nature biotechnology

ARTICLES

https://doi.org/10.1038/s41587-020-0412-8

Continuous evolution of SpCas9 variants compatible with non-G PAMs

Shannon M. Miller[1,2,3,6], Tina Wang[1,2,3,6], Peyton B. Randolph[1,2,3], Mandana Arbab[1,2,3], Max W. Shen[1,2,3,4], Tony P. Huang[1,2,3], Zaneta Matuszek[1,2,3,5], Gregory A. Newby[1,2,3], Holly A. Rees[1,2,3] and David R. Liu[1,2,3*]

The targeting scope of *Streptococcus pyogenes* Cas9 (SpCas9) and its engineered variants is largely restricted to protospacer-adjacent motif (PAM) sequences containing G bases. Here we report the evolution of three new SpCas9 variants that collectively recognize NRNH PAMs (where R is A or G and H is A, C or T) using phage-assisted non-continuous evolution, three new phage-assisted continuous evolution strategies for DNA binding and a secondary selection for DNA cleavage. The targeting capabilities of these evolved variants and SpCas9-NG were characterized in HEK293T cells using a library of 11,776 genomically integrated protospacer–sgRNA pairs containing all possible NNNN PAMs. The evolved variants mediated indel formation and base editing in human cells and enabled A•T-to-G•C base editing of a sickle cell anemia mutation using a previously inaccessible CACC PAM. These new evolved SpCas9 variants, together with previously reported variants, in principle enable targeting of most NR PAM sequences and substantially reduce the fraction of genomic sites that are inaccessible by Cas9-based methods.

The CRISPR–Cas9 system has recently transformed the life sciences by enabling a wide range of targeted genome manipulation methods[1,2]. Cas9 is limited, however, by its requirement for a PAM to bind a DNA sequence. SpCas9, the most widely used and well-characterized Cas9 homolog[1], recognizes an NGG PAM immediately 3′ of the target DNA sequence[3]. To expand the range of targetable genomic loci, researchers have used naturally occurring Cas9 orthologs with different PAM specificities[4]. The vast majority of these natural homologs, however, are less characterized, less active and/or more stringent in their PAM requirements than SpCas9.

Both *Staphylococcus aureus* Cas9 (SaCas9)[5] and SpCas9 (refs. [6–8]) have been evolved or engineered to increase their PAM targeting scope. Although these efforts have expanded SpCas9's compatibility from NGG PAM sites to most NG PAM sites[6,7], locations in the genome lacking G bases remain difficult to access. Restrictions on Cas9 targeting are especially problematic for precision genome editing techniques that require strict placement of the Cas9 in relation to the desired edit, such as homology-directed repair (HDR)[9], predictable template-free end-joining[10,11] and base editing[12]. Base editing is particularly sensitive to Cas9 PAM compatibility: activity for SpCas9-derived base editors is optimal when the PAM is located ~13–17 nucleotides away from the target base[12]. In addition, for any given base edit, it may be desirable to test multiple protospacers to maximize on-target activity while minimizing unwanted editing[13–16].

Here we report the directed evolution of three new SpCas9 variants capable of recognizing NRRH, NRTH and NRCH PAMs. These variants were evolved through phage-assisted non-continuous evolution (PANCE) and three new phage-assisted continuous evolution (PACE) selection strategies for SpCas9 binding to specific sequences with non-G PAMs. We characterized these three new variants, as well as SpCas9-NG[7], a previously reported engineered SpCas9 that recognizes NG PAMs, on 92 endogenous human genomic target sites and a library of 11,776 integrated target sites. The new variants reported here, together with previously reported NG PAM-compatible SpCas9 variants, greatly expand the potentially accessible PAM sequence space of SpCas9 to most NR PAMs.

Results

Phage-assisted evolution of SpCas9 toward non-G PAM sequences. PACE, a method for the rapid, directed evolution of biomolecules[17], has been successfully applied to evolve proteins, including polymerases[17–22], proteases[23,24], genome-editing proteins[6,25], antibody-like proteins[26,27], insecticidal proteins[26], aminoacyl-tRNA synthetases[28,29]and methanol dehydrogenases[30]. We previously evolved SpCas9 variants with broadened PAM compatibility in PACE using a bacterial one-hybrid protein:DNA binding selection[6] in which binding of catalytically inactivated SpCas9 (dSpCas9) fused to a subunit of the *Escherichia coli* RNA polymerase (RNAP; ω–dSpCas9) to a target protospacer-PAM sequence recruits endogenous RNAP to drive gene III (gIII) transcription from an accessory plasmid (AP) (Fig. 1a). Only selection phage (SP) carrying ω–dSpCas9 variants recognizing the target PAM sequence produce infectious progeny phage and replicate during PACE[6]. Evolving SpCas9 against a mixture of trinucleotide PAMs using this selection led to xCas9, which binds some NG PAMs but very few non-G PAMs[6]. We hypothesized that the use of a mixture of many PAMs during evolution would greatly reduce the selection pressure for binding to any specific PAM. Therefore, we reasoned that selecting for binding to several specific PAM sequences might yield SpCas9 variants with stronger recognition of non-canonical PAMs.

We focused our evolution efforts on the NAN subset of PAM sequence space, which is largely inaccessible by commonly used Cas9 variants. Because both SpCas9 and xCas9 in phage-based

[1]Merkin Institute of Transformative Technologies in Healthcare, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [2]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [3]Howard Hughes Medical Institute, Harvard University, Cambridge, MA, USA. [4]Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA. [6]These authors contributed equally: Shannon M. Miller, Tina Wang. *e-mail: drliu@fas.harvard.edu

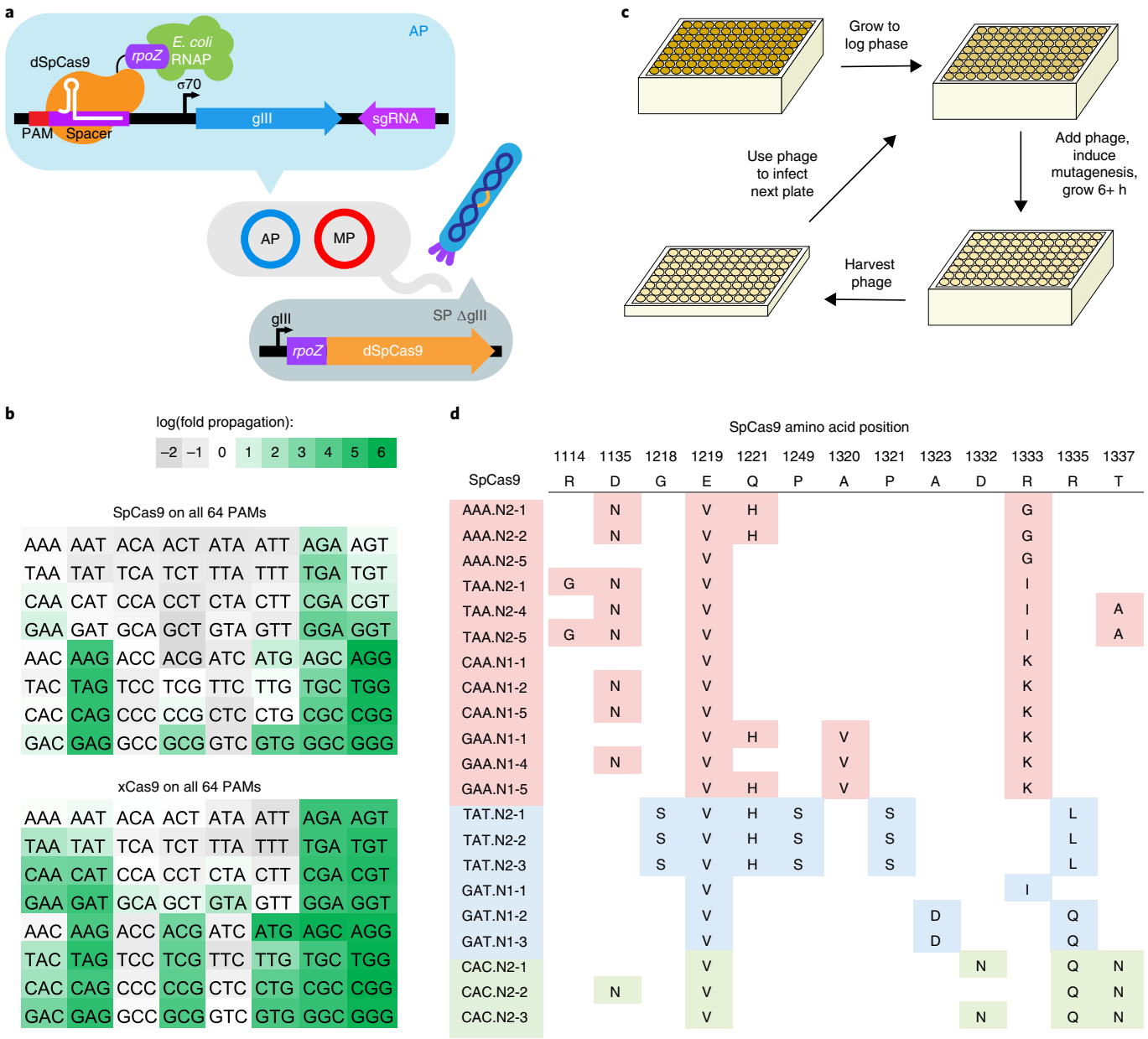
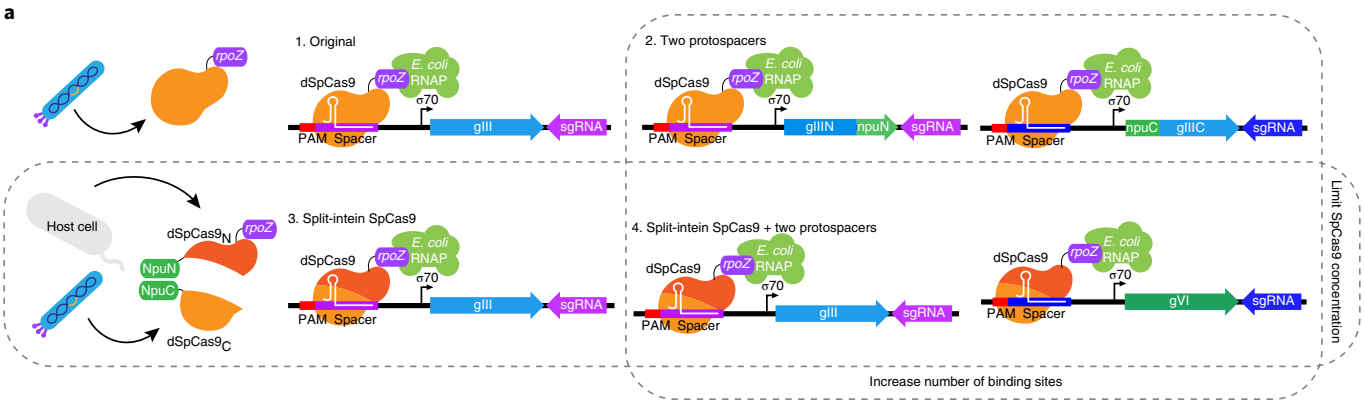
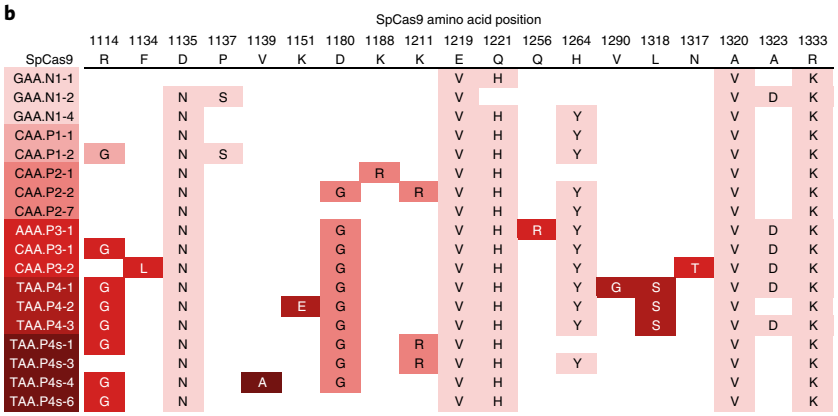


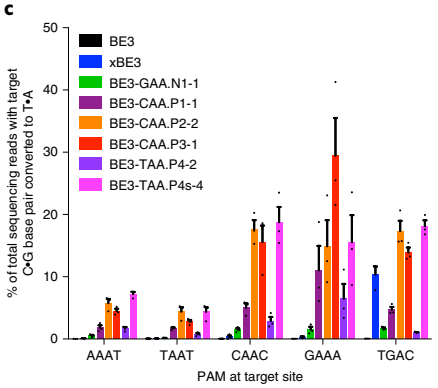
Fig. 1 | PANCE of SpCas9 binding activity on non-G PAMs. a, Original selection scheme for Cas9 DNA binding, in which ω-dSpCas9 expressed by ΔgIII SP binds to a designated protospacer and PAM sequence upstream of gIII on an AP in host *E. coli* cells. Host cells and infecting SP are continuously mutagenized by a mutagenesis plasmid (MP). **b**, Binding activity of SpCas9 and xCas9 on all 64 possible NNN PAMs, determined by fold propagation of SP expressing ω-dSpCas9 (top) or ω-dxCas9 (bottom) on host cells containing APs bearing these PAM sequences upstream of gIII. **c**, Schematic overview of the PANCE workflow. Host cells containing an AP and MP are grown to log phase in a deep-well plate or tube before being infected with SP. Mutagenesis is induced and SP are allowed to propagate for 6–18 h before cells are pelleted and the SP-containing supernatant is collected. The SP pool is then used to infect host cells in the next iteration of PANCE. **d**, Consensus mutations arising from evolution of ω-dSpCas9 (N1) or ω-dxCas9 (N2) on NAA (red), NAT (blue) and NAC (green) PAM sequences.

