

# **HHS Public Access**

Author manuscript *Nat Biotechnol.* Author manuscript; available in PMC 2020 August 10.

Published in final edited form as:

Nat Biotechnol. 2020 April; 38(4): 471-481. doi:10.1038/s41587-020-0412-8.

# Continuous evolution of SpCas9 variants compatible with non-G PAMs

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

<sup>1</sup>Merkin Institute of Transformative Technologies in Healthcare, The Broad Institute of Harvard and MIT, Cambridge, MA, USA

<sup>2</sup>Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA

<sup>3</sup>Howard Hughes Medical Institute, Harvard University, Cambridge, MA, USA

<sup>4</sup>Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>5</sup>Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA

<sup>6</sup>These authors contributed equally

# Abstract

The targeting scope of *Streptococcus pyogenes* Cas9 (SpCas9) and its engineered variants is largely restricted to protospacer-adjacent motif (PAM) sequences containing Gs. Here, we report the evolution of three new SpCas9 variants that collectively recognize NRNH PAMs (where R = Aor G and H = A, C, or T) using phage-assisted non-continuous evolution (PANCE), three new phage-assisted continuous evolution (PACE) 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 mediate indel formation and base editing in human cells and enable the A•T-to-G•C base editing of a sickle-cell anemia mutation using a previously inaccessible CACC PAM. These new evolved SpCas9s, together with previously reported variants, in principle enable targeting the majority of

\*Corresponding author.

Author Contributions

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial\_policies/license.html#terms

S.M.M., T.W., D.R.L. designed research. S.M.M., T.W., P.B.R. performed evolution. S.M.M., T.W., P.B.R, T.P.H characterized variants in bacteria. S.M.M., T.W., P.B.R., M.A., Z.M. conducted human cell experiments M.A., M.S. designed and constructed the integrated target site human cell library. M.S. wrote custom software. S.M.M., T.W., M.A., M.S. analyzed library data. S.M.M., 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., T.P.H. designed and performed the HbS editing experiments. S.M.M., T.W., 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, Pariwise Plants, Beam Therapeutics, and Prime Medicine, companies that use genome editing technologies.

NR PAM sequences and substantially reduce the fraction of genomic sites that are inaccessible by Cas9-based methods.

# **Editorial summary**

PAM sequences without Gs can be edited with SpCas9 variants that were continuously evolved in the laboratory.

The CRISPR-Cas9 system has recently transformed the life sciences by enabling a wide range of targeted genome manipulation methods<sup>1,2</sup>. Cas9 is limited, however, by its requirement for a protospacer-adjacent motif (PAM) in order to bind a DNA sequence. *Streptococcus pyogenes* Cas9 (SpCas9), the most widely used and well-characterized Cas9 homolog<sup>1</sup>, recognizes an NGG PAM immediately 3' of the target DNA sequence<sup>3</sup>. To expand the range of targetable genomic loci, researchers have used naturally occurring Cas9 orthologs with different PAM specificities<sup>4</sup>. 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)<sup>5</sup> and SpCas9<sup>6–8</sup> 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<sup>6,7</sup>, 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)<sup>9</sup>, predictable template-free end-joining<sup>10,11</sup>, and base editing<sup>12</sup>. 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 (nt) away from the target base<sup>12</sup>. In addition, for any given base edit, it may be desirable to test multiple protospacers to maximize on-target activity while minimizing unwanted editing<sup>13–16</sup>.

Here we report the directed evolution of three new SpCas9 variants capable of recognizing NRRH, NRTH, and NRCH PAMs, respectively, where R = A or G, and H = A, C, or T. 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<sup>7</sup>, 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 the majority of NR PAMs.

# Results

# Phage-assisted evolution of SpCas9 toward non-G PAM sequences

Phage-assisted continuous evolution (PACE), a method for the rapid directed evolution of biomolecules<sup>17</sup>, has been successfully applied to evolve proteins including

polymerases<sup>17–22</sup>, proteases<sup>23,24</sup>, genome-editing proteins<sup>6,25</sup>, antibody-like proteins<sup>26,27</sup>, insecticidal proteins<sup>26</sup>, aminoacyl-tRNA synthetases<sup>28,29</sup>, and methanol dehydrogenases<sup>30</sup>. We previously evolved SpCas9 variants with broadened PAM compatibility in PACE using a bacterial one-hybrid protein:DNA binding selection<sup>6</sup>, where *E. coli* RNA polymerase (RNAP) subunit fused dSpCas9 ( $\omega$ -dSpCas9) binding to a target protospacer-PAM sequence recruits endogenous RNAP to drive gene III (gIII) transcription from an accessory plasmid (AP) (Fig. 1a). Only SP carrying  $\omega$ -dSpCas9 variants recognizing the target PAM sequence produce infectious progeny phage and replicate during PACE<sup>6</sup>. 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<sup>6</sup>. We hypothesized that the use of a mixture of many PAMs during evolution greatly reduced 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. Since both SpCas9 and xCas9 in phage-based DNA-binding experiments showed low activity on NAH PAM sites (Fig. 1b), we began by evolving either SpCas9 or xCas9 for binding to each of the 12 possible NAH PAM target sequences in parallel using phage-assisted non-continuous evolution (PANCE)<sup>29,30</sup> (Fig. 1c). While slower than PACE, PANCE enables weakly active variants to replicate, and can be performed in parallel<sup>30</sup>. After 19 rounds of PANCE (total net phage dilution of 10<sup>38</sup>-fold) on each of the 12 NAH PAM targeted for evolution (Fig. 1d, 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 (Fig. 1d and 2b). Despite acquiring multiple consensus mutations (D1135N, E1219V, Q1221H, H1264Y, A1320V, 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 stringencies and 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 dual-AP system (Fig. 2a and Supplementary Fig. 1a)<sup>27</sup>. 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; Fig. 1d and 2b) were subjected to dual-AP PACE targeting a CAA PAM, resulting in five additional consensus mutations (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<sup>31</sup>. Since both

the promoter and ribosome-binding site for  $\omega$ -dSpCas9 reside on the SP,  $\omega$ -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<sup>32</sup> (dSpCas9<sub>C</sub>) on the evolving SP, and provided the  $\omega$ -N-terminal portion (residues 1–573) fused to NpuN ( $\omega$ -dSpCas9<sub>N</sub>) in restricted amounts by a complementary plasmid (CP) in the 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<sub>C</sub> 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<sup>33</sup>, from the SP for use as a second phage selection marker in PACE. This allowed us to combine both selection strategies described above by requiring split-dSpCas9 to bind each of two distinct protospacers in order 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, these 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, R1333K) (Fig. 2b, d; 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<sup>8</sup> after PANCE to remove nuclease-inactivating mutations that may have contributed to reduced CBE activity of some variants isolated during NAA PAM evolution (Supplementary Fig. 1e–f; Supplementary Note 1). These experiments resulted in the isolation of TAT.P5–1 (R1114G, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, E1253K, P1321S, D1332G, 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, Supplementary Table 2).

