014).

15. M. C. Canver, E. C. Smith, F. Sher, L. Pinello, N. E. Sanjana, O. Shalem, D. D. Chen, P. G. Schupp, D. S. Vinjamur, S. P. Garcia, S. Luc, R. Kurita, Y. Nakamura, Y. Fujiwara, T. Maeda, G.-C. Yuan, F. Zhang, S. H. Orkin, D. E. Bauer, BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*. **527**, 192–7 (2015).

16. J. Shi, E. Wang, J. P. Milazzo, Z. Wang, J. B. Kinney, C. R. Vakoc, Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol*. **33**, 661–7 (2015).

17. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. **533**, 420–4 (2016).

18. N. M. Gaudelli, A. C. Komor, H. A. Rees, M. S. Packer, A. H. Badran, D. I. Bryson, D. R. Liu, Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*. **551**, 464–471 (2017).

19. Y. Kan, B. Ruis, T. Takasugi, E. A. Hendrickson, Mechanisms of precise genome editing using oligonucleotide donors. *Genome Res*. **27**, 1099–1111 (2017).

20. C. Anders, O. Niewoehner, A. Duerst, M. Jinek, Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*. **513**, 569–73 (2014).

21. B. P. Kleinstiver, V. Pattanayak, M. S. Prew, S. Q. Tsai, N. T. Nguyen, Z. Zheng, J. K. Joung, High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*. **529**, 490–5 (2016).

22. H. Nishimasu, X. Shi, S. Ishiguro, L. Gao, S. Hirano, S. Okazaki, T. Noda, O. O. Abudayyeh, J. S. Gootenberg, H. Mori, S. Oura, B. Holmes, M. Tanaka, M. Seki, H. Hirano, H. Aburatani, R. Ishitani, M. Ikawa, N. Yachie, F. Zhang, O. Nureki, Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*. **361**, 1259–1262 (2018).

23. J. H. Hu, S. M. Miller, M. H. Geurts, W. Tang, L. Chen, N. Sun, C. M. Zeina, X. Gao, H. A. Rees, Z. Lin, D. R. Liu, Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*. **556**, 57–63 (2018).

24. C. Anders, K. Bargsten, M. Jinek, Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. *Mol Cell*. **61**, 895–902 (2016).

25. S. Hirano, H. Nishimasu, R. Ishitani, O. Nureki, Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. *Mol Cell*. **61**, 886–94 (2016).

26. See Supplementary Materials

27. K. Suzuki, Y. Tsunekawa, R. Hernandez-Benitez, J. Wu, J. Zhu, E. J. Kim, F. Hatanaka, M. Yamamoto, T. Araoka, Z. Li, M. Kurita, T. Hishida, M. Li, E. Aizawa, S. Guo, S. Chen, A. Goebl, R. D. Soligalla, J. Qu, T. Jiang, X. Fu, M. Jafari, C. R. Esteban, W. T. Berggren, J. Lajara, E. Nuñez-Delicado, P. Guillen, J. M. Campistol, F. Matsuzaki, G.-H. Liu, P. Magistretti, K. Zhang, E. M. Callaway, K. Zhang, J. C. I. Belmonte, In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*. **540**, 144–149 (2016).

28. H. A. Rees, D. R. Liu, Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet*. **19**, 770–788 (2018).

29. L. W. Koblan, J. L. Doman, C. Wilson, J. M. Levy, T. Tay, G. A. Newby, J. P. Maianti, A. Raguram, D. R. Liu, Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol*. **36**, 843–846 (2018).

30. I. M. Slaymaker, L. Gao, B. Zetsche, D. A. Scott, W. X. Yan, F. Zhang, Rationally engineered Cas9 nucleases with improved specificity. *Sci New York N Y*. **351**, 84–8 (2015).

31. S. Q. Tsai, Z. Zheng, N. T. Nguyen, M. Liebers, V. V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, A. J. Iafrate, L. P. Le, M. J. Aryee, J. K. Joung, GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol*. **33**, 187–97 (2014).

32. W. Balemans, M. Ebeling, N. Patel, E. V. Hul, P. Olson, M. Dioszegi, C. Lacza, W. Wuyts, J. V. D. Ende, P. Willems, A. F. Paes-Alves, S. Hill, M. Bueno, F. J. Ramos, P. Tacconi, F. G. Dikkers, C. Stratakis, K. Lindpaintner, B. Vickery, D. Foernzler, W. V. Hul, Increased bone density in sclerosteosis is due to the deficiency of a novel secreted protein (SOST). *Hum Mol Genet*. **10**, 537–543 (2001).