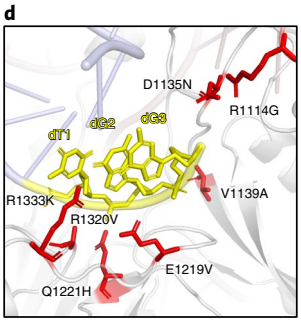
DNA-binding experiments showed low activity on NAH PAM sites (Fig. 1b), we began by evolving SpCas9 and xCas9 for binding to each of the 12 possible NAH PAM target sequences in parallel using PANCE[29,30] (Fig. 1c). Although slower than PACE, PANCE enables weakly active variants to replicate and can be performed in parallel[30]. After 19 rounds of PANCE (total net phage dilution of 10^{38}-fold) on each of the 12 NAH PAMs in parallel, we observed distinct sets of mutations depending on the third base of the NAH PAM targeted for evolution (Fig. 1d and Supplementary Table 1). Given this early divergence, we divided the evolution of SpCas9 into three separate trajectories, each aimed at recognizing a NAN PAM with no required G: NAA, NAT and NAC.

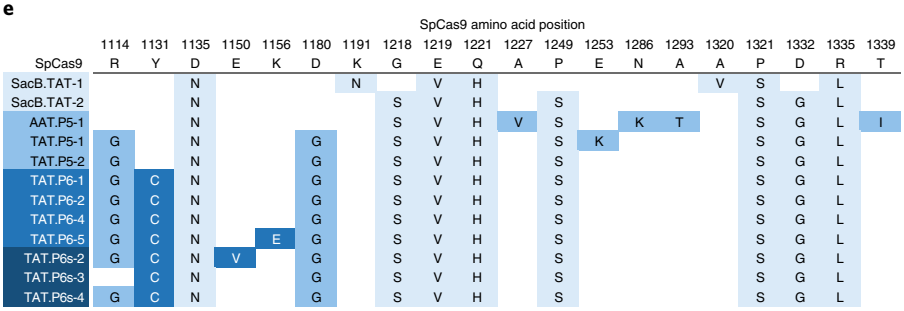
We first pursued the NAA PAM trajectory (Figs. 1d and 2b). Despite acquiring multiple consensus mutations (D1135N, E1219V, Q1221H, H1264Y, A1320V and R1333K) in the PAM-interacting domain (PID, residues 1099–1368), the NAA-targeted PANCE-evolved variants exhibited low activity when converted into cytosine base editors (CBEs) and tested on sites containing NAA PAMs in human cells (clone GAA.N1-1; Fig. 2c). We hypothesized that evolving increased binding activity might require higher selection

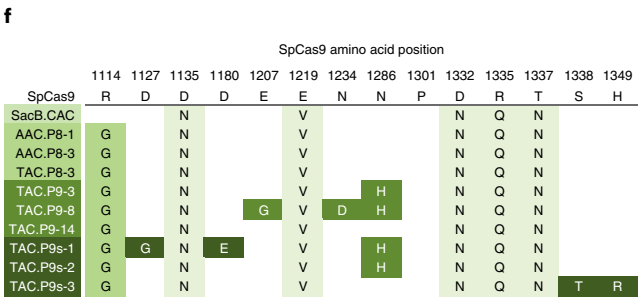


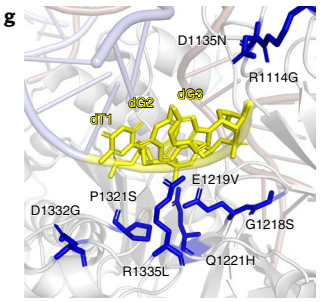












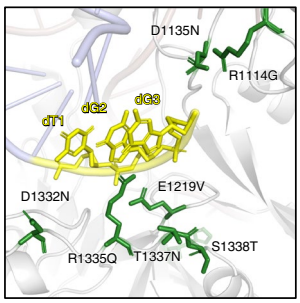


Fig. 2 | Selection improvement enables evolution of SpCas9 variants with robust activity on non-G PAMs. a, Four selections for Cas9:DNA binding in PACE. (1) Original selection scheme. (2) Dual-AP selection, in which ω-dSpCas9 binds two distinct protospacer-PAM sequences to drive either half of a split-intein pIII. (3) Use of split-intein Cas9 limits total Cas9 concentration in host cells, thus avoiding saturation of protospacer-PAM binding sites. Residues 574–1368 of SpCas9 fused to NpuC ($dSpCas9_C$) are expressed by the SP and ω-dSpCas9 (1–573) fused to NpuN ($ω-dSpCas9_N$) is encoded by a low-copy CP in host cells. (4) Combination of the selection principles from (2) and (3) through use of gVI as an additional PACE-compatible selection marker for phage propagation and ΔgIIIΔgVI SP. **b**, Mutations from further PACE of original PANCE mutants evolved to bind NAA PAMs. Mutations in darker red were acquired in later PACE generations. **c**, Comparison of human cell base editing efficiency of the evolved clones shown in **b** in HEK293T cells on NRA PAM sites. Bars represent mean and s.e.m. of $n=3$ independent biological replicates, with individual values shown as dots. **d**, Mutated residues in the PID of TAA.P4s-4 mapped onto the SpCas9 crystal structure (4UN3). **e**, Mutations from further PACE of original PANCE mutants evolved to bind NAT PAMs. Mutations in darker blue were acquired in later PACE generations. We note that the Y1131C mutation was found to be inactivating (Supplementary Fig. 2a). **f**, Mutations from further PACE of original PANCE mutants evolved to bind NAC PAMs. Mutations in darker green were acquired in later PACE generations. **g**, Mutated residues in the PID of TAT.P5-1 mapped onto the SpCas9 crystal structure (4UN3). **h**, Mutated residues in the PID of TAC.P9s-3 mapped onto the SpCas9 crystal structure (4UN3).

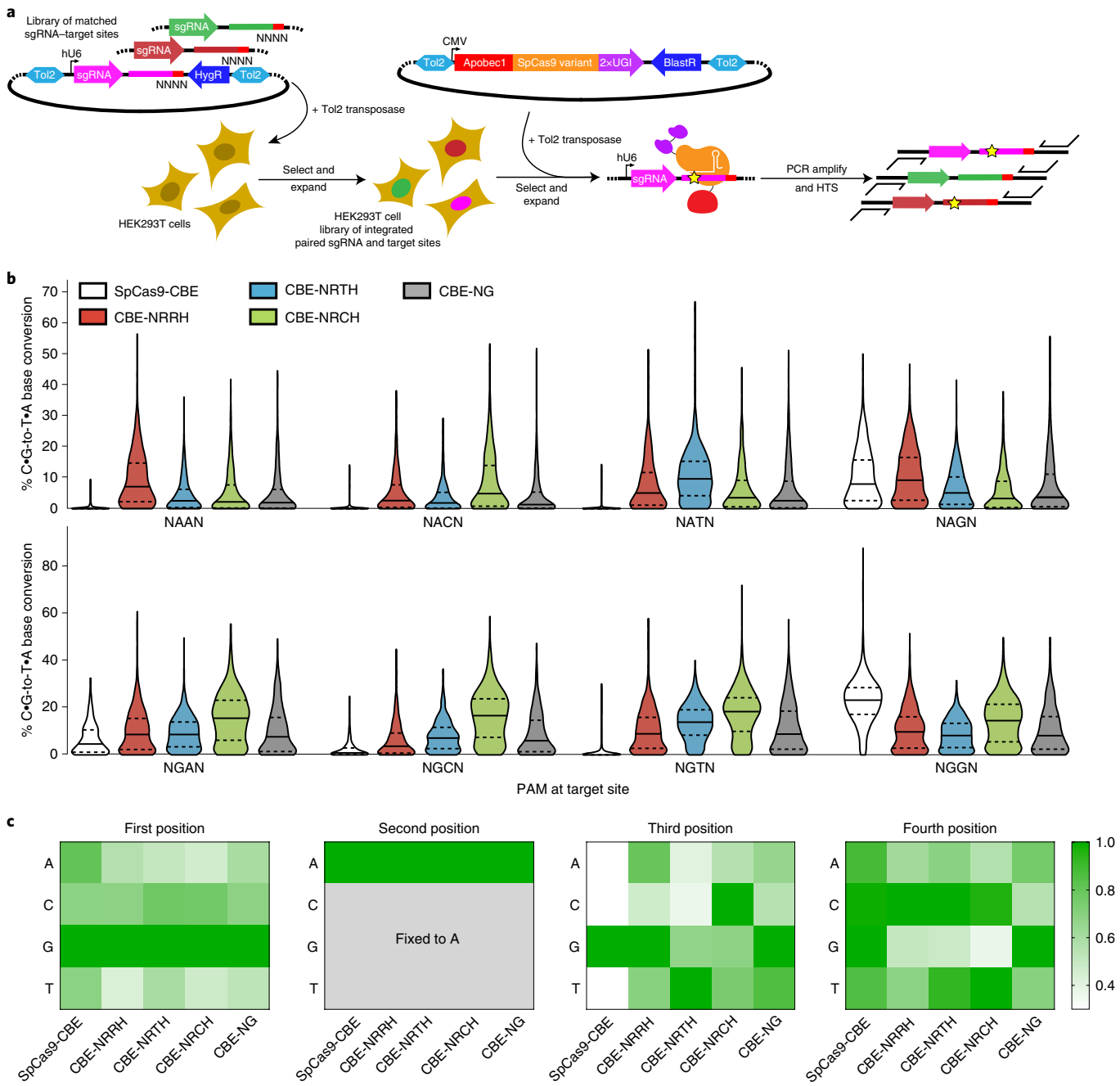


Fig. 3 | Comprehensive characterization of PAM preferences using a genomically integrated human cell target sequence library. a, Schematic overview of human cell base editing library experiments. A library of 11,776 matched sgRNA and protospacer target sites spanning all NNNN PAMs was genomically integrated in HEK293T cells. Library cells were transfected with and selected for genomic integration of plasmids encoding CBE variants. After antibiotic selection, the integrated sgRNA and protospacer site was amplified by PCR for HTS analysis. **b**, Violin plots of base editing activity on the 11,776-member NNNN PAM library in HEK293T cells, with positions 2 and 3 of all NRN PAMs defined. For each construct, the editing across all sites containing the designated PAM over two independent biological replicates is shown, with solid lines indicating median and dotted lines indicating first and third quartiles for $n = 666–763$ target sites per PAM (see Supplementary Table 6 for exact values). **c**, Relative editing activities on the subset of NANN PAMs in the library for SpCas9-CBE, CBE-NRRH, CBE-NRTH, CBE-NRCH and CBE-NG by nucleotide at the first, third or fourth position of the PAM. Full data on the genomically integrated 11,776-member library can be found in Supplementary Table 5.

stringencies, and we developed three new selection strategies to achieve this.

First, we required that the evolving SpCas9 also bind a second, distinct protospacer by leveraging a split-intein pIII (Fig. 2a and Supplementary Fig. 1a)[27]. Here, evolving variants must bind each of two protospacer-PAM sequences to produce both split-intein pIII halves (Supplementary Fig. 1a). Two variants from PANCE (GAA.N1-1 and GAA.N1-4; Figs. 1d and 2b) were subjected to dual-AP PACE targeting a CAA PAM, resulting in five additional consensus variants (A10T, I322V, S409I, E427G and G715C; Supplementary Table 2) that together improved CBE activity in human cells 4.2 fold on average when compared to GAA.N1-1 (CAA.P1-1; Fig. 2c).

Second, we limited the amount of functional SpCas9 produced in host cells to reduce saturation of binding to protospacer-PAM sites by variants with modest affinity[31]. Because both the promoter and ribosome-binding site for ω–dSpCas9 reside on the SP, ω–dSpCas9 expression is subject to selection in PACE and falls outside of the researcher's control. Therefore, we developed a split-SpCas9 strategy in which we fused the C-terminal segment of dSpCas9 (residues 574–1368) to the NpuC split intein[32] (dSpCas9$_C$) on the evolving SP and provided the ω–N-terminal portion (residues 1–573) fused to NpuN (ω–dSpCas9$_N$) in restricted amounts by a complementary plasmid (CP) in host cells (Fig. 2a and Supplementary Fig. 1b). Four mutations (A10T, I322V, S409I and E427G) in residues 1–573 of CAA.P1-1 found to improve SpCas9 binding in phage-based assays were included in the CP supporting all subsequent evolution efforts (Supplementary Fig. 1d). Split-SpCas9 PACE of CAA.P1-1 dSpCas9$_C$ on AAA or CAA PAM targets led to variants CAA.P2-2 and CAA.P3-1, both of which acquired new mutations in the PID (D1180G/K1211R and R1114G/D1180G, respectively) and showed more than double the CBE activity of CAA.P1-1 (Fig. 2c).

Third, we removed gene VI (gVI), which is essential for phage propagation[33], from the SP for use as a second phage selection marker in PACE. This allowed us to combine the selection strategies described above by requiring split-dSpCas9 to bind each of two distinct protospacers to express both gIII and gVI (Fig. 2a and Supplementary Fig. 1c). CAA.P2-1, CAA.P2-2 and CAA.P3-1 were subjected to this highest-stringency selection in PACE, resulting in additional PID mutations (Fig. 2b). However, the resulting evolved variants showed little CBE activity in human cells (Fig. 2c). To remove non-beneficial mutations, we performed DNA shuffling of the C-terminal portion (residues 574–1368) of these evolved variants with that of wild-type SpCas9, leading to the isolation of TAA.P4s-4 (R1114G, D1135N, V1139A, D1180G, E1219V, Q1221H, A1320V and R1333K) (Fig. 2b,d and Supplementary Table 2), which demonstrated an average 1.2-fold increase in human cell CBE activity relative to CAA.P3-1 across all HAA PAM sites tested (Fig. 2c).

We applied the above selection strategies to evolution along the NAT and NAC PAM trajectories but also incorporated a DNA cleavage selection[8] after PANCE to remove nuclease-inactivating mutations that might have contributed to the reduced CBE activity of some variants isolated during NAA PAM evolution (Supplementary Fig. 1e,f and Supplementary Note 1). These experiments resulted in the isolation of TAT.P5-1 (R1114G, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, E1253K, P1321S, D1332G and R1335L) (Fig. 2e,g) from the NAT PAM trajectory and TAC.P9s-3 (R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N, S1338T and H1349R) (Fig. 2f,h) from the NAC PAM trajectory (Supplementary Note 1 and Supplementary Table 2).

