Directed evolution of Cas9 for mammalian genome editing

Directed evolution of Cas9 for mammalian genome editing

A dissertation presented

by

Johnny Hao Hu

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

April 2018

Dissertation Advisor: David Ruchien Liu Johnny Hao Hu

Directed evolution of Cas9 for mammalian genome editing

Abstract

The ability to manipulate the genome has been a longstanding goal since the discovery of the central dogma of molecular biology. Such a technology would not only allow us to correct deleterious mutations for human therapeutics but would also be greatly enabling for understanding the function of a given gene. The simplicity of the Cas9 system, an RNA-guided endonuclease, has led to widespread adoption in many animal and plant species. The ability to separate the DNA-binding activity of Cas9 from its nuclease activity has also given us the ability to tether a number of DNA effector domains to Cas9 to further expand its repertoire of DNA modifications. Nonetheless, a number of key limitations still exist for the system such as the requirement that a protospacer adjacent motif be present at the target site. For the most commonly used Cas9 from *Streptococcus pyogenes* (*Sp*Cas9), the required PAM sequence is NGG. No natural or engineered Cas9 variants that have been shown to function efficiently in mammalian cells offer a PAM less restrictive than NGG.

Here we use phage-assisted continuous evolution (PACE) to evolve an expanded PAM *Sp*Cas9 variant (xCas9) that can recognize a broad range of PAM sequences including NG, GAA and GAT. The PAM compatibility of xCas9 is the broadest reported, to our knowledge among Cas9 proteins that are active in mammalian cells, and supports applications in human cells including targeted transcriptional activation, nuclease-mediated gene disruption, and cytidine and adenine base editing. Notably, despite its broadened PAM compatibility, xCas9 has much greater DNA specificity than *Sp*Cas9, with substantially lower genome-wide off-target activity at all NGG target sites tested, as well as minimal off-target activity when targeting genomic sites with non-NGG PAMs. These findings expand the DNA targeting scope of CRISPR systems and establish that there is no trade-off between Cas9 editing efficiency, PAM

compatibility and DNA specificity. Comparing the mutations in xCas9 with *Sp*Cas9 gives us additional clues to the underlying mechanism of DNA recognition and cutting employed by the Cas9 enzyme and general principles for designing new Cas9 variants. PACE has also been proven to be a versatile system for Cas9 evolution and could play a major role in the goal of building a suite of Cas9 variants that would allow us to programmatically modify any genomic sequence at will with high specificity.

Table of Contents

Acknowledgements

I would like to thank my advisor Professor David Liu for his support and mentorship over the years and for providing the research environment where a project such as this is possible. This project, in particular, drew on a number of difference subfields and is built upon the work of many former Liu Lab members who have pioneered new techniques that were crucial. I would also like to thank David for pushing me to do more than I thought possible and staying up late with us to put the manuscript together. I have learned much from you not only in thinking about science but also in the value of pursuing a passion deeply.

I would also like to thank my co-advisor George Church and his lab for all of their help and support. You have always asked us to think big George and I hope that we can live up to your lofty dreams. There are few places as exciting and stimulating as the Church Lab where nothing is quite impossible.

To my lab mentors and colleagues, I will be eternally gratefully not only for your scientific help and advice but also for your friendship. I have learned many techniques from my rotation mentors, Basil Hubbard and Kevin Esvelt, and from other former members, Ahmed Badran and David Thompson. I also could not have completed this project without the help of my coauthors, all of whom were necessary getting this project to completion. I would especially like to thank Shannon Miller for carrying out that final long sprint with me at the end, Maarten Geurts for spending a year with us away from his home in Europe, and Christina Zeina, for spending the past few summers in lab.

From the community at Harvard, where I have spent nearly my entire adult life so far, I have gathered knowledge and wisdom both inside and outside of the classroom. Cabot House has been my home for home for much of that time, first as a student and then as a tutor. I cannot imagine anywhere else I would rather have spent my time at Harvard, and I know that the people there from Rakesh and Stephanie to the students to the house staff have shaped me

into the person that I am today and have made me a better person. I will continue to write my own story.

To my fiancée, Amy Zhang, I am sure that I would not be where I am without you and I know that we will support each other for the rest of our lives. I look forward to our many adventures together.

And finally to my parents, Margaret Huang and Zhengwei Hu and my sister, Alison, as always you have been there encouraging me and offering advice. Word cannot capture the gratitude I owe to my family.

List of Figures

List of Tables

List of Abbreviations

Accessory plasmid (AP)
Adenine base editor (ABE)
Atomic force microscopy (AFM)
Base editor (BE)
Chromatin immunoprecipitation (ChIP)
Clustered regularly interspaced short palindromic repeats (CRISPR)
CRISPR RNA (crRNA)
Double stranded break (DSB)
Embryonic stem (ES)
Fluorescence resonance energy transfer (FRET)
Phage-assisted continuous evolution (PACE)
High-throughput, genome-wide translocation sequence (HTGTS)
Homologous recombination (HR)
Mutagenesis plasmid (MP)
Protospacer adjacent motif (PAM)
Selection phage (SP)
Single guide RNA (sgRNA)
Specificity inference for TAL-effector design
Staphylococcus aureus Cas9 (*Sa*Cas9)
Streptococcus pyogenes Cas9 (*Sp*Cas9)
Trans-activating crRNA (tracrRNA)
Transcription activator-like effector nuclease (TALEN)
Truncated sgRNA (trugRNA)
Uracil N-glycosylase inhibitor (UGI)
Zinc-finger nuclease (ZFN)

Chapter 1:

Introduction

Copyright Disclosure

Portions of this Introduction appear in or are adapted from the following publications:

Hu, J. H., Davis, K. M., Liu, D. R. Chemical Biology Approaches to Genome Editing: Understanding, Controlling, and Delivering Programmable Nucleases. Cell Chemical Biology (2016). doi:10.1016/j.chembiol.2015.12.009.

Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., Liu, D. R. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* (2018). doi:10.1038/nature26155.

Overview

Gene editing holds the promise of transforming both human therapeutics and research into the causative variants of genetic diseases. Indeed the idea of modifying the human gene to treat inherited monogenic disorders has been proposed since at least the 1970s[1]. The development of recombinant DNA technology during the same time by Herbert Boyer and the University of California at San Francisco and Stanley Cohen at Stanford University[2] and DNA sequencing technologies by such pioneers as Frederick Sanger[3], Ray Wu[4], Walter Gilbert[5], and Allan Maxam[5], greatly advanced the ability of researchers to interrogate biological questions, create new tools for molecular biology, and develop new classes of drugs. Since then the ability to sequence whole human genomes through high-throughput sequencing[6], has allowed to map entire genomes and define the genetic variation between species, within a species, and between healthy and malignant cells. The ability to read DNA has nowadays gone beyond the research lab, and with exponential decreases in cost[7] has been made available to the general public for diagnostics or to trace genetic lineage. With this increase in DNA sequencing has also come a push for the development of tools that would enable writing to the genome.

The recognition that double-stranded breaks (DSBs) could stimulate homologous recombination[8,9] led to the search for enzymes that could programmatically be used for inducing such breaks at specific sites in the genome. From megaendonucleases to zinc-finger nucleases (ZFNs) to transcription activator-like effector nucleases (TALENs) to Cas9 and other Clustered

Regularly Interspaced Short Palindromic Repeats (CRISPR) systems, each advancement has brought about another wave of innovations and challenges unique to each system. The ability to fuse DNA effectors to the DNA-binding domain of these systems has also led to the development of everything from transcriptional activators and repressors to "base editors" that can modify single nucleotides without double-stranded DNA breaks. The ability to multiplex and the rapid decrease in the cost of not only sequencing but also oligonucleotide synthesis means that we have been to interrogate biology questions at an unprecedented scale. The rapid advancement of the field also promises to bring new technologies online that will lead us closer to the goal of a universal DNA editor.

Development of programmable nucleases for genome editing by inducing double-stranded breakage

Elucidation of homologous recombination for gene replacement

Gene replacement has been available for several decades due to the elucidation of homologous recombination (HR) as a genetic modification tool. Procedures for gene replacement in *Saccaryomyces cerevisiae* has been available since 1979[10], and making targeted gene replacement in mice[11] has been made possible with the availability of embryonic stem (ES) cells that can be manipulated in culture. Selection against random integration has to be applied with the use of selectable markers, but the pluripotency of the ES cells allows them to populate all cell lineages after injection into early embryos. The absolute frequency of HR is however quite low at once in every 10^4 to 10^7 cells[12]. The recognition that DSBs could stimulate recombination by homologous donor DNA by several orders of magnitude was a key insight that then led to the exploration of different enzymes that could make targeted DSBs[12].

Meganucleases were the first class of enzymes used to verify that DSB stimulation would lead to higher rates of recombination. Their long recognition sites make them suitable for mammalian genome editing, but their lack of programmability make them difficult to use generally.

While some limited success has been seen with designing meganucleases to recognize new sequences, the intimate connection between the recognition and cleavage domains makes it challenging to modify one without breaking the other. Better directed evolution and protein design tools could still mean that meganucleases have a role in certain applications but recent systems that separate the DNA binding and cleavage domains have proven to be much easier to reprogram.

Zinc-finger nucleases and Transcription activator-like effector nuclease

Zinc-finger proteases were the first programmable nucleases described in the mid-1980s[12-15]. The recognition that the natural type IIS restriction enzyme, FokI, has physically separateable binding and cleaving activities and that the cleavage domain has no apparent sequence specificity[16] meant that it could be used with zinc fingers for DNA editing. FokI activity occurs when dimerized so by designing a pair of ZFNs to bind close to each other, DNA cleavage will occur. Each ZFN contains between three and six individual zinc finger repeats that recognize between 9 and 18 basepairs. The majority of these engineered zinc finger arrays are based on the zinc finger domain of the murine transcription factor Zif268[15], but the inherent difficult in designing and validating such ZFNs has limited their adoption. Nonetheless, ZFNs have been pivotal in verifying the mechanisms by which we can introduce point mutations, deletions, insertions, duplications, and translocations in the genome. Furthermore, ZFNs designed to disrupt the expression of the HIV co-receptor *CCR5*[17] as well as clinical trials designed to evaluate the safety of gene editing for hemophilia A and B and mucopolysaddharidosis type I and II have all been initiated.

TALENs rely on a similar principal of tethering a DNA-binding domain with a FokI nuclease domain[18]. The DNA-binding domain is composed of highly conserved repeats derived from transcription activator-like effectors (TALEs)—proteins used by *Xanthamonas* bacteria to alter transcription of genes in host plant cells[18]. Due to the simplicity of the protein-DNA code that

was elucidated for TALENs and the modular nature of the TALE domains, designing a new TALEN to target a gene of interest was simpler than previous methods. Furthermore, the method was used for many different model organisms, including fruit flies[19], roundworms[20], zebrafish[21], frogs[22], rats[23], and pigs[24], as well as human cells[25]. Challenges, however have remained, such as the difficulty of cloning TALE repeat arrays with nearly identical repeat sequences[26].

Development of the CRISPR-Cas9 System

As with other programmable nucleases, the CRISPR-Cas9 system was identified by combing nature for inspiration. As early as 1987, researchers had found series of short direct repeats interspersed with short sequences in bacterial genomes[27]. In 2005, these short repeats were found to be frequently of plasmid or viral origins[28], and at the same time, the *cas* (CRISPR-associated) genes were found to contain putative nuclease and helicase domains. Together, this information led to the proposal that CRISPR-Cas is an adaptive immune system against foreign DNA that uses antisense RNAs as memory signatures of past invasions. This hypothesis was verified in 2007 by two researchers at Danisco, who provided the first evidence of adaptive immunity upon infection of *Streptococcus thermophilus* with lytic phage[29]. Soon thereafter, mature CRISPR RNAs (crRNAs) were shown to serve as the guide that targets the Cas protein to the pathogen[30]. An additional RNA sequence, the trans-activating crRNA (tracrRNA), was later found to be essential for crRNA maturation by ribonuclease III and Cas9[31], which itself had been identified by bioinformatics analysis as a large multifunctional protein with two putative nuclease domains. Overall, adaptive immunity requires (1) insertion of a short sequence of the invading DNA as a spacer sequence, (2) transcription of the precursor crRNA that matures to generate the crRNA and pairs with the tracrRNA, and (3) crRNA-directed cleavage of the foreign nucleic acid by Cas9[32].

By 2012, the basic components of CRISPR-Cas9 had been elucidated and the *S. pyogenes* CRISPR-Cas9 was found to be an RNA-guided DNA endonuclease. The dual

tracrRNA:crRNA duplex was later fused into a single guide RNA (sgRNA) that includes the 20-nucleotide sequence at the 5' end that binds to the target sequence via Watson-Crick base pairing as well as the 3' end that binds to the Cas9 protein[33]. Thus, by changing the 20-nucleotide spacer sequence we are able to reprogram the Cas9 protein to any corresponding 20-nucleotide sequence in the genome, termed the protospacer. Unlike ZFNs and TALENs, which require substantial protein modification, Cas9 can theoretically be targeted by changing the RNA sequence provided that there is a protospacer adjacent motif (PAM) sequence next to the protospacer. The PAM, which is recognized by the Cas9 protein itself, plays an important role in distinguishing self and non-self DNA in bacteria but also restricts the sequence space that can be targeted by Cas9.

Mechanism of CRISPR-Cas9 Target Search and Recognition

Structure of *S. pyogenes* Cas9

While a number of different Cas9 enzymes have been identified, the most studied and widely used is the one from *Streptococcus pyogenes* (*Sp*Cas9). A large 1,368 amino acid multidomain and multifunctional enzyme, *Sp*Cas9 contains two distinct nuclease domains, an HNH domain that cuts the strand complementary to the sgRNA (target strand) and a RuvC domain that cuts the DNA strand opposite the complementary strand (nontarget strand)[34]. Mutating either the HNH (H840A) or the RuvC domain (D10A) converts Cas9 into a DNA nickase (nCas9) while mutating both leads to a nuclease dead Cas9 (dCas9). In addition to the nuclease (NUC) lobe, Cas9 also contains an alpha-helical recognition (REC) lobe and a more variable C-terminal lobe (CTD)[34].

Comparison of the apo-Cas9 and sgRNA-bound Cas9 shows significant conformational changes upon sgRNA binding[35]. Extensive interactions between the Cas9 and the bound sgRNA have been demonstrated, and Cas9 preorders the 10-nt RNA seed sequence on the 3' end required for initial DNA interrogation in an A-form conformation. In addition, two residues known

to be important for PAM sequence recognition, R1333 and R1335 that are disordered in the apo structure are prepositioned prior to making contact with target DNA. The 5' 10-nt nonseed RNA, while disordered in the sgRNA-bound crystal structure, is shown to lie in the cavity between the HNH and RuvC domains, where it is protected from degradation[35]. Thus, the Cas9-sgRNA complex is engaged to search for the target sequence.

DNA search and PAM recognition

Cas9 target recognition requires both a matching PAM sequence and complementarity between the spacer and the protospacer. The PAM sequence is crucial for discriminating between self and non-self DNA and single mutations in the PAM can disable Cas9 cleavage activity in vitro and leads to bacteriophage evasion. Single-molecule experiments have shown that PAM interrogation is the first step in DNA recognition and occurs before spacer-DNA matching occurs. Cas9 DNA dwell times are also quite short when a PAM sequence is not present and target recognition occurs through three-dimensional collisions. PAM recognition leads to local DNA melting at the PAM-adjacent nucleation site and RNA-DNA hybridization[36].

DNA curtain assays performed by Doudna, Greene, and their respective coworkers on Cas9:gRNA suggest that Cas9 does not find its target by facilitated diffusion[37]. These experiments used lipid-tethered DNA molecules aligned at a diffusion barrier, which upon fluid flow, align to one another and form a curtain of DNA molecules[38,39]. Cas9 programmed with a gRNA matching a site present in the DNA molecules were injected into the chamber, buffer flow was terminated, and reactions were visualized in real time. The data revealed long-lived binding events at the target site and transient binding events at other sites. Neither one-dimensional sliding nor hopping was observed; rather, all target recognition events appear to occur directly through three-dimensional collisions. The shorter-lived, non-specific binding events exhibited complex dissociation kinetics and indicated that Cas9 undergoes at least two and possibly more intermediates on the pathway towards cognate target recognition. Regions of the DNA with a

higher abundance of the NGG PAM recognized by *Sp*Cas9 exhibited higher Cas9:gRNA occupancy, suggesting that Cas9 might bind specifically to PAM sites when searching for potential targets. Indeed, while Cas9 must transiently sample DNA lacking PAMs, these rapid binding events are not detectable in either single-molecule or bulk binding experiments[37]. PAM recognition leads to a destabilization of the adjacent sequence followed by formation of the RNA-DNA heteroduplex. Only upon binding to a bona fide target locus do the nuclease domains in Cas9 activate.

The structure of the *Sp*Cas9 nickase bound to the PAM sequence has elucidated the direct structural contacts between the protein and the DNA[36]. For *Sp*Cas9, the canonical PAM is NGG, where N is any of the four bases. The PAM duplex was found to be nestled in a positively charged groove between the REC and NUC lobes with the PAM-containing non-target strand mostly in the CTD. The 5' N of the PAM did not make interact with the Cas9 and remained basepaired to its complementary base. The GG sequence in the PAM, however, formed base-specific hydrogen-bonding interactions with R1333 and R1335[36]. The findings agree with previous studies that found that PAM recognition occurred through the nontemplate strand and that template strand mismatches are tolerated. Noncanonical PAMs, which can show some activity with *Sp*Cas9 has been shown to be recognized through an induced fit mechanism in which recognition of noncanonical PAMs induce a subtle distortion in the DNA backbone of the PAM duplex[36]. Further structural analysis of engineered *Sp*Cas9 will lead to a greater understanding of the mechanism of PAM recognition and important resides within Cas9 that program PAM recognition.

DNA cleavage by Cas9

The completion of Cas9 binding to DNA is required but does not always lead to DNA cleavage. ChIP-sequencing suggests that while the Cas9-sgRNA complex can bind to many off-target sites only a small fraction trigger off-target cleavage[40-42]. Bulk fluorescence resonance energy transfer (FRET) experiments suggest that the RNA-DNA hybridization drives

conformational changes in the HNH domain, which itself allosterically controls cleavage by the RuvC domain[43]. Structural analysis has shown a large conformation change in the HNH that brings the active site closer to the cleavage site. Furthermore, two hinge regions connecting the HNH domain to the RuvC domain undergo a folding-unfolding rearrangement that engages the RuvC domain for cleavage[43]. Cleavage of the DNA for *Sp*Cas9 usually occurs around 3 bp upstream of the PAM sequence.

Studies by Marszalek and coworkers also used atomic force microscopy (AFM) to shed light on the conformational changes that Cas9 undergoes upon binding to DNA[44]. The results provide evidence of a "conformational gating" mechanism in which interactions between the gRNA and the 14th-17th nucleotide region of the target DNA drives a Cas9 conformational change that leads to DNA cleavage. Both Cas9 and catalytically dead Cas9 (dCas9) were found to undergo progressive structural transitions as they bind to DNA. dCas9 molecules bound to sgRNA are predominantly smaller and egg-shaped but increase in both height and volume as they bind to target sequences. Catalytically active Cas9 undergoes a similar increase in size although is slightly smaller, possibly due to a final conformational change upon cleavage of the DNA[44]. Using both structural data and size-exclusion chromatography, Doudna and coworkers also observed similar changes in the Cas9 conformation upon binding to the target DNA site[35].