33. M. Schuelke, K. R. Wagner, L. E. Stolz, C. Hübner, T. Riebel, W. Kömen, T. Braun, J. F. Tobin, S.-J. Lee, Myostatin Mutation Associated with Gross Muscle Hypertrophy in a Child. *New Engl J Med*. **350**, 2682–2688 (2004).

34. J. J. Cox, J. Sheynin, Z. Shorer, F. Reimann, A. K. Nicholas, L. Zubovic, M. Baralle, E. Wraige, E. Manor, J. Levy, C. G. Woods, R. Parvari, Congenital insensitivity to pain: novel SCN9A missense and in-frame deletion mutations. *Hum Mutat*. **31**, E1670-86 (2010).

35. J. Flannick, G. Thorleifsson, N. L. Beer, S. B. R. Jacobs, N. Grarup, N. P. Burtt, A. Mahajan, C. Fuchsberger, G. Atzmon, R. Benediktsson, J. Blangero, D. W. Bowden, I. Brandslund, J. Brosnan, F. Burslem, J. Chambers, Y. S. Cho, C. Christensen, D. A. Douglas, R. Duggirala, Z. Dymek, Y. Farjoun, T. Fennell, P. Fontanillas, T. Forsén, S. Gabriel, B. Glaser, D. F. Gudbjartsson, C. Hanis, T. Hansen, A. B. Hreidarsson, K. Hveem, E. Ingelsson, B. Isomaa, S. Johansson, T. Jørgensen, M. E. Jørgensen, S. Kathiresan, A. Kong, J. Kooner, J. Kravic, M. Laakso, J.-Y. Lee, L. Lind, C. M. Lindgren, A. Linneberg, G. Masson, T. Meitinger, K. L. Mohlke, A. Molven, A. P. Morris, S. Potluri, R. Rauramaa, R. Ribel-Madsen, A.-M. Richard, T. Rolph, V. Salomaa, A. V. Segrè, H. Skärstrand, V. Steinthorsdottir, H. M. Stringham, P. Sulem, E. S. Tai, Y. Y. Teo, T. Teslovich, U. Thorsteinsdottir, J. K. Trimmer, T. Tuomi, J. Tuomilehto, F. Vaziri-Sani, B. F. Voight, J. G. Wilson, M. Boehnke, M. I. McCarthy, P. R. Njølstad, O. Pedersen, G.-T. Consortium, T.-G. Consortium, L. Groop, D. R. Cox, K. Stefansson, D. Altshuler, Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet*. **46**, 357–63 (2014).

36. T. and H. W. G. of the E. S. P. Institute National Heart, Lung, and Blood, J. Crosby, G. M. Peloso, P. L. Auer, D. R. Crosslin, N. O. Stitziel, L. A. Lange, Y. Lu, Z. Tang, H. Zhang, G. Hindy, N. Masca, K. Stirrups, S. Kanoni, R. Do, G. Jun, Y. Hu, H. M. Kang, C. Xue, A. Goel, M. Farrall, S. Duga, P. A. Merlini, R. Asselta, D. Girelli, O. Olivieri, N. Martinelli, W. Yin, D. Reilly, E. Speliotes, C. S. Fox, K. Hveem, O. L. Holmen, M. Nikpay, D. N. Farlow, T. L. Assimes, N. Franceschini, J. Robinson, K. E. North, L. W. Martin, M. DePristo, N. Gupta, S. A. Escher, J.-H. Jansson, N. V. Zuydam, C. N. A. Palmer, N. Wareham, W. Koch, T. Meitinger, A. Peters, W. Lieb, R. Erbel, I. R. Konig, J. Kruppa, F. Degenhardt, O. Gottesman, E. P. Bottinger, C. J. O'Donnell, B. M. Psaty, C. M. Ballantyne, G. Abecasis, J. M. Ordovas, O. Melander, H. Watkins, M. Orho-Melander, D. Ardissino, R. J. F. Loos, R. McPherson, C. J. Willer, J. Erdmann, A. S. Hall, N. J. Samani, P. Deloukas, H. Schunkert, J. G. Wilson, C. Kooperberg, S. S. Rich, R. P. Tracy, D.-Y. Lin, D. Altshuler, S. Gabriel, D. A. Nickerson, G. P. Jarvik, L. A. Cupples, A. P. Reiner, E. Boerwinkle, S. Kathiresan, Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *New Engl J Medicine*. **371**, 22–31 (2014).