Previous efforts to engineer or evolve SpCas9 to accept alternative PAMs have focused on modifying the PID[7,8], which mediates PAM specificity[34]. Our experiments, however, mutagenized either the entire sequence of SpCas9 or residues 574–1368 (SpCas9$_C$), resulting in the accumulation of 3–15 mutations outside the PID (Supplementary Tables 1 and 2). Although mutations within the nuclease domains could impair SpCas9 DNA cleavage activity[35], mutations in the helical domain might benefit SpCas9 DNA binding and unwinding. We added the evolved PIDs to a fixed N-terminal sequence that included the ω–dSpCas9$_N$ mutations (A10T, I322V, S409I and E427G) as well as R654L and R753G, which enriched in multiple independent PACE experiments (Supplementary Fig. 2b and Supplementary Table 2). The addition of all six N-terminal domain mutations to CBEs containing the PIDs of TAA.P4s-4, TAT.P5-1 and TAC.P9s-3 improved CBE activity in human cells an average of 1.5-fold across 4–5 tested sites compared to variants containing the PID mutations alone (Supplementary Fig. 2c). Therefore, we added these six N-terminal mutations to the PID mutations from TAA.P4s-4, TAT.P5-1 and TAC.P9s-3, resulting in final variants SpCas9-NRRH, SpCas9-NRTH and SpCas9-NRCH, respectively (Supplementary Table 2 and Supplementary Notes 1 and 4).

PAM compatibility of evolved variants and SpCas9-NG. We profiled the PAM preferences of SpCas9-NRRH, SpCas9-NRTH, SpCas9-NRCH, SpCas9-NG and SpCas9 in both bacterial PAM depletion assays and a mammalian cell library with genomically integrated target sites and sgRNAs (Fig. 3, Supplementary Figs. 3 and 4 and Supplementary Note 2). For profiling in mammalian cells, we designed a library of 11,776 unique target sequence–sgRNA pairs using 46 distinct protospacers derived from sequences found in the human genome, each with different sequence contexts surrounding a fixed C at position 6, counting the PAM as positions 21–23. Each protospacer was adjacent to a PAM sequence of all possible NNNN tetranucleotides and was additionally flanked with primer-binding sites for high-throughput sequencing (HTS) analysis (Fig. 3a). We genomically integrated this library into HEK293T cells and then treated the library cells with each of the above five SpCas9 variants as optimized CBEs (in the BE4max architecture[36], hereafter referred to as CBE) (Fig. 3a).

HTS of the target library sequences after treatment revealed that, compared to SpCas9-CBE, the evolved variants showed an increased preference for A at the second PAM position and mediated the highest editing activity when the third position matched the nucleotide (A, C or T) on which the variant was evolved (Fig. 3b,c and Supplementary Fig. 4a). CBE-NRRH also preferred a G at the third position (Fig. 3b,c and Supplementary Fig. 4a,d). Interestingly, C at the third position was disfavored for all variants except CBE-NRCH, supporting the idea that CBE-NRCH has a distinctly evolved preference for a C at this position (Fig. 3b,c). CBE-NG was also active on NA PAMs; however, it edited these sites with lower efficiency than our evolved variants (Fig. 3b and Supplementary Fig. 4d). Finally, we observed activity from SpCas9-CBE on its known secondary NAG PAM (Fig. 3b and Supplementary Fig. 4a). CBE-NRCH also performed well across all NGH PAMs, editing with efficiencies similar to or greater than those of CBE-NG at these sites in the library setting (Fig. 3b and Supplementary Fig. 4a). No notable changes in base editing product purity were observed across our SpCas9 variants (Supplementary Fig. 4e).

On the subset of ~3,000 library sequences containing NANN PAMs, our variants showed the highest activity when a non-G was present in the fourth PAM position (Fig. 3c). In contrast, CBE-NG was most active when this position contained a G and, to a lesser extent, an A or T (Fig. 3c). SpCas9-CBE displayed no preference for any base at this position (Fig. 3c). Finally, both CBE-NG and our variants performed best when a G was present at position 1 of the PAM and worst when a T was at this position (Fig. 3b,c and Supplementary Fig. 4a). Taken together, these results suggest that CBE-NRRH, CBE-NRTH and CBE-NRCH are active on NRRH, NRTH and NRCH PAMs, respectively.

Evolved SpCas9 nucleases generate indels at endogenous human genomic loci. Next, we assessed the activity of our variants and SpCas9-NG as nucleases in HEK293T cells at 48 endogenous sites spanning all possible NANH PAMs (Supplementary Table 9). Generally, each of our variants displayed robust indel formation activity on sites containing a PAM that it was evolved to recognize. SpCas9-NRRH generated indels at an average of 27% ± 5.0% and 27% ± 4.4% of the sites containing NAAH and NAGH PAMs, respectively, whereas SpCas9-NRTH mediated an average of 27% ± 4.7% indel formation on NATH PAM sites and SpCas9-NRCH averaged 24% ± 3.9% indel formation on NACH PAM sites (Fig. 4a and Supplementary Fig. 5a). In contrast, SpCas9-NG showed much lower indel formation at these sites, displaying on average 14% ± 2.7%, 15% ± 3.2%, 3.9% ± 0.7% and 7.9% ± 1.8% indel

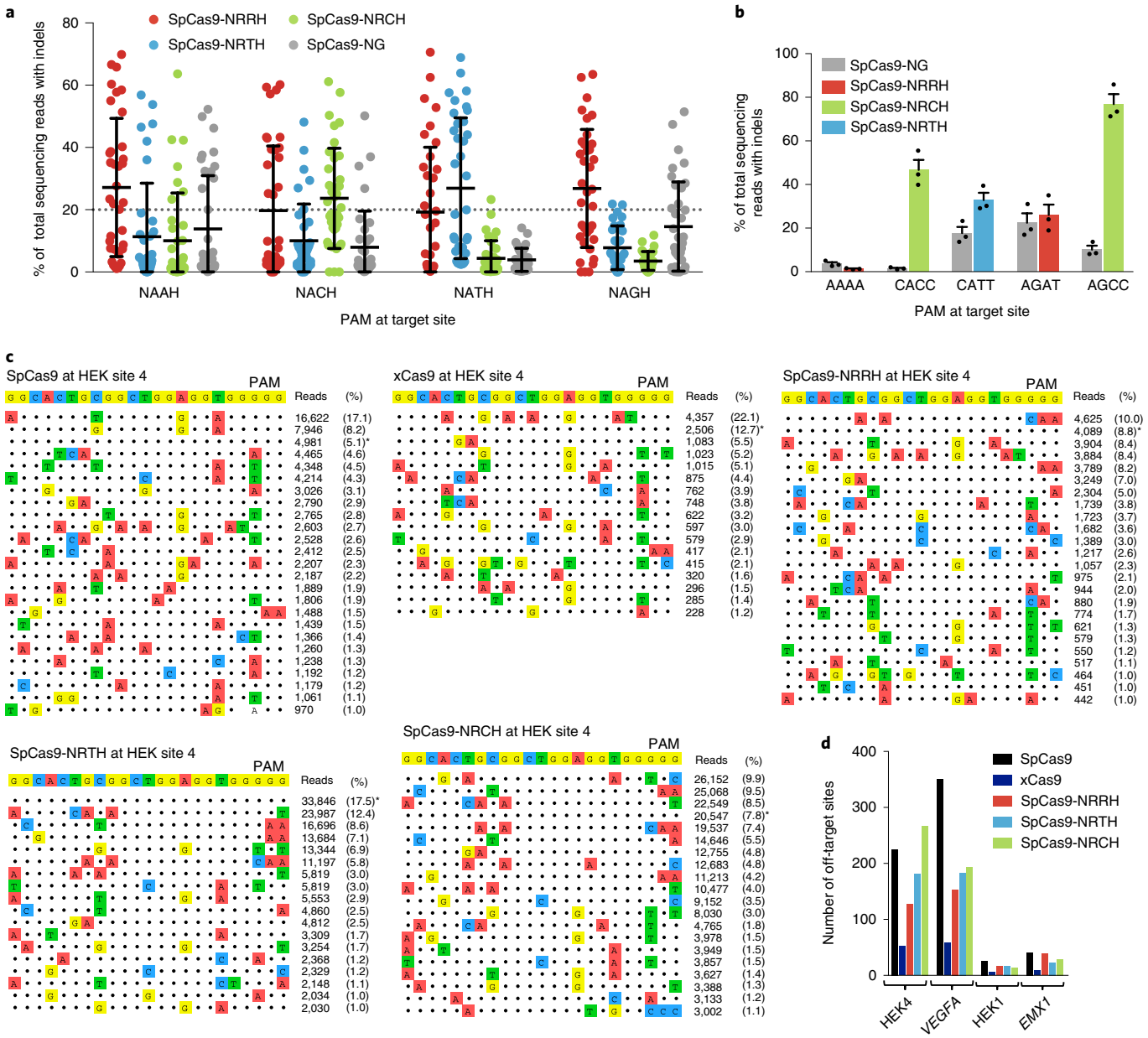


Fig. 4 | Mammalian cell indel formation and DNA specificity of evolved Cas9 variants. a, Summary of indel formation efficiencies in HEK293T cells across 48 endogenous human sites containing NANH PAMs for SpCas9-NRRH, SpCas9-NRTH, SpCas9-NRCH and SpCas9-NG. The mean and s.d. of all individual values of $n = 3$ independent biological replicates are plotted. **b**, Indel formation in primary human fibroblasts across given endogenous human sites containing NR PAMs for SpCas9-NRRH, SpCas9-NRTH, SpCas9-NRCH and SpCas9-NG. Bars represent the mean and s.e.m. of $n = 3$ independent biological replicates, with individual values shown as dots. **c**, GUIDE-seq on-target reads (indicated by an asterisk) and off-target reads for SpCas9, xCas9 and evolved variants SpCas9-NRRH, SpCas9-NRTH and SpCas9-NRCH at HEK site 4. Off-target reads that represented less than 1% of total reads are not shown but are available in Supplementary Table 4. **d**, DNA targeting specificity of SpCas9, xCas9 and evolved variants SpCas9-NRRH, SpCas9-NRTH and SpCas9-NRCH resulting from GUIDE-seq analysis on HEK site 4, *VEGFA* site 2, HEK site 1 and *EMX1* in U2OS cells as determined by the number of off-target sites detected. See Supplementary Fig. 6 for additional GUIDE-seq results.

formation on NAAH, NAGH, NATH and NACH PAMs, respectively (Fig. 4a). We also tested 12 endogenous sites containing NANG PAMs. Consistent with the library results (Fig. 3 and Supplementary Fig. 4), SpCas9-NG displayed higher editing at these sites, resulting in average editing of 13% ± 1.7% on NAAG PAM sites, 20% ± 2.5% on NAGG PAM sites, 23% ± 3.6% on NATG PAM sites and 14% ± 3.0% on NACG PAM sites (Supplementary Fig. 5a). Finally, we observed minimal indel formation by xCas9 at NAN sites, as expected (Supplementary Fig. 5b). These results suggest that SpCas9-NRRH, SpCas9-NRTH and SpCas9-NRCH support higher editing efficiencies at NANH PAM sites than previously reported SpCas9 variants.

We also tested the indel formation activity of our evolved variants and SpCas9-NG on endogenous sites containing NGN PAMs in HEK293T cells. Treatment with SpCas9-NG led to robust indel formation on most NGN PAMs examined (averaging 41% ± 6.2%, 60% ± 6.5%, 62% ± 3.2% and 42% ± 3.0% on NGA, NGG, NGT and NGC PAM sites, respectively). SpCas9-NRTH, however,