Characterization and Application of CRISPR-Cas9

Off-target profiling of genome-editing proteins

The ability of programmable DNA-binding proteins and nucleases to distinguish on-target from off-target DNA sequences is crucial for their application as research tools and potential human therapeutics. The specificity of ZFNs, TALENs, and Cas9 has been studied using a variety of methods both *in vitro* and in cells. Due to their concatameric nature, ZFNs and TALENs can theoretically be expanded to extend DNA specificity. However, expanding the number of zinc-

finger or TALE repeats also increases the likelihood that a partial match with an off-target sequence is sufficient to support DNA binding.

In vitro and cell-culture studies on TALEN specificity have not only substantiated off-target activity but also elucidated design principles for improving specificity. Testing three distinct sequences, *CCR5A, CCR5B,* or *ATM*, Liu and coworkers profiled the DNA cleavage specificity of 30 unique heterodimeric TALEN pairs with different C-terminal, N-terminal, and FokI domain variants using an *in vitro* selection method that evaluates the ability of each TALEN to cleave DNA at each of 10^{12} different potential substrate sequences that are related to the on-target sequence[45]. The results of the *in vitro* specificity profile correctly predicted off-target DNA substrates in the human genome for the *CCR5A*- and *ATM*-targeting TALENs that were detectably modified by these TALENs in human cells. More recently, a modified high-throughput, genome-wide translocation sequencing (HTGTS) method was used by Alt and coworkers to detect DNA double-stranded breaks induced by TALENs in a genome-wide fashion[46]. Two TALENs were studied that cleave *C-MYC* or *ATM*. A large number of off-target sites were detected for both of these targets—522 off-target sites for *ATM* and 384 off-target sites for *C-MYC*. Many highly enriched TALEN off-targets were pseudo-palindromic sequences that correspond to variants of the recognition site of a single TALEN monomer, suggesting that the binding energy of these single monomers were enough to bind and activate DNA cutting[46]. Other large-scale studies such as one conducted by Bulyk and coworkers quantitatively assayed the DNA-binding specificities of 21 representative TALEs to ~5,000 to 20,000 unique DNA sequences per protein using custom-designed protein-binding microarrays[47]. Using the data from the study, the Specificity Inference for TAL-Effector Design (SIFTED) tool was created that models the intrinsic specificity of each RVD and considers a variety of repeat context features, including the number of repeats in the protein, each repeat's position within the repeat array, and the immediately neighboring N- and C-terminal repeat types. Further analysis by Baldwin, Segal, and their respective coworkers on TALE specificity also

determined that RVD choice can affect affinity to the DNA target by four orders of magnitude[48]. Additionally, base mismatches at the 5′ end of the target site had more disruptive effects on affinity than those at the 3′ end.

Researchers have developed and applied similar methods for analyzing the DNA specificity of Cas9. Initial studies from several groups showed that some single-base mutations could be tolerated, especially in nucleotides distal from the PAM sequence[33,49,50]. Early results suggested that a "seed sequence," the 8-10 nucleotides immediately adjacent to the PAM, was the primary determinant of Cas9 DNA recognition specificity[33,51,52]. Later studies have shown, however, that while more sensitive to mutations than at distal positions, off-target substrates can still contain mutations in the seed sequence[53-56]. Further studies by Liu and coworkers used libraries that enabled the broad discovery of off-target sequences both *in vitro* and in human cells and concluded that off-target substrates cleaved by Cas9 could contain even five mismatches[53]. These methods use focused deep sequencing to identify indel mutations at potential off-target sites but also likely missed other off-target substrates as they assume that off-target sequences are closely related to the on-target sequence. The ability of Cas9 to bind and cleave at DNA sites that contain insertions (DNA bulge) or deletions (RNA bulge) relative to the gRNA also increase the difficulty of predicting off-target sites[57-59]. Several groups have used ChIP-seq to find a large number of potential off-target binding sites of dCas9, but very few of these sites appear to represent off-target sites of cleavage by active Cas9 nuclease[40-42,60]. As cells are cross-linked post-transfection, false positives could arise due to transient interactions between Cas9 and genomic DNA[41]. Whole-genome sequencing has been used to identify off-target alterations in edited single-cell clones[61-63], but is impractical to perform at the depth necessary to detect low-frequency genome-modification events needed to detect many off-target substrates that are only weakly modified. For example, detecting less than 1% modification would require sequencing thousands of whole genomes using such an approach.

More recently, researchers have developed methods to identify genomic double-stranded breaks in human cells in an unbiased manner that does not presume that off-target substrates are similar to on-target sequences. These methods were used to profile TALENs, Cas9, and the Cas9 D10A paired nickase as well as modified gRNAs. Using blunt-double-stranded oligodeoxynucleotides (dsODNs) as tags that are inserted into the genome upon double-stranded breaks and repair through an end-joining process consistent with NHEJ, Joung and coworkers used a method called genome-wide, unbiased identification of double-stranded breaks enabled by sequencing (GUIDE-seq) to characterize off-target sequences of 13 Cas9:sgRNA complexes in two human cell lines, U2OS and HEK293[64]. Consistent with previous results, as many as six mismatches were found in the protospacer sequence for off-target sites. Noncanonical PAMs such as NAG, NGA, NAA, NGT, NGC, and NCG, were also all found among sites that were cleaved by Cas9. Greater numbers of mismatches decreased the likelihood of being cleaved, mismatches near the 5′ end of the target site were more tolerated than at the 3′ end, although mismatches 1-4 bp away from the PAM were somewhat better tolerated than those 5-8 bp away, and wobble mismatches were found to be more highly tolerated than non-wobble mismatches. Interestingly, the number of off-target sites identified for each Cas9:sgRNA varied greatly from zero to > 150[64]. Similarly, Alt and coworkers used the HTGTS method to find variable off-target modification frequencies for four *RAG1*-targeting Cas9:sgRNA complexes, with two of the four showing no detectable off-target activity[46].

A separate method, Digenome-seq, developed by Kim and coworkers also observed that off-target activity was highly variable depending on the targeted site[65]. In Digenome-seq, human genomic DNA is digested using Cas9 nuclease *in vitro*. The resulting DNA is then subjected to whole genome sequencing and Cas9 cleavage sites were identified by looking for sequence reads that are vertically aligned that represent DNA fragments with identical 5′ ends. Using this method, for one Cas9:sgRNA that targeted *VEGF*, a total of 80 potential off-target sites were identified. All

80 sites and an additional 28 sites representing off-target sites that differ from the on-target sites by three or fewer nucleotides were further investigated by targeted deep sequencing. Four previously identified off-target sites were validated and four new off-target sites captured by Digenome-seq also showed off-target cleavage. These off-target sites contained up to six nucleotide mismatches, highlighting the large number of mismatches between the guide RNA and DNA substrate that in some cases can be tolerated by Cas9[65].

Zhang and coworkers also interrogated the genome-wide off-target cutting activity of both *S. pyogenes* and *S. aureus* Cas9[59]. In the BLESS (direct *in situ* breaks labeling, enrichment on streptavidin and next-generation sequencing) method[66], cells are transfected with either *S. pyogenes* Cas9 or *S. aureus* Cas9, and fixed. Free genomic DNA ends are captured using biotinylated adaptors and analyzed by deep sequencing. A sequencing pipeline filters out the DSB clusters and assigns DSB scores by comparing the count of DSBs in the experimental and negative control samples. Top-ranked off-target sites identified by BLESS were verified using targeted deep sequencing to detect indel formation. Only sites that contained a PAM and homology to the guide sequence exhibited indels. Consistent with similar genome-wide off-target studies, BLESS revealed additional off-target sites not previously predicted by simple sequence similarity to the target or by ChIP[59]. Additional off-target sites contained not only Watson-Crick base-pairing mismatches to the guide but also insertion and deletion mismatches between the gRNA and the target DNA, consistent with other reports[57-59].

Protein engineering and evolution to improve genome-editing specificity

In light of the imperfect specificity of wild-type or canonical programmable nucleases, researchers have sought to improve specificity using several strategies including protein engineering and evolution. Soon after the discovery of the catalytic domains of Cas9, two mutations in the conserved HNH and RuvC nuclease domains, D10A and H840A, respectively, were used to abolish the nuclease activity of Cas9 to form catalytically dead Cas9 (dCas9)[33].

Since each domain mediates the cleavage of one DNA strand, either mutation by itself creates a Cas9 nickase that only cleaves a single strand of DNA. Church and coworkers and Zhang and coworkers independently demonstrated the use of pairs of the D10A Cas9 nickase (Cas9n) to induce double-stranded genomic breaks by using two gRNAs to target the Cas9n to nearby sites offset up to 100 bp[55,56]. The requirement of two independent nicking events relatively close to each other on opposite strands in order to cause a double-stranded break increases the specificity of paired Cas9 nickases relative to wild-type Cas9. Deep sequencing of known off-target sites in cells either transfected with wild-type Cas9 or Cas9n showed 50- to 1,500-fold greater specificity of the nickase compared to that of wild-type Cas9[55].

Fusion of dCas9 to FokI (creating "fCas9"), similar to the strategy used in ZFNs and TALENs, also increases DNA cleavage specificity relative to wild-type Cas9 as shown independently by Liu and coworkers, Joung and coworkers, and Mahfouz and coworkers[67-69]. By fusing FokI to the N-terminus of dCas9, fCas9 requires the assembly of two FokI monomers at the target site to induce double-stranded DNA cleavage. The spacing requirements of fCas9 are more stringent than what is required for Cas9n, with optimal activity achieved when the two target sites are either ~15 or 25 base pairs apart[45]. As a result, the specificity of fCas9 was shown to be higher than that of paired nickases at loci with highly similar off-target sites[67-69].

The engineering of monomeric forms of programmable nucleases has also been used to substantially improve their DNA specificity. Following the use of the *in vitro* specificity profiling method described above on ZFNs, TALENs, and Cas9, Liu and coworkers hypothesized that excess DNA-binding energy erodes the specificity of genome-editing agents, and speculated that reducing the strength of non-specific DNA-binding interactions could improve the specificity of these nucleases by decreasing off-target DNA binding while still allowing the nuclease to bind on-target sequences[45,53,70]. Indeed, the researchers found that changing cationic Lys or Arg residues that do not make base-specific contacts with DNA to neutral Gln residues could increase the DNA

cleavage specificity of TALENs in cells by ~10- to 100-fold, without necessarily decreasing activity for on-target sequences[45]. Given that wild-type Cas9 with full-length sgRNA can cleave a wide variety of sequences and has likely evolved to tolerate some degree of mutation among natively targeted viral DNA sequences, it is likely that the Cas9:sgRNA–DNA binding interaction is also stronger than is necessary for nuclease function, and that reducing this excess DNA-binding energy can decrease off-target activity while maintaining on-target activity in a manner similar to the improved specificity of the engineered TALENs. Indeed, Joung and coworkers demonstrated improvements in Cas9 specificity with truncated sgRNAs (see below)[71], and Zhang and coworkers recently identified specificity-enhancing mutations in Cas9 that neutralize positively-charged residues in the non-target strand groove[72]. By generating *S. pyogenes* Cas9 mutants consisting of individual alanine substitutions at 32 positively-charged residues, the authors were able to identify mutations that could lead to higher specificity without decreasing on-target cleavage efficiency. Top single amino acid mutants were combined and tested to identify three Cas9 mutants with both high efficiency of cutting at the on-target site and improved specificity.

Laboratory evolution efforts on TALENs and Cas9 have also revealed mutations within each protein that can alter or enhance DNA specificity. Through a positive selection on an on-target site and negative selection against cleaving a known off-target site, Liu and coworkers used phage-assisted continuous evolution (PACE) to generate more specific TALE proteins that target the human *ATM* locus[73]. Interestingly, specificity profiling against 10^{12} other off-target sites that were not used in the negative selection revealed that the evolved TALENs showed broad specificity improvements across virtually all recognized DNA positions, indicating that the specificity of the TALENs was broadly improved, rather than only improved against the selected off-target sequence[73]. While evolving *S. pyogenes* Cas9 variants that accept non-canonical PAM sequences, Joung and coworkers discovered a D1135E mutation that improves discrimination between NGG and NGA PAM sequences[74]. This single mutation led to increased specificity at 19

of the 22 off-target sites tested using deep-sequencing of known off-target sites. GUIDE-seq using three different sgRNAs also suggested that these evolved Cas9 variants show general improvements in genome-wide specificity relative to wild-type Cas9[74].

Guide RNA modifications to improve genome-editing specificity

Engineering guide RNAs can also improve Cas9 specificity by increasing the requirements for interacting with the DNA target. Some of these modifications lead to less activity overall, highlighting a tradeoff between activity and specificity. For example, early sgRNA architectures varied the length of the sgRNA backbone, extending the Cas9-binding sequence by 10 bases to form a more active variant[33,55,56]. Comparing the shorter, less-active variant to the longer, more-active variant commonly used, on-target modification can be substantially lower with the shorter sgRNA, but off-target modification is also minimized and can drop below the limit of detection at some sites in which the more active sgRNA exhibits off-target activity[55].

Using sgRNA variants with shorter regions of target complementarity can also improve specificity, in most cases without sacrificing on-target genome-editing efficiencies. Hypothesizing that off-target effects of Cas9 can be minimized by decreasing the length of the sgRNA-DNA interface and that 5′-end sgRNA:DNA interactions might compensate for mismatches at other positions along the interface, Joung and co-workers tested sgRNAs with progressively shorter regions of complementarity for a target site in an *EGFP* reporter[71]. The researchers found that truncated sgRNAs (trugRNAs) with 17 or 18 nucleotides of target complementarity, instead of the canonical 20 nucleotides, functioned as efficiently as canonical sgRNAs for different *EGFP* targets. On sites from three human genes, *VEGFA, EMX1,* and *CLTA*, five of seven tested trugRNAs induced Cas9-mediated indel formations at comparable rates as full-length sgRNAs. Comparing the specificity of these trugRNAs versus full-length sgRNAs, Cas9 with trugRNAs generally appear to induce either very low or undetectable levels of mutations at potential off-target sites that differ by one or two mismatches as compared to full-length sgRNAs that can induce cutting

at off-target sites bearing up to four or five mismatches. trugRNAs were also tested with paired Cas9 nickases at four off-target sites of the *VEGF* gene and found that the combination of the two approaches led to undetectable off-target mutation rates, whereas either approach by itself induced detectable off-target mutations[71].

Further studies by Church, Weiss, and their respective coworkers using truncated guide RNAs also showed that even shorter guides, while not competent to mediate DNA cleavage, can still be used with dCas9 for DNA binding[75]. This finding establishes that Cas9 DNA binding and cleavage can be decoupled, and that the cleavage does not necessarily follow DNA binding even when the nuclease domains within Cas9 are present. By using Cas9, Cas9-VPR (consisting of the tripartite transcriptional activator VP64-p65-Rta), and dCas9-VPR with truncated gRNAs with 20-, 16-, or 14-nt complementarity, Weiss, Church, and coworkers tested the ability of each truncated gRNA to either bind or cleave DNA. Cas9-VPR was able to induce target chromosomal gene expression with either 16-nt or 14-nt gRNA, but not 20-nt gRNA, as the 20-nt gRNA led to DNA cleavage. Cas9-VPR with 14-nt gRNAs was able to activate expression of all targets tested to at least 40% of the expression level of Cas9-VPR with full-length gRNAs. No mutagenesis was observed using the 14-nt gRNA, while indels were observed in some cases using the 16-nt gRNAs, and in all cases using the 20-nt gRNAs[75]. In addition to establishing the ability to use a single Cas9 construct fused to a transcriptional activation domain for gene activation, repression, and cleavage, the use of shortened gRNAs also illuminate the relationship between gRNA-DNA interaction and Cas9 effector function.

Interestingly, the addition of two guanine nucleotides to the 5′ end of gRNA can also increase Cas9:sgRNA specificity[65,76]. Kim and coworkers found that these "ggX_{20}" gRNAs, where the lowercase g represents a mismatched guanine, increase specificity up to 660-fold compared with conventional gRNAs on the basis of on-target to off-target indel frequencies[65]. The extra, mismatched guanines did not lead to substantially lower activity when compared to a regular 20-

nt gRNA with either a mismatched or matched guanine at the 20th position in one cell line tested, but did lead to lower activity in a second tested cell line. These observations are consistent with a model in which the extra guanines weaken RNA:DNA binding energetics, which may or may not affect on-target activity depending on the degree of excess binding energy between the unmodified gRNA and DNA[65].

Natural and engineered Cas9 PAM variants

The PAM requirement of Cas9 limits the number of sites that can be targeted by Cas9, thereby reducing the applicability of the system for research and human therapeutics. The NGG PAM requirement of canonical SpCas9, which occurs on average only once in every ~16 randomly chosen genomic loci, greatly limits the targeting scope of Cas9 especially for applications that require precise Cas9 positioning, such as base editing (reviewed later), which requires a PAM ~15±2 nucleotides from the target base[77,78], and some forms of homology-directed repair, which are most efficient when DNA cleavage occurs ~10-20 base pairs away from a desired alteration[79,80]. These requirements limit the fraction of genomic DNA that can be targeted with CRISPR systems and highlight the need for more general genome editing tools.

To address this limitation, researchers have harnessed natural CRISPR nucleases with different PAM requirements and engineered existing systems to accept variants of naturally recognized PAMs. Other natural CRISPR nucleases shown to function efficiently in mammalian cells include *Staphylococcus aureus* Cas9 (SaCas9)[59], *Acidaminococcus sp.* Cpf1[81], *Lachnospiraceae bacterium* Cpf1[81], *Campylobacter jejuni* Cas9[82], *Streptococcus thermophilus* Cas9[83], and *Neisseria meningitides* Cas9[84]. None of these mammalian cell-compatible CRISPR nucleases, however, offer a PAM that occurs as frequently as that of SpCas9.

Kleinsteiver and coworkers also reprogrammed *Sp*Cas9 using a combination of directed mutagenesis, bacterial selection-based directed evolution, and combinatorial design. They were able to reprogram *Sp*Cas9 to accept a NGA PAM and NGCG PAM specifically, without activity on

other PAMs. The NGA PAM *Sp*Cas9 contained three mutations (D1135V/R1335Q/T1337R) and was termed the VQR variant while the NGCG PAM *Sp*Cas9 contained four mutations (D1135V/G1218R/R1335E/T1337R) and was termed the VRER variant. In addition, in another study, *Sa*Cas9 was shown to have broadened PAM with three mutation (E782K/N968K/R1015H). Specificity of the broadened *Sa*Cas9 appeared to be comparable to the wild-type variant or slightly less specific at certain sits. While these CRISPR nucleases engineered to accept additional PAM sequences[74,85] also expand the scope of genomic targets available for Cas9-mediated manipulation, many target sequences remain inaccessible.

Genomic Modifications and Regulation using CRISPR-Cas9

As the DNA binding and nuclease domains of Cas9 are separate and the nuclease domain can be inactivated, a number of different DNA modifying enzymes have been built with dCas9 as a programmable DNA-targeting enzyme. Thus, beyond targeted gene disruption and genetic replacement, Cas9 has also been used for transcriptional activation and repression[86], epigenetic modification[86], and direct conversion of a target base pair to a different base pair[77,78] in a broad range of organisms and cell types[87].

Of note, as DSB mediated repair often leads to unwanted indel formation, direct conversion from one base to another could provide a route to programmable genome without unwanted side product. By tethering a single-stranded DNA specific deaminase to dCas9, our lab and others have shown that we are able to catalyze direct base conversion specifically at the site bound by dCas9. These deaminases are active due to the generation of an exposed single-stranded R-loop 13-17 basepairs upstream of the PAM.