Previous efforts to engineer or evolve SpCas9 to accept alternative PAMs have focused on modifying the PID<sup>7,8</sup>, which mediates PAM specificity<sup>34</sup>. Our experiments, however, mutagenized either the entire sequence of SpCas9 or residues 574–1368 (SpCas9<sub>C</sub>), resulting in the accumulation of three to 15 mutations outside the PID (Supplementary Tables 1 and 2). Although mutations within the nuclease domains could impair SpCas9 DNA cleavage activity<sup>35</sup>, mutations in the helical domain may benefit SpCas9 DNA binding/unwinding. We transplanted the evolved PIDs onto a fixed N-terminal sequence that included the  $\omega$ -dSpCas9<sub>N</sub> mutations (A10T, I322V, S409I, E427G) and R654L and R753G, which enriched in multiple independent PACE experiments (Supplementary Fig. 2b and

Supplementary Table 2). The addition of all six NTD 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 Note 1).

# 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 and a mammalian cell genomically integrated target site-sgRNA library (Fig. 3, Supplementary Fig. 3–4, 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 is adjacent to a PAM sequence of all possible NNNN tetranucleotides and is additionally flanked with primer binding sites for high-throughput sequencing (HTS) analysis (Fig. 3a). We genomically integrated this library into HEK293T cells, then treated the library cells with each of the above five SpCas9 variants as optimized CBEs (in the BE4max architecture<sup>36</sup>, hereafter referred to as CBE) (Fig. 3a).

HTS of the target library sequences after treatment revealed that, compared with 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, Supplementary Fig. 4a). CBE-NRRH also preferred a G at the third position (Fig. 3b–c, Supplementary Fig. 4a, d). Interestingly, C at the third position is disfavored for all variants except CBE-NRCH, supporting that CBE-NRCH has distinctly evolved preference for a C at this position (Fig. 3b–c). CBE-NG was also active on NA PAMs; however, its editing was lower than that of our evolved variants (Fig. 3b, Supplementary Fig. 4d). Finally, we observed activity from SpCas9-CBE on its known secondary NAG PAM (Fig. 3b, Supplementary Fig. 4a). CBE-NRCH also performs well across all NGH PAMs, editing with efficiencies comparable to or greater than those of CBE-NG at these sites in the library setting (Fig. 3b). 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 contains a G and, to a lesser extent, 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 one of the PAM and worst when a T was at this position (Fig. 3c, Supplementary Fig. 4d). 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. Generally, each of our variants displayed robust indel formation activity on sites containing a PAM it was evolved to recognize. SpCas9-NRRH generated an average of 27±5.0% and 27±4.4% indels at sites containing NAAH and NAGH PAMs, respectively, while 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 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  $13\pm1.7\%$  average editing on NAAG PAM sites, 20±2.5% on NAGG PAM sites, 23±3.6% on NATG 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  $41\pm6.2\%$ ,  $60\pm6.5\%$ ,  $62\pm3.2\%$ , and  $42\pm3.0\%$ , on NGA, NGG, NGT, and NGC PAM sites, respectively). SpCas9-NRTH, however, showed higher activity than SpCas9-NG at NGT PAMs, averaging  $68\pm4.0\%$  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 ( $67\pm2.0\%$ ) and NGCH ( $43\pm1.4\%$ ) PAMs, respectively, when compared to SpCas9-NG (NGTH:  $61\pm1.7\%$ , NGCH:  $32\pm1.5\%$ ). 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-NRRH, and SpCas9-NRTH resulted in  $47\pm8.0\%$ ,  $33\pm5.7\%$ , and  $1.3\pm0.3\%$  average indel formation at three sites containing CACC, CATT, and AAAA PAMs, respectively (Fig. 4b), while SpCas9-NG averaged  $1.5\pm0.5\%$ ,  $18\pm5.0\%$ , and  $3.6\pm1.3\%$  average indel formation at the same sites (Fig. 4b). SpCas9-NRRH and SpCas9-NRCH also induced  $26\pm8.2\%$  and  $77\pm7.9\%$  indel formation at sites containing AGAT and AGCC PAMs, respectively, whereas SpCas9-NG averaged  $22\pm7.5\%$  and  $10\pm3.1\%$  indel formation at these two sites (Fig. 4b). These results suggest that our evolved variants are able to edit genomic sites in primary human cells following mRNA delivery, although editing efficiencies remain subject to variability from loci-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-targets edits<sup>5,7</sup>, we characterized the DNA specificity of our variants by performing GUIDE-seq<sup>37</sup> on SpCas9, SpCas9-NRRH, SpCas9-NRCH, and SpCas9-NRTH in U2OS cells on four characterized on-target genomic sites<sup>6,7,8,37</sup>. For comparison, we also analyzed xCas9, which has greatly reduced off-target activity compared to SpCas9<sup>6,38–40</sup>. 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, though xCas9 was the most highly specific variant tested (Fig. 4c-d and Supplementary Fig. 6). Among 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 (Supplementary Fig. 6d). 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 compared to that of SpCas9 on NGG PAM sites. Similar to high-fidelity SpCas9 variants<sup>6,40–43</sup>, SpCas9-NRRH, SpCas9-NRTH, and SpCas9-NRCH exhibit greater sensitivity to mismatches in the 5' sgRNA protospacer commonly tolerated by wild-type SpCas9 (Supplementary Fig. 4b-c and Supplementary Note  $3)^{44}$ .

# Evolved SpCas9s 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 and Supplementary Fig. 7a). As with their nuclease forms, each of the three evolved CBE variants showed the highest average activity on sites containing the PAM it was evolved to recognize. Thus CBE-NRRH and CBE-NRTH showed the highest activity on cytosines within the canonical base editing window at NAAH and NATH PAM sites, averaging 14±2.4%, 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  $13.0\pm2.0\%$  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 due 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 function 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<sup>36</sup> 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 are 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 SpCas9s enable base editing of a previously inaccessible pathogenic SNP