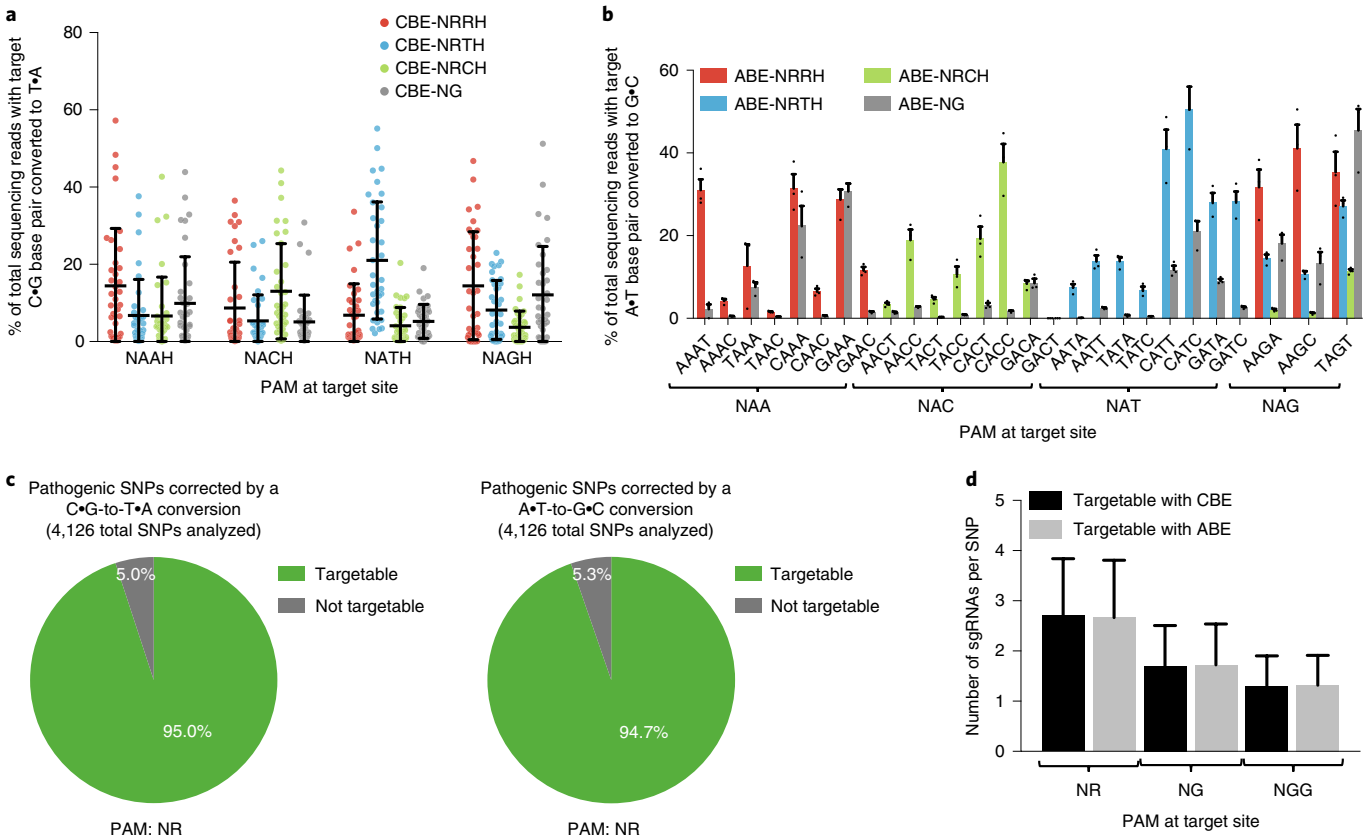


Fig. 5 | Mammalian cytosine and adenine base editing activity and scope of evolved variants and SpCas9-NG. a, Summary of cytosine base editing in HEK293T cells across 48 endogenous human sites containing NANH PAMs for CBE-NRRH, CBE-NRTH, CBE-NRCH and CBE-NG. The mean and s.d. of all individual values of $n = 3$ independent biological replicates are shown. **b**, Adenine base editing in HEK293T cells across 27 endogenous human sites containing NANN PAMs for ABE-NRRH, ABE-NRTH, ABE-NRCH and ABE-NG. Bars represent the mean and s.e.m. of three independent biological replicates, shown as dots. **c**, Fraction of pathogenic SNPs in the ClinVar database that, in principle, can be corrected by a C•G-to-T•A (left) or A•T-to-G•C (right) base conversion using NR PAMs. **d**, Mean number of possible sgRNAs capable of targeting each pathogenic SNP in the ClinVar database using NR, NG or NGG PAMs. The mean and s.e.m. of the number of targeting sgRNAs are shown for $n = 3{,}919$, $n = 12{,}095$, $n = 1{,}154$, $n = 9{,}740$, $n = 1{,}132$ and $n = 3{,}841$ individual ClinVar entries for CBE/NR, ABE/NR, CBE/NG, ABE/NG, CBE/NGG and ABE/NGG, respectively.

showed higher activity than SpCas9-NG at NGT PAMs, averaging 68% ± 4.0% indel formation (Supplementary Fig. 5c). Consistent with our PAM characterization results, we also observed a preference for H at PAM position 4 for SpCas9-NRTH and SpCas9-NRCH on NGTH and NGHH sites, respectively (Supplementary Fig. 5c). In particular, SpCas9-NRTH and SpCas9-NRCH supported higher average editing efficiencies on sites with NGTH (67% ± 2.0%) and NGCH (43% ± 1.4%) PAMs, respectively, compared to SpCas9-NG (NGTH, 61% ± 1.7%; NGCH, 32% ± 1.5%). These findings suggest that, at NGNH PAM sites, the SpCas9 variants evolved in this work offer similar or higher editing efficiencies than previously reported variants.

Finally, we characterized the ability of the evolved SpCas9 variants to edit five endogenous loci in primary human fibroblasts. Nucleofection of mRNA encoding SpCas9-NRCH, SpCas9-NRTH and SpCas9-NRRH resulted in 47% ± 8.0%, 33% ± 5.7% and 1.3% ± 0.3% average indel formation at three sites containing CACC, CATT and AAAA PAMs, respectively (Fig. 4b), whereas SpCas9-NG resulted in 1.5% ± 0.5%, 18% ± 5.0% and 3.6% ± 1.3% average indel formation at the same sites (Fig. 4b). SpCas9-NRRH and SpCas9-NRCH also induced 26% ± 8.2% and 77% ± 7.9% indel formation at sites containing AGAT and AGCC PAMs, respectively, whereas SpCas9-NG averaged 22% ± 7.5% and 10% ± 3.1% indel formation at these two sites (Fig. 4b). These results suggest that our evolved variants were able to edit genomic sites in primary human cells after mRNA delivery, although editing efficiencies remain subject to variability from locus-specific determinants.

DNA specificity of evolved SpCas9 nucleases. As broadening the PAM targeting capabilities of some Cas9 variants has been shown to increase the proportion of genomic off-target edits[5,7], we characterized the DNA specificity of our variants by performing GUIDE-seq[37] on SpCas9, SpCas9-NRRH, SpCas9-NRCH and SpCas9-NRTH in U2OS cells on four characterized on-target genomic sites[6–8,37]. For comparison, we also analyzed xCas9, which has greatly reduced off-target activity compared to SpCas9 (refs. [6,38–40]). On all four sites examined, our evolved variants displayed higher on-target activity and similar or fewer numbers of detected off-target sites compared to SpCas9, although xCas9 was the most highly specific variant tested (Fig. 4c,d and Supplementary Fig. 6). For the four sites tested, SpCas9, SpCas9-NRRH, SpCas9-NRCH, SpCas9-NRTH and xCas9 averaged 81%, 64%, 61%, 55% and 53% off-target reads among all GUIDE-seq reads, respectively (Fig. 4c,d and Supplementary Fig. 6). The three new variants also primarily displayed off-target activity at sites containing PAMs consistent with their evolved preferences (Fig. 4c and Supplementary Fig. 6a–c). Taken together, these results suggest that, even though the evolved variants can access a broader set of off-target sequences, they have similar or higher overall DNA specificity and on-target activity than SpCas9 on NGG PAM sites. Similarly to high-fidelity SpCas9 variants[6,40–43],

SpCas9-NRRH, SpCas9-NRTH and SpCas9-NRCH exhibited greater sensitivity to mismatches in the 5′ end of the sgRNA spacer, which are commonly tolerated by wild-type SpCas9 (Supplementary Fig. 4b,c and Supplementary Note 3)[44].

Evolved SpCas9 variants support cytosine and adenine base editing. Expanding the targeting scope of base editing is a major motivation behind the development of SpCas9 variants with diversified PAM compatibilities. We tested the evolved Cas9 variants for CBE activity at the 48 endogenous NANH PAM sites examined above for indel formation (Fig. 5a, Supplementary Fig. 7a and Supplementary Table 9). As with their nuclease forms, each of the three evolved CBE variants showed the highest average activity on sites containing the PAM that it was evolved to recognize. Thus, CBE-NRRH and CBE-NRTH showed the highest activity on C bases within the canonical base editing window at NAAH and NATH PAM sites, averaging 14% ± 2.4% and 21% ± 2.4% C•G-to-T•A conversion, respectively (Fig. 5a). CBE-NRCH on NACH PAM sites was slightly less efficient, editing with an average of 13.0% ± 2.0% base conversion (Fig. 5a). Finally, CBE-NRRH also edited NAGH sites with 17% ± 2.6% average base conversion (Fig. 5a). Average CBE editing efficiency across all 48 sites was lower than that of indel formation, likely owing to increased requirements for efficient base editing such as sequence context, position of the C within the editing window and formation of an R-loop that is conformationally accessible to the cytidine deaminase domain. These editors also functioned on sites with NGN PAMs, with average efficiencies of 17% ± 2.3%, 9.1% ± 3.0%, 19% ± 2.9% and 20% ± 4.0% for CBE-NRRH, CBE-NRTH, CBE-NRCH and CBE-NG, respectively (Supplementary Fig. 7b).

We also generated ABEmax[36] variants (hereafter referred to as ABE) from SpCas9-NRRH, SpCas9-NRTH, SpCas9-NRCH and SpCas9-NG and measured adenine base editing at 54 endogenous loci in HEK293T cells. We observed that the newly evolved variants were also compatible with adenine base editing at sites containing NANH and NGN PAMs (Fig. 5b and Supplementary Fig. 7c). ABE-NRRH, ABE-NRTH, ABE-NRCH and ABE-NRRH edited most efficiently at NAAH, NATH, NACH and NAGH PAM sites, with 16% ± 2.6%, 24% ± 2.9%, 13% ± 2.2% and 26% ± 3.5% average A•T-to-G•C base conversion, respectively (Fig. 5b).

Evolved SpCas9 variants enable base editing of a previously inaccessible pathogenic single-nucleotide polymorphism. Base editing requires that the target base be located within the CBE or ABE editing window (approximately protospacer positions 4–8 for CBE and positions 4–7 for ABE, counting the PAM as positions 21–23)[12]. SpCas9-NRRH, SpCas9-NRCH and SpCas9-NRTH, together with previously reported NG-PAM-compatible SpCas9 variants[6,7], expand the targeting scope of SpCas9 to most genomic sites containing NR PAMs and greatly increase the fraction of known human pathogenic single-nucleotide polymorphisms (SNPs) that can theoretically be corrected by base editing. Of all pathogenic SNPs in the ClinVar database[45] that can be corrected by C•G-to-T•A conversion, 95% are, in principle, now targetable with CBEs derived from SpCas9-NRRH, SpCas9-NRCH, SpCas9-NRTH or SpCas9-NG/xCas9 (Fig. 5c). Likewise, 95% of pathogenic SNPs in ClinVar that are correctable upon A•T-to-G•C conversion might, in principle, be targeted with ABEs derived from the same set of SpCas9 variants (Fig. 5c). These new variants also approximately double the number of possible protospacers that position a target SNP within the optimal window for base editing compared to SpCas9 (Fig. 5d). Because 38% and 47% of coding SNPs are accompanied by a non-silent A or C bystander nucleotide within the canonical ABE or CBE editing window, respectively[12], the expansion of accessible base editing sites to include NR PAMs frequently enables multiple targeting strategies to optimize editing of the desired base, increasing the likelihood of target SNP-selective base editing by 1.4- to 1.6-fold (Supplementary Fig. 7d).

To demonstrate the utility of SpCas9 variants evolved in this study in a disease-relevant context, we targeted the HbS allele of the β-globin (*HBB*) gene that is the most common cause of sickle cell anemia[46]. HbS arises from a GAG-to-GTG (glutamic acid-to-valine) mutation at HBB amino acid 6 that cannot be reverted through current base editing technologies. However, this SNP can be converted with adenine base editing to GCG (alanine) through A•T-to-G•C conversion on the opposite strand (Fig. 6a) to yield the *HBB* E6A genotype, known as the Makassar allele (HbG), which is thought to be non-pathogenic in both homozygous and heterozygous individuals[47–49]. Unfortunately, the only NGG and NGN PAMs available at this site place the target A at protospacer positions 2 and 9, respectively, which are outside the editing window for ABE[12]. However, two alternative protospacers using CATG and CACC PAMs place the target A at positions 4 and 7, respectively, with a bystander A present at position 6 or 9 leading to a silent CCT-to-CCC (proline-to-proline) mutation. Thus, we tested the ability of our evolved variants, along with SpCas9-NG, to convert the sickle cell SNP to the Makassar mutation using these two non-G PAM protospacer sites.

We transfected ABE-NRRH, ABE-NRTH, ABE-NRCH or ABE-NG into HEK293T cells homozygous for the HbS allele (Supplementary Fig. 8a). Whereas ABEs derived from the SpCas9 variants evolved in this study supported substantial (14–55%) A•T-to-G•C conversion using sgRNAs targeting either the CATG or CACC PAM, ABE-NG edited efficiently (40% ± 0.2%) only with the protospacer containing a CATG PAM (Fig. 6b,c), perhaps owing to the presence of a G at the fourth position of this PAM. Unfortunately, editing using the CATG PAM protospacer occurred primarily at the silent bystander base (position 6), with the target A (position 4) edited with less than 10% efficiency across all four ABEs tested (Fig. 6b and Supplementary Fig. 8b).

Targeting the HbS mutation using the CACC PAM protospacer, however, was much more efficient. As expected, ABE-NRCH showed the highest editing activity, with 41% ± 3.8% base conversion at the target A (position 7) and 13% ± 3.2% base conversion at the silent bystander A (position 9) (Fig. 6c and Supplementary Fig. 8c). ABE-NRRH and ABE-NRTH achieved 29% ± 4.3% and 14% ± 2.8% conversion, respectively (Fig. 6c and Supplementary Fig. 8c). In comparison, ABE-NG showed negligible (1.0% ± 0.5%) target base conversion activity at this site (Fig. 6c and Supplementary Fig. 8c). Collectively, these results demonstrate that the newly evolved SpCas9 variants enable efficient base editing of previously inaccessible pathogenic SNPs using non-G PAMs and highlight the utility of evaluating multiple protospacer and PAM sequences when targeting a particular SNP.

Discussion

Using phage-assisted evolution (PACE and PANCE), we identified three new variants of SpCas9 that recognize NRRH, NRCH and NRTH PAM sequences. Generating SpCas9 variants active on non-G PAMs required improved selection strategies for evolving Cas9:DNA binding, such as increasing the number of target DNA protospacer-PAM binding sites by using an additional PACE-compatible selection circuit that expresses gVI, implementing a dual-AP split-gIII strategy and limiting the total concentration of functional SpCas9 in the host cell through split-intein SpCas9. These new selections should be applicable to other Cas9 orthologs to enable the evolution of Cas9 variants targeting additional PAM sequences that remain difficult to access.