The first demonstrated deaminase that was active with Cas9 was a cytosine deaminse that catalyzed C•G to T•A base editing. APOBEC1 was tethered to dCas9 for the first generation of “base editors” termed BE1. As the uracil repair pathways impeded efficient base conversion, a uracil N-glycosylase inhibitor (UGI) was tethered to the C-terminus of dCas9 to form the BE2

architecture. The third generation BE3 restored the catalytic activity at H840 to create a Cas9 nickase. The nick of the G-containing strand increases templated repair of the U-containing strand and increases base editing efficiency. Further enhancements of the system have also now created narrow window base editors that reduces the editing window from around five nucleotides to two. Furthermore, a BE4 architecture is also available that increases product purity by increasing the proportion of C to T base conversion events.

In addition to the cytosine base editor, an adenosine base editor (ABE) has also been developed by our lab. Since no known DNA adenosine deaminase exists, a bacterial selection was used to evolve the *E. coli* TadA tRNA. After seven rounds of selection, a number of ABE variants were found to be active in mammalian cells. Since their introduction, base editors have been used for a number of applications across many different contexts from agricultural gene editing to cellular recording.

Summary

As genome editing tools have progressed their programmability and ease of use has increased tremendously to the point where they have become commonplace amongst many research labs and have even progressed to clinical trials. Cas9, as an RNA-guided endonuclease, has revolutionized our ability to write to the genome, complementing the rapid increase in DNA sequencing that has revolutionized genomics. Challenges to general usage of the technology remains, chief of which is the PAM requirement that limits the sequence space that is targetable by Cas9. In the following chapter we describe the evolution of a broadened PAM *Sp*Cas9 that greatly increases the number of targetable sites by Cas9, is widely compatible with different DNA effectors, and is also more specific than wild-type *Sp*Cas9.

Chapter 2:

Phase-assisted continuous evolution of Cas9 and characterization in bacterial cells

Copyright Disclosure

Portions of this chapter appear in or are adapted from the following publications:

Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., Liu, D. R. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* (2018). doi:10.1038/nature26155.

Attributions

The contributors to this work include Johnny H. Hu, Liwei Chen, Ning Sun, Christina M. Zeina, and David R. Liu. J.H.H. designed the research, performed PACE, characterized variants in bacteria. L.C. assisted with cloning and PACE. N.S. optimized Cas9 PACE. C.Z. assisted with cloning and PACE. D.R.L designed and supervised the research.

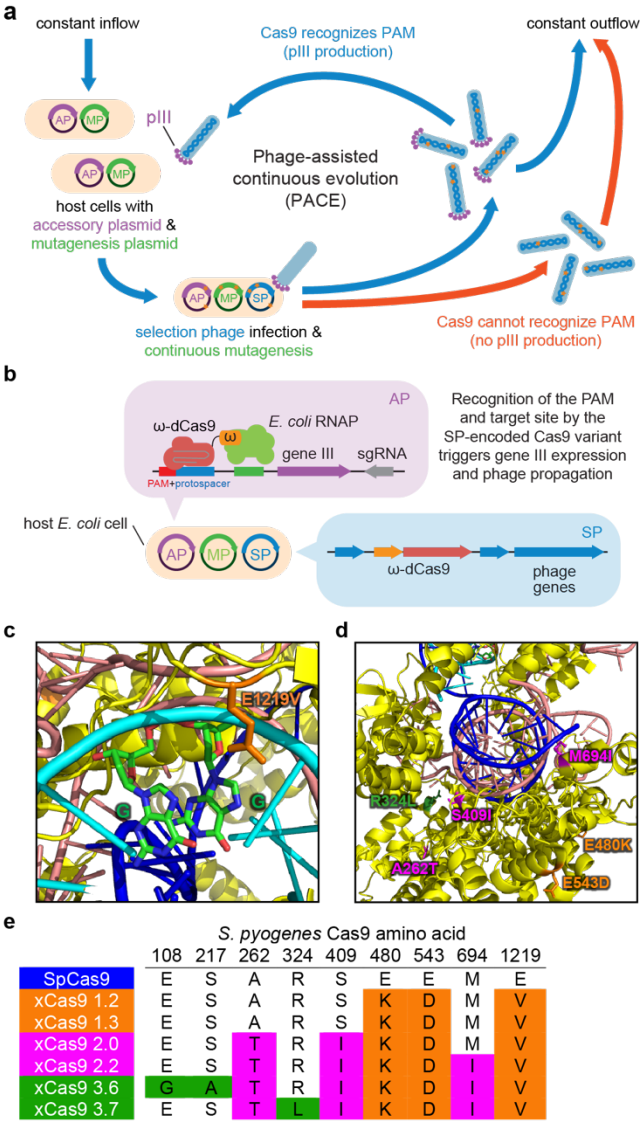
Introduction

As noted in the previous chapter, the NGG PAM requirement of canonical SpCas9, which occurs on average only once in every ~16 randomly chosen genomic loci, greatly limits the targeting scope of Cas9 especially for applications that require precise Cas9 positioning, such as base editing, which requires a PAM ~15±2 nucleotides from the target base[77,78], and some forms of homology-directed repair, which are most efficient when DNA cleavage occurs ~10-20 base pairs away from a desired alteration[79,80]. Here we used phage-assisted continuous evolution (PACE) to rapidly generate Cas9 variants that accept an expanded range of PAM sequences. During PACE, host *E. coli* cells continuously dilute an evolving population of bacteriophages (selection phage, SP). Since dilution occurs faster than cell division but slower than phage replication, only the SP, and not the host cells, can accumulate mutations[88]. Each SP carries a gene to be evolved instead of a phage gene (gene III) that is required for the production of infectious progeny phage. SP containing desired gene variants trigger host-cell gene III expression from the accessory plasmid (AP) and the production of infectious SP that propagate the desired variants. Phage encoding inactive variants do not generate infectious progeny and

are rapidly diluted out of the culture vessel (Fig. 1a). As phage replication can occur in as little as 10 minutes, PACE enables hundreds of generations of directed evolution to occur per week without researcher intervention[73,88-92].

Figure 1 | Phage-assisted continuous evolution (PACE) of Cas9 variants with broadened PAM compatibility. a, PACE takes place in a fixed-volume "lagoon" that is continuously diluted with fresh host *E. coli* cells. Upon infection, each selection phage (SP) that encodes a Cas9 variant capable of binding the target PAM and protospacer on the accessory plasmid (AP) induces expression of gene III, resulting in infectious progeny phage that propagate the active Cas9 variant in subsequent host cells. **b,** Anatomy of a phage-infected host cell during PACE. The host cell carries the AP, which links Cas9 target DNA binding to phage propagation, and the mutagenesis plasmid (MP), which elevates mutagenesis during PACE. **c, d,** The crystal structure of SpCas9 with the location of xCas9 mutations shown. **e,** Genotypes of some evolved xCas9 variants, colored by evolution stage. See Supplementary Table 5 for 95 xCas9 variant genotypes.

Figure 1 (Continued)



Results

Evolution of Cas9 variants with expanded PAM compatibility

To link Cas9 DNA recognition to phage propagation during PACE, we developed a bacterial one-hybrid selection[73,93,94] in which the SP encodes a catalytically dead SpCas9 (dCas9) fused to the ω subunit of bacterial RNA polymerase. When this fusion binds an AP-encoded sgRNA and a PAM and protospacer upstream of gene III in the AP, RNA polymerase recruitment causes gene III expression and phage propagation (Fig. 1b). We envisioned installing a library of all 64 possible NNN PAM sequences at the target protospacer in the AP, so that SP encoding Cas9 variants with broader PAM compatibility would replicate in a larger fraction of host cells and thus experience a fitness advantage.

We optimized the relationship between Cas9 DNA binding and gene expression (Fig. 2a-c). These studies revealed that (*i*) a fusion of the orientation N–ω–dCas9–C, (*ii*) a simple Ala–Ala fusion linker, and (*iii*) placement of the protospacer on the reverse complement strand 45 bp upstream of the -35 box together resulted in the strongest guide RNA-dependent gene expression activation of 13-fold (Fig. 2a-c). Together, these results establish a linkage between Cas9 DNA binding activity and gene expression in a selection system suitable for PACE.

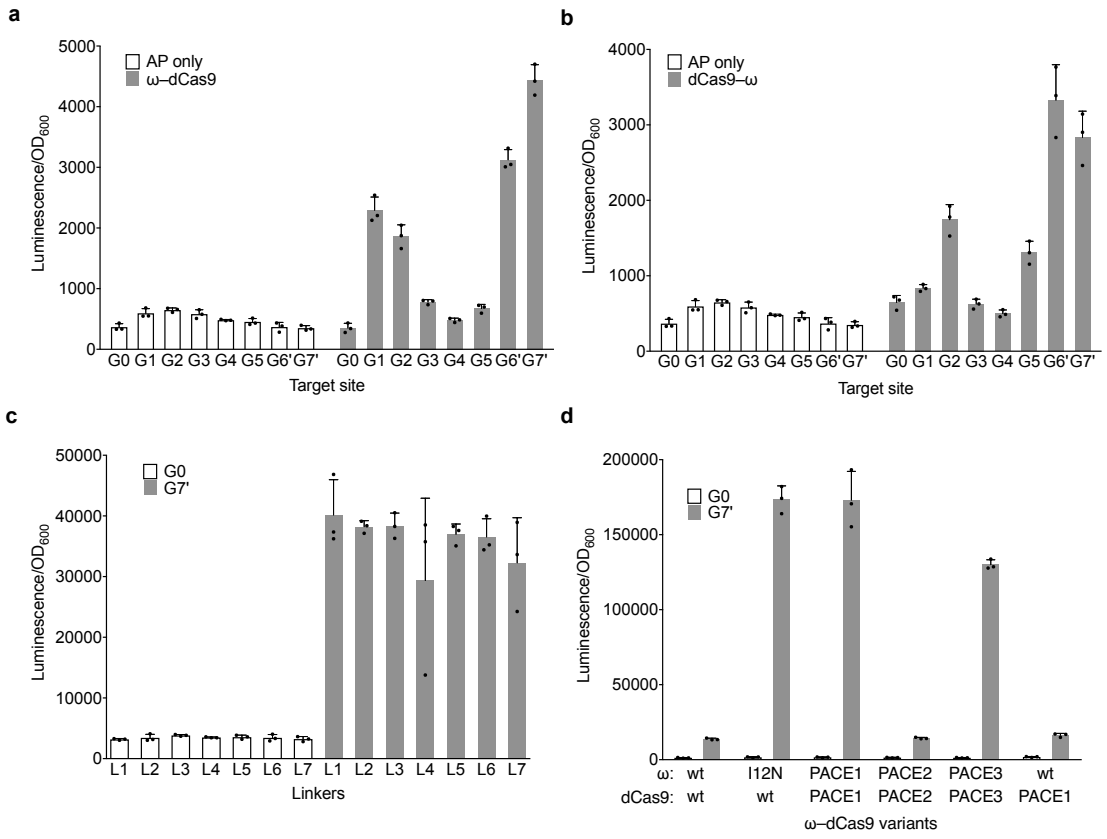
Using this selection, we first allowed an SP encoding the ω–dCas9 fusion to self-optimize on host cells containing an AP with a canonical NGG PAM, resulting in enrichment of an I12N mutation in the ω subunit. Adding this single mutation to ω–dCas9 boosted activation from 13-fold to over 100-fold (Fig. 2d), representing early-stage optimization of ω–dCas9 during PACE prior to evolution for broadened PAM compatibility.

To evolve Cas9 variants with expanded PAM compatibility we generated three AP libraries, each containing a different sgRNA (Supplementary Table 4) and corresponding protospacer upstream of an NNN PAM library, where N is an equimolar mixture of all four DNA bases. This design imposes selection pressure to recognize many different PAM sequences, as well as to

maintain compatibility with different target DNA sequences. All three AP libraries were introduced into host *E. coli* cells harboring the mutagenesis plasmid MP6[88]. The resulting host cells were incubated overnight with SP containing ω(I12N)–dCas9. This phage-assisted non-continuous evolution (PANCE) system[90,91] preferentially replicates Cas9 variants that bind a greater variety of PAM sequences, similar to PACE, but with lower stringency since there is no outflow of phage.

Figure 2 | Optimization of Cas9 PACE. Luciferase expression in *E. coli was* used as a proxy of gene III expression during efforts to link Cas9 binding to gene expression for PACE. **a, b,** Seven guide RNAs targeting the luciferase reporter (G1-G7', see Supplementary Table 1), as well as a scrambled guide RNA negative control (G0) were tested without dCas9 (white bars) and with ω–dCas9 (a) or dCas9–ω (b) fusions (grey bars). **c,** Tests of seven different linkers between ω and dCas9. See Supplementary Table 2 for linker sequences. **d,** Evolution of ω–dCas9 on an NGG PAM site in PACE yielded variants (PACE1, PACE2, and PACE3) that were tested in comparison with canonical ("wt") ω–dCas9, ω tethered to PACE1 dCas9, and the I12N ω mutant tethered to canonical dCas9. Values and error bars reflect the mean and s.d. of n=3 biologically independent samples.

Figure 2 (Continued)



After 24 days of serial overnight propagation and 1:1000 dilution, we isolated five Cas9 clones for sequencing and characterization (xCas9 1.0-1.4). Notable recurring mutations include E480K, E543D, and E1219V (Fig. 1e and Supplementary Table 5). E1219 is close in the SpCas9 crystal structure to R1333 and R1335 (Fig. 1c), two residues known to play a critical role in PAM recognition[36]. The mixture of phage from the final PANCE pool were further evolved for 72 h in PACE on host cells containing the same AP libraries harboring NNN PAM sequences. Among individual Cas9 clones emerging from PACE (xCas9 2.0-2.6), E480K, E543D, and E1219V were present in all sequenced phage, along with additional mutations seen in multiple clones, including A262T, K294R, S409I, and M694I (Fig. 1e and Supplementary Table 5). Finally, the resulting phage were continuously evolved in PACE for an additional 72 h on host cells containing three protospacer-sgRNA pairs and HHH PAM libraries, where H is A, C, or T, to favor Cas9 variants with activity on non-NGG PAMs.

PAM depletion assay to characterize the evolved variants

Fourteen resulting evolved Cas9 variants (xCas9 3.0-3.13) containing consensus mutations (Fig. 1e and Supplementary Table 5) emerged from two apparent evolution trajectories. While all phage shared E480K, E543D, and E1219V core mutations, xCas9 3.0-3.5 also all contained K294R and Q1256K mutations while xCas9 3.6-3.13 all contained A262T, S409I, and M694I mutations, with some of the latter also containing R324L. The Cas9 crystal structure predicts that R324L, S409I, and M694I lie near the DNA-sgRNA interface (Fig. 1d) and could play a role in mediating DNA sequence recognition and the switching of Cas9 from the open to the closed conformation upon target recognition[43,95]. Since the entire Cas9 gene was subject to mutagenesis during PACE, the mechanism of xCas9 may differ from that of engineered Cas9 variants that primarily mutated DNA-contacting residues[74,85,96].

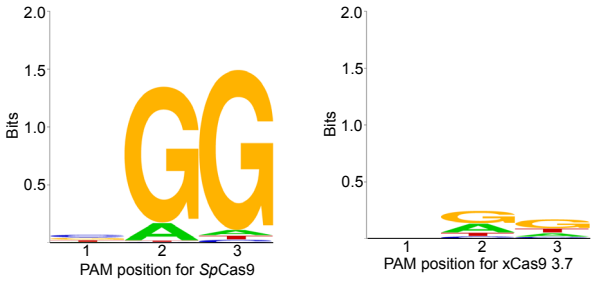
We characterized evolved xCas9 variants in several contexts. We first restored the catalytic residues Asp 10 and His 840 to test if xCas9 nucleases can cleave DNA even though

they were evolved only for DNA binding. The xCas9 3.0-3.13 clones were tested in a PAM depletion assay[74,85] in which they were given the opportunity to cleave a library of plasmids containing a protospacer and all possible NNN PAM sequences in an antibiotic resistance gene in bacterial cells. Plasmid cleavage results in the loss of spectinomycin resistance. This PAM depletion assay revealed that xCas9 3.0-3.3 and 3.5-3.9 cleave DNA site with NG, NNG, GAA, GAT, and CAA PAMs (Fig. 3). The clone with the highest PAM depletion score, xCas9 3.7, depleted NG, NNG, GAA, GAT, and CAA PAMs ≥ 100-fold compared to the starting library, while xCas9 3.6 showed the second highest average PAM depletion score.

Figure 3 | PAM profiling of xCas9 variants. a, A plasmid library containing a protospacer with all possible NNN PAM sequences and a spectinomycin resistance gene was electroporated into *E. coli* along with a plasmid expressing SpCas9 or the xCas9 variant shown in separate experiments. PAMs that are cleaved are depleted from the library when plated on media containing spectinomycin. HTS of the library before versus after selection enables quantification of the change in library composition, resulting in a sequence logo[97] for the PAM preference of SpCas9 (left) and xCas9 3.7 (right). **b, c,** PAM depletion scores of Cas9 variants from spectinomycin selection in *E. coli*, calculated as described previously[70], with 1.0 representing complete cleavage of that PAM sequence. Scores for NGN, NNG, GAA, GAT, and CAA are shown in (b) while the rest of the PAM sequences are shown in (c).