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)<sup>12</sup>. SpCas9-NRRH, SpCas9-NRCH, and SpCas9-NRTH, together with previously reported NG PAM compatible SpCas9s<sup>6,7</sup>, expand the targeting scope of SpCas9 to the majority of genomic sites containing NR PAMs and greatly increase the fraction of known human pathogenic SNPs that can theoretically be corrected by base editing. Among all pathogenic SNPs in the ClinVar database<sup>45</sup> that can be corrected by C•Gto-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 in principle may 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). Since 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<sup>12</sup>, 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 SNPselective 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  $\beta$ -globin (*HBB*) that is the most common cause of sickle-cell anemia<sup>46</sup>. *HbS* arises from a GAG (Glu) to GTG (Val) mutation at HBB amino acid 6 that cannot be reverted through current base editing technologies. However, this SNP can be converted with ABE to a GCG (Ala) 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<sup>47–49</sup>. Unfortunately, the only NGG or NGN PAMs available at this site place the target A at either protospacer position 2 or 9, respectively, outside the editing window for ABE<sup>12</sup>. However, two alternative protospacers using a CATG or CACC PAM place the target A at either position 4 or 7, respectively, with a bystander A present at either position 6 or 9 leading to a silent CCT to CCC (Pro to Pro) 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). While 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 sites, ABE-NG edited efficiently (40±0.2%) only with the protospacer containing a CATG PAM (Fig. 6b–c), perhaps due 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, 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  $41\pm3.8\%$  base conversion at the target A (position 7), and  $13\pm3.2\%$  at the silent bystander A (position 9) (Fig. 6c, Supplementary Fig. 8c). ABE-NRRH and ABE-NRTH achieved  $29\pm4.3\%$  and  $14\pm2.8\%$  conversion, respectively (Fig. 6c, Supplementary Fig. 8c). In comparison, ABE-NG showed negligible ( $1.0\pm0.5\%$ ) target base conversion activity at this site (Fig. 6c, 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 desired 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 gene VI, implementing a dual-AP split-gene III strategy, and limiting the total concentration of functional SpCas9 in the host cell through a 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 to potential modes of PAM interaction. SpCas9-NRRH acquired a mutation at R1333, which in SpCas9 contacts the first guanine in its canonical PAM, but not R1335, which contacts the second guanine in NGG (Fig. 2b, d). The R1333K mutation likely allows SpCas9-NRRH to accept both A and G at the second PAM nucleotide, while the preservation of R1335 may explain why this variant recognizes both NAA and NAG. On the other hand, SpCas9-NRTH preserves R1333 but replaces R1335 with Leu (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, Supplementary Fig. 3, 4a–b). Finally, SpCas9-NRCH evolved altered interactions at both R1335 and T1337 (Fig. 2f, h); T1337N in particular may form contacts with a fourth PAM nucleotide to compensate for weakened binding interactions with the NAC target PAM.

We also observed a number of 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<sup>38</sup>. Mutations at D1135 have been previously reported<sup>7,8</sup> and may 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, Supplementary Tables 1 and 2). These mutations, when added to the N-terminal region of NRRH and NRCH, improve their recognition of non-G PAMs in base editing experiments (Supplementary Fig. 2c) and may 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 specificities 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 display a different and complementary fourthposition PAM preference (H) compared to SpCas9-NG (G). While further investigation is required to explain the fourth base preferences of these mutants, crystal structures of SpCas9-NG and other evolved SpCas9s (VRER/VRQR) suggest that the T1337R mutation in these variants may 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 other CRISPR proteins that access smaller subsets of NR PAM space are not included in this table but have also been shown to mediate genome editing in human cells52-60.

The variants evolved in this study, together with SpCas9-NG or 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  $\beta$ -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 may 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<sup>61,62</sup>, HDR<sup>9</sup>, and predictable template-free genome editing<sup>10,11</sup>.

# Methods

# **General methods**

Antibiotics (Gold Biotechnology) were used at the following working concentrations: carbenicillin 50 µg/mL, spectinomycin 50 µg/mL, chloramphenicol 25 µg/mL, kanamycin 50

µg/mL, tetracycline 10 µg/mL, streptomycin 50 µg/mL. 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) were used for all PCRs. Unless otherwise noted, plasmids and SPs were cloned by USER assembly<sup>26</sup>. 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) prior to Sanger sequencing. Strain S2060<sup>25</sup> was used in all luciferase, phage propagation, and plaque assays, and in all PACE experiments. A description of plasmids, SP, primers, and protospacer sequences used in this work is provided in Table S4.

# Phage propagation assay

Chemicompetent S2060 cells were transformed with the AP(s) of interest as previously described<sup>27</sup>. Overnight cultures of single colonies grown in 2xYT media supplemented with maintenance antibiotics were diluted 1000-fold into DRM media with maintenance antibiotics and grown at 37 °C with shaking at 230 RPM to  $OD_{600} \sim 0.4-0.6$ . Cells were then infected with SP at an initial titer of  $5 \times 10^4$  pfu/mL. Cells were incubated for another 16–18 h at 37 °C with shaking at 230 RPM, then centrifuged at 8000 *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 2xYT media supplemented with maintenance antibiotics were diluted 1000-fold into fresh 2xYT media with maintenance antibiotics and grown at 37 °C with shaking at 230 RPM to  $OD_{600} \sim 0.6-0.8$  before use. SP were serially diluted 100-fold (4 dilutions total) in H<sub>2</sub>O. 150 µL of cells was added to 10 µL of each phage dilution and to this 1 mL of liquid (55°C) top agar (2xYT media + 0.5% agar) supplemented with 2% Bluo-gal (Gold Biotechnology) was added and mixed by pipetting up and down once. This mixture was then immediately pipetted onto one quadrant of a quartered Petri dish already containing 2 mL of solidified bottom agar (2xYT media + 1.5% agar, no antibiotics). Plates were incubated at 37 °C for 16–18 h.

#### Phage-assisted noncontinuous evolution

Phage-assisted noncontinous evolution (PANCE) was performed as previously reported with the following modifications<sup>29,30</sup>. SP using the SP41 backbone encoding either  $\omega$ -dSpCas9 or  $\omega$ -dxCas9 were evolved on host S2060 cells transformed with pSM072a APs containing the G7' protospacer sequence followed by each NAN PAM and DP6<sup>63</sup> (to give 32 total evolving populations). To reduce the likelihood of contamination with gIII-encoding recombined SP, phage stocks were purified as previously described<sup>26</sup>. 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<sup>6</sup> pfu/mL. The first 10 passages were run as follows: 6 h propagation with 10 mM arabinose and 50 ng/mL anhydrotetracycline (aTc) followed by 18 h propagation with 10 mM arabinose only (no

drift induced). The next 4 passages were run as follows: 6 h propagation with 10 mM arabinose and 25 ng/mL anhydrotetracycline (aTc) followed by 18 h propagation with 10 mM arabinose only (no drift induced). The final 5 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 (see below).

#### qPCR estimation of phage titer

Phage titers with estimated using a modified version of a previously reported qPCR-based procedure<sup>64</sup>. SP pools from PANCE passages (25–50  $\mu$ L) were first heated at 80 °C for 30 min to destroy polyphage. Polyphage genomes were then degraded by adding 5  $\mu$ L of heated SP to 45  $\mu$ L of 1x DNase I buffer containing 0.5  $\mu$ L DNase I (New England Biolab) and incubated at 37°C for 20 min followed by 95°C for 20 min. Next, qPCR reaction mixtures were prepared by mixing 20  $\mu$ L H<sub>2</sub>O, 5  $\mu$ L 5X Phusion HF Buffer (dye-free; Thermo), 0.5  $\mu$ L dNTPs (10 mM; New England Biolabs), 0.25  $\mu$ L Sybr green (100X; Thermo), 0.25  $\mu$ L each primer (5'-CACCGTTCATCTGTCCTCTTT and 5'-

CGACCTGCTCCATGTTACTTAG), and 1.5  $\mu$ L 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.