Our initial experiments evolving SpCas9 for binding on all 16 individual NAN PAMs identified three distinct groups of consensus mutations corresponding to NAA, NAT and NAC PAM targets. These mutations give insight into potential modes of PAM interaction. SpCas9-NRRH acquired a mutation at R1333, which, in

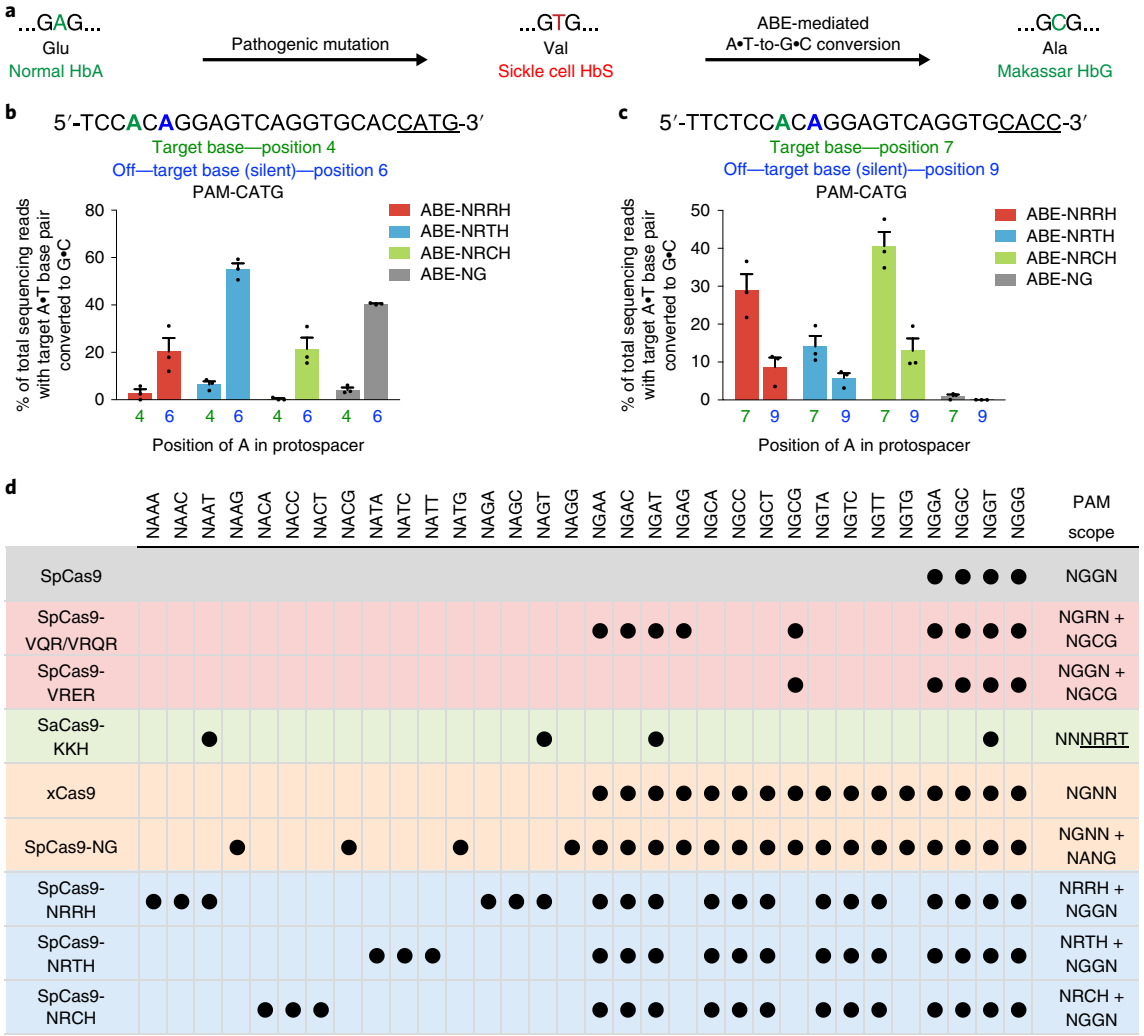


Fig. 6 | Evolved SpCas9 variants enable correction of pathogenic SNPs with non-G PAMs. a, Overview of the adenine base editing strategy for correcting the sickle cell hemoglobin (HbS) SNP. In HbS, the glutamic acid (GAG codon) at position 6 of normal β-globin (*HBB*) is mutated to a valine (GTG codon). Targeting this SNP with adenine base editing on the reverse strand enables a valine-to-alanine (GTG-to-GCG) base conversion, leading to the rare Makassar β-globin variant (HbG) that is thought to be non-pathogenic. **b**, Adenine base editing in HEK293T cells engineered with the HbS mutation using a CATG PAM by ABE-NRRH, ABE-NRTH, ABE-NRCH and ABE-NG. This PAM places the target A at position 4 and an off-target A at position 6, leading to a silent proline-to-proline (CCT-to-CCC) mutation when converted to a G. **c**, Adenine base editing in HEK293T cells engineered with the *HbS* mutation using a CACC PAM by ABE-NRRH, ABE-NRTH, ABE-NRCH and ABE-NG. This PAM places the target A at position 7 and an off-target A, which leads to a silent proline-to-proline (CCT-to-CCC) mutation at position 9. For **b** and **c**, bars represent the mean and s.e.m. of $n = 3$ independent biological replicates, with individual values shown as dots. **d**, Recommended Cas9 variants for accessing all possible PAMs within the NRNN PAM space. Only Cas9 variants that require recognition of three or fewer defined nucleotides in their PAMs are listed. The variants evolved and characterized in this study are highlighted in blue.

SpCas9, contacts the first G in its canonical PAM, but not R1335, which contacts the second G in NGG (Fig. 2b,d). The R1333K mutation likely allows SpCas9-NRRH to accept both A and G at the second PAM nucleotide, whereas the preservation of R1335 might explain why this variant recognizes both NAA and NAG. On the other hand, SpCas9-NRTH preserves R1333 but replaces R1335 with leucine (Fig. 2e,g). Interestingly, SpCas9-NRTH shows a strong preference for T in the third PAM position and appears to have lost some recognition of the wild-type NGG PAM (Fig. 3d and Supplementary Figs. 3 and 4a,b). Finally, SpCas9-NRCH evolved altered interactions at both R1335 and T1337 (Fig. 2f,h); T1337N in particular might form contacts with a fourth PAM nucleotide to compensate for weakened binding interactions with the NAC target PAM.

We also observed several additional substitutions that likely modulate more general interactions with the target and non-target DNA, including R1114G, E1219V, Q1221H and D1135N (Fig. 2b,d–h). Residue E1219 forms hydrogen bonds with R1335 in SpCas9, and mutations at this position are thought to destabilize the interaction between R1335 and the third PAM base[38]. Mutations at D1135 have been previously reported[7,8] and might modulate interactions with the sugar-phosphate backbone of the non-target DNA strand; R1114G and Q1221H could alter similar interactions. Finally, we observed mutations in the helical domain of Cas9 that arose in several independently evolving populations (Supplementary Fig. 2b and Supplementary Tables 1 and 2). These mutations, when added to the N-terminal region of NRRH and NRCH, improved their recognition of non-G PAMs in base editing experiments (Supplementary Fig. 2c) and might contribute to increasing the overall DNA binding or unwinding activity of these variants.

In addition to testing on 92 individual sites in human cells and performing bacterial PAM library depletion experiments, we also

characterized the PAM compatibilities of the evolved variants and SpCas9-NG using a 11,776-member protospacer and NNNN PAM library genomically integrated into HEK293T cells. The very large number and diversity of sites in this library allowed us to comprehensively profile the editing activity of these proteins using all NNNN PAMs in a genomic context and further illuminated their sequence preferences, for instance demonstrating that our variants displayed a different and complementary fourth-position PAM preference (H) compared to SpCas9-NG (G). Although further investigation is required to explain the fourth-base preferences of these mutants, crystal structures of SpCas9-NG and other evolved SpCas9 variants (VRER and VRQR) suggest that the T1337R mutation in these variants might create a direct interaction with the fourth-position G[7,50,51]. In light of these sequence preferences, we suggest testing the three major variants reported here along with SpCas9-NG when optimizing targeting efficiency on sites with NR PAMs. We recommend SaCas9 and SpCas9 variants to test for targeting any NRNN PAM in Fig. 6d. We note that variants with access to smaller subsets of NR PAM space are not included in this table but have also been shown to mediate genome editing in human cells[52–60].

The variants evolved in this study, together with SpCas9-NG and xCas9, substantially expand the utility of SpCas9 for disease-relevant genome-editing applications. Access to a broad range of PAMs is especially essential for base editing, as illustrated by experiments targeting the sickle cell mutation of human β-globin, in which high levels of target base conversion and low levels of bystander editing were made possible by our variants using an adjacent CACC PAM (Fig. 6c). That the sickle cell mutation lies in the optimal ABE window for both sgRNAs tested, but is efficiently edited using only one sgRNA, demonstrates the benefits of accessing multiple protospacer sequences for a single target. Our observation of substantial genome editing in primary human cells suggests that these variants might prove useful in a range of cell types (Fig. 4b). Although indel formation, cytosine base editing and adenine base editing are shown in this work, we anticipate that our evolved variants should also be compatible with other Cas9-mediated genome editing technologies, including transcriptional modulation[61,62], HDR[9] and predictable template-free genome editing[10,11].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-020-0412-8.

Received: 27 September 2019; Accepted: 7 January 2020;
Published online: 10 February 2020

References

1. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**, 20–36 (2017).
2. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR–Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **20**, 490–507 (2019).
3. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
4. Cebrian-Serrano, A. & Davies, B. CRISPR–Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. *Mamm. Genome* **28**, 247–261 (2017).
5. Kleinstiver, B. P. et al. Broadening the targeting range of *Staphylococcus aureus* CRISPR–Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33**, 1293–1298 (2015).
6. Hu, J. H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
7. Nishimasu, H. et al. Engineered CRISPR–Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).
8. Kleinstiver, B. P. et al. Engineered CRISPR–Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
9. Paquet, D. et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).
10. Shen, M. W. et al. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646–651 (2018).
11. Iyer, S. et al. Precise therapeutic gene correction by a simple nuclease-induced double-stranded break. *Nature* **568**, 561–565 (2019).
12. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–778 (2018).
13. Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
14. Jin, S. et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292–295 (2019).
15. Xin, H., Wan, T. & Ping, Y. Off-targeting of base editors: BE3 but not ABE induces substantial off-target single nucleotide variants. *Signal Transduct. Target. Ther.* **4**, 9 (2019).
16. Lee, H. K. et al. Targeting fidelity of adenine and cytosine base editors in mouse embryos. *Nat. Commun.* **9**, 4804 (2018).
17. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
18. Dickinson, B. C., Leconte, A. M., Allen, B., Esvelt, K. M. & Liu, D. R. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proc. Natl Acad. Sci. USA* **110**, 9007–9012 (2013).
19. Carlson, J. C., Badran, A. H., Guggiana-nilo, D. A. & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat. Chem. Biol.* **10**, 216–222 (2014).
20. Pu, J., Disare, M. & Dickinson, B. C. Evolution of C-terminal modification tolerance in full-length and split T7 RNA polymerase biosensors. *Chembiochem* **20**, 1547–1553 (2019).
21. Pu, J., Kentala, K. & Dickinson, B. C. Multidimensional control of Cas9 by evolved RNA polymerase-based biosensors. *ACS Chem. Biol.* **13**, 431–437 (2018).
22. Pu, J., Zinkus-Boltz, J. & Dickinson, B. C. Evolution of a split RNA polymerase as a versatile biosensor platform. *Nat. Chem. Biol.* **13**, 432–438 (2017).
23. Packer, M. S., Rees, H. A. & Liu, D. R. Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nat. Commun.* **8**, 956 (2017).
24. Dickinson, B. C., Packer, M. S., Badran, A. H. & Liu, D. R. A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nat. Commun.* **5**, 5352 (2014).
25. Hubbard, B. P. et al. Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat. Methods* **12**, 939–942 (2015).
26. Badran, A. H. et al. Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance. *Nature* **533**, 58–63 (2016).
27. Wang, T., Badran, A. H., Huang, T. P. & Liu, D. R. Continuous directed evolution of proteins with improved soluble expression. *Nat. Chem. Biol.* **14**, 972–980 (2018).
28. Bryson, D. I. et al. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **13**, 1253–1260 (2017).
29. Suzuki, T. et al. Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.* **13**, 1261–1266 (2017).
30. Roth, T. B., Woolston, B. M., Stephanopoulos, G. & Liu, D. R. Phage-assisted evolution of *Bacillus methanolicus* methanol dehydrogenase 2. *ACS Synth. Biol.* **8**, 796–806 (2019).
31. Karvelis, T. et al. Rapid characterization of CRISPR–Cas9 protospacer adjacent motif sequence elements. *Genome Biol.* **16**, 253 (2015).
32. Shah, N. H., Dann, G. P., Vila-Perelló, M., Liu, Z. & Muir, T. W. Ultrafast protein splicing is common among cyanobacterial split inteins: implications for protein engineering. *J. Am. Chem. Soc.* **134**, 11338–11341 (2012).
33. Brödel, A. K., Jaramillo, A. & Isalan, M. Engineering orthogonal dual transcription factors for multi-input synthetic promoters. *Nat. Commun.* **7**, 13858 (2016).
34. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
35. Jiang, F. & Doudna, J. A. CRISPR–Cas9 structures and mechanisms. *Annu. Rev. Biophys.* **46**, 505–529 (2017).
36. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).
37. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR–Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
38. Guo, M. et al. Structural insights into a high fidelity variant of SpCas9. *Cell Res.* **29**, 183–192 (2019).
39. Kim, S., Bae, T., Hwang, J. & Kim, J. S. Rescue of high-specificity Cas9 variants using sgRNAs with matched 5′ nucleotides. *Genome Biol.* **18**, 218 (2017).
40. Lee, J. K. et al. Directed evolution of CRISPR–Cas9 to increase its specificity. *Nat. Commun.* **9**, 3048 (2018).

41. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
42. Chen, J. S. et al. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
43. Kleinstiver, B. P. et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
44. Zhang, D. et al. Perfectly matched 20-nucleotide guide RNA sequences enable robust genome editing using high-fidelity SpCas9 nucleases. *Genome Biol.* **18**, 191 (2017).
45. Landrum, M. J. et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980–D985 (2014).
46. Rees, D. C., Williams, T. N. & Gladwin, M. T. Sickle-cell disease. *Lancet* **376**, 2018–2031 (2010).
47. Blackwell, R. Q. et al. Hemoglobin G Makassar: β6 Glu→Ala. *Biochim. Biophys. Acta.* **214**, 396–401 (1970).
48. Viprakasit, V., Wiriyasateinkul, A., Sattayasevana, B., Miles, K. L. & Laosombat, V. Hb G-Makassar [β6(A3)Glu→Ala; codon 6 (GAG→GCG)]: molecular characterization, clinical, and hematological effects. *Hemoglobin* **26**, 245–253 (2002).
49. Sangkitporn, S. et al. Hb G Makassar (β6: Glu-Ala) in a Thai family. *J. Med. Assoc. Thai* **85**, 577–582 (2002).
50. Hirano, S., Nishimasu, H., Ishitani, R. & Nureki, O. Structural basis for the altered PAM specificities of engineered CRISPR–Cas9. *Mol. Cell* **61**, 886–894 (2016).
51. Anders, C., Bargsten, K. & Jinek, M. Structural plasticity of PAM recognition by engineered variants of the RNA-guided endonuclease Cas9. *Mol. Cell* **61**, 895–902 (2016).
52. Chatterjee, P., Jakimo, N. & Jacobson, J. M. Minimal PAM specificity of a highly similar SpCas9 ortholog. *Sci. Adv.* **4**, eaau0766 (2018).
53. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
54. Esvelt, K. M. et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10**, 1116–1121 (2013).
55. Edraki, A. et al. A compact, high-accuracy Cas9 with a dinucleotide PAM for in vivo genome editing. *Mol. Cell* **73**, 714–726 (2019).
56. Hirano, H. et al. Structure and engineering of *Francisella novicida* Cas9. *Cell* **164**, 950–961 (2016).
57. Harrington, L. B. et al. A thermostable Cas9 with increased lifetime in human plasma. *Nat. Commun.* **8**, 1424 (2017).
58. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system. *Cell* **163**, 759–771 (2015).
59. Hou, Z. et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc. Natl Acad. Sci. USA* **110**, 15644–15649 (2013).
60. Kim, E. et al. In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.* **8**, 14500 (2017).
61. Mali, P. et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).
62. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR–Cas9 complex. *Nature* **517**, 583–588 (2015).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

General methods. Antibiotics (Gold Biotechnology) were used at the following working concentrations: carbenicillin, 50 μg ml^{-1}; spectinomycin, 50 μg ml^{-1}; chloramphenicol, 25 μg ml^{-1}; kanamycin, 50 μg ml^{-1}; tetracycline, 10 μg ml^{-1}; streptomycin, 50 μg ml^{-1}. Nuclease-free water (Qiagen) was used for PCR reactions and cloning. For all other experiments, water was purified using a MilliQ purification system (Millipore). Unless otherwise noted, Phusion U Hot Start or Phusion Hot Start II DNA polymerase (Thermo Fisher Scientific) was used for all PCRs. Unless otherwise noted, plasmids and SPs were cloned by USER assembly[26]. Genes were obtained as synthesized gene fragments from Twist Bioscience. Plasmids were cloned and amplified using either Mach1 (Thermo Fisher Scientific) or Turbo (New England Biolabs) cells. Plasmid or SP DNA was amplified using the Illustra Templiphi 100 Amplification Kit (GE Healthcare Life Sciences) before Sanger sequencing. Strain S2060 (ref. [25]) was used in all luciferase, phage propagation and plaque assays and in all PACE experiments. A description of the plasmids, SP, primers and protospacer sequences used in this work is provided in Supplementary Tables 7–9.

Phage propagation assay. Chemicompetent S2060 cells were transformed with the AP(s) of interest as previously described[27]. Overnight cultures of single colonies grown in 2× YT medium supplemented with maintenance antibiotics were diluted 1,000 fold into DRM medium with maintenance antibiotics and grown at 37 °C with shaking at 230 RPM to OD_{600} ~0.4–0.6. Cells were then infected with SP at an initial titer of 5×10^4 plaque-forming units (p.f.u.) ml^{-1}. Cells were incubated for another 16–18 h at 37 °C with shaking at 230 r.p.m. and then centrifuged at 8,000*g* for 2 min. The supernatant containing phage was removed and stored at 4 °C until use.

Plaque assay. Chemicompetent S2060 cells were transformed with the AP(s) of interest. Overnight cultures of single colonies grown in 2× YT medium supplemented with maintenance antibiotics were diluted 1,000-fold into fresh 2× YT medium with maintenance antibiotics and grown at 37 °C with shaking at 230 r.p.m. to OD_{600} ~0.6–0.8 before use. SPs were serially diluted 100-fold (four dilutions total) in water. Then, 150 μl of cells was added to 10 μl of each phage dilution, and to this 1 ml of liquid (55 °C) top agar (2× YT medium and 0.5% agar) supplemented with 2% Bluo-gal (Gold Biotechnology) was added and mixed by pipetting up and down once. This mixture was immediately pipetted onto one quadrant of a quartered Petri dish already containing 2 ml of solidified bottom agar (2× YT medium and 1.5% agar, no antibiotics). Plates were incubated at 37 °C for 16–18 h.

PANCE. PANCE was performed as previously reported with the following modifications[29,30]. SPs using the SP41 backbone encoding either ω–dSpCas9 or ω–dxCas9 were evolved on host S2060 cells transformed with pSM072a APs containing the G7′ protospacer sequence followed by each NAN PAM and DP6 (ref. [63]; to give 32 total evolving populations). To reduce the likelihood of contamination with gIII-encoding recombined SP, phage stocks were purified as previously described[26]. With the exception of the first passage, all passages were infected with a 1:100 dilution of SP isolated from the previous passage. For the first passage, host cells were infected with SP to give a final titer of 10^6 p.f.u. ml^{-1}. The first ten passages were run as follows: 6-h propagation with 10 mM arabinose and 50 ng ml^{-1} anhydrotetracycline (aTc) followed by 18-h propagation with 10 mM arabinose only (no drift induced). The next four passages were run as follows: 6-h propagation with 10 mM arabinose and 25 ng ml^{-1} aTc followed by 18-h propagation with 10 mM arabinose only (no drift induced). The final five passages were allowed to propagate for 18 h with 10 mM arabinose only (no drift induced). The SP titers were estimated using a qPCR-based method.

qPCR estimation of phage titer. Phage titers were estimated using a modified version of a previously reported qPCR-based procedure[64]. SP pools from PANCE passages (25–50 μl) were first heated at 80 °C for 30 min to destroy polyphage. Polyphage genomes were then degraded by adding 5 μl of heated SP to 45 μl of 1× DNase I buffer containing 0.5 μl DNase I (New England Biolabs) and incubated at 37 °C for 20 min followed by 95 °C for 20 min. Next, qPCR reaction mixtures were prepared by mixing 20 μl water, 5 μl 5× Phusion HF Buffer (dye-free, Thermo Fisher Scientific), 0.5 μl dNTPs (10 mM, New England Biolabs), 0.25 μl SYBR green (100×, Thermo Fisher Scientific), 0.25 μl of each primer (5′-CACCGTTCATCTGTCCTCTTT-3′ and 5′-CGACCTGCTCCATGTTACTTAG-3′; Supplementary Table 8) and 1.5 μl DNase-treated SP pool. qPCR conditions were as follows: 98 °C for 2 min, 40 cycles of 98 °C for 10 s, 60 °C for 20 s and 72 °C for 15 s and 12 °C for 30 s. Titers were calculated using a titration curve of SP standard of known titer.

PACE. Unless otherwise noted, the PACE apparatus, including host cell strains, lagoons, chemostats and media, were all used as previously described[17,26]. To reduce the likelihood of contamination with gIII-encoding recombined SP, phage stocks were purified as previously described[26]. Continuous dilution was performed using Masterflex L/S Digital Drive pumps (Cole-Parmer) fitted with Masterflex L/S Multichannel pump heads (Cole-Parmer).

Chemically competent S2060 cells were transformed with AP(s), CP (if using) and MP6 or DP6 (ref. [63]), plated on 2× YT medium and 1.5% agar supplemented with 25 mM glucose (to prevent induction of mutagenesis) and maintenance antibiotics and grown at 37 °C for 18–20 h. Four colonies were picked into 1 ml DRM each in a 96-well deep-well plate, and this was diluted fivefold eight times serially into DRM. The plate was sealed with a porous sealing film and grown at 37 °C with shaking at 230 r.p.m. for 16–18 h. Dilutions with OD_{600} ~0.4–0.8 were then used to inoculate a chemostat containing 80 ml DRM. The chemostat was grown to OD_{600} ~0.8–1.0 and then continuously diluted with fresh DRM at a rate of ~1.5 chemostat volumes per hour. The chemostat was maintained at a volume of 60–80 ml.

Before SP infection, lagoons were continuously diluted with culture from the chemostat at 1 lagoon volume per hour and pre-induced with 10 mM arabinose for at least 2 h. If DP6 was used, the lagoons were also pre-induced with aTc. Lagoons were infected with SP at a starting titer of 10^6 p.f.u. ml^{-1} and maintained at a volume of 15 ml. Samples (500 μl) of the SP population were taken at indicated times from lagoon waste lines. These were centrifuged at 8,000*g* for 2 min, and the supernatant was stored at 4 °C. Lagoon titers were determined by plaque assays using S2060 cells transformed with pJC175e. For Sanger sequencing of lagoons, single plaques were PCR amplified using primers AB1793 (5′-TAATGGAAACTTCCTCATGAAAAAGTCTTTAG-3′) and AB1396 (5′-ACAGAGAGAATAACATAAAAACAGGGAAGC-3′) (Supplementary Table 8), both of which anneal to regions of the M13 phage backbone flanking the evolving gene of interest. Generally, eight plaques were picked and sequenced per lagoon.

NAA evolution. P1: Host cells transformed with pTW168a3, pTW169a3 and MP6 were maintained in a 40-ml chemostat. The lagoons were cycled at 1 volume per hour with 10 mM arabinose for 4 h before infection with phage with the SP47 backbone containing the CTD mutations from GAA.N1-2 and GAA.N1-4. The lagoon dilution rate was increased to 1.5 volume per h at 19 h and 2.0 volume per h at 44 h or 68 h. The experiment ended at 72 h.

P2: Host cells transformed with pTW199b3, pTW221b5 and MP6 were maintained in a 40-mL chemostat. The lagoon was cycled at 1 volume per hour with 10 mM arabinose for 4 h before infection with phage with the SP56 backbone containing residues 574–1368 of CAA.P1-2 fused to NpuC. Upon infection, lagoon dilution rates were decreased to 0.5 volume per h. The lagoon dilution rate was increased to 1.5 volumes per hour at 96 h, 2.0 volumes per hour at 136 h, and 3.0 volumes per hour at 168 h. The experiment ended at 192 h.

P3: Host cells transformed with pTW199b1, pTW221b1 or pTW221b3 and MP6 were maintained in a 40-ml chemostat. The lagoon was cycled at 1 volume per hour with 10 mM arabinose for 4 h before infection with the pool from P2. Upon infection, lagoon dilution rates were decreased to 0.5 volumes per hour. The lagoon dilution rate was increased to 1.0 volume per hour at 19 h, 2.0 volumes per hour at 43 h and 2.5 volumes per hour at 68 h. The experiment ended at 72 h.

P4: Host cells transformed with pTW199b2, pTW170b2, pTW221b5 and MP6 were maintained in a 40-ml chemostat. The lagoon was cycled at 1 volume per hour with 10 mM arabinose for 4 h before infection with phage with the SP58 backbone containing CAA.P2-1, CAA.P2-2 and CAA.P3-1 fused to NpuC. Upon infection, lagoon dilution rates were decreased to 0.5 volumes per hour. The lagoon dilution rate for lagoon 1 was increased to 1.0 volume per hour at 19 h, 1.5 volumes per hour at 43 h and 2.0 volumes per hour at 68 h before washing out. The lagoon dilution rate for lagoon 2 was increased to 0.75 volumes per hour at 19 h, 1.0 volume per hour at 43 h, 1.5 volumes per hour at 68 h, 2.0 volumes per hour at 92 h and 3.0 volumes per hour at 114 h. The experiment ended at 114 h.

NAT evolution. P5: Host cells transformed with pTW199b9 (lagoons 1 and 2) or pTW199b10 (lagoons 3 and 4), pTW221b5 and MP6 were maintained in a 40-ml chemostat. The lagoon was cycled at 1 volume per hour with 10 mM arabinose for 4 h before infection with phage with the SP56 backbone containing residues 574–1368 of SacB.TAT-1, SacB.TAT-2 and SpCas9-NG fused to NpuC. Upon infection, lagoon dilution rates were decreased to 0.5 volumes per hour. The lagoon dilution rate for lagoons 1 and 2 was increased to 1.0 volume per hour at 43 h, 1.5 volumes per hour at 68 h and 3.0 volumes per hour at 138 h. The lagoon dilution rate for lagoons 3 and 4 was increased to 1.0 volume per hour at 19 h, 1.5 volumes per hour at 43 h, 2.0 volumes per hour at 68 h and 3.0 volumes per hour at 138 h. The experiment ended at 164 h.

P6: Host cells transformed with pTW199b10, pTW170b10, pTW221b5 and MP6 were maintained in a 40-ml chemostat. The lagoon was cycled at 1 volume per hour with 10 mM arabinose for 4 h bfore infection with phage with the SP58 backbone containing residues 574–1368 of TAT.P5-1 and TAT.P5-2 fused to NpuC. Upon infection, lagoon dilution rates were decreased to 0.5 volumes per hour. The lagoon dilution rate was increased to 1.0 volume per hour at 19 h, 1.5 volumes per hour at 40 h, 2.0 volumes per hour at 68 h and 3.0 volumes per hour at 92 h. The experiment ended at 114 h.

NAC evolution. P7a: Host cells transformed with pTW199b5 (lagoons 1 and 2) or pTW199b7 (lagoons 3 and 4), pTW221b5 and MP6 were maintained in a 40-ml chemostat. The lagoon was cycled at 1 volume per hour with 10 mM arabinose

for 4 h before infection with phage with the SP56 backbone containing residues 574–1368 of SacB.CAC fused to NpuC. Upon infection, lagoon dilution rates were decreased to 0.5 volume per h. The lagoon dilution rate was increased to 1.0 volume per hour at 19 h and 1.5 volumes per hour at 40 h, before washing out by 66 h.