Figure 3 (Continued)

a



b

	xCas9 3.0	xCas9 3.1	xCas9 3.2	xCas9 3.3	xCas9 3.4	xCas9 3.5	xCas9 3.6	xCas9 3.7	xCas9 3.8	xCas9 3.9	xCas9 3.10	xCas9 3.11	xCas9 3.12	xCas9 3.13	*Sp*Cas9
AAG	0.93	0.94	0.97	0.97	0.00	0.96	0.98	0.99	0.96	0.95	0.94	0.91	0.96	0.96	0.97
ACG	0.90	0.97	0.98	0.99	0.00	0.97	0.96	0.99	0.91	0.85	0.95	0.26	0.17	0.00	0.00
AGA	0.92	0.96	0.97	0.98	0.00	0.97	0.97	0.99	0.91	0.92	0.93	0.00	0.29	0.43	0.33
AGC	0.86	0.83	0.91	0.88	0.06	0.87	0.92	0.82	0.01	0.91	0.87	0.34	0.42	0.57	0.53
AGG	0.92	0.99	0.97	0.99	0.03	0.99	0.99	1.00	0.88	0.92	0.95	0.28	0.97	0.99	1.00
AGT	0.92	0.90	0.94	0.94	0.00	0.91	0.95	0.96	0.91	0.85	0.92	-0.02	0.63	0.00	0.00
ATG	0.91	0.93	0.96	0.98	0.00	0.95	0.97	0.99	0.94	0.92	0.79	0.55	0.59	0.43	0.42
CAA	0.35	0.50	0.85	0.80	0.00	0.71	0.84	0.78	0.89	0.83	0.49	0.31	-0.01	0.00	0.00
CAG	0.93	0.91	0.96	0.96	0.03	0.95	0.97	0.99	0.71	0.49	0.30	0.96	0.89	0.93	0.94
CCG	0.69	0.61	0.85	0.85	0.00	0.78	0.88	0.81	0.76	0.80	0.81	0.24	0.00	0.00	-0.01
CGA	0.89	0.98	0.98	0.99	0.00	0.98	0.96	0.99	0.77	0.87	0.81	0.25	0.25	0.50	0.31
CGC	0.93	0.88	0.94	0.93	0.00	0.92	0.95	0.91	0.92	0.91	0.81	0.75	0.81	0.72	0.78
CGG	0.94	1.00	0.98	1.00	0.03	0.99	0.99	1.00	0.83	0.82	0.96	0.97	0.95	0.97	1.00
CGT	0.92	0.96	0.97	0.98	0.00	0.95	0.96	0.99	0.94	0.86	0.94	-0.01	0.28	-0.01	0.00
CTG	0.81	0.70	0.85	0.88	0.00	0.83	0.91	0.76	0.82	0.89	0.85	0.26	0.29	0.52	0.29
GAA	0.90	0.86	0.93	0.91	0.00	0.89	0.91	0.85	0.89	0.83	0.78	0.49	0.00	0.04	-0.01
GAG	0.92	0.98	0.98	0.99	0.00	0.99	0.99	1.00	0.96	0.95	0.90	0.94	0.95	0.98	0.99
GAT	0.87	0.83	0.91	0.89	0.00	0.85	0.91	0.90	0.81	0.90	0.09	0.00	0.08	-0.01	-0.01
GCG	0.92	0.92	0.95	0.98	0.00	0.97	0.97	0.99	0.94	0.92	0.89	0.35	0.47	0.53	0.54
GGA	0.90	0.99	0.97	0.99	0.06	0.99	0.99	0.99	0.67	0.95	0.97	0.96	0.94	0.98	0.98
GGC	0.92	0.90	0.94	0.93	0.00	0.93	0.96	0.94	0.83	0.73	0.91	0.87	0.95	0.74	0.95
GGG	0.95	1.00	0.99	1.00	0.00	0.99	0.99	1.00	0.92	0.90	0.93	0.94	0.96	0.99	1.00
GGT	0.91	0.96	0.97	0.99	0.04	0.98	0.98	1.00	0.91	0.93	0.66	0.92	0.94	0.94	0.96
GTG	0.91	0.94	0.96	0.99	0.02	0.97	0.98	1.00	0.84	0.94	0.96	0.89	0.89	0.62	0.88
TAG	0.91	0.90	0.94	0.95	0.00	0.93	0.96	0.97	0.79	0.94	0.09	0.88	0.88	0.89	0.94
TCG	0.90	0.88	0.93	0.94	0.01	0.90	0.94	0.97	0.30	0.76	0.89	-0.01	-0.07	-0.03	-0.02
TGA	0.90	0.98	0.98	0.99	0.10	0.99	0.98	0.99	0.96	0.82	0.94	0.44	0.68	0.73	0.75
TGC	0.92	0.89	0.93	0.91	0.00	0.91	0.93	0.87	0.85	0.85	0.91	0.57	0.84	0.75	0.77
TGG	0.92	1.00	0.98	1.00	0.00	0.99	0.99	1.00	0.95	0.91	0.91	0.90	0.95	0.95	1.00
TGT	0.91	0.95	0.97	0.97	0.01	0.95	0.95	0.99	0.90	0.89	0.95	0.00	-0.01	-0.01	-0.01
TTG	0.93	0.85	0.90	0.93	0.00	0.87	0.93	0.92	0.88	0.84	0.89	-0.01	-0.01	0.00	0.00

PAM depletion score

c

	xCas9 3.0	xCas9 3.1	xCas9 3.2	xCas9 3.3	xCas9 3.4	xCas9 3.5	xCas9 3.6	xCas9 3.7	xCas9 3.8	xCas9 3.9	xCas9 3.10	xCas9 3.11	xCas9 3.12	xCas9 3.13	*Sp*Cas9
ATA	-0.02	-0.02	-0.03	-0.03	0.10	-0.03	-0.03	-0.03	-0.01	-0.01	-0.02	0.05	-0.01	0.00	0.00
ATC	-0.02	-0.02	-0.03	-0.03	0.03	-0.03	-0.03	-0.04	-0.03	-0.03	-0.06	-0.01	0.00	0.00	-0.01
ATT	-0.02	-0.02	-0.02	-0.02	0.00	-0.02	-0.02	-0.03	-0.01	-0.04	-0.01	-0.01	0.00	-0.01	-0.01
CAC	-0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.68	0.39	0.23	0.00	-0.01	0.00	0.00	0.00
CAT	0.77	0.83	0.94	0.87	0.87	0.84	0.88	0.87	0.95	0.98	0.95	0.94	0.94	0.00	0.85
CCA	-0.02	-0.02	-0.03	-0.03	0.00	-0.03	-0.03	-0.03	-0.03	-0.04	-0.01	0.00	-0.02	-0.01	-0.01
CCC	-0.01	-0.02	-0.02	-0.02	0.01	-0.02	-0.02	-0.02	-0.03	-0.01	-0.01	0.12	-0.01	-0.01	-0.01
CCT	-0.01	-0.01	-0.01	-0.01	0.00	-0.01	-0.01	-0.02	-0.01	0.04	-0.02	0.17	0.00	-0.01	0.00
CTA	-0.01	-0.01	-0.01	-0.01	0.04	-0.01	-0.01	-0.01	0.00	-0.01	-0.01	0.50	0.32	0.07	0.00
CTC	-0.02	-0.02	-0.02	-0.02	0.00	-0.02	-0.02	-0.03	-0.02	-0.04	-0.01	-0.01	-0.02	0.00	-0.01
CTT	-0.01	-0.01	-0.01	-0.01	0.06	-0.02	-0.02	0.00	-0.01	-0.01	-0.01	-0.01	0.00	0.00	0.00
GAC	0.00	0.26	0.79	0.66	0.00	0.58	0.81	0.76	0.74	0.75	0.64	-0.02	0.00	-0.01	-0.01
GCA	-0.01	-0.01	-0.01	-0.01	0.00	-0.01	-0.01	-0.01	-0.01	0.00	0.00	0.31	-0.02	0.00	-0.01
GCC	-0.01	-0.01	-0.01	-0.01	0.04	-0.01	-0.02	-0.01	-0.03	-0.01	-0.01	-0.01	0.72	-0.01	0.00
GCT	-0.01	-0.01	-0.02	-0.02	0.00	-0.02	-0.02	-0.01	-0.02	0.00	-0.04	-0.01	0.11	0.00	-0.01
GTA	-0.01	-0.01	-0.01	-0.01	0.05	-0.01	0.04	0.36	0.28	0.08	0.20	-0.01	0.72	0.09	0.00
GTC	-0.01	-0.01	-0.01	-0.01	0.00	-0.01	-0.01	-0.01	0.00	-0.03	-0.01	0.00	0.00	0.00	0.00
GTT	-0.01	-0.01	-0.02	-0.02	0.00	-0.01	-0.02	0.00	-0.01	-0.02	0.00	-0.01	0.38	0.00	0.00
TAA	-0.01	-0.01	0.39	0.03	0.02	0.07	0.43	0.37	0.68	0.55	0.47	0.20	0.10	-0.01	-0.01
TAC	-0.05	-0.05	-0.06	-0.06	0.02	-0.06	-0.06	-0.08	-0.04	-0.06	-0.03	0.00	-0.03	-0.01	-0.02
TAT	-0.02	-0.01	0.22	0.00	0.00	0.00	0.19	0.62	0.39	0.53	0.02	-0.01	-0.01	-0.01	-0.01
TCA	-0.02	-0.02	-0.02	-0.02	0.00	-0.02	-0.02	-0.02	-0.03	-0.01	-0.01	-0.02	-0.01	-0.01	-0.01
TCC	-0.02	-0.02	-0.03	-0.02	0.00	-0.03	-0.03	-0.02	-0.02	-0.05	-0.03	0.25	0.00	0.00	-0.01
TCT	-0.02	-0.02	-0.02	-0.02	0.01	-0.02	-0.03	-0.02	-0.02	-0.02	-0.03	-0.01	0.00	0.00	-0.01
TTA	-0.01	-0.01	-0.02	-0.02	0.17	-0.02	-0.02	-0.04	-0.01	-0.04	-0.01	0.00	0.28	-0.01	0.00
TTC	-0.01	-0.01	-0.02	-0.01	0.02	-0.02	-0.02	-0.03	-0.01	0.08	-0.01	0.00	0.09	-0.01	0.00
TTT	-0.01	-0.01	-0.01	-0.01	0.03	-0.01	-0.01	-0.03	-0.02	-0.01	-0.01	-0.01	-0.01	0.00	0.00

PAM depletion score

Discussion

PACE provides a generalizable system for the evolution of Cas9 without the need for structural information and without the need to prioritize certain regions of the protein over others. Previous Cas9 evolutions have relied on targeted mutagenesis and rational design using the crystal structure, but these studies have usually focused on residues or domains known to interact with the PAM sequence. By allowing PACE to mutate the entire protein, we allowed the evolution to access unexplored region of *Sp*Cas9. Using this selection provides both opportunities and challenges, however, and indeed our early PAM evolutions trying to shift the PAM to a NAG PAM did not encounter success and will be revisited in future evolutions as that is still an inaccessible PAM for *Sp*Cas9.

A single mutation, the I12N mutation, in the ω subunit that arose in the initial PACE run that boosted the activation from ~10-fold to ~100-fold was also crucial in carrying out the later PACE experiments as the higher dynamic range led to much great selection pressure for viable variants. We have seen similar albeit different mutations arise in the ω subunit for other one-hybrid PACE selections so this result was not unexpected. While the mechanism for the increase in activation has not been elucidated we speculate that there is a delicate balance between the need for the ω subunit to recruit the other subunits of the RNA polymerase and the disengagement of those subunits to activate transcription. The fact that this particular mutation is different from previous evolutions is most likely due to the different orientation that the ω is forced to adopt being tethered to dCas9.

The choice of the selection system also presented both advantages and disadvantages. As the one-hybrid system has been verified for previous PACE systems and is generally highly sensitive it provided a viable starting point. Different schemes also considered included a cutting selection that would require Cas9 to cleave pIII-neg[98], dominant negative pIII that poisons phage propagation. Such a system would have to be fine-tuned to give enough of a response after Cas9

cutting to lead to phage propagation while still limiting phage production when Cas9 is not active. One of the largest risk of the binding selection used is that mutations could have arisen to inactivate the Cas9 nuclease domains. While reversion analysis could have decoupled the activity being selected for from any inactivating mutations, it is possible that the two effects would have difficult or impossible to disentangle. As shown by the results of our PAM depletion assay, however, such an effort was not required as most of the xCas9 variants were active for cutting activity.

While the characterization experiments in bacteria were encouraging, we also needed to show broadened PAM in mammalian cell assays to have a greater impact beyond just bacterial genome modifications. While we take for granted now that Cas9 is active universally in nearly all cell types, there are many examples of other systems that have not ported over to mammalian cells such as the λ red recombineering system. Even for Cas9 systems, many systems that have been characterized to work in bacterial cells show little to no activity in mammalian cells[99]. Thus, verifying xCas9 activity in mammalian cells was crucial for making broader claims about its applicability.

Methods

General methods and cloning. DNA sequences used in this work are listed in the Supplementary Information. PCR was performed using Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs) or Phusion U Green Multiplex PCR Master Mix (Thermo Fisher Scientific). PACE plasmids and phage were constructed by USER cloning or Gibson cloning (New England Biolabs). Cas9 genes and plasmid backbones for PACE were obtained from previously reported plasmids[73,90] available from Addgene. Plasmids encoding dxCas9–VPR[100], xCas9 nucleases, xCas9–BE3[77], xCas9–BE4[101], and xCas9–ABE[78] were constructed by replacing SpCas9 using Gibson cloning. Plasmids for sgRNA expression were constructed using one-piece blunt-end ligation of a PCR product containing a variable 20-nt sequence

corresponding to the desired sgRNA targeted site. Primers and templates used in the synthesis of all sgRNA plasmids used in this work are listed in Supplementary Tables 8, 9, 11, 12, 15, and 16. All guide RNAs used in this study were transcribed from a U6 promoter, and natively started with a G at the 5' end to avoid possible losses in activity caused by a mismatched 5' guide terminus. PCR was performed using Q5 Hot Start High-Fidelity Polymerase (New England Biolabs) with the phosphorylated primers and the plasmid pFYF1320 (EGFP sgRNA expression plasmid) as a template according to the manufacturer's instructions. PCR products were analyzed by agarose gel electrophoresis, the band of the expected molecular weight was cut out, and the DNA was extracted using a Zymoclean Gel DNA Recovery Kit (Zymo Research) and ligated using T4 DNA Ligase (New England Biolabs) according to the manufacturer's instructions. DNA vector amplification was carried out using Mach1 competent cells (Thermo Fisher Scientific). All mammalian ABE constructs, sgRNA plasmids and bacterial constructs were transformed and stored as glycerol stocks at -80 °C in Mach1 $T1^R$ Competent Cells (Thermo Fisher Scientific), which are *recA*$^-$. Molecular biology grade Hyclone water (GE Healthcare Life Sciences) was used in all assays and PCR reactions. All vectors used in evolution experiments and mammalian cell assays were purified using ZympPURE Plasmid Midiprep (Zymo Research Corporation), which includes endotoxin removal. Antibiotics were purchased from Gold Biotechnology.

PAM depletion assay. Electrocompetent NEB 10-beta cells (New England Biolabs) were electroporated with two plasmids. The first plasmid expresses Cas9 (inducible by anhydrotetracycline, ATc), the sgRNA (inducible by arabinose), and a spectinomycin resistance gene. The second plasmid contains the target protospacer and a kanamycin resistance gene. After incubation with SOC outgrowth medium (New England Biolabs) for 1 hour the bacteria were plated on agar plates containing both spectinomycin and kanamycin along with ATc and arabinose inducers. After incubating overnight, the bacterial cells on the agar plates were scraped and the plasmids extracted using the ZymoPURE Plasmid Midiprep Kit (Zymo Research). The

resulting post-selection DNA included all of the protospacer plasmids not cleaved by Cas9. The same region around the protospacer in the pre-selection library and the post-selection DNA was then amplified separately using NEBNext High-Fidelity PCR Polymerase (New England Biolabs) with flanking HTS primer pairs listed in the Supplementary Table 7. Illumina barcoding PCR reaction was assembled with NEBNext High-Fidelity PCR Polymerase. PCR products were purified by electrophoresis with a 2% agarose gel using a QIAquick Gel Extraction Kit, eluting with 15 µL of water. DNA concentration was quantified with the KAPA Library Quantification Kit-Illumina (KAPA Biosystems) and sequenced on an Illumina MiSeq instrument according to the manufacturer's protocols.

Chapter 3:

Characterization of xCas9 for transcriptional activation, DNA cleavage, base editing, and targeting specificity in mammalian cells

Copyright Disclosure

Portions of this chapter appear in or are adapted from the following publications:

Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., Liu, D. R. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* (2018). doi:10.1038/nature26155.

Attributions

The contributors to this work include Johnny H. Hu, Shannon M. Miller, Maarten H. Geurts, Weixin Tang, Xue Gao, Zhi Lin, and David R. Liu. J.H.H. designed the research, conducted human cell experiments, analyzed data, and performed off-target analysis. S.M.M. performed human cell experiments, analyzed data. M.H.G. performed human cell experiments and data analysis. W.T. performed human cell experiments and cloning. X.G. assisted with off-target analysis. H.A.R. assisted with indel and base editing analyses. Z.L. performed human cell experiments. D.R.L designed and supervised the research and wrote the manuscript.

Introduction

To test if mutations evolved during PACE in bacteria are compatible with xCas9 function in mammalian cells, we characterized xCas9 variants for their activity and PAM compatibility in human cells in four contexts: transcriptional activation, genomic DNA cutting, cytidine base editing, and adenine base editing. These assays were chosen to showcase a broad range of xCas9 activities in contexts that would be relevant for other researchers conducting mammalian genome modifications with Cas9. When possible, endogenous genomic sites were targeted and high-throughput sequencing is used to read out the changes to the genome directly. The different Cas9 effector constructs and off-target sequencing methods are reviewed in the Chapter I introduction.

Results

Transcriptional activators by xCas9 variants

To test for transcriptional activation, catalytically dead versions of xCas9 were fused to the transcriptional activator VP64–p65–Rta (dxCas9–VPR)[100]. Plasmids encoding dxCas9–VPR, a GFP reporter downstream of a target protospacer, and a corresponding sgRNA were co-transfected into HEK293T cells[100]. Target gene transcriptional activation was measured by cellular GFP fluorescence after three days. Three different target site PAM sets were tested: a single reporter with an NGG PAM, a reporter library containing a NNN PAM library, and a reporter library containing a NNNNN PAM library. In addition, two different protospacer sequences, reporter 1 and reporter 2, were tested with their corresponding sgRNAs.

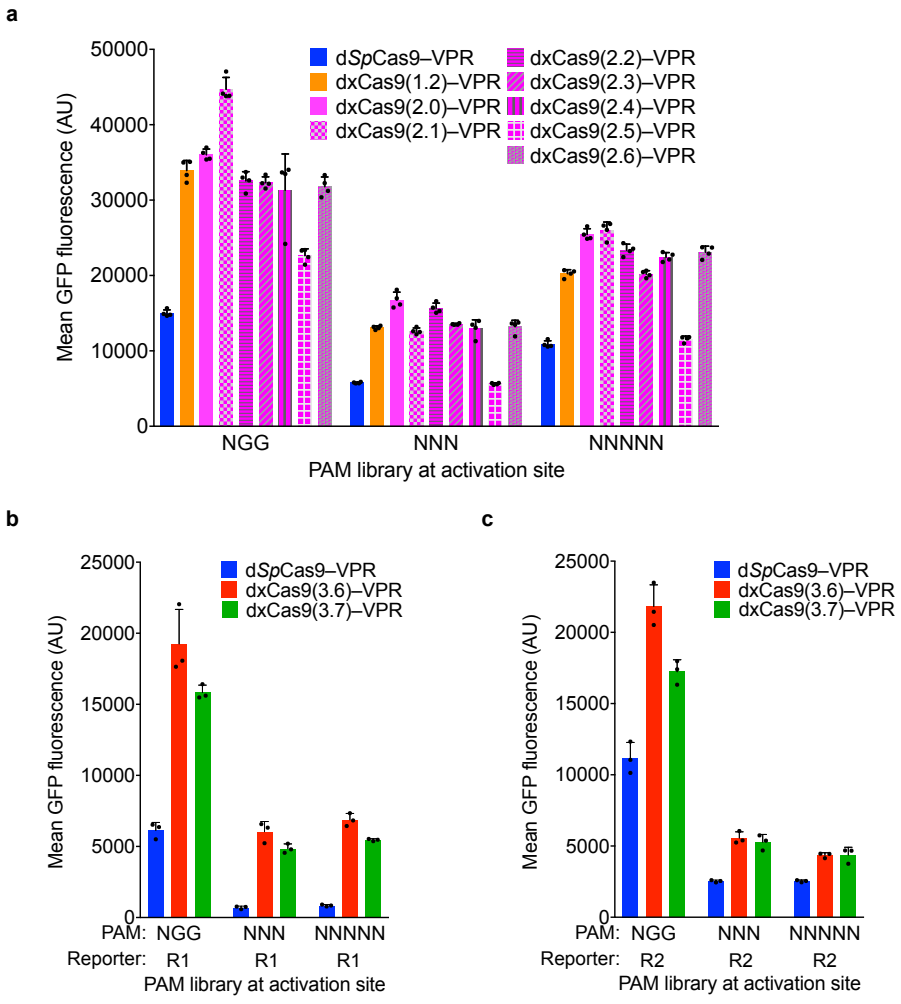


Figure 4 (Continued)

Figure 4 | Transcriptional activation of reporter site PAM libraries with xCas9. Transcriptional activation by dSpCas9–VPR and dxCas9–VPR variants, transfected as plasmids, on GFP reporter plasmids containing different PAM sites in HEK293T cells. **a,** Earlier generations of xCas9 variants were tested on the R1 site with the NGG, NNN, or NNNNN PAM libraries. Values and error bars reflect the mean and s.d. of n=3 biologically independent samples. **b, c,** Transcriptional activators dxCas9(3.6)–VPR and dxCas9(3.7)–VPR were tested on two different protospacer reporters, R1 in (a) and R2 in (b), containing adjacent NGG, NNN, or NNNNN PAM libraries. Values and error bars reflect the mean and s.d. of n=3 biologically independent samples.