#### Phage-assisted continuous evolution

Unless otherwise noted, PACE apparatus, including host cell strains, lagoons, chemostats, and media, were all used as previously described<sup>17,26</sup>. To reduce the likelihood of contamination with gIII-encoding recombined SP, phage stocks were purified as previously described<sup>26</sup>. 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 S2060s were transformed with AP(s), CP (if using), and MP6 or DP6<sup>63</sup>, plated on 2xYT media + 1.5% agar supplemented with 25 mM glucose (to prevent induction of mutagenesis) in addition to 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 5-fold eight times serially into DRM. The plate was sealed with a porous sealing film and grown at 37 °C with shaking at 230 RPM for 16–18 h. Dilutions with OD<sub>600</sub> ~ 0.4–0.8 were then used to inoculate a chemostat containing 80 mL DRM. The chemostat was grown to OD<sub>600</sub> ~ 0.8–1.0, then continuously diluted with fresh DRM at a rate of ~1.5 chemostat volumes/h. The chemostat was maintained at a volume of 60–80 mL.

Prior to SP infection, lagoons were continuously diluted with culture from the chemostat at 1 lagoon volume/h 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  $10^6$  pfu/mL 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 8000 *g* for 2 min, and the supernatant 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) and AB1396 (5'-

ACAGAGAGAATAACATAAAAACAGGGAAGC), 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/h with 10 mM arabinose for 4 h prior to 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/h at 19 h and 2.0 volumes/h at 44 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/h with 10 mM arabinose for 4 h prior to 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/h. The lagoon dilution rate was increased to 1.5 volume/h at 96 h, 2.0 volumes/h at 136 h, and 3.0 at 168h. 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/h with 10 mM arabinose for 4 h prior to infection with the pool from P2. Upon infection, lagoon dilution rates were decreased to 0.5 volume/h. The lagoon dilution rate was increased to 1.0 volume/h at 19 h, 2.0 volumes/h at 43 h, and 2.5 volumes/h at 68h. 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/h with 10 mM arabinose for 4 h prior to 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 volume/h. The lagoon dilution rate for lagoon 1 was increased to 1.0 volume/h at 19 h, 1.5 volumes/h at 43 h, and 2.0 volumes/h at 68 h before washing out. The lagoon dilution rate for lagoon 2 was increased to 0.75 volume/h at 19 h, 1.0 volumes/h at 43 h, 1.5 volumes/h at 68 h, 2.0 volumes/h at 92 h, and 3.0 volumes/h 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/h with 10 mM arabinose for 4 h prior to 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 volume/h. The lagoon dilution rate for lagoons 1 and 2 was increased to 1.0 volume/h at 43 h, 1.5 volumes/h at 68 h, and 3.0 volumes/h at 138 h. The lagoon dilution rate for lagoons 3 and 4 was increased to

1.0 volume/h at 19 h, 1.5 volumes/h at 43 h, 2.0 volumes/h at 68 h, and 3.0 volumes/h 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/h with 10 mM arabinose for 4 h prior to 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 volume/h. The lagoon dilution rate was increased to 1.0 volume/h at 19 h, 1.5 volumes/h at 40 h, 2.0 volumes/h at 68 h, and 3.0 volumes/h 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/h with 10 mM arabinose for 4 h prior to 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/h. The lagoon dilution was increased to 1.0 volume/h at 19 h, 1.5 volumes/h 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/h with 10 mM arabinose for 4 h prior to infection with the surviving phage pool from P7a. Upon infection, lagoon dilution rates were decreased to 0.5 volume/h. The lagoon dilution rate for lagoons 1 and 3 was increased to 1.0 volume/h at 43 h, and 1.5 volumes/h at 68 h. The lagoon dilution rate for lagoons 2 and 4 was increased to 1.0 volume/h at 43 h, and 2.0 volumes/h 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/h with 10 mM arabinose for 4 h prior to infection with the surviving phage pool from P7b. Upon infection, lagoon dilution rates were decreased to 0.5 volume/h. The lagoon dilution rate for lagoons 1 and 3 was increased to 1.0 volume/h at 19 h, 2.0 volumes/h at 43 h, and 3.0 volumes/h at 116h. The lagoon dilution rate for lagoons 2 and 4 was increased to 1.0 volume/h at 19 h, 1.5 volumes/h at 43 h, 2.0 volumes/h at 116 h, and 3.0 volumes/h at 140h. 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/h with 10 mM arabinose for 4 h prior to 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 volume/h. The lagoon dilution rate was increased to 1.0 volume/h at 19 h, 1.5 volumes/h at 40 h, 2.0 volumes/h at 68 h, and 3.0 volumes/h 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 (NExT) method<sup>65,66</sup>. SP libraries encoding evolved  $\omega$ –SpCas9<sub>C</sub> were amplified by PCR using primers 5'- GCCAGTTACAAAATAAACAGCC and 5'-CACCGTTCATCTGTCCTCTTT to provide template for the initial uracil incorporation PCR. Clonal SP encoding wild-type  $\omega$ –SpCas9<sub>C</sub> was similarly amplified to provide a wild-type template. Separate uracil incorporation PCRs were performed using primers 5'-GACTCCCTGCAAGCCTCAG and 5'-ACAGAGAGAATAACATAAAACAGGGAAGC on both wild-type and evolved library templates, and the resulting amplicons were combined in a 1:4 ratio of wild-type:library before digesting with USER enzyme as previously described<sup>65</sup>. Digested fragments were purified using the QiaxII Gel Extraction kit (Qiagen) according to manufacturer 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, 50 cycles of: [92 °C for 30 s, 22–72 °C +1 °C/cycle for 60 s, and 72 °C for 60 s +5 s per cycle], and finally 72 °C for 10 min. The reassembled SpCas9<sub>C</sub> was then amplified with primers 5'-

ATGTTGAAAATCTCCTTCUAGATCAGTCTCCACCGAGCTGAG and 5'-ATCGCCAGCAAUTGTTTCGACTCTGTTGAAATCAG and cloned through USER assembly into either SP or plasmids.