P7b: Host cells transformed with pTW199b5 (lagoons 1 and 2) or pTW199b6 (lagoons 3 and 4), pTW221b5 and MP6 were maintained in a 40-ml chemostat. The lagoon was cycled at 1 volume per hour with 10 mM arabinose for 4 h before infection with the surviving phage pool from P7a. Upon infection, lagoon dilution rates were decreased to 0.5 volumes per hour. The lagoon dilution rate for lagoons 1 and 3 was increased to 1.0 volume per hour at 43 h and 1.5 volumes per hour at 68 h. The lagoon dilution rate for lagoons 2 and 4 was increased to 1.0 volume per hour at 43 h and 2.0 volumes per hour at 68 h. All lagoons became contaminated with gIII-recombination phage. Lagoons 1 and 2 were collected and sequenced at 48 h, and lagoons 3 and 4 were collected at 64 h (not sequenced).

P8: Host cells transformed with pTW199b5 (lagoons 1 and 2) or pTW199b6 (lagoons 3 and 4), pTW221b5 and MP6 were maintained in a 40-ml chemostat. The lagoon was cycled at 1 volume per hour with 10 mM arabinose for 4 h before infection with the surviving phage pool from P7b. Upon infection, lagoon dilution rates were decreased to 0.5 volumes per hour. The lagoon dilution rate for lagoons 1 and 3 was increased to 1.0 volume per hour at 19 h, 2.0 volumes per hour at 43 h and 3.0 volumes per hour at 116 h. The lagoon dilution rate for lagoons 2 and 4 was increased to 1.0 volume per hour at 19 h, 1.5 volumes per hour at 43 h, 2.0 volumes per hour at 116 h and 3.0 volumes per hour at 140 h. The experiment ended at 166 h.

P9: Host cells transformed with pTW199b6, pTW170b6, pTW221b5 and MP6 were maintained in a 40-ml chemostat. The lagoon was cycled at 1 volume per hour with 10 mM arabinose for 4 h before infection with phage with the SP58 backbone containing residues 574–1368 of AAA.P8-1, AAC.P8-3 and TAC.P8-3 fused to NpuC. Upon infection, lagoon dilution rates were decreased to 0.5 volumes per hour. The lagoon dilution rate was increased to 1.0 volume per hour at 19 h, 1.5 volumes per hour at 40 h, 2.0 volumes per hour at 68 h and 3.0 volumes per hour at 92 h. The experiment ended at 114 h.

DNA shuffling. DNA shuffling was performed using a modified version of the nucleotide exchange and excision technology method[65,66]. SP libraries encoding evolved ω–SpCas9$_C$ were amplified by PCR using primers 5′-GCCAGTTACAA AATAAACAGCC-3′ and 5′-CACCGTTCATCTGTCCTCTTT-3′ (Supplementary Table 8) to provide template for the initial uracil incorporation PCR. Clonal SP encoding wild-type ω–SpCas9$_C$ was similarly amplified to provide a wild-type template. Separate uracil incorporation PCRs were performed using primers 5′-GACTCCCTGCAAGCCTCAG-3′ and 5′-ACAGAGAGAATAACATAAAA ACAGGGAAGC-3′ (Supplementary Table 8) on both wild-type and evolved library templates, and the resulting amplicons were combined in a 1:4 ratio of wild-type to library before digesting with USER enzyme as previously described[65]. Digested fragments were purified using the QiaxII Gel Extraction Kit (Qiagen) according to the manufacturer's protocol. The purified fragments were reassembled by internal primer extension, as described previously. The reassembly PCR was performed as follows: 94 °C for 3 min and 50 cycles of 92 °C for 30 s, 22–72 °C and 1 °C per cycle for 60 s, 72 °C for 60 s and 5 s per cycle, and finally 72 °C for 10 min. The reassembled SpCas9$_C$ was then amplified with primers 5′-ATGTTGAAAATCTCCTTCUAGATCAGTCTCCACCGAGCTGAG-3′ and 5′-ATCGCCAGCAAUTGTTTCGACTCTGTTGAAATCAG-3′ (Supplementary Table 8) and cloned through USER assembly into either SP or plasmids.

PAM depletion assay. DH10β cells (New England Biolabs) were transformed with plasmids (pTW222) encoding nuclease-active SpCas9 variants under the control of the arabinose promoter and an sgRNA targeting the HEK3 protospacer sequence. Cells were then made electrocompetent[67] and electroporated with a PAM depletion plasmid library (pTPH308) that contained the HEK3 protospacer sequence followed by a randomized NNNNN PAM, as well as a kanamycin resistance gene. After electroporation, cells were immediately incubated with SOC outgrowth medium (New England Biolabs) for 1 h before 10 mM arabinose was added to induce SpCas9 expression for 1 h, 3 h or overnight. Cells were then plated on 2× YT medium and 1.5% agar supplemented with carbenicillin and kanamycin (both 50 μg ml^{-1}). The plates were incubated at 37 °C for 16–18 h, and the resultant colonies were scraped and plasmids were extracted using the Qiagen Plasmid Plus Midiprep Kit (Qiagen) following the manufacturer's protocol. The protospacer-PAM sequence of the extracted plasmids as well as the input PAM depletion plasmid library was amplified with primers 5′- ACACTCTTTCCC TACACGACGCTCTTCCGATCTNNNNCAATACGCAAACCGCCTCTC-3′ and 5′- TGGAGTTCAGACGTGTGCTCTTCCGATCTCTTGTCTGT AAGCGGATGC-3′ (Supplementary Table 8). The resulting amplicons were then amplified using Illumina barcoding primers, and the product was purified by gel extraction using a QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's protocol. DNA concentration was quantified with the KAPA Library Quantification Kit-Illumina (Kapa Biosystems), and the resultant libraries were sequenced using an Illumina MiSeq high-throughput sequencer. Sequencing data were analyzed using a previously reported Python script[8]. Depletion scores for any given PAM were calculated by taking the ratio of the frequency of that PAM in the input library to its frequency after selection and then multiplying this ratio by an arbitrary scaling factor.

Cloning of nuclease-active SpCas9 libraries for SacB plasmid cleavage selection. Post-selection SpCas9 SP pools were subcloned into the pTW222 vector, which encodes nuclease-active SpCas9 variants under the control of the arabinose promoter and an sgRNA targeting the HEK3 protospacer sequence on the SacB selection plasmid using an USER enzyme mixture supplemented with T7 ligase (New England Biolabs) as follows: 10 μl 2× T7 ligase buffer, 100,000 U T7 ligase, 1 μl USER enzyme mixture and 1 μl DpnI were mixed with 0.5 pmol of each DNA fragment, and nuclease-free water was added to a final volume of 20 μl total. The reaction was incubated at 37 °C for 1 h, heated to 80 °C and slowly (1 °C s^{-1}) cooled to 12 °C, purified using the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's protocol and transformed into electrocompetent DH10β cells (New England Biolabs). Cells were immediately plated after electroporation on 2× YT medium and 1.5% agar supplemented with carbenicillin (50 μg ml^{-1}) and incubated at 37 °C for 16–18 h. The resultant colonies were then scraped and plasmids were extracted using the Qiagen Plasmid Plus Midiprep Kit (Qiagen) following the manufacturer's protocol.

SacB plasmid cleavage selection. DH10β cells (New England Biolabs) transformed with pTW218, which encodes *Bacillus subtilis* levansucrase (SacB), the HEK3 protospacer with the target PAM, sfGFP and a kanamycin resistance cassette, were made electrocompetent and transformed with the pTW222-based SpCas9 library plasmid (2 μl per 25 μl cells). After electroporation, cells were immediately incubated with SOC outgrowth medium (New England Biolabs) for 1 h in SOC before addition of carbenicillin (50 μg ml^{-1}) and arabinose (10 mM). Cells were incubated for 1–18 h before being plated on 2× YT medium and 1.5% agar supplemented with carbenicillin (50 μg ml^{-1}) and sucrose (100 mM). Plates were incubated at 37 °C for 16–18 h, and the SpCas9 variants encoded by surviving colonies were characterized by Sanger sequencing.

HEK 293T transfection and genomic DNA extraction. HEK 293T cells (American Type Culture Collection (ATCC), CRL-3216) maintained in DMEM plus GlutaMAX (Thermo Fisher Scientific) supplemented with 10% (vol/vol) FBS at 37 °C with 5% CO_2 were seeded on 96-well plates (Corning). Then, 16–20 h after seeding, cells were transfected at approximately 80–85% confluency with 200 ng of SpCas9 or base editor plasmid, 40 ng of sgRNA plasmid and 0.5 μl of Lipofectamine 2000 (Thermo Fisher Scientific). Cells were cultured for 3 d after transfection before the medium was removed, cells were washed with 1× PBS (Thermo Fisher Scientific) and genomic DNA was extracted by the addition of 30 μl lysis buffer (10 mM Tris-HCl, pH 7.0, 0.05% SDS, 25 μg ml^{-1} proteinase K) (Thermo Fisher Scientific). Genomic DNA was stored at −20 °C until further use.

Cas9 mRNA in vitro transcription. A fragment containing a T7 promoter driving expression of poly-adenylated Cas9 transcript was isolated from purified plasmid using SpeI restriction digestion and purified with the MinElute PCR Purification Kit (Qiagen). mRNA was transcribed using the HiScribe T7 ARCA mRNA Kit (NEB) and purified using lithium chloride precipitation according to the manufacturer's instructions.

Human primary fibroblast nucleofection and genomic DNA extraction. Next, 2.5×10^5 primary human fibroblasts (GM04541, Coriell) cultured in DMEM with GlutaMAX (Thermo Fisher Scientific) supplemented with 20% (vol/vol) FBS at 37 °C with 5% CO_2 were transfected with 50 pmol of sgRNA (Synthego) and 1 μg of in vitro-transcribed SpCas9 mRNA. Twenty microliters of P2 primary cell solution (Lonza) was used with program DS-150 on a Lonza Nucleofector 4-D, and cells were plated on 12-well plates. The medium was changed after 24 h of transfection. Cells were cultured for 5 d after transfection before the medium was removed, cells were washed with 1× PBS (Thermo Fisher Scientific) and genomic DNA was extracted by the addition of 200 μl lysis buffer (10 mM Tris-HCl, pH 7.0, 0.05% SDS, 25 μg ml^{-1} proteinase K) (Thermo Fischer Scientific). Genomic DNA was stored at −20 °C until further use.

HTS of genomic DNA. Genomic sites were amplified with primers targeting the region of interest (Supplementary Table 8) and the appropriate universal Illumina forward and reverse adapters. Primer pairs for the first round of PCR (PCR 1) either were reported previously[68] or are listed in Supplementary Table 4. Twenty-five-microliter-scale PCR 1 reactions used 1.25 μl each of 10 μM forward and reverse primers and 0.5 μl genomic DNA extract in 25 μl of Phusion HS II PCR mix prepared according to the manufacturer's protocol (Thermo Fisher Scientific). PCR 1 conditions were as follows: 98 °C for 2 min and then 30 cycles of 98 °C for 15 s, 61 °C for 20 s and 72 °C for 15 s, followed by a final 72 °C extension for 2 min. PCR products were verified by comparison with DNA standards (Quick-Load 2-Log Ladder, New England Biolabs) on a 2% agarose gel supplemented with ethidium bromide. Unique Illumina barcoding primers were subsequently appended to each PCR 1 sample in a second PCR reaction (PCR 2). PCR 2 reactions used 1.25 μl each of 10 μM forward and reverse Illumina barcoding primers and 1 μl of unpurified PCR 1 reaction product in 25 μl of Phusion HS II PCR mix prepared according

to the manufacturer's protocol (Thermo Fisher Scientific). PCR 2 conditions were as follows: 98 °C for 2 min and 12 cycles of 98 °C for 15 s, 61 °C for 20 s and 72 °C for 20 s, followed by a final 72 °C extension for 2 min. PCR products were pooled and purified by electrophoresis with a 2% agarose gel using the QIAquick Gel Extraction Kit (Qiagen) eluting with 15 µl water. DNA concentration was quantified with the KAPA Library Quantification Kit-Illumina (Kapa Biosystems) and sequenced on an Illumina MiSeq instrument (paired-end reads: R1, 280 cycles; R2, 0 cycles) according to the manufacturer's protocols.

General HTS data analysis. Sequencing reads were demultiplexed using the MiSeq Reporter (Illumina), and fastq files were analyzed using Crispresso2 (ref. [69]).

GUIDE-seq. Next, 3×10^5 U2OS cells maintained in McCoy's 5A medium (Thermo Fisher Scientific) supplemented with 10% (vol/vol) FBS at 37 °C with 5% CO_2 were transfected with 750 ng of Cas9 plasmid, 250 ng of sgRNA plasmid and 5 pmol of the GUIDE-seq dsODN[37]. A control sample with GFP plasmid (Lonza) instead of Cas9 plasmid was also included. Twenty microliters of solution SE (Lonza) was used with program DN-100 on a Lonza Nucleofector 4-D, and cells were plated on 24-well plates. Three wells of transfected cells were used per condition for GUIDE-seq processing. Genomic DNA was extracted using the Quick-DNA Miniprep Plus Kit (Zymo Research) following the manufacturer's protocol. The DNA was sheared to an average length of 500 bp using a Covaris S200 focused ultrasonicator following the manufacturer's protocol. End repair, dA tailing, adaptor ligation, tag-specific CPR1 and PCR 2 were carried out using the primers and methods described previously[37]. DNA concentration was quantified with the KAPA Library Quantification Kit-Illumina (Kapa Biosystems) and normalized to 4 nM; a maximum of six samples were sequenced on an Illumina MiSeq instrument according to the manufacturer's protocols for Nextera paired-end sequencing (150|150|8|8). Samples were analyzed using GUIDE-seq analysis software (http://www.jounglab.org/guideseq).