Most early-stage xCas9 variants outperformed wild-type SpCas9 on sites with NGG PAMs, as well as with NNN and NNNNN PAM libraries (Fig. 4a). For reporter 1 and reporter 2, respectively, xCas9 3.7 achieved 2.8- and 1.5-fold higher mean fluorescence for the NGG PAM, 7.9- and 2.1-fold higher mean fluorescence for the NNN PAM library, and 5.2- and 1.7-fold higher mean fluorescence for the NNNNN PAM library compared with SpCas9 (Fig. 4b, c). The similar performances on NNN and NNNNN PAM libraries suggest that xCas9 3.7 did not evolve strong sequence preferences at nucleotides immediately downstream of the NNN PAM. The xCas9 3.6 variant showed similar results to those of xCas9 3.7 in this assay (Fig. 4b, c).

To dissect activity on individual PAM sequences, we tested dxCas9–VPR transcriptional activators on individual target sites containing each of the 64 possible three-nucleotide PAM sequences (Fig. 4a and Extended Data Figs A1-3) in HEK293T cells. Consistent with the PAM library results, dxCas9(3.7)–VPR showed broad improvements in transcriptional activity relative to dSpCas9–VPR across many individual non-NGG PAMs. Transcriptional activation by dxCas9(3.7)–VPR at sites containing NGT, NGA, NGC, NNG, GAA, and GAT PAMs averaged 56–91% of the average activity of dxCas9(3.7)–VPR on the four NGG PAM sites (Fig. 4a and Extended Data Fig. A2). The performance of xCas9 3.6 transcriptional activators was similar to

that of xCas9 3.7 (Extended Data Fig. A3). To test if broadened transcriptional activation by xCas9 is limited to reporter plasmids that may be only partially chromatinized, we also tested the ability of dxCas9(3.7)–VPR to activate transcription of six endogenous genomic loci in human cells and observed 3.3-fold average improved activation of the two NGG PAM sites and 39-fold average improved activation among the three NGN PAM sites, but no improvement on the tested NNG site, relative to dSpCas9–VPR (Extended Data Fig. A2e). Overall, these results establish that xCas9 is compatible with the dCas9–VPR architecture and can serve as potent transcriptional activators in human cells at a substantially expanded set of PAMs. Based on their strong performance in the PAM depletion assay and as transcriptional activators, we chose xCas9 3.7 and xCas9 3.6 for further characterization.

DNA cutting by xCas9 variants

To test targeted genomic DNA cleavage in human cells we expressed xCas9 3.7 and 3.6 nuclease in a HEK293T cell line with a genomically integrated GFP gene and measured the loss of GFP fluorescence reflecting DNA cleavage and indel-mediated disruption of the target site. Two NGG PAM sites, three NGT sites, and one NGC, NGA, GAT, NCG, and NTG site are present in the *GFP* sequence; all were tested with SpCas9, xCas9 3.7, and xCas9 3.6. For NGG PAM sites, xCas9 3.7 modestly outperformed SpCas9, resulting in 46±2.0% compared to 33±3.4% GFP disruption, respectively (mean ± SD of three independent replicates, Fig. 5b). For all tested non-NGG PAM sites, xCas9 3.7 showed substantially higher (1.6- to 5.1-fold) average apparent cleavage activity than SpCas9 (Fig. 5b). At all tested sites, xCas9 3.6 showed comparable or slightly lower GFP disruption percentages than xCas9 3.7 (Extended Data Fig. A4a). Neither xCas9 3.7 nor 3.6 increased GFP loss relative to SpCas9 for either NNG PAM site tested, suggesting that the transcriptional activation observed at some NNG PAM sites by dxCas9–VPR activators, and the strong NNG PAM signal in the bacterial PAM depletion assay, do not

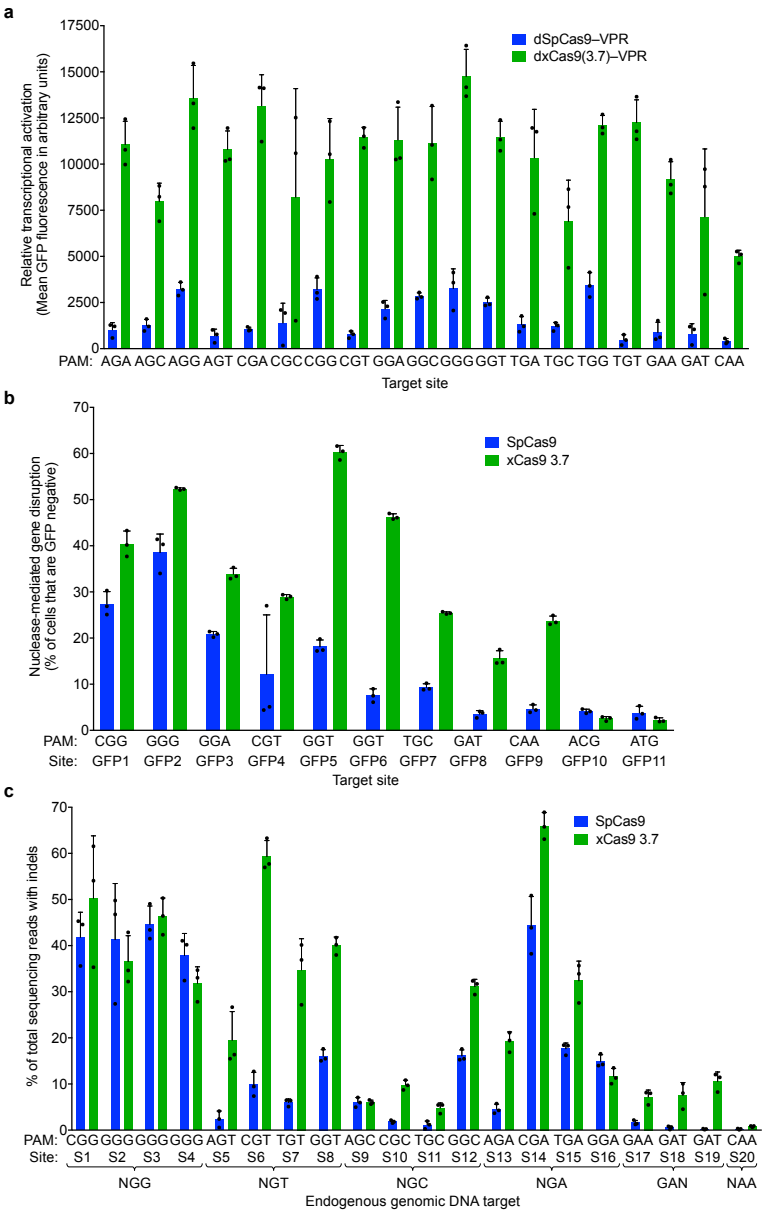
necessarily translate to DNA cleavage at all NNG sites in mammalian cells. Target site dependence is also well-known for SpCas9[102].

To further characterize DNA cleavage in human cells by xCas9 variants, we targeted endogenous genomic sites in HEK293T cells and measured indel formation by high-throughput sequencing (HTS). Twenty endogenous sites were tested covering four NGG PAM sites, all 12 possible NGT, NGC, and NGA PAMs, and GAT, GAA, and CAA PAMs (Fig. 5c). On the four NGG PAM sites tested, xCas9 3.7 showed comparable activity to SpCas9, averaging 41±6.4% indels compared to 41±6.0% for SpCas9 (Fig. 5c). All four NGT PAM sites showed much higher indel formation with xCas9 3.7 than with SpCas9, averaging 38±4.1% indels compared to 8.6±1.5% for SpCas9, a 4.5-fold increase. The four NGA PAM sites averaged 32±2.4% indels for xCas9 and 20±2.7% for SpCas9, a 1.6-fold increase, consistent with previous reports that NGA can serve as a secondary PAM for SpCas9[103]. While indel frequencies at the endogenous NGC PAM sites were more variable, ranging from 4.8–31%, xCas9 3.7 averaged 13±0.90% indels compared to 6.3±0.77% for SpCas9, a 2.1-fold increase. Among the three GAA and GAT sites tested, SpCas9 showed virtually no activity, averaging 1.4±1.3% indel formation, while xCas9 3.7 averaged 7.2±2.8% indel formation, a 5.2-fold increase. The xCas9 3.6 variant showed comparable or slightly lower indel frequencies for all sites tested (Extended Data Fig. A4b). Negative control experiments lacking sgRNA plasmid resulted in no indels above background (Extended Data Fig. A5). Taken together, these results indicate that xCas9 3.7 nuclease mediates target gene disruption at NGG PAM sites with comparable efficiencies as wild-type SpCas9, but cleaves NG, GAA, and GAT PAM sites with substantially higher efficiencies than SpCas9. The greater PAM-dependent and protospacer-dependent variability of xCas9 nuclease-mediated gene disruption relative to transcriptional activation (Fig. 2 and Extended Data Figs A1-A3) may reflect more extensive requirements for DNA cleavage than DNA binding[43,64], or differences in the chromatin

state of plasmid (Fig. 5a and Extended Data Figs A1, A2a-A2d, A3) versus genomic targets (Fig. 5b, c, and Extended Data Figs A2e, A4)

Figure 5 | Transcriptional activation and genomic DNA cleavage by evolved xCas9 3.7 in human cells. a, Transcriptional activation by dSpCas9–VPR and dxCas9(3.7)–VPR targeting GFP reporter plasmids containing the same protospacer but different PAM sites in HEK293T cells. **b,** Genomic DNA cleavage in HEK293-GFP cells containing a genomically integrated GFP reporter gene by SpCas9 or xCas9 3.7. After 5 days, the cells were analyzed for loss of GFP fluorescence by flow cytometry. **c,** DNA cleavage of endogenous genomic DNA sites with NGG and non-NGG PAMs by SpCas9 and xCas9 3.7 in HEK293T cells. Indel rates were measured by HTS 5 days after plasmid transfection. Values and error bars reflect the mean and s.d. of n=3 biologically independent samples. Target sites are in Supplementary Tables 8 (R1 site), 11, and 12.

Figure 5 (Continued)



Base editing by xCas9 variants

Base editing is a newer genome editing approach that uses a catalytically impaired Cas9 fused to a natural or laboratory-evolved nucleobase deaminase enzyme and, in some cases, a DNA glycosylase inhibitor to directly convert a target C•G to T•A, or a target A•T to G•C, without introducing double-stranded DNA breaks or requiring homology-directed repair[77,78,104]. The suitability of a target site for base editing is highly dependent on the presence of a suitably positioned PAM, which must exist within a narrow window downstream of the target base pair (typically 15±2 nucleotides). The broad PAM compatibility of xCas9 variants thus has the potential to expand the DNA targeting scope of base editors (Fig. 6c).

To evaluate C•G-to-T•A base editing activity of xCas9 variants, we substituted SpCas9 with xCas9 3.7 and 3.6 in the third-generation (BE3) base editor architecture[77]. Both xCas9–BE3s and SpCas9–BE3 were separately transfected into mammalian cells to compare editing efficiency on all 20 sites tested above for endogenous genomic DNA cleavage (Fig. 6a). At the four NGG PAM target sites tested, xCas9(3.7)–BE3 averaged 37±10% C•G to T•A conversion, while SpCas9–BE3 averaged 28±5.2% (Fig. 6a). At NG, GAA, and GAT PAM sites tested, xCas9(3.7)–BE3 resulted in substantially improved base editing, averaging 24±5.4% editing at NGT PAM sites, an 9.5-fold increase over that of SpCas9; 16±3.5% editing at NGA sites, a 3.5-fold increase over SpCas9; 6.2±0.34% editing at NGC sites, a 13-fold increase over SpCas9; 10±0.75% editing on the GAA PAM site, a > 50-fold increase over SpCas9; and 12±1.5% editing on the GAT sites, a > 100-fold increase over SpCas9 (Fig. 6a). The base editing efficiencies of xCas9(3.6)–BE3 were comparable to, or slightly worse than, those of xCas9(3.7)–BE3 (Extended Data Fig. A4c). We also tested cytosine base editing at an additional 15 endogenous genomic sites within the *FANCF* gene and observed similar large improvements for xCas9(3.7)–BE3 over SpCas9–BE3 (Fig. 7a). Overall, these results indicate that xCas9 variants are compatible with the BE3 architecture, and enable cytidine base editing of target sites that cannot be accessed by SpCas9–BE3 or, with the

exception of NGA PAM sites[104], by any other previously reported base editors. We also tested xCas9 3.7 in the BE4 architecture designed to reduce undesired byproducts[101], and observed fewer indels and higher product purities, although with slightly lower editing efficiencies (Fig. 7b-d).

The recent development of an adenine base editor (ABE) enables programmable installation of A•T to G•C mutations[78]. No ABEs have been reported yet that can target non-NGG PAM sites, limiting its targeting scope. We replaced SpCas9 in ABE 7.10[78] with xCas9 3.7 and 3.6, and assayed the resulting xCas9–ABEs in HEK293T cells at the seven endogenous genomic sites tested above that contain an A in the targeting window of ABE (positions 4-8, counting the PAM as positions 21-23). At all seven of these sites, xCas9(3.7)–ABE resulted in higher base editing efficiencies than the original SpCas9–ABE (Fig. 7b). Average base editing efficiency at the NGG PAM site tested increased from 48±2.1% to 69±3.7%. At the GAT PAM site tested, xCas9(3.7)–ABE resulted in 16±1.5% base editing, while SpCas9–ABE yielded no detectable editing (≤ 0.1%), representing a > 100-fold increase. On the two NGC and three NGA sites tested, xCas9(3.7)–ABE averaged 21±2.5% and 43±1.5% base editing, respectively, while SpCas9–ABE averaged 7.0±1.3% and 22±1.2%, respectively. Base editing by xCas9(3.7)–ABE was comparable to or higher than that of xCas9(3.6)–ABE (Extended Data Fig. A4d). Collectively, these results establish that xCas9–ABE mediates adenine base editing at sites that cannot currently be accessed otherwise.

Figure 6 | Cytidine and adenine base editing by xCas9. a, The C•G to T•A conversion frequencies for 20 endogenous genomic loci in HEK293T cells at the most efficiently edited base 3 days after plasmid transfection are shown. **b,** The A•T to G•C conversion frequencies for seven endogenous genomic loci in HEK293T cells at the most efficiently edited base 5 days after plasmid transfection are shown. Values and error bars in (a) and (b) reflect the mean and s.d. of n=3 biologically independent samples. See Supplementary Table 14 for complete HTS results across the protospacer. **c,** Fraction of T•A to C•G pathogenic SNPs in ClinVar[105] that, in principle, can be corrected by SpCas9–BE3 (left), xCas9(3.7)–BE3 (middle), or xCas9(3.7)–BE3 + all BE3 variants reported to date (right). **d,** Fraction of G•C to A•T pathogenic SNPs in ClinVar that, in principle, can be corrected by SpCas9–ABE (left) or xCas9(3.7)–ABE (right).

Figure 6 (Continued)

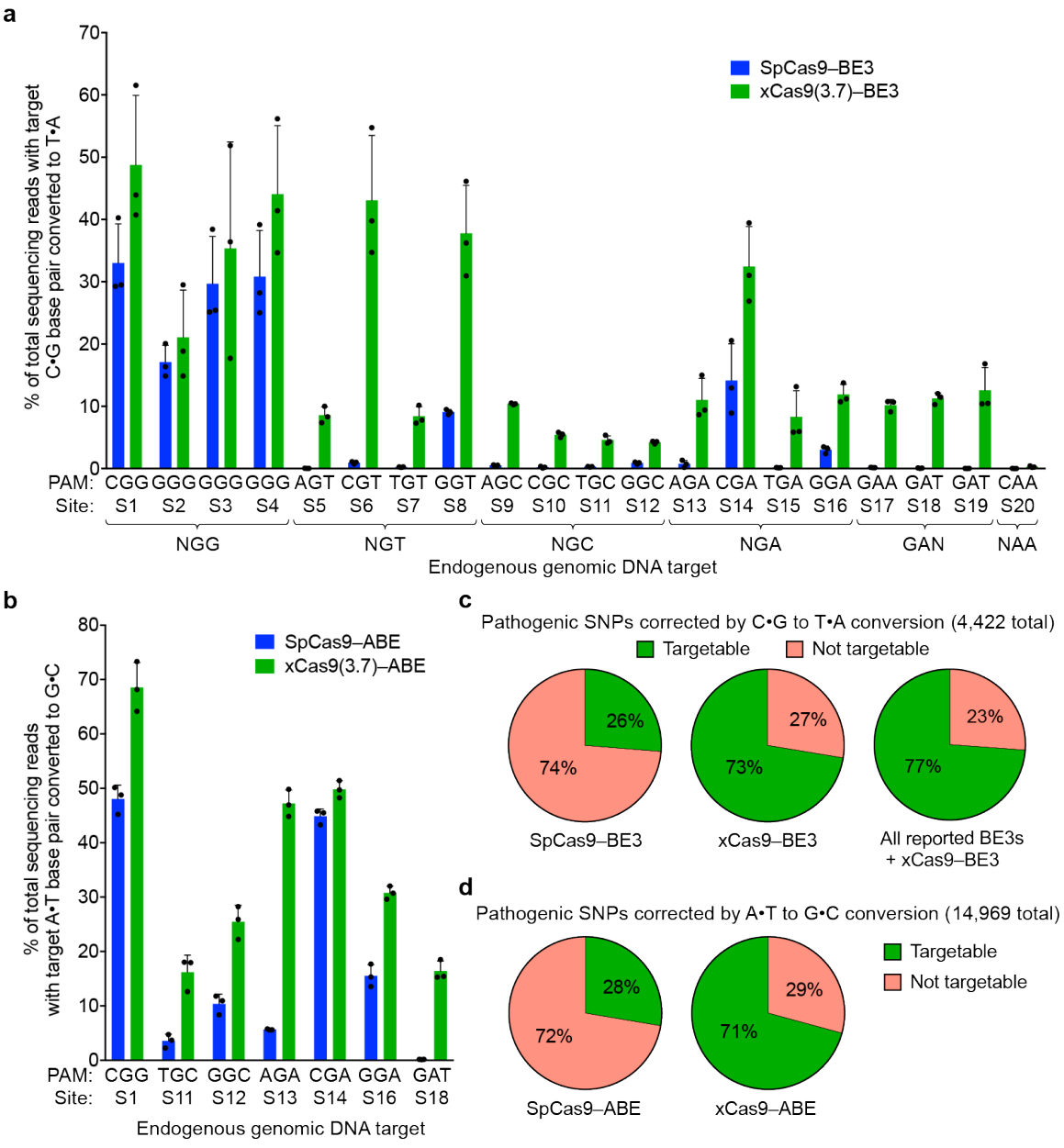
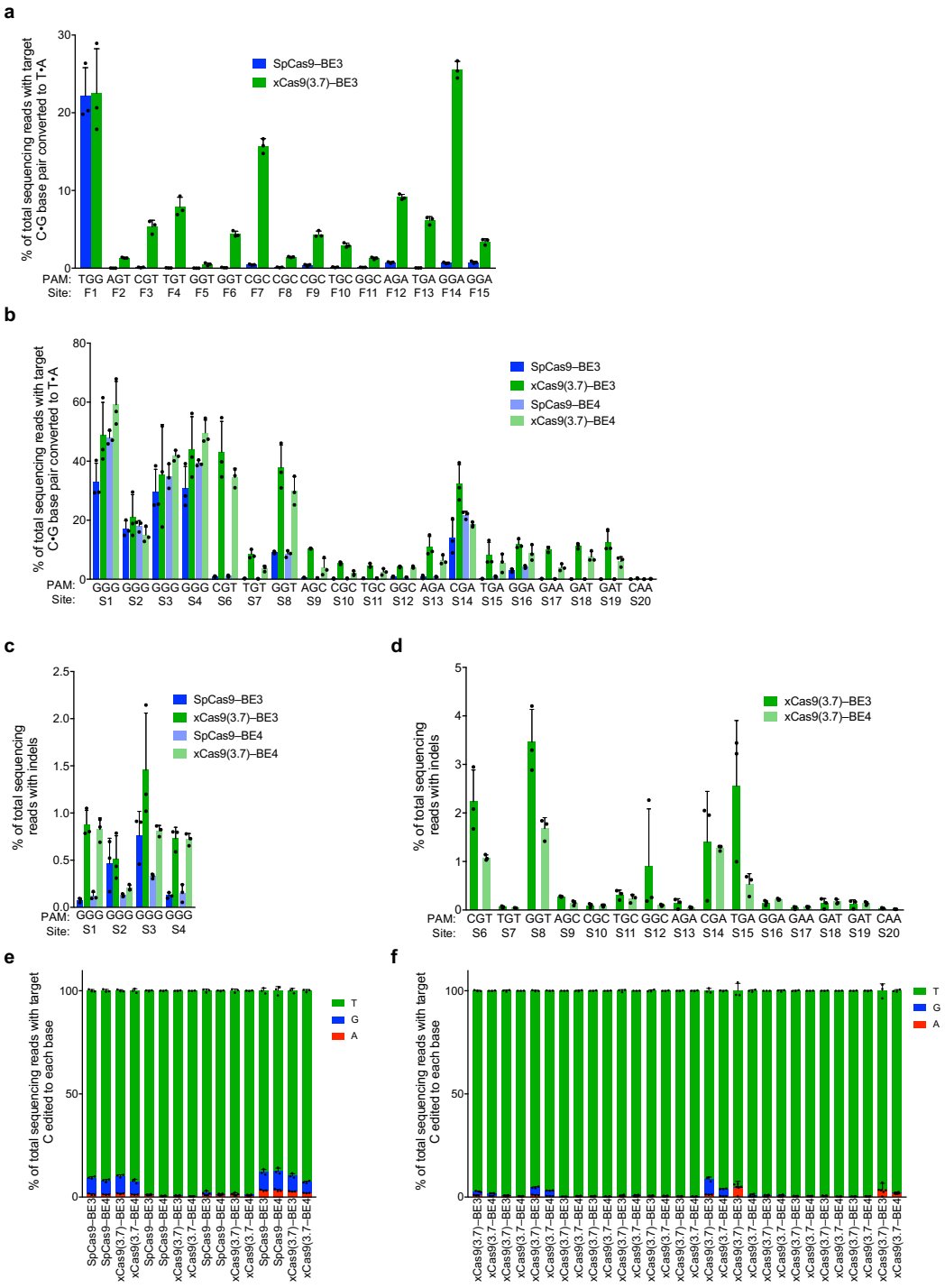