# PAM depletion assay

DH10 $\beta$  cells (New England Biolabs) were transformed with plasmids (pTW222) encoding nuclease-active SpCas9 variants under control of the arabinose promoter and an sgRNA targeting the HEK3 protospacer sequence. Cells were then made electrocompetent<sup>67</sup> and electroporated with a PAM depletion plasmid library (pTPH308), which contains 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 media (New England Biolabs) for 1 h before 10 mM arabinose was added to induce SpCas9 expression for 1 or 3 h or overnight. Cells were then plated on 2xYT media + 1.5% agar supplemented with carbenicillin and kanamycin (both 50 µg/mL). The plates were incubated at 37 °C for 16–18 h, and the resultant colonies scraped and plasmids extracted using the Qiagen Plasmid Plus Midiprep Kit (Qiagen) following manufacturer protocol. The protospacer-PAM sequence of the extracted plasmids as well as the input PAM depletion plasmid library were amplified with primers 5'-

ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAATACGCAAACCGCCTCT C and 5'- TGGAGTTCAGACGTGTGCTCTTCCGATCTCTTGTCTGTAAGCGGATGC. The resulting amplicons were then amplified using Illumina barcoding primers and the product purified by gel extraction using a QIAquick Gel Extraction Kit (Qiagen) according to manufacturer protocol. DNA concentration was quantified with the KAPA Library Quantification Kit-Illumina (KAPA Biosystems) and the resultant libraries sequenced using an Illumina MiSeq high-throughput sequencer. Sequencing data was analyzed using a previously reported Python script<sup>8</sup>. 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 postselection 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 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  $\mu$ L 2X T7 ligase buffer, 100,000 U T7 ligase, 1  $\mu$ L USER enzyme mixture, and 1  $\mu$ L DpnI was mixed with 0.5 pmol of each DNA fragment, and nuclease-free water was added to a final volume of 20  $\mu$ L total. The reaction was incubated at 37°C for 1hr, heated to 80 °C, and slowly (1 °C /sec) cooled to 12 °C, purified using the QIAquick PCR purification kit (Qiagen) according to manufacturer protocol, and transformed into electrocompetent DH10 $\beta$  cells (New England Biolabs). Cells were immediately plated following electroporation on 2xYT media + 1.5% agar supplemented with carbenicillin (50  $\mu$ g/mL) and incubated at 37 °C for 16–18 h. The resultant colonies were then scraped and plasmids extracted using the Qiagen Plasmid Plus Midiprep Kit (Qiagen) following manufacturer protocol.

# SacB plasmid cleavage selection

DH10 $\beta$  cells (New England Biolabs) transformed with pTW218, which encodes *B. subtillis* 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 media (New England Biolabs) for 1 h in SOC before addition of carbenicillin (50 µg/mL) and arabinose (10 mM). Cells were incubated for 1–18 h before being plated on 2xYT media + 1.5% agar supplemented with carbenicillin (50 µg/mL) and sucrose (100 mM). Plates were incubated at 37 °C for 16–18 h, and the SpCas9 variants encoded by surviving colonies characterized by Sanger sequencing.

# HEK 293T transfection and genomic DNA extraction

HEK 293T cells (ATCC CRL-3216) maintained in Dulbecco's Modified Eagle's Medium plus GlutaMax (Thermo Fisher Scientific) supplemented with 10% (v/v) FBS at 37 °C with 5% CO<sub>2</sub> were seeded on 96-well plates (Corning). 16–20 h after seeding, cells were transfected at approximately 80–85% confluency as with 200 ng of SpCas9 or Base Editor plasmid, 40 ng of sgRNA plasmid, and 0.5  $\mu$ L of Lipofectamine 2000 (Thermo Fischer Scientific). Cells were cultured for 3 days post-transfection before the media was removed, cells were washed with 1 x PBS solution (Thermo Fisher Scientific), and genomic DNA was extracted by addition of 30  $\mu$ L lysis buffer (10mM Tris-HCl, pH 7.0, 0.05% SDS, 25  $\mu$ g/mL Proteinase K (Thermo Fischer Scientific). Genomic DNA was stored at –20 °C until further use.

# Cas9 mRNA in vitro transcription

A fragment containing a T7 promoter driving poly-adenylated Cas9 expression was isolated from purified plasmid using SpeI restriction digestion and purified with a MinElute PCR purification Kit (Qiagen). mRNA was transcribed using HiScribe T7 ARCA mRNA Kit (NEB) and purified using LiCl precipitation according to manufacturer's instructions.

# Human primary fibroblast nucleofection and genomic DNA extraction

 $2.5 \times 10^5$  human primary fibroblasts (Coriell GM04541) cultured in Dulbecco's Modified Eagle's Medium with GlutaMax (Thermo Fisher Scientific) supplemented with 20% (v/v) FBS at 37 °C with 5% CO<sub>2</sub> were transfected with 50 pmol of sgRNA (Synthego) and 1 µg of *in vitro* transcribed SpCas9 mRNA. 20 µL 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. Media was changed after 24 hours post-transfection. Cells were cultured for 5 days posttransfection before media was removed, cells were washed with 1 x PBS solution (Thermo Fisher Scientific), and genomic DNA was extracted by addition of 200 µL lysis buffer (10mM Tris-HCl, pH 7.0, 0.05% SDS, 25 µg/mL Proteinase K (Thermo Fischer Scientific). Genomic DNA was stored at -20 °C until further use.

#### High-throughput sequencing of genomic DNA

Genomic sites were amplified with primers targeting the region of interest and the appropriate universal Illumina forward and reverse adapters. Primer pairs for the first round of PCR (PCR 1) were either reported previously<sup>68</sup> or are listed in Table S4. 25-µL 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: 98 °C for 2 min, then 30 cycles of [98 °C for 15 s, 61 °C for 20 s, 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  $\mu$ L of unpurified PCR 1 reaction product in 25  $\mu$ L of Phusion HS II PCR mix prepared according to the manufacturer's protocol (Thermo Fisher Scientific). PCR 2 conditions: 98 °C for 2 min, then 12 cycles of [98 °C for 15 s, 61 °C for 20 s, 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 a QIAquick Gel Extraction Kit (Qiagen Inc.) eluting with 15 µL H<sub>2</sub>O. DNA concentration was quantified with the KAPA Library Quantification Kit-Illumina (KAPA Biosystems) and sequenced on an Illumina MiSeq instrument (paired-end read - 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<sup>69</sup>.