Human cell 11,776-member integrated target site and sgRNA library. *Library cloning.* A specified pool of 12,000 oligonucleotides 200 nucleotides in length was synthesized by Twist Bioscience, pairing an SpCas9 gRNA spacer with its 61-bp target site. We included 46 spacers, each preceding every possible 4-nucleotide combination to analyze any NNNN PAM. Libraries were then cloned into a backbone (71485, Addgene) containing a U6 promoter to facilitate gRNA expression, a hygromycin resistance cassette and flanking Tol2 transposon sites to facilitate integration into the genome[70]. Cloning was performed as previously reported[10] with the following modifications. The synthetic library was amplified with NEBNext Ultra II Q5 Master Mix (New England Biolabs) polymerase using a 2-min initial denaturing and extension time, using primers 'oligonucleotide library forward' (5′-TTTTTGTTTTCTGTGTTCCG TTGTCCGTGCTGTAACGAAAGGATGGGTGCGACGCGTCAT-3′) and 'oligonucleotide library reverse' (5′-GTTGATAACGGACTAGCCT TATTTAAACTTGCTATGCTGTTTCCAGCATAGCTCTTAAAC-3′). The donor template for circular assembly was amplified from an SpCas9 sgRNA expression plasmid (71485, Addgene)[70] using primers 'circular donor forward' (5′-GTTTAAGAGCTATGCTGGAAACAGC-3′) and 'circular donor reverse' (5′- ACTGCACCGCGTCGCACCCATCCTTTCGTTACAGCACGGACAACG GAACACAGACAAAACAAAAAAGCACCGACTC-3′) to amplify the sgRNA hairpin and terminator and extended further with a linker region meant to separate the gRNA expression cassette from the target site in the final library. Library and donor templates were ligated by Gibson Assembly (New England Biolabs) for 1 h at 50 °C, and unligated fragments were digested with Plasmid Safe ATP-Dependent DNase (Lucigen) for 1 h at 37 °C. Assembled circularized sequences were linearized by digestion with SspI (New England Biolabs) for ≥3 h at 37 °C. The linearized fragment was further amplified using primers 'plasmid insert forward' (5′- TAACTTGAAAGTATTTCGATTTCTTGGCTTTATATA TCTTGTGGAAAGGACGAAACACCG-3′) and 'plasmid insert reverse' (5′-TTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTC TGCTCGAAGCGGCCGTACCTCTAGAACACTCTTTCCCTACACGAC-3′) for the addition of overhangs complementary to the 5′ and 3′ regions of the gRNA expression plasmid (71485, Addgene)[70] previously digested with BbsI and XbaI (New England Biolabs), to facilitate gRNA expression, hygromycin selection and integration of the library into the genome of human cells. Assembled plasmids were purified by isopropanol precipitation with GlycoBlue Coprecipitant (Thermo Fisher Scientific), reconstituted in water and transformed into DH10β electrocompetent cells (New England Biolabs). The plasmid library was isolated by Maxiprep plasmid purification (Zymo Research). Library integrity was verified by restriction digest with SapI (New England Biolabs) for 1 h at 37 °C, and sequence diversity was validated by HTS as described below.

Tol2 BE4 cassette cloning. For library experiments, the BE4 base editor plasmid was expressed from a modified BE4max backbone (112093, Addgene)[36] containing a blasticidin resistance cassette and flanked by 5′ and 3′ Tol2-transposon sequences, hereafter p2T-CMV-BE4max-BlastR. These plasmids were assembled with the In-Fusion Cloning Kit (Takara Bio) through insertion of a PCR-amplified BE4 variant after digestion of the backbone with NotI and AgeI.

Cell culture. HEK293T cells were purchased from the ATCC and cultured as recommended by the ATCC. Cell lines were authenticated by the suppliers and tested negative for mycoplasma. For stable Tol2 transposon plasmid integration, cells were transfected using Lipofectamine 3000 (Thermo Fisher Scientific) with a 1:2 ratio of Tol2 transposase plasmid[71] to transposon-containing plasmid, for a total of 80 µg of DNA per 15-cm plate. For library integration, two 15-cm plates with $>2\times10^7$ initial cells were used, and cells were maintained on 17.5 µg ml^{-1} hygromycin starting 1 d after transfection, for a minimum of 2 weeks. Library cells were validated by HTS as described below. For library editing experiments, two 15-cm plates with $>2\times10^7$ initial cells were used, and cells were maintained on 3 µg ml^{-1} blasticidin starting 1 d after transfection.

High-throughput DNA sequencing. Genomic DNA was collected from cells after 4 d of selection. Library samples were prepped for HTS in two steps as previously reported[10] with the following modifications. PCR1 was performed with NEBNext Ultra II Q5 Master Mix polymerase (New England Biolabs) using an Illumina TruSeq P5 index primer (5′-AATGATACGGCGACCACCGAGATCTACAC NNNNNNNN ACACTCTTTCCCTACACGAC-3′) and 'sequencing reverse 1' (5′-GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCT GTGGAAAGGA CGAAACACCG-3′) for 12 cycles using a 2-min initial denaturing and extension time. Products were purified and cycle number for PCR 2 was determined by qPCR to avoid over-amplification of the product. PCR 2 was performed with 'sequencing forward 2' (5′-AATGATACGGCGACCACCGAGATCTACAC-3′) and an Illumina TruSeq P7 index primer (5′-CAAGCAGAAGACGGCAT ACGAGAT NNNNNNNN GTGACTGGAGTTCAGACGTGTGCTCTTC-3′). Purified products were validated and quantified by Tape Station (Agilent) and pooled for HTS. Pools were quantified by KAPA Illumina Library Quantification Kit (Kapa Biosystems) and sequenced on an Illumina NextSeq instrument.

Data processing of library HTS data. Sequencing reads were assigned to library target sites by locality-sensitive hashing as previously described[10]. Then, 61-nucleotide target contexts that were intentionally designed to be highly similar to each other (varying only in the NNNN PAM region) were designed with a 5-nucleotide barcode to assist accurate assignment. Sequence alignment was performed using Smith–Waterman with the following parameters: match, +1; mismatch, −1; indel start, −5; indel extend, 0. Nucleotides with a PHRED score below 30 were assumed to be the wild-type nucleotide. Aligned reads with no indels were retained for analysis. Events were defined as C-to-C or C-to-T substitutions at all Cs in the target site. A single read constituted an observation of a single event. A given read with multiple substrate nucleotides was considered to have an edit if at least one nucleotide was edited. Events in treatment conditions were adjusted by control conditions as follows: any editing event with a frequency above 5% in controls was filtered in treatment as treatments underwent an additional selection step beyond control conditions. Treatment events with frequencies that could be explained by control events were filtered at a false discovery rate of 0.05 in a statistical procedure considering the difference between binomial distributions. The frequencies of all control events were subtracted from the frequencies of corresponding treatment events to a minimum of 0. Frequencies of treatment events that could be explained by a binomial distribution parameterized by the Illumina substitution sequencing error estimated as each nucleotide's PHRED score were filtered at a false discovery rate of 0.05. C-to-T events occurring between protospacer positions −9 to 20 were retained, where position 1 was the 5′-most nucleotide of the protospacer and 0 was used to refer to the position between −1 and 1. The fraction of reads with any C-to-T event was tabulated for each target site and retained for downstream analysis.

Generation of HEK 293T cells containing the sickle cell anemia mutation. HEK 293T cells maintained in DMEM plus GlutaMAX (Thermo Fisher Scientific) supplemented with 10% (vol/vol) FBS at 37 °C with 5% CO_2 at ~60% confluency in a 48-well plate were transfected with 300 ng of a plasmid for human expression of Cas9(D10A), 100 ng of sgRNA plasmid (spacer: GTAACGGCAGACTTC TCCTC) and 200 ng of ssODN (ACTTCATCCACGTTCACCTTGCCCCAC AGGGCAGTAACGGCAGACTTCTCCACAGGAGTCAGATGCACCATGGTG TCTGTTTGAGGTTGCTAGTGAACAC). The DNA-containing solution was complexed with 1.5 µl Lipofectamine 2000 to a total volume of 25 µl in Opti-MEM (Thermo Fisher Scientific) before addition to cells. The medium was changed 3 d after transfection. Four days after transfection, cells were trypsinized using TrypLE solution (Thermo Fisher Scientific) and suspended in 1.5 ml of medium. A Beckman Coulter Astrios fluorescence-activated cell sorting instrument was used to isolate single cells into individual wells of two 96-well plates (100 µl of medium per well). After seven more days of growth, cells were dispersed in 10 µl of TrypLE solution, pipetted gently to disperse clumps and left to reattach to the same plate in 100 µl of fresh media. After 7 more days of growth, cells were split 1:1 into fresh wells, with half of the cells harvested for sequencing. An Illumina MiSeq was used to deep sequence the *HBB* locus using primers 5′-ACACTCTTTCC CTACACGACGCTCTTCCGATCTNNNNAGGGTTGGCCAATCTACTCCC-3′ and 5′-TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTTCTCTGTCTCC ACATGCC-3′. At this point, four clones harbored the sickle cell HbS mutation, but none above a frequency of 60%, indicating that the polyploid cell line was

not entirely converted. The above process of transfection, sorting and growth was repeated with one such clone, after which we identified a clone with 100% conversion to the mutant HbS allele. Cells were sequence verified for the sickle cell mutation using an Illumina MiSeq as described above.

Conversion of the sickle cell anemia mutation to the *HBB* Makassar allele. HEK293T cells homozygous for the sickle cell anemia mutation were seeded in 48-well plates and transfected at ~60% confluency with 150 ng of ABE plasmid pSM069 and 50 ng of sgRNA plasmid encoding spacer sequences HbS-CATG and HbS-CACC (Supplementary Table 9) complexed with 1 µl of Lipofectamine 3000. Three days after transfection, the medium was aspirated and exchanged for fresh medium. Five days after transfection, genomic DNA was extracted for PCR and deep sequencing of the target locus. The *HBB* primers listed above were used for amplification of genomic DNA and sequencing of the sickle cell mutation site using an Illumina MiSeq.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

HTS data were deposited in the Sequence Read Archive database (PRJNA596996). Plasmids encoding SpCas9-NRRH, SpCas9-NRTH and SpCas9-NRCH nucleases, BE4 and ABE are available through Addgene. A subset of the selection plasmids used in this study are available through Addgene. Other materials are available from the corresponding authors upon reasonable request.

Code availability

Custom scripts used to analyze the human cell target site library are available at https://github.com/maxwshen-PAMvar-processing.

References

63. Badran, A. H. & Liu, D. R. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* **6**, 8425 (2015).
64. Peng, X., Nguyen, A. & Ghosh, D. Quantification of M13 and T7 bacteriophages by TaqMan and SYBR green qPCR. *J. Virol. Methods* **242**, 100–107 (2018).
65. Gaudelli, N. M. et al. Programmable base editing of T to G C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
66. Müller, K. M. et al. Nucleotide exchange and excision technology (NExT) DNA shuffling: a robust method for DNA fragmentation and directed evolution. *Nucleic Acids Res.* **33**, e117 (2005).
67. Dower, W. J., Miller, J. F. & Ragsdale, C. W. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* **16**, 6127–6145 (1988).
68. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
69. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
70. Arbab, M., Srinivasan, S., Hashimoto, T., Geijsen, N. & Sherwood, R. I. Cloning-free CRISPR. *Stem Cell Rep.* **5**, 908–917 (2015).
71. Urasaki, A., Morvan, G. & Kawakami, K. Functional dissection of the Tol2 transposable element identified the minimal *cis*-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics* **174**, 639–649 (2006).

Acknowledgements

We thank T.R. Blum and A. Raguram for helpful discussions. This work was supported by National Institutes of Health grants R01 EB027793, U01 AI142756, RM1 HG009490 and R35 GM118062, the Howard Hughes Medical Institute, the Bill and Melinda Gates Foundation and the St. Jude Collaborative Research Consortium. S.M.M. and M.S. were supported by a National Science Foundation Graduate Research Fellowship. T.W. was supported by a Ruth L. Kirchstein National Research Service Awards Postdoctoral Fellowship (F32GM119228). M.A. was supported by an NWO Rubicon Fellowship. G.A.N was supported by a Helen Hay Whitney Postdoctoral Fellowship.

Author contributions

S.M.M., T.W. and D.R.L. designed the research. S.M.M., T.W. and P.B.R. performed evolution. S.M.M., T.W., P.B.R. and T.P.H. characterized variants in bacteria. S.M.M., T.W., P.B.R., M.A. and Z.M. conducted human cell experiments. M.A. and M.S. designed and constructed the integrated target site human cell library. M.S. wrote custom software. S.M.M., T.W., M.A. and M.S. analyzed library data. S.M.M. and T.W. performed GUIDE-seq experiments. G.N. and H.A.R. designed and performed the sickle cell anemia site experiments. S.M.M., T.W., P.B.R. and T.P.H. designed and performed the HbS editing experiments. S.M.M., T.W. and D.R.L. wrote the manuscript.

Competing interests

The authors declare competing financial interests: S.M.M., T.W. and D.R.L. have filed patent applications on this work. D.R.L. is a consultant and co-founder of Editas Medicine, Pairwise Plants, Beam Therapeutics and Prime Medicine, companies that use genome-editing technologies.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41587-020-0412-8.

Correspondence and requests for materials should be addressed to D.R.L.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): David R. Liu

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Illumina Miseq Control software (3.1) was used on the Illumina Miseq sequencers to collect the high-throughput sequencing data
Data analysis	Crispresso2 was used to analyze HTS data for quantifying editing activity at genomic sites. Custom scripts were used to process and analyze library data, and are available at https://github.com/maxwshen/pamvar-processing. Weblogo 3 was used for sequence logos and is available at http://weblogo.threeplusone.com/create.cgi. GUIDEseq scripts are available at https://github.com/aryeelab/guideseq

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

High-throughput sequencing data will be available from the NCBI Sequence Read Archive database under accession code PRJNA596996. Plasmids encoding SpCas9-

NRRH, SpCas9-NRTH, and SpCas9-NRCH nucleases, BE4 and ABE will be available through Addgene. A subset of selection plasmids used in this study will be available through Addgene. Other materials are available upon reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were determined based on literature precedence for directed evolution.
Data exclusions	No data was excluded
Replication	Excepting the evolutions (where replication is not standard practice) and some phage experiments (noted in the figure legends), all experiments were repeated at least once. All attempts at replication were successful.
Randomization	Bacteria and mammalian cells used in this study were grown under identical conditions; no randomization was used.
Blinding	Bacteria and mammalian cells used in this study were grown under identical conditions; blinding was not used

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All materials are readily available.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T (ATCC), U2OS (ATCC), Human Fibroblasts (Coriell GM04541)
Authentication	Cells were authenticated by the supplier using STR analysis
Mycoplasma contamination	All cell lines tested negative for mycoplasma
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used