Figure 7 | Cytidine base editing at 15 additional genomic sites and xCas9 base editing with the BE4 architecture. a, Base editing by SpCas9–BE3 and xCas9(3.7)–BE3 at 15 sites within the *FANCF* gene in HEK293T cells. The C•G to T•A conversion frequency at the most efficiently edited base 3 days after plasmid transfection is shown. **b**, Test of xCas9 3.7 in the BE4 architecture[101] on the same sites tested in Fig. 3. The C•G to T•A conversion frequency in HEK293T cells at the most efficiently edited base 3 days after plasmid transfection is shown. **c, d,** Indel frequency following treatment with BE3 or BE4 variants targeting sites with NGG PAMs (c) and non-NGG PAMs (d). **e, f,** Product distribution among edited DNA sequence reads (reads in which the target C is mutated) following treatment with BE3 or BE4 variants targeting sites with NGG PAMs (e) and non-NGG PAMs (f). Since SpCas9 has minimal activity on non-NGG PAM sites, only xCas9(3.7)–BE3 and xCas9(3.7)–BE4 data is compared on non-NGG PAM sites. Values and error bars reflect the mean and s.d. of n=3 biologically independent samples. Target sites are in Supplementary Table 12.

Figure 7 (Continued)



xCas9 targeting specificity

In order to test the off-target activity of xCas9 variants, we performed GUIDE-seq, an unbiased genome-wide off-target analysis[64], on xCas9 3.7, xCas9 3.6, and SpCas9 in HEK293T and U2OS cells. Remarkably, for all five endogenous genomic NGG PAM sites tested in HEK293T cells and for both NGG PAM sites tested in U2OS cells, GUIDE-seq analysis revealed that xCas9 3.7 and 3.6 resulted in *much lower* off-target activity than SpCas9, as reflected by both the number of detected off-target sites, as well as the total modification frequency at each detected off-target site (Fig. 8 and Extended Data Fig. A6). For example, at the *EMX1* target site in HEK293T cells, SpCas9 showed 5,649 on-target reads and a total of 1,132 off-target reads, while xCas9 3.7 showed 6,874 on-target reads but zero off-target reads (Fig. 8b). In U2OS cells, the off-target:on-target ratio for the same site was 0.65 for SpCas9 with 6,328 on-target reads, and 0.015 for xCas9 3.7 with 22,539 on-target reads, representing a 43-fold reduction in off-target modification (Extended Data Fig. A6c). Likewise, for HEK sites 1, 2, and 3, xCas9 3.7 resulted in ≥ 100-fold lower off-target to on-target modification ratios (0.023, < 0.001, and 0.028, respectively) compared to those of SpCas9 (0.28, 0.14, and 0.33, respectively) (Fig. 8 and Extended Data Fig. A6). For the known highly promiscuous target sites HEK site 4 and *VEGFA*[64], the off-target:on-target ratios were 9.4 and 2.0, respectively, for SpCas9, but only 1.0 and 0.48 for xCas9 3.7, a 4.2- to 9.4-fold improvement (Extended Data Fig. A6). We observed these large improvements in DNA specificity for xCas9 3.7 and 3.6 even though, consistent with the above findings, xCas9 variants showed a much broader range of PAM sequences among detected off-target sites (Fig. 8 and Extended Data Fig. A6). These GUIDE-seq results were verified by HTS of many individual on-target and off-target sites from the genomic DNA of treated cells (Fig. 9). In addition, where base editing compatible sites exist, they were also verified to have low off-target editing for xCas9 (Fig. 10).

We also evaluated the off-target DNA specificity of xCas9 3.7 at two non-NGG PAM (GAA and CGT) sites with both SpCas9 and xCas9 3.7 in HEK293T cells. As expected, GUIDE-seq did not yield any on-target reads for SpCas9 at either of these non-NGG PAM sites, while xCas9 3.7 had 3,627 on-target reads for the GAA PAM site and 3,055 on-target reads for the CGT PAM site (Fig. 8e, f). Importantly, neither site was accompanied by any detected off-target GUIDE-seq reads for xCas9 3.7, although potential off-target reads were detected for SpCas9 (Fig. 8e, f). Collectively, these findings reveal that xCas9 3.7 and 3.6 offer greatly reduced off-target activity compared with wild-type SpCas9, despite their broader PAM compatibility. These results also establish that there is no necessary trade-off between Cas9-mediated editing efficiency, PAM compatibility, and DNA specificity, a key finding as natural and engineered genome editing agents advance into widespread applications including human clinical trials.

Figure 8 | Off-target editing analysis of xCas9. a, GUIDE-seq[64] was performed on SpCas9 and xCas9 3.7 nucleases. Six endogenous genomic sites with NGG PAMs were tested in HEK293T or U2OS cells. The percentage of off-target sequencing reads relative to total reads are shown**. b-d,** All GUIDE-seq on-target reads (*) and off-target reads for three sites in HEK293T cells are shown for SpCas9 and xCas9 3.7. See Extended Data Fig. A6 for additional GUIDE-seq results. **e, f,** GUIDE-seq results for two endogenous genomic non-NGG PAM sites in HEK293T cells. No on-target GUIDE-seq reads (*) were detected for SpCas9 at either of these non-NGG sites. See Extended Data Fig. A6 for GUIDE-seq analysis of xCas9 3.6 and Fig. 9 for HTS validation of GUIDE-seq results. Target sequences are listed in Supplementary Table 15.

Figure 8 (Continued)

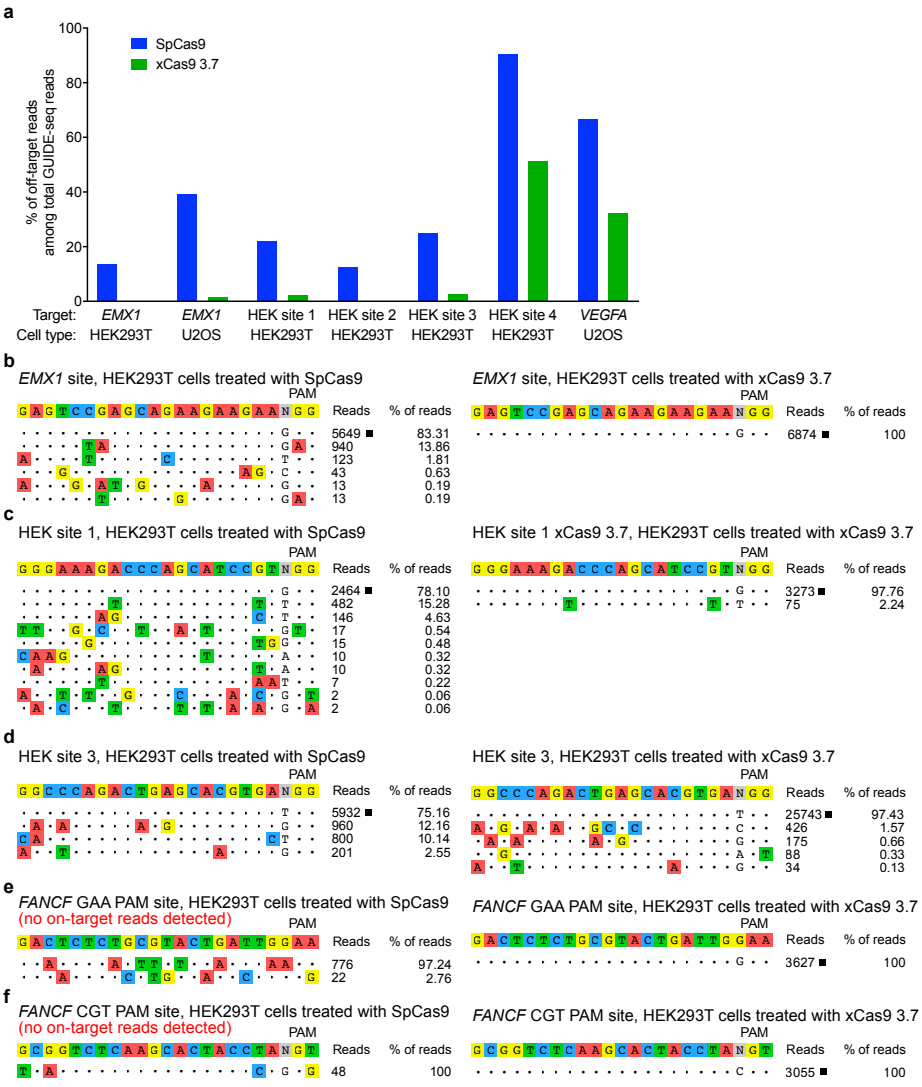


Figure 9 | Validation by high-throughput sequencing of GUIDE-seq results. The most frequent off-target sites identified by GUIDE-seq were verified by HTS of genomic DNA following treatment of HEK293T cells with SpCas9 or xCas9 3.7 (**a, b**), or following treatment with SpCas9 or xCas9 3.6 (**c**, **d**). New sites with non-NGG PAMs that were identified as xCas9 off-target sites were also analyzed in (a-d). Values and error bars reflect the mean and s.d. of n=3 biologically independent samples. Target sequences are in Supplementary Table 16.

Figure 9 (Continued)

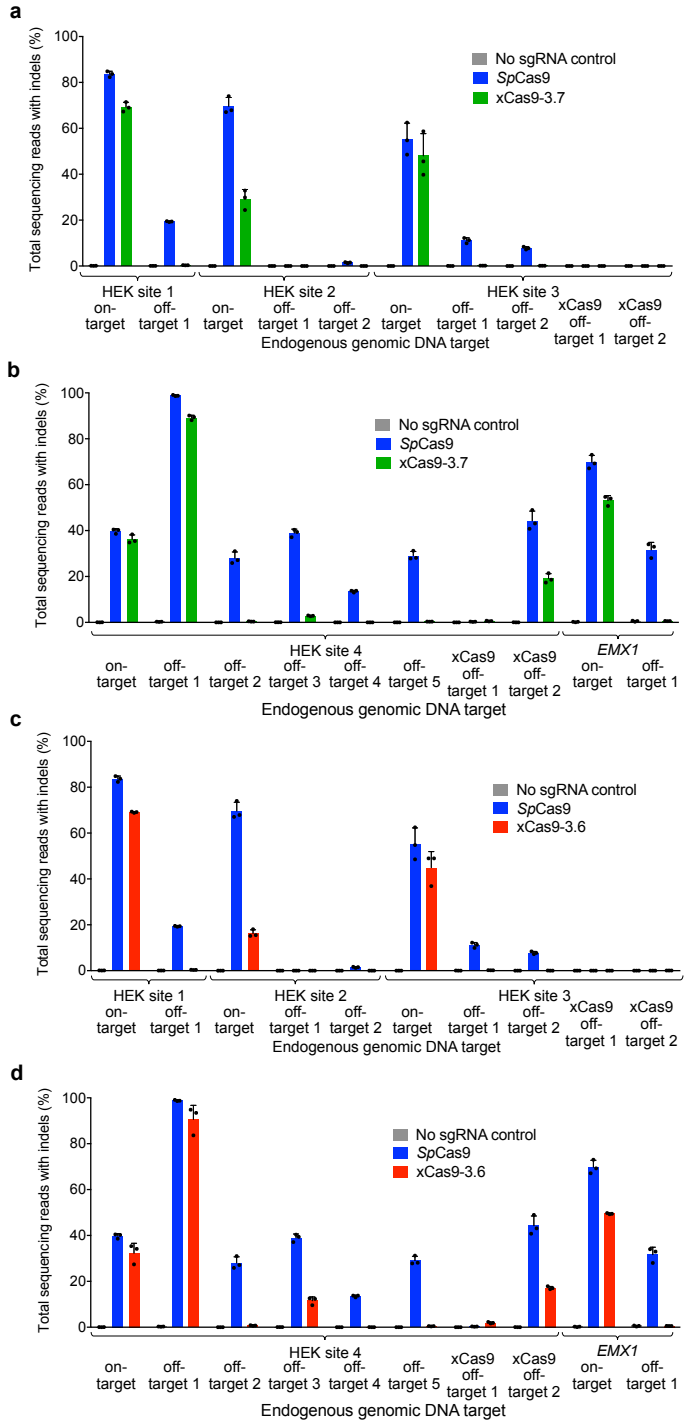
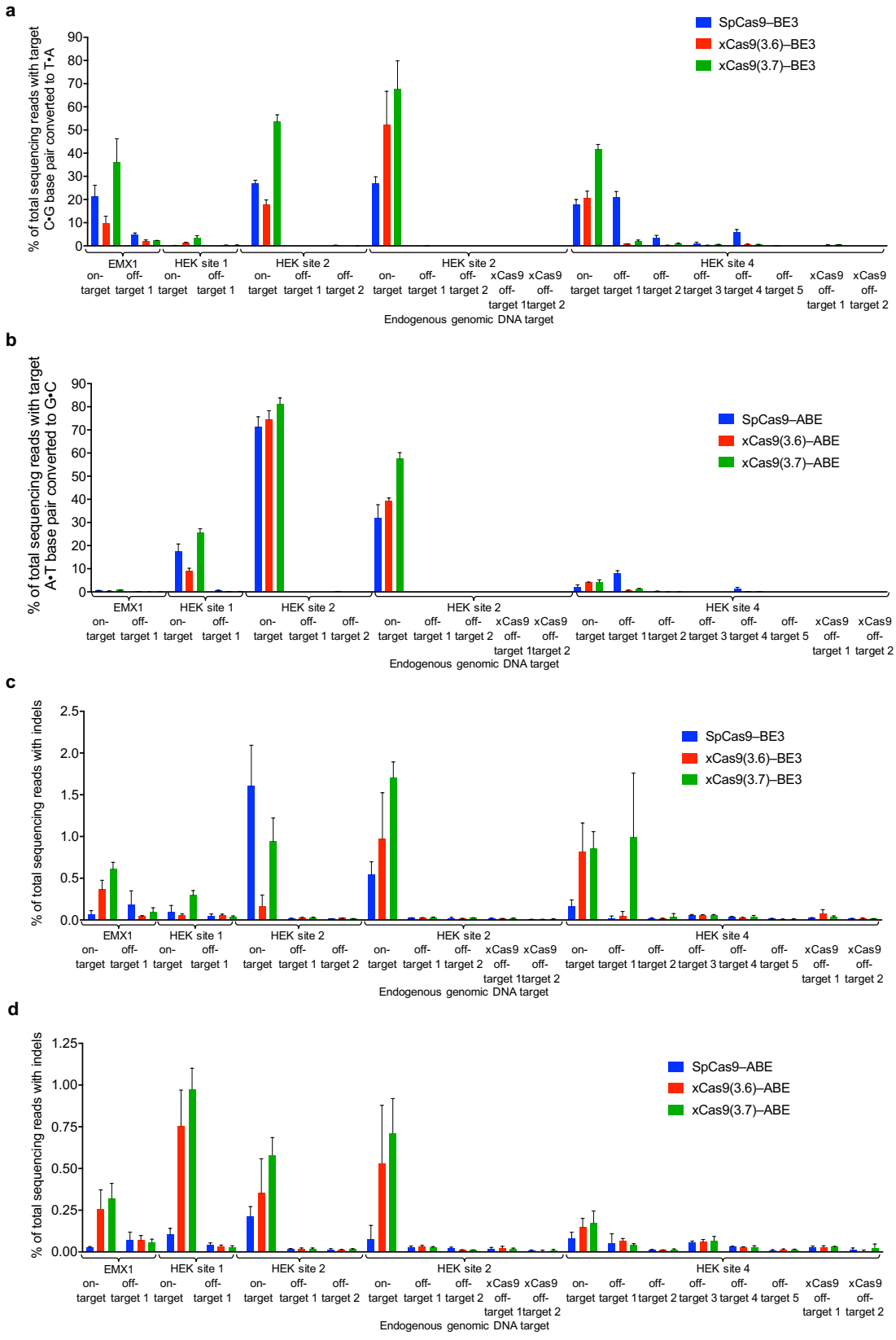


Figure 10 | Base editing of off-target sites. Off-target sites identified by GUIDE-seq were tested for base editing with either *Sp*Cas9–BE3, xCas9(3.6)–BE3, or xCas9(3.6)–BE4 **(a,c)** or *Sp*Cas9–ABE, xCas9(3.6)–ABE, or xCas9(3.6)–ABE **(b,d).** The percentage of converted bases is shown **(a,b)** as well as the indel percentages **(c,d).** Values and error bars reflect the mean and s.d. of n=3 biologically independent samples. Target sequences are in Supplementary Table 16.