## GUIDE-seq

 $3 \times 10^5$  U2OS cells maintained in McCoy's 5A Medium (Thermo Fisher Scientific) supplemented with 10% (v/v) FBS at 37 °C with 5% CO<sub>2</sub> were transfected with 750 ng of Cas9 plasmid, 250 ng of sgRNA plasmid, and 5 pmol of the GUIDE-seq dsODN<sup>37</sup>. A control sample with GFP plasmid (Lonza) instead of Cas9 plasmid was also included. 20 µL 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 of 500 bp using a Covaris S200 focused ultrasonicator using the manufacturer's protocol. End repair, dA-tailing, adaptor ligation, tag-specific CPR1 and PCR2 were carried out using the primers and methods described previously<sup>37</sup>. DNA concentration was quantified with the KAPA Library Quantification Kit-Illumina (KAPA Biosystems), normalized to 4 nM, and 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 nt in length was synthesized by Twist Bioscience, pairing an SpCas9 gRNA protospacer with its 61-bp target site. We include 46 protospacers, each preceding every possible 4-nt combination to analyze any 4N PAM. Libraries were then cloned into a backbone (Addgene 71485) containing a U6 promoter to facilitate gRNA expression, a hygromycin resistance cassette, and flanking Tol2 transposon sites to facilitate integration into the genome<sup>70</sup>. Cloning was performed as previously reported<sup>10</sup> 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′-TTTTGTTTTCTGTGTTCCGTTGTCCGTGCTGTAACGAAAGGATGGGTGCGACGC GTCAT) and "oligonucleotide library reverse" (5′-

GTTGATAACGGACTAGCCTTATTTAAACTTGCTATGCTGTTTCCAGCATAGCTCTTA AAC). The donor template for circular assembly was amplified from an SpCas9 sgRNA expression plasmid (Addgene 71485)<sup>70</sup> using primers "circular donor forward" (5'-GTTTAAGAGCTATGCTGGAAACAGC) and "circular donor reverse" (5'-ACTGCACCGCGTCGCACCCATCCTTTCGTTACAGCACGGACAACGGAACACAGAC AAAACAAAAAAGCACCGACTC) 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'-

TAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAAC ACCG) and "plasmid insert reverse" (5'-

TTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGCTCGAAGCGGCCGT ACCTCTAGAACACTCTTTCCCTACACGAC) for the addition of overhangs complementary to the 5' and 3' regions of the gRNA expression plasmid (Addgene 71485)<sup>70</sup> 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) and reconstituted in water and transformed into DH10 $\beta$  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 (Addgene 112093)<sup>36</sup> 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 (Takarabio) through insertion of a PCR amplified BE4 variant after digestion of the backbone with NotI and AgeI.

**Cell culture**—HEK293T cells were purchased from ATCC and cultured as recommended by 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 1:2 amounts of Tol2 transposase plasmid<sup>71</sup> and transposon-containing plasmid, for a total of 80 µg of DNA per 15 cm plate. For library integration, two 15-cm plates with  $> 2 \times 10^7$  initial cells were used and cells were maintained on 17.5 µg/mL hygromycin starting one day after transfection, for a minimum of two weeks. Library cells were validated by HTS as described below. For library editing experiments, two 15-cm plates with  $> 2 \times 10^7$  initial cells were used and cells were maintained on 3 µg/mL blasticidin starting one day after transfection.

**High-throughput DNA sequencing**—Genomic DNA was collected from cells after 4 days of selection. Library samples were prepped for high throughput sequencing in two steps as previously reported<sup>10</sup> 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 NNNNNNN ACACTCTTTCCCTACACGAC) and "sequencing reverse 1" (5'-

GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCT

GTGGAAAGGACGAAACACCG) for 12 cycles using a 2 min initial denaturing and extension time. Products were purified and cycle number for PCR2 was determined by qPCR to avoid over amplification of the product. PCR2 was performed with "sequencing forward 2" (5'- AATGATACGGCGACCACCGAGATCTACAC) and an Illumina TruSeq P7 index primer (5'- CAAGCAGAAGACGGCATACGAGAT NNNNNNN GTGACTGGAGTTCAGACGTGTGCTCTTC). Purified products were validated and quantified by Tape Station (Agilent) and pooled for high throughput sequencing. Pools were quantified by KAPA Illumina library quantification kit (KAPA), 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<sup>10</sup>. 61 nt target contexts that were intentionally designed to be highly similar to each other (varying only in the 4N PAM region) were designed with a 5 nt barcode to assist accurate assignment. Sequence alignment

was performed using Smith-Waterman with the parameters: match +1, mismatch -1, indel start -5, indel extend 0. Nucleotides with 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 constitutes an observation of a single event. A given read with multiple substrate nucleotides is considered to have an edit if at least one nucleotide is edited. Events in treatment conditions were adjusted by control conditions as follows: any editing event with frequency above 5% in controls were filtered in treatment as treatments undergo an additional selection step beyond control conditions. Treatment events with frequencies that could be explained by control events were filtered at 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 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 false discovery rate of 0.05. C-to-T events occurring between protospacer positions -9 to 20 were retained, where position 1 is the 5'-most nucleotide of the protospacer, and 0 is 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 Dulbecco's Modified Eagle's Medium plus GlutaMax (Thermo Fisher Scientific) supplemented with 10% (v/v) FBS at 37 °C with 5% CO<sub>2</sub> 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 (protospacer:

GTAACGGCAGACTTCTCCTC), and 200 ng of ssODN

(ACTTCATCCACGTTCACCTTGCCCCACAGGGCAGTAACGGCAGACTTCTCCACAG GAGTCAGATGCACCATGGTGTCTGTTTGAGGTTGCTAGTGAACAC). The DNAcontaining solution was complexed with 1.5  $\mu$ L Lipofectamine 2000 to a total volume of 25  $\mu$ L in Opti-MEM (Thermo Fisher Scientific) before addition to cells. Media was changed three days post-transfection. Four days post-transfection, cells were trypsonized using TrypLE solution (Thermo Fisher Scientific) and suspended in 1.5mL of media. A Beckman-Coulter Astrios FACS instrument was used to isolate single cells into each well of 2 96-well plates (100  $\mu$ L of media per well). After 7 days of growth, cells were dispersed in 10  $\mu$ L of TrypLE solution, pipetted gently to disperse clumps, and left to reattach to the same plate in 100  $\mu$ L 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 deepsequence the HbB locus using primers 5'-

ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGGGTTGGCCAATCTACTC CC and 5'-

TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTTCTGTCTCCACATGCC. At this point, 4 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

HEK 293T 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 protospacer sequences HbS-CATG and HbS-CACC (Table S4) complexed with 1  $\mu$ L of Lipofectamine 3000. Three days post-transfection, media was aspirated and exchanged for fresh media. 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.

#### Data availability

High-throughput sequencing data have been deposited in the NCBI Sequence Read Archive database (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.

# **Code availability**

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

# Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

# Acknowledgments

We thank T. R. Blum and A. Raguram for helpful discussions. This work was supported by US NIH R01 EB027793, U01 AI142756, RM1 HG009490, R35 GM118062, HHMI, the Bill and Melinda Gates Foundation, and the St. Jude Collaborative Research Consortium. S.M.M. and M.S. were supported by an NSF 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.