37. A. R. Harper, S. Nayee, E. J. Topol, Protective alleles and modifier variants in human health and disease. *Nat Rev Genetics*. **16**, 689–701 (2015).

38. M. I. G. C. Investigators, N. O. Stitziel, H.-H. Won, A. C. Morrison, G. M. Peloso, R. Do, L. A. Lange, P. Fontanillas, N. Gupta, S. Duga, A. Goel, M. Farrall, D. Saleheen, P. Ferrario, I. König, R. Asselta, P. A. Merlini, N. Marziliano, M. F. Notarangelo, U. Schick, P. Auer, T. L. Assimes, M. Reilly, R. Wilensky, D. J. Rader, G. K. Hovingh, T. Meitinger, T. Kessler, A. Kastrati, K.-L. Laugwitz, D. Siscovick, J. I. Rotter, S. L. Hazen, R. Tracy, S. Cresci, J. Spertus, R. Jackson, S. M. Schwartz, P. Natarajan, J. Crosby, D. Muzny, C. Ballantyne, S. S. Rich, C. J. O'Donnell, G. Abecasis, S. Sunaev, D. A. Nickerson, J. E. Buring, P. M. Ridker, D. I. Chasman, E. Austin, I. J. Kullo, P. E. Weeke, C. M. Shaffer, L. A. Bastarache, J. C. Denny, D. M. Roden, C. Palmer, P. Deloukas, D.-Y. Lin, Z. Tang, J. Erdmann, H. Schunkert, J. Danesh, J. Marrugat, R. Elosua, D. Ardissino, R. McPherson, H. Watkins, A. P. Reiner, J. G. Wilson, D. Altshuler, R. A. Gibbs, E. S. Lander, E. Boerwinkle, S. Gabriel, S. Kathiresan, Inactivating mutations in NPC1L1 and protection from coronary heart disease. *New Engl J Medicine*. **371**, 2072–82 (2014).

39. B. P. Kleinstiver, A. A. Sousa, R. T. Walton, Y. E. Tak, J. Y. Hsu, K. Clement, M. M. Welch, J. E. Horng, J. Malagon-Lopez, I. Scarfò, M. V. Maus, L. Pinello, M. J. Aryee, J. K. Joung, Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat Biotechnol*. **37**, 276–282 (2019).

40. K. Clement, H. Rees, M. C. Canver, J. M. Gehrke, R. Farouni, J. Y. Hsu, M. A. Cole, D. R. Liu, J. K. Joung, D. E. Bauer, L. Pinello, CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol*. **37**, 224–226 (2019).

41. Scripts for analyzing High-Throughput PAM Determination Assay (HT-PAMDA) experimental data for CRISPR enzymes; available through Zenodo, doi: 10.5281/zenodo.3710516.

42. N. Rohland, D. Reich, Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res*. **22**, 939–46 (2012).

43. L. Gao, D. B. T. Cox, W. X. Yan, J. C. Manteiga, M. W. Schneider, T. Yamano, H. Nishimasu, O. Nureki, N. Crosetto, F. Zhang, Engineered Cpf1 variants with altered PAM specificities. *Nat Biotechnol*. **35**, 789–792 (2017).

44. B. P. Kleinstiver, M. S. Prew, S. Q. Tsai, N. T. Nguyen, V. V. Topkar, Z. Zheng, J. K. Joung, Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol*. **33**, 1293–1298 (2015).

45. M. Guo, K. Ren, Y. Zhu, Z. Tang, Y. Wang, B. Zhang, Z. Huang, Structural insights into a high fidelity variant of SpCas9. *Cell Res*. **29**, 183–192 (2019).

46. Y. Wu, J. Zeng, B. P. Roscoe, P. Liu, Q. Yao, C. R. Lazzarotto, K. Clement, M. A. Cole, K. Luk, C. Baricordi, A. H. Shen, C. Ren, E. B. Esrick, J. P. Manis, D. M. Dorfman, D. A. Williams, A. Biffi, C. Brugnara, L. Biasco, C. Brendel, L. Pinello, S. Q. Tsai, S. A. Wolfe, D. E. Bauer, Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat Med*. **25**, 776–783 (2019).