Figure 10 (Continued)



Discussion

These results, together with the success of multiple independent efforts to create high-fidelity Cas9 variants[72,95,106], suggest that the DNA promiscuity of wild-type Cas9, which likely evolved to impede viral evasion, can readily be overcome by protein engineering or evolution. That xCas9 exhibits much higher DNA specificity than SpCas9 even though it was not explicitly selected for this property suggests that the off-target activity of wild-type SpCas9 may lie at a narrow fitness peak suitable for defending the much smaller bacterial genome but not optimal for genome editing in mammalian cells. These observations are therefore consistent with a model in which native SpCas9 is poised to become more specific, rather than less specific, upon mutation.

To our knowledge, the targeting scope of xCas9 is the broadest among Cas9 variants known to function efficiently in mammalian cells. Evolved xCas9 variants are also the first to offer improvements in targeting scope, activity, and DNA specificity in a single entity relative to wild-type SpCas9. While the efficacy of xCas9 on non-NGG PAMs varies based on application—here, transcriptional activation, DNA cleavage, or base editing—and on target site, the ability to access some NG, GAA, and GAT PAM sequences greatly expands the breadth of targets available for site-sensitive genome editing applications. Indeed, compared to SpCas9–BE3, xCas9(3.7)–BE3 increases the percentage of 4,422 pathogenic SNPs in the ClinVar database[105] that in principle could be targeted by C•G to T•A base editing from 26% to 73% (Fig. 6c). Likewise, xCas9(3.7)–ABE increases the fraction of 14,969 pathogenic ClinVar SNPs that could be targeted by A•T to G•C base editing from 28% to 71% (Fig. 6d). We anticipate that xCas9 and additional CRISPR enzyme variants with broadened PAM compatibilities may also expand the scope of other forms of nucleic acid editing, CRISPR-based screens, and epigenetic modification.

Methods

Cell culture. HEK293T (ATCC CRL-3216) and U2OS (ATCC-HTB-96) were maintained in Dulbecco's Modified Eagle's Medium plus GlutaMax (Thermo Fisher Scientific) supplemented with 10% (v/v) fetal bovine serum (FBS) at 37 °C with 5% CO_2.

Transfections. HEK293T cells were seeded on 48-well poly-D-lysine-coated BioCoat plates (Corning) and transfected at approximately 85% confluency. For genomic DNA cutting or base editing, 750 ng of Cas9 or BE3 and 250 ng of sgRNA expression plasmids were transfected using 1.5 µL of Lipofectamine 2000 (Thermo Fisher Scientific) per well according to the manufacturer's protocol. For GFP activation, 200 ng of dCas9–VPR plasmid, 50 ng of sgRNA expression plasmid, 60 ng of GFP reporter plasmid, and 30 ng of iRFP expression plasmid were transfected using 1.5 µL of Lipofectamine 2000 (Thermo Fisher Scientific) per well according to the manufacturer's protocol. Endogenous gene activation was done similarly but with 200 ng of dCas9–VPR plasmid and 50 ng of sgRNA expression plasmid only.

GFP transcriptional activation assay. Transfected HEK293T cells were trypsinized and resuspended in Dulbecco's Modified Eagle's Medium plus GlutaMax (Thermo Fisher Scientific) supplemented with 10% (v/v) fetal bovine serum (FBS). The cells were kept on ice and flow cytometry was performed using a LSRFortessa from BD Biosciences. Events were gated for iRFP positive cells to analyze transfected cell. The percentage of GFP-positive cells and the intensity of GFP fluorescence from each cell was collected.

RNA expression quantification for endogenous transcriptional activation assay. RNA was extracted from HEK293T cells using the Quick-RNA Plus Kit (Zymo Research). cDNA was synthesized using the iScript cDNA Synthesis Kit (Bio-Rad) and qPCR was performed on a Bio-Rad CFX96 Real-Time PCR Detection System using Q5 Polymerase (NEB) and SYBR Green (Lonza). Primers for qPCR are listed in Supplementary Table 10.

High-throughput DNA sequencing of genomic DNA samples. Transfected cells were harvested after 3 days (BE3 and BE4) or 5 days (DNA cutting and ABE). Media was removed and cells were washed with 1x PBS solution (Thermo Fisher Scientific). Genomic DNA was extracted by addition of 100 µL freshly prepared lysis buffer (10 mM Tris-HCl, pH 7.0, 0.05% SDS, 25 µg/mL Proteinase K (Thermo Fisher Scientific) directly into each well of the tissue culture plate. The plate was incubated at 37 °C for 1 h. The genomic DNA mixture was transferred to a 96-well PCR plate and incubated at 80 °C for 15 min to denature enzymes. Genomic regions of interest were amplified by PCR with flanking HTS primer pairs listed in the Supplementary Table 13. Each 25 µL PCR reaction was assembled with Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific) according to manufacturer's instructions using 1.0 µM of each forward and reverse primer and 1 µL of genomic DNA extract. PCR reactions were carried out as follows: 95°C for 3 min, then 30 cycles of [98°C for 30 s, 60°C for 20 s, and 72°C for 1 min], followed by a final 72 °C extension for 5 min. PCR products were verified by comparison to DNA standards (1-kb Plus DNA Ladder) on a 2% agarose gel with ethidium bromide. Each 25-µL Illumina barcoding PCR reaction was assembled with Phusion DNA polymerase according to manufacturer's instructions using 0.5 µM of each unique forward and reverse Illumina barcoding primer pair and 1 µL of unpurified genomic amplification PCR reaction mixture. The barcoding reactions were carried out as follows: 98°C for 2 min, then 8 cycles of [98°C for 12 s, 61°C for 25 s, and 72°C for 30 s], followed by a final 72 °C extension for 1.5 min. PCR products were purified by electrophoresis with a 2% agarose gel using a QIAquick Gel Extraction Kit, eluting with 15 µL of water. DNA concentration was determined with the KAPA Library Quantification Kit-Illumina (KAPA Biosystems) and sequenced on an Illumina MiSeq instrument according to the manufacturer's protocols. Analysis was carried out using previously published Matlab code[78], provided in Supplementary Notes 1 and 2.

Analysis of human disease-associated mutations in ClinVar database. Bioinformatic analysis of the ClinVar database was carried out in a manner similar to previously described analysis[104]. The code is provided in Supplementary Note 3.

GUIDE-Seq. HEK293T cells were transfected with 750 ng of the Cas9 plasmid, 250 ng of the gRNA plasmid, and 20 pmol of GUIDE-seq dsODN. U2OS cells were transfected with 750 ng of the Cas9 plasmid, 250 ng of the gRNA plasmid, and 100 pmol of GUIDE-seq dsODN. For both cell types, 20 µL of Solution SE (Lonza) was used along with a Lonza Nucleofector 4-D. Program CM-137 was used for HEK293T cells while program DN-100 was used for U2OS cells. 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 S220 focused ultrasonicator as previously described[64]. End repair, dA-tailing, adapter ligation, tag-specific PCR1, and tag-specific PCR2 were carried out using the primers and methods previously described[64]. DNA concentration was quantified with the KAPA Library Quantification Kit-Illumina (KAPA Biosystems) and sequenced on an Illumina MiSeq instrument according to the manufacturer's protocols. Analysis was carried out using the previously published Python code[64].

Data availability. High-throughput sequencing data have been deposited in the NCBI Sequence Read Archive database under accession code SRP130166. Plasmids encoding xCas9 3.7 and 3.6 transcriptional activators, nucleases, BE3, BE4, and ABE variants will be available from Addgene.

Chapter 4:

Conclusion and future directions

Conclusion

The evolution of xCas9 has given us a new tool that should be widely useful for genome editing as it greatly expands the sequence space that can be targeted for genome editing. We showed xCas9 compatibility in bacterial and mammalian cells for NG, GAA, and GAT PAMs for Cas9 DNA cleavage, base editing, and transcriptional activation. The ability to broaden the PAM with improvements in target specificity was a surprise to not only us but many in the field as more sites are accessible to xCas9 than *Sp*Cas9. Indeed, the GUIDE–seq data matches this observation as there is the appearance of some non-NGG sites in the off–target reads, but the overall specificity is still greatly improved for all of the targets that we tested due to the lower number of reads at the existing off-target sites. We believe that xCas9 will be enabling for targeted genome editing where precise positioning of the Cas9 is required such as for base editing a single nucleotide polymorphism. Future base editors or other technologies based on Cas9 such as recombinases could also be compatible with xCas9.

While experiments to clarify the role of each mutation are still ongoing, the crystal structure of *Sp*Cas9[107] suggests that the Glu to Val mutation at residue 1219 plays an important role in PAM discrimination. The conversion of an acidic amino acid to a non-polar amino acid could relax the spatial constraints on the DNA in the PAM groove (Fig. 1c), especially due to the competing negative charges of the DNA backbone with the glutamic acid at residue 1219. The smaller size of the valine sidechain could also play a role in allowing for a less restrictive placement of the PAM bases in the binding pocket. This study as well as others[74,85] that have reprogrammed the PAM have given us evidence that even a small number of amino acid changes can have a drastic effect on the ability of *Sp*Cas9 to recognize different PAM sequences.

The structural basis for xCas9's improved specificity is harder to tease apart without further experiments but other studies have suggested various mechanisms that may also be applicable to xCas9[72,95,106]. Given the additional properties of xCas9 to the other higher specificity

Cas9s, however, there is the possibility that xCas9 specificity could be mediated by other means than for the reporter higher specificity *Sp*Cas9 variants. One possible reason for xCas9's ability to better discriminate between on-target and off-target sequences is that *Sp*Cas9 has excess binding energy to most on-target sequences. By decreasing the binding energy for xCas9, we have decreased Cas9's ability to bind to off-target sequences without taking a noticeable decrease in on-target activity due to the headroom that exists for *Sp*Cas9.

In addition to the creating a new Cas9 variant, we have also validated the applicability of PACE for the evolution of Cas9 for mammalian genome editing. The technique provides a powerful platform to allow us to shift the sequence space accepted by Cas9 and could also be applied to select for other desired properties as described below.

Future Directions

The evolution of xCas9 raises many interesting basic biology questions about Cas9 in general that still needs to be explored. Most simply, reversion analysis will start dissecting the contribution of each mutation. Further mechanistic experiments such as *in vitro* cleavage and binding kinetics will measure differences between the properties of *Sp*Cas9 and xCas9 while the crystal structure of xCas9 will reveal the residues that play a role in both canonical and non-canonical PAM recognition. How different the conformations are that xCas9 adopts to recognize NGG PAMs and non-NGG PAMs will interesting both mechanistically but also could provide clues as to how to engineer further PAM variants. Single molecule imaging studies looking at residence time of xCas9 on its substrate, its search pattern on DNA, etc. will also reveal how xCas9 locates its target and if the sampling of more PAM sites changes the kinetics of its DNA sampling.

Characterization of xCas9 and other Cas9s will also continue both on different targets and in different cell lines. As activity is known to be very variable on different targets, a sgRNA library will be constructed that can, in a high throughput manner, allow us to assay both *Sp*Cas9 and xCas9 for activity on a variety of targets throughout the genome. Genomic features that lead to

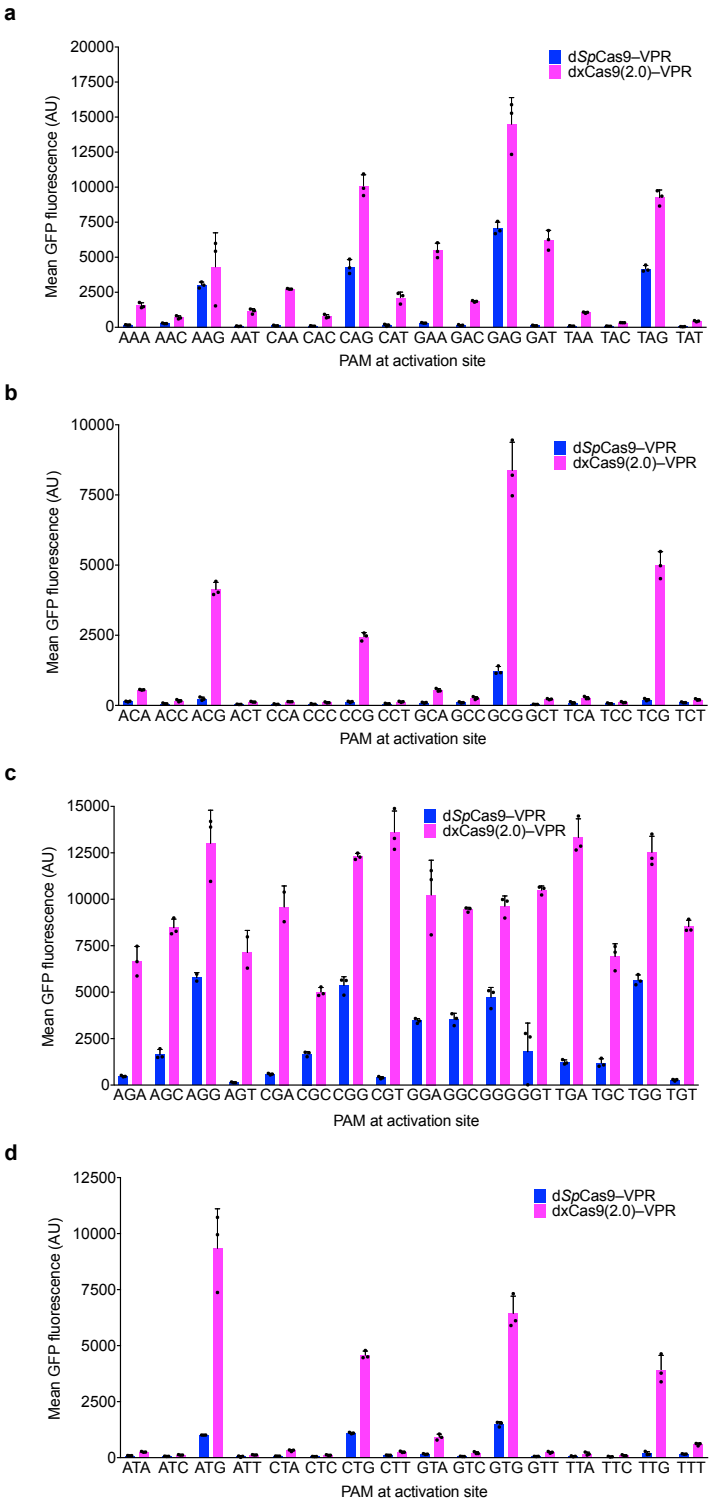
different levels of observed activity still need to be elucidated. Indeed, as we saw during the characterization of xCas9, different contexts gave different results in terms of the ability to recognize different PAM, and methylation state and the packaging of DNA into chromatin in mammalian cells likely affects the ability of Cas9 to access and edit the targeted sequence. Any new Cas9 variant then has to be thoroughly vetted in both prokaryotic and eukaryotic cells for endogenous sites in contexts that are as close as possible to the contexts in which other researchers will be using the technology. In addition, as new DNA effectors are developed for Cas9, their ability to integrate with xCas9 will also be tested.

Further evolution of new Cas9 PAM variants will be pursued. In addition to attempting to broaden the PAM even further, we also believe that providing a suite of different Cas9 variants that can cover the remaining PAMs would be broadly useful for the field as we march steadily towards coverage of the entire sequence space. Small Cas9 variants with troublesome PAMs such as *S.* aureus Cas9 could also be evolved to make them more amenable as a general tool for genome editing. Other undesirable properties of Cas9 such as troublesome off-target sites could be selected against using a negative selection with pIII-neg. For therapeutic applications, where off-target activity could be potentially harmful, PACE provides a general platform for both selecting for the on-target site and against off-target sites. Indeed, for therapeutic applications of Cas9 where the enzyme needs to be optimized to target a single disease site, engineered variants that are highly specialized for that particular site could be required for both optimal activity and safety.

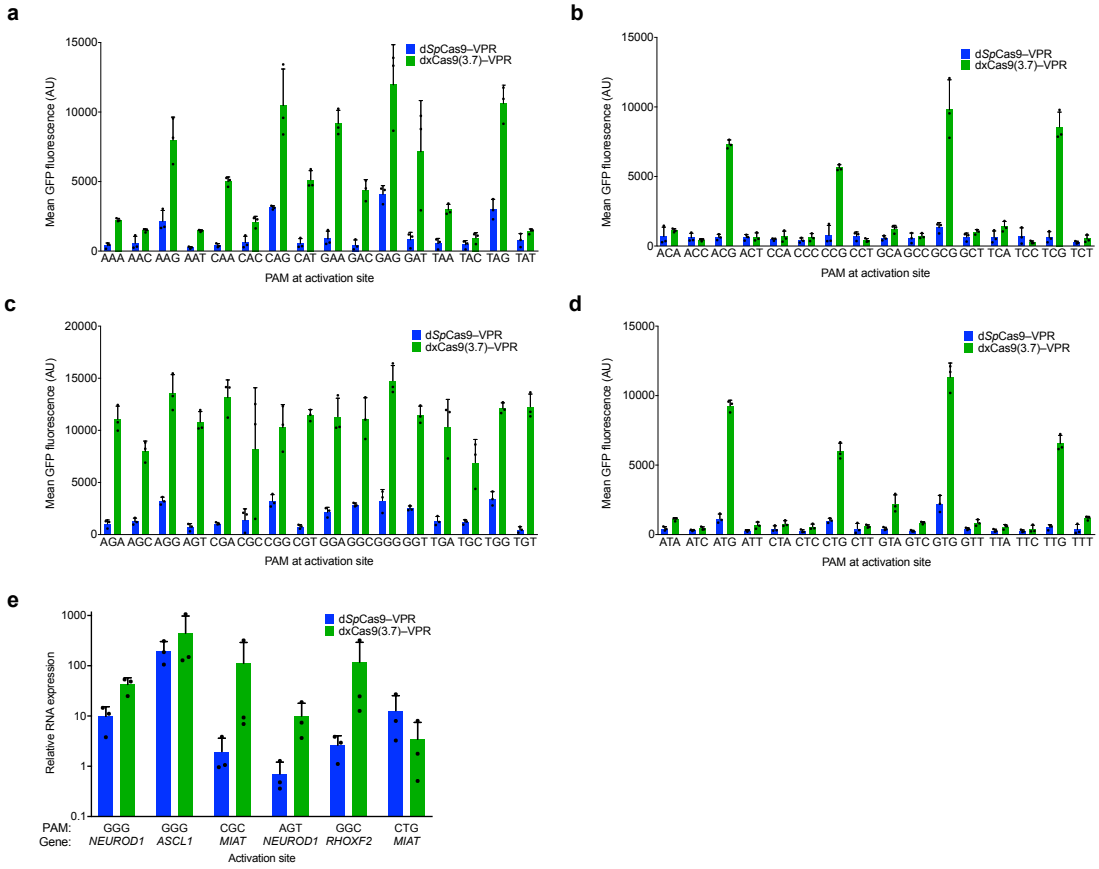
The furious pace of progress of the genome editing field will, no doubt, lead to further refinement of xCas9 and future versions that go beyond the capabilities that we have currently engineered. As protein design and evolution become easier to access and more mainstream researchers will hopefully have access to a number of different Cas9 variants tailored to their specific needs or will have the ability to create their own design Cas9 simply and quickly.