# REFERENCES

- Komor AC, Badran AH & Liu DR CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell 168, 20–36 (2017). [PubMed: 27866654]
- Pickar-Oliver A & Gersbach CA The next generation of CRISPR–Cas technologies and applications. Nat. Rev. Mol. Cell Biol 20, 490–507 (2019). [PubMed: 31147612]
- Jinek M et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (80-.). 337, 816–821 (2012).
- 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). [PubMed: 28634692]

- 5. Kleinstiver BP et al. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat. Biotechnol 33, 1293–1298 (2015). [PubMed: 26524662]
- 6. Hu JH et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature 556, 57–63 (2018). [PubMed: 29512652]
- Nishimasu H et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science (80-.). 361, 1259–1262 (2018).
- Kleinstiver BP et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature 523, 481–485 (2015). [PubMed: 26098369]
- 9. Paquet D et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature 533, 125–129 (2016). [PubMed: 27120160]
- Shen MW et al. Predictable and precise template-free CRISPR editing of pathogenic variants. Nature 563, 646–651 (2018). [PubMed: 30405244]
- Iyer S et al. Precise therapeutic gene correction by a simple nuclease-induced double-stranded break. Nature 568, 561–565 (2019). [PubMed: 30944467]
- Rees HA & Liu DR Base editing: precision chemistry on the genome and transcriptome of living cells. Nat. Rev. Genet 19, 770–778 (2018). [PubMed: 30323312]
- Zuo E et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. Science (80-. ). 364, 289–292 (2019).
- Jin S et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. Science (80-. ). 364, 292–295 (2019).
- Xin H, Wan T & Ping Y Off-Targeting of Base Editors: BE3 but not ABE induces substantial offtarget single nucleotide variants. Signal Transduct. Target. Ther 4, 9 (2019). [PubMed: 30993016]
- Lee HK et al. Targeting fidelity of adenine and cytosine base editors in mouse embryos. Nat. Commun 9, 4804 (2018). [PubMed: 30442934]
- Esvelt KM, Carlson JC & Liu DR A system for the continuous directed evolution of biomolecules. Nature 472, 499–503 (2011). [PubMed: 21478873]
- Dickinson BC, Leconte AM, Allen B, Esvelt KM & Liu DR Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. Proc. Natl. Acad. Sci 110, 9007–9012 (2013). [PubMed: 23674678]
- Carlson JC, Badran AH, Guggiana-nilo DA & Liu DR Negative selection and stringency modulation in phage-assisted continuous evolution. Nat. Chem. Biol 10, 216–222 (2014). [PubMed: 24487694]
- Pu J, Disare M & Dickinson BC Evolution of C-Terminal Modification Tolerance in Full-Length and Split T7 RNA Polymerase Biosensors. ChemBioChem 20, 1547–1553 (2019). [PubMed: 30694596]
- Pu J, Kentala K & Dickinson BC Multidimensional Control of Cas9 by Evolved RNA Polymerase-Based Biosensors. ACS Chem. Biol 13, 431–437 (2018). [PubMed: 28809467]
- 22. Pu J, Zinkus-Boltz J & Dickinson BC Evolution of a split RNA polymerase as a versatile biosensor platform. Nat. Chem. Biol 13, 432–438 (2017). [PubMed: 28192413]
- 23. Packer MS, Rees HA & Liu DR Phage-assisted continuous evolution of proteases with altered substrate specificity. Nat. Commun 8, 956 (2017). [PubMed: 29038472]
- Dickinson BC, Packer MS, Badran AH & Liu DR A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. Nat. Commun 5, 5352 (2014). [PubMed: 25355134]
- Hubbard BP et al. Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. Nat. Methods 12, 939–942 (2015). [PubMed: 26258293]
- Badran AH et al. Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. Nature 533, 58–63 (2016). [PubMed: 27120167]
- 27. Wang T, Badran AH, Huang TP & Liu DR Continuous directed evolution of proteins with improved soluble expression. Nat. Chem. Biol 14, 972–980 (2018). [PubMed: 30127387]
- Bryson DI et al. Continuous directed evolution of aminoacyl-tRNA synthetases. Nat. Chem. Biol 13, 1253–1260 (2017). [PubMed: 29035361]

- 29. Suzuki T et al. Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. Nat. Chem. Biol 13, 1261–1266 (2017). [PubMed: 29035363]
- Roth TB, Woolston BM, Stephanopoulos G & Liu DR Phage-Assisted Evolution of Bacillus methanolicus Methanol Dehydrogenase 2. ACS Synth. Biol 8, 796–806 (2019). [PubMed: 30856338]
- 31. Karvelis T et al. Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. Genome Biol. 16, 253 (2015). [PubMed: 26585795]
- Shah NH, Dann GP, Vila-Perelló M, Liu Z & Muir TW Ultrafast protein splicing is common among cyanobacterial split inteins: Implications for protein engineering. J. Am. Chem. Soc 134, 11338–11341 (2012). [PubMed: 22734434]
- Brödel AK, Jaramillo A & Isalan M Engineering orthogonal dual transcription factors for multiinput synthetic promoters. Nat. Commun 7, 13858 (2016). [PubMed: 27982027]
- 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). [PubMed: 25079318]
- Jiang F & Doudna JA CRISPR–Cas9 Structures and Mechanisms. Annu. Rev. Biophys 46, 505– 529 (2017). [PubMed: 28375731]
- 36. Koblan LW et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nat. Biotechnol 36, 843–846 (2018). [PubMed: 29813047]
- Tsai SQ et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol 33, 187–197 (2015). [PubMed: 25513782]
- Guo M et al. Structural insights into a high fidelity variant of SpCas9. Cell Res. 29, 183–192 (2019). [PubMed: 30664728]
- 39. Kim S, Bae T, Hwang J & Kim JS Rescue of high-specificity Cas9 variants using sgRNAs with matched 5' nucleotides. Genome Biol. 18, 218 (2017). [PubMed: 29141659]
- Lee JK et al. Directed evolution of CRISPR-Cas9 to increase its specificity. Nat. Commun 9, 3048 (2018). [PubMed: 30082838]
- Slaymaker IM et al. Rationally engineered Cas9 nucleases with improved specificity. Science (80-.). 351, 84–88 (2016).
- Chen JS et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature 550, 407–410 (2017). [PubMed: 28931002]
- Kleinstiver BP et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide offtarget effects. Nature 529, 490–495 (2016). [PubMed: 26735016]
- 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). [PubMed: 29020979]
- 45. Landrum MJ et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 42, D980–D985 (2014). [PubMed: 24234437]
- 46. Rees DC, Williams TN & Gladwin MT Sickle-cell disease. Lancet 376, 2018–2031 (2010). [PubMed: 21131035]
- Quentin Blackwell R, Oemijati S, Pribadi W, Weng MI & Liu CS Hemoglobin G Makassar: β6 Glu→Ala. BBA - Protein Struct. 214, 396–401 (1970).
- 48. Viprakasit V, Wiriyasateinkul A, Sattayasevana B, Miles KL & Laosombat V Hb G-Makassar [β6(A3)Glu→Ala; codon 6 (GAG→GCG)]: Molecular characterization, clinical, and hematological effects. Hemoglobin 26, 245–253 (2002). [PubMed: 12403489]
- 49. Sangkitporn S, Rerkamnuaychoke B, Sangkitporn S, Mitrakul C & Sutivigit Y Hb G Makassar (beta 6: Glu→ Ala) in a Thai Family. 85, 577–582 (2002).
- 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). [PubMed: 26990991]
- 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). [PubMed: 26990992]
- Chatterjee P, Jakimo N & Jacobson JM Minimal PAM specificity of a highly similar SpCas9 ortholog. Sci. Adv 4, eaau0766 (2018). [PubMed: 30397647]
- Cong L et al. Multiplex genome engineering using CRISPR/Cas systems. Science (80-. ). 339, 819–823 (2013).

- 54. Esvelt KM et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat. Methods 10, 1116–1121 (2013). [PubMed: 24076762]
- 55. Edraki A et al. A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for In Vivo Genome Editing. Mol. Cell 73, 714–726.e4 (2019). [PubMed: 30581144]
- 56. Hirano H et al. Structure and Engineering of Francisella novicida Cas9. Cell 164, 950–961 (2016). [PubMed: 26875867]
- 57. Harrington LB et al. A thermostable Cas9 with increased lifetime in human plasma. Nat. Commun. 8, 1424 (2017). [PubMed: 29127284]
- Zetsche B et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell 163, 759–771 (2015). [PubMed: 26422227]
- 59. Hou Z et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc. Natl. Acad. Sci 110, 15644–15649 (2013). [PubMed: 23940360]
- Kim E et al. In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. Nat. Commun 8, 14500 (2017). [PubMed: 28220790]
- Mali P et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol 31, 833–838 (2013). [PubMed: 23907171]
- Konermann S et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 517, 583–588 (2015). [PubMed: 25494202]
- Badran AH & Liu DR Development of potent in vivo mutagenesis plasmids with broad mutational spectra. Nat. Commun 6, 8425 (2015). [PubMed: 26443021]
- 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).
- Gaudelli NM et al. Programmable base editing of T to G C in genomic DNA without DNA cleavage. Nature 551, 464–471 (2017). [PubMed: 29160308]
- 66. Müller KM 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). [PubMed: 16061932]
- 67. Dower WJ, Miller JF & Ragsdale CW High efficiency transformation of E.coli by high voltage electroporation. Nucleic Acids Res. 16, 6127–6145 (1988). [PubMed: 3041370]
- Komor AC, Kim YB, Packer MS, Zuris JA & Liu DR Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420–424 (2016). [PubMed: 27096365]
- 69. Clement K et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat. Biotechnol 37, 224–226 (2019). [PubMed: 30809026]
- Arbab M, Srinivasan S, Hashimoto T, Geijsen N & Sherwood RI Cloning-free CRISPR. Stem Cell Reports 5, 908–917 (2015). [PubMed: 26527385]
- 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). [PubMed: 16959904]

Author Manuscript



Figure 1. Phage-assisted non-continuous evolution (PANCE) of SpCas9 binding activity on non-G PAMs.

(a) Original selection scheme for Cas9 DNA binding in which  $\omega$ -dSpCas9 expressed by gIII selection phage (SP) binds to a designated protospacer and PAM sequence upstream of gIII on an accessory plasmid (AP) in host *E. coli* cells. Host cells and infecting SP are continuously mutagenized by a mutagenesis plasmid (MP). (b) Binding activity of SpCas9 or xCas9 on all 64 possible NNN PAMs, determined by fold propagation of SP expressing  $\omega$ -dSpCas9 (top) or  $\omega$ -dxCas9 (bottom) on host cells containing APs bearing these PAM sequences upstream of gIII. (c) Schematic overview of 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 hours 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  $\omega$ -dSpCas9 (N1) or  $\omega$ -dxCas9 (N2) on NAA (red), NAT (blue), or NAC (green) PAM sequences.

Miller et al.



# Figure 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 where  $\omega$ -dSpCas9 binds two distinct protospacer-PAM 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<sub>C</sub>) is expressed by the SP and  $\omega$ -dSpCas9(1–573) fused to NpuN ( $\omega$ dSpCas9<sub>N</sub>) is encoded by a low-copy complimentary plasmid (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 gIII gVI 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 evolved clones shown in (b) in HEK293T cells on NRA PAM sites. Bars represent mean and standard error (SEM) 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 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 TAC-P9s-3 mapped onto the SpCas9 crystal structure (4UN3).

Miller et al.





(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/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 quartile for n=666 to

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, and 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.

Miller et al.

Page 30



# Figure 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 (H=non-G) PAMs for SpCas9-NRRH, -NRTH, -NRCH, and SpCas9-NG. Mean and standard deviation (SD) of all individual values of n=3 independent biological replicates are plotted. (b) Indel formation in primary human fibroblasts across give endogenous human sites containing NR PAMs for SpCas9-NRRH, -NRTH, -NRCH, and SpCas9-NG. Bars represent mean and standard deviation (SD) 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, -NRTH-, and NRCH at HEK site 4. Off-target reads that represent 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, -NRTH-, and NRCH resulting from GUIDE-seq analysis on HEK site 4, *VEGFA* site 2, HEK site 1, and *EMX1* in U2OS cells as determined by number of off-target sites detected. See Supplementary Fig. 6 for additional GUIDE-seq results.



Figure 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 (H=non-G) PAMs for CBE-NRRH, CBE-NRTH, CBE-NRCH, and CBE-NG. Mean and SEM of three 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 mean and SEM 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. Mean and SEM of the number of targeting sgRNA are shown for n=3,919, n=12,095, n=1,154, n=9,740, n=1,132, or n=3,841 individual ClinVar entries for CBE/NR, ABE/NR, CBE/NG, ABE/NG, CBE/NGG, and ABE/NGG, respectively.



Figure 6. Evolved SpCas9 variants enable correction of pathogenic SNPs with non-G PAMs. (a) Overview of adenine base editing strategy for correcting the sickle hemoglobin (*HbS*) SNP. In HbS, the Glu (GAG codon) at position 6 of normal  $\beta$ -globin (*HBB*) is mutated to a Val (GTG codon). Targeting this SNP with adenine base editing on the reverse strand enables a Val to Ala (GTG to GCG) base conversion, leading to the rare Makassar  $\beta$ -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 that leads to a silent Pro (CCT) to Pro (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 pro (CCT) to pro (CCC) mutation, at position 9. For (b) and (c), bars represent mean and SEM of n=3 independent biological replicates, with individual values shown as dots. (d) Recommended Cas9 variants for accessing all possible PAMs within NRNN PAM space. Only Cas9s 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.