Appendix I:

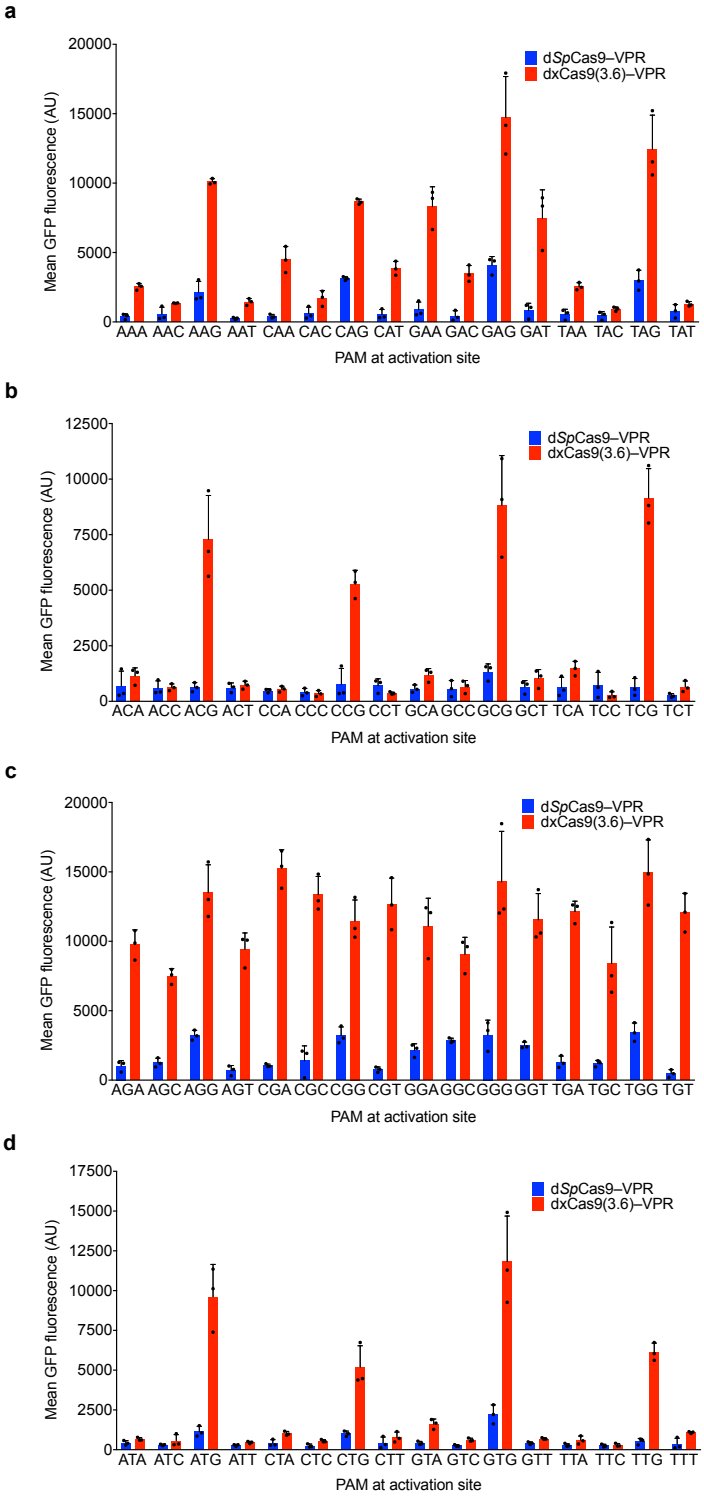
Extended Data Figures for Chapter 3



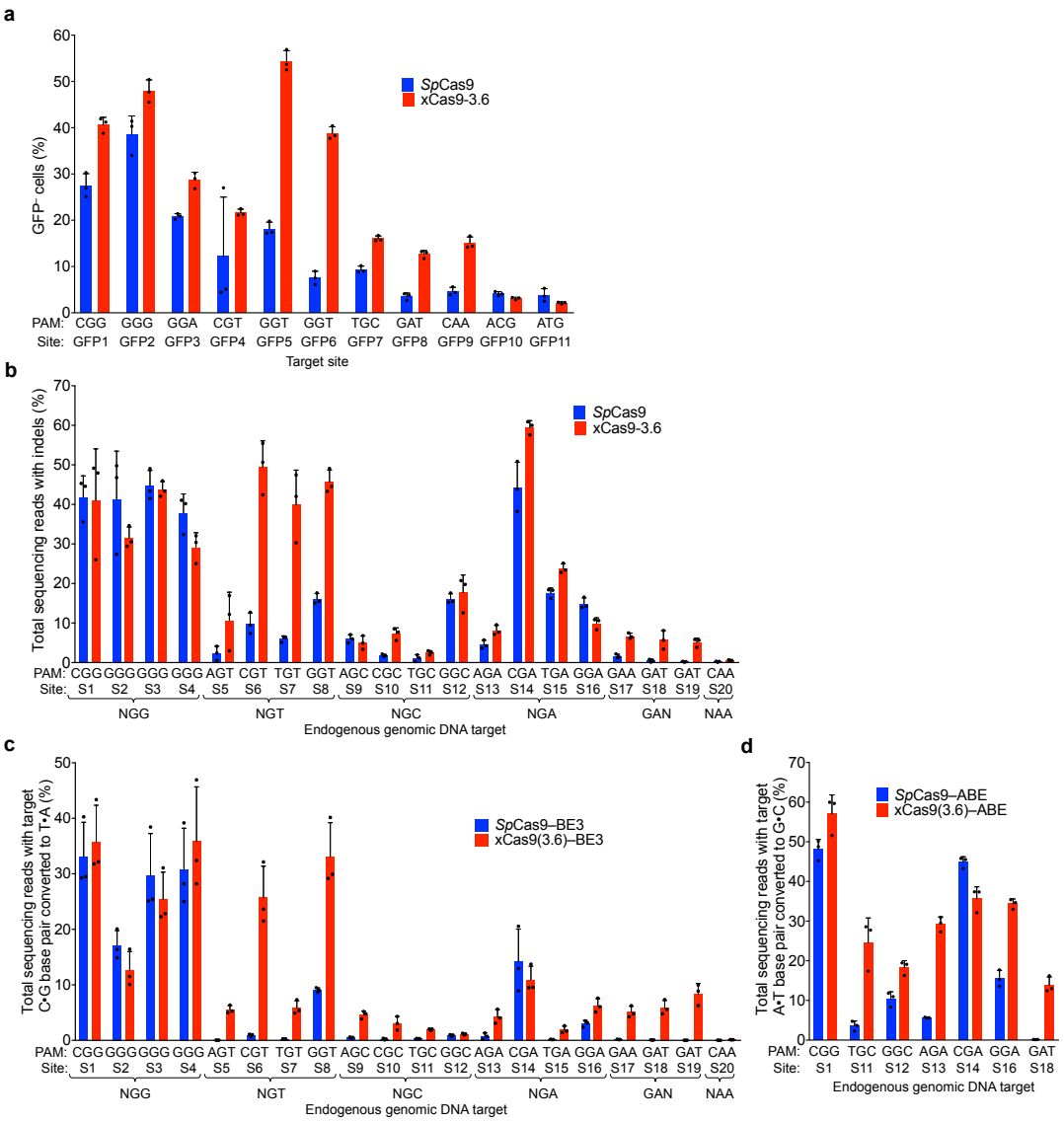
Extended Data Figure A1 | Transcriptional activation with xCas9 2.0. a-d, The transcriptional activator dxCas9(2.0)–VPR was tested on the R1 protospacer (Extended Data Fig. 3 and Supplementary Table 8) with each of the 64 possible NNN PAMs (NAN, NCN, NGN, and NTN) in HEK293T cells. Values and error bars reflect the mean and s.d. of n=3 biologically independent samples.



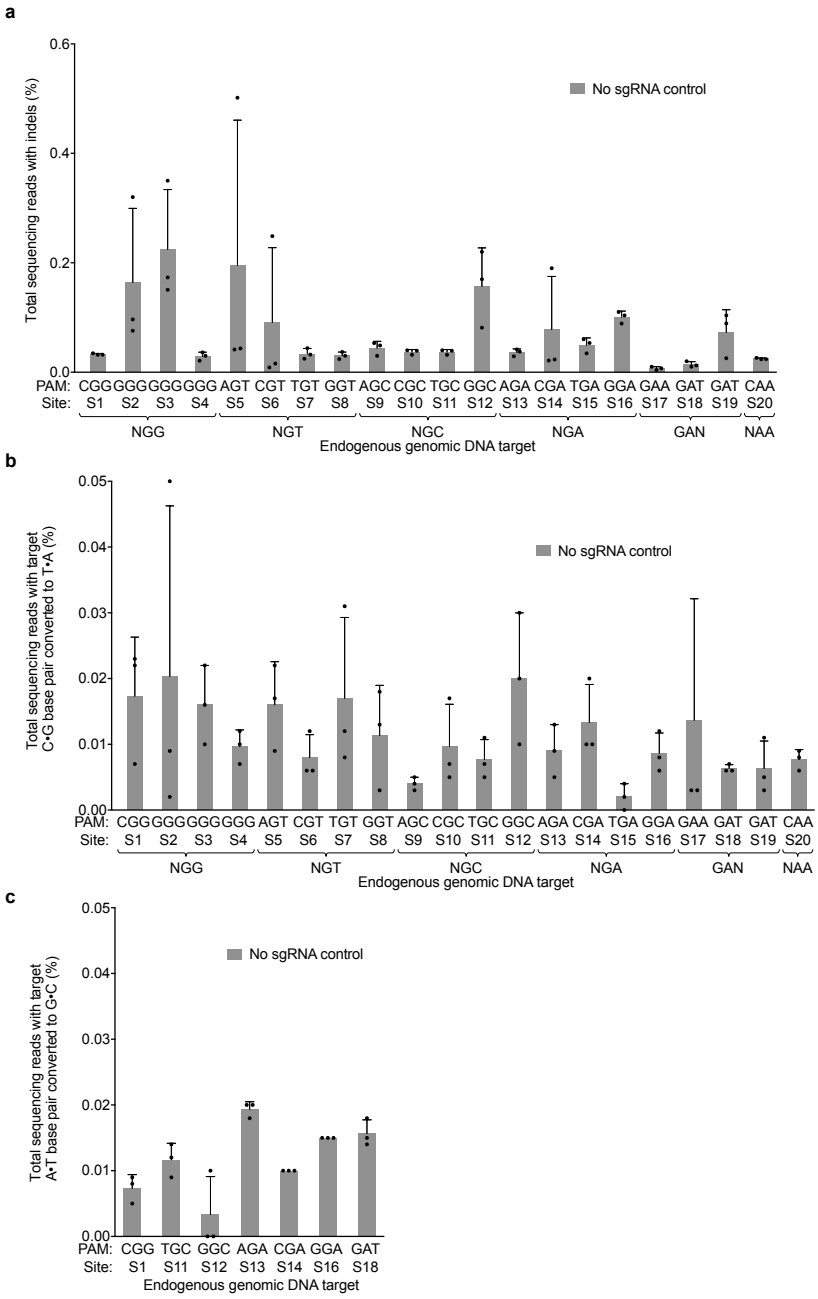
Extended Data Figure A2 | Transcriptional activation with xCas9 3.7 on all 64 NNN PAM sites and endogenous gene activation in human cells. a-d, The transcriptional activator dxCas9(3.7)–VPR was tested on the R1 protospacer (Extended Data Fig. 3 and Supplementary Table 8) with each of the 64 possible NNN PAMs (NAN, NCN, NGN, and NTN) in HEK293T cells. **e,** Endogenous gene activation was tested using both dSpCas9–VPR and dxCas9(3.7)–VPR to activate expression of the *NEUROD1, ASCL1, MIAT,* or *RHOXF2* at six total sites. RNA expression as measured by RT-qPCR was compared to background expression levels for each gene (measured in the control with no sgRNA) and was normalized to *ACTB* expression. Values and error bars reflect the mean and s.d. of n=3 biologically independent samples. Target sites are in Supplementary Table 9.



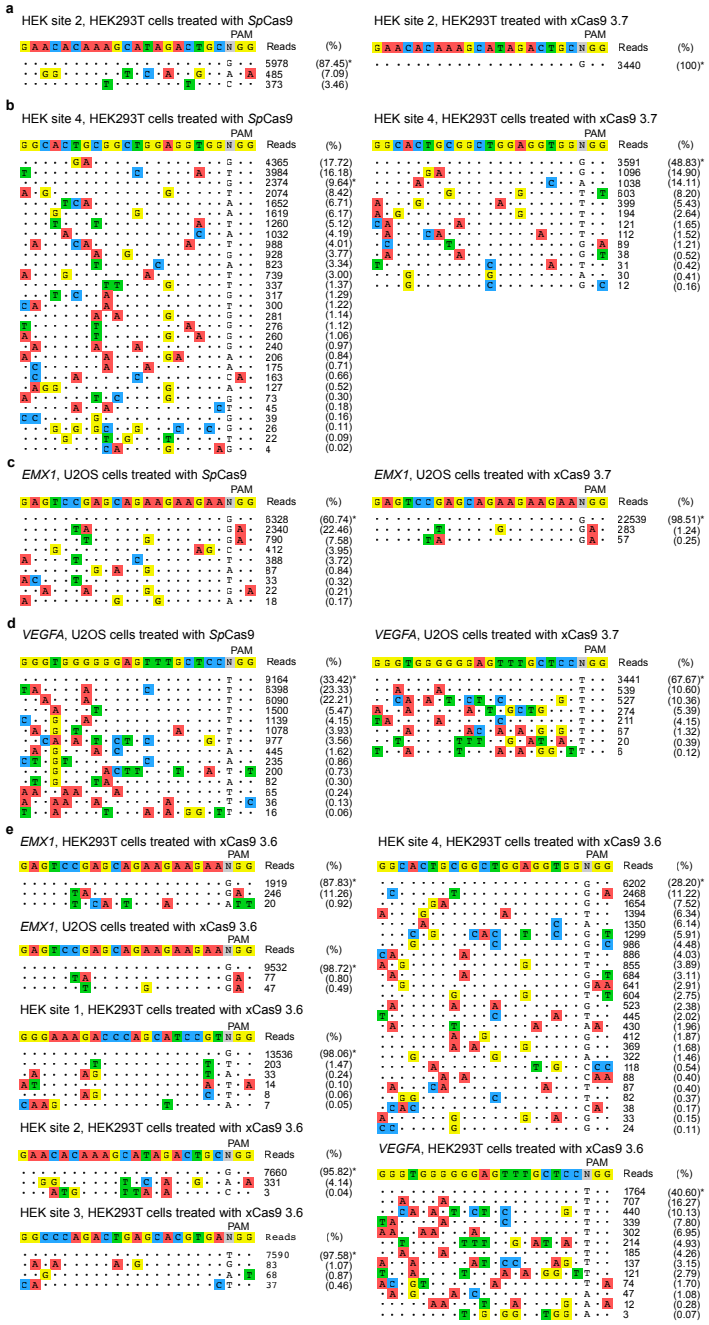
Extended Data Figure A3 | Transcriptional activation with xCas9 3.6 on all 64 NNN PAM sites. a-d, The transcriptional activator dxCas9(3.6)–VPR was tested on the R1 protospacer (Extended Data Fig. 3 and Supplementary Table 8) with each of the 64 possible NNN PAMs (NAN, NCN, NGN, and NTN) in HEK293T cells. Values and error bars reflect the mean and s.d. of n=3 biologically independent samples.



Extended Data Figure A4 | Genomic DNA cleavage and base editing by evolved xCas9 3.6. **a,** Genomic DNA cleavage in HEK293-GFP cells containing a genomically integrated GFP gene by SpCas9 or xCas9 3.6, transfected as plasmids. After 5 days, the cells were analyzed for loss of GFP fluorescence by flow cytometry. Sequences for all target sites are listed in Supplementary Table 11. **b,** DNA cleavage of endogenous genomic DNA sites with a variety of NGG and non-NGG PAMs by SpCas9 and xCas9 3.6 in HEK293T cells. Indel rates were measured by HTS 5 days after plasmid transfection. Sequences for all target sites are listed in Supplementary Table 12. **c,** 20 sites containing NG, GAA, GAT, or CAA PAM sites were tested for C•G to T•A base editing in HEK293T cells by SpCas9–BE3 or xCas9(3.6)–BE3. The C•G to T•A conversion frequency by HTS at the most efficiently edited base 3 days after plasmid transfection is shown. **d,** Of the 20 sites in (c), seven contained an A in the canonical window for ABE editing[78] and were tested for A•T to G•C base editing by SpCas9–ABE and xCas9(3.6)–ABE. The A•T to G•C conversion frequency by HTS at the most efficiently edited base 5 days after plasmid transfection is shown. Values and error bars reflect the mean and s.d. of n=3 biologically independent samples. Complete HTS results across the protospacer are provided in Supplementary Table 14.



Extended Data Figure A5 | Negative controls lacking guide RNA for nuclease and base editing experiments. To verify genomic DNA cleavage and base editing results, the same sites were sequenced after treatment with SpCas9 nuclease, SpCas9-BE3, or SpCas9-ABE but without any sgRNA. **a**, Indel rates at endogenous target sites 5 days after treatment of HEK293T cells with SpCas9. **b**, Target C•G to T•A conversion 3 days after treatment of HEK293T cells with SpCas9-BE3. **c,** Target A•T to G•C base conversion 5 days after treatment of HEK293T cells with SpCas9-ABE. Values and error bars reflect the mean and s.d. of n=3 biologically independent samples. Complete HTS results across the protospacer are provided in Supplementary Table 14.



Extended Data Figure A6 | Additional characterization of xCas9 3.7 and xCas9 3.6 by GUIDE-seq. In addition to the GUIDE-seq data shown in Fig. 4, two additional sites in HEK293T cells (**a, b**) and two sites in U2OS cells (**c, d**) were analyzed after treatment with SpCas9 and xCas9 3.7. **e,** All six GUIDE-seq sites with an NGG PAM that were tested with SpCas9 and xCas9 3.7 in HEK293T and U2OS cells were also tested with xCas9 3.6. On-target reads (*) and off-target reads for all sites are shown. Target sequences are listed in Supplementary Table 15.

Appendix II:

Supplementary Information for Chapter 2

Supplementary Table 1. Target sites used to test sgRNAs for activation of gene III expression in PACE.

Label	Sequence	PAM	Distance from PAM to -35 signal
G0	GTGAACCGCATCGAGCTGAA	n/a	Scrambled
G1	AACACGCACGGTGTTACATT	AGG	10 bp
G2	TAGGGACCCTACAACACGCA	CGG	22 bp
G3	CGTTTTGCTGAGGAGACTTA	GGG	44 bp
G4	CTGGGCCTTTCGTTTTGCTG	AGG	50 bp
G5	AAAGGCTCAGTCGAAAGACT	GGG	68 bp
G6'	GTAACACCGTGCGTGTTGTA	CGG	36 bp
G7'	GTCTCCTCAGCAAAACGAA	CGG	65 bp

Supplementary Table 2. Linker optimization of ω–dCas9 constructs for PACE from Extended Data Figure 1c.

Label	Sequence	Length
L1	AA	2
L2	EQKLISEEDL	10
L3	GGSGGSGGSGGSGGS	15
L4	GHGTGSTGSGSS	12
L5	MSRPDPA	7
L6	GSAGSAAGSGEF	12
L7	SGSETPGTSESATPES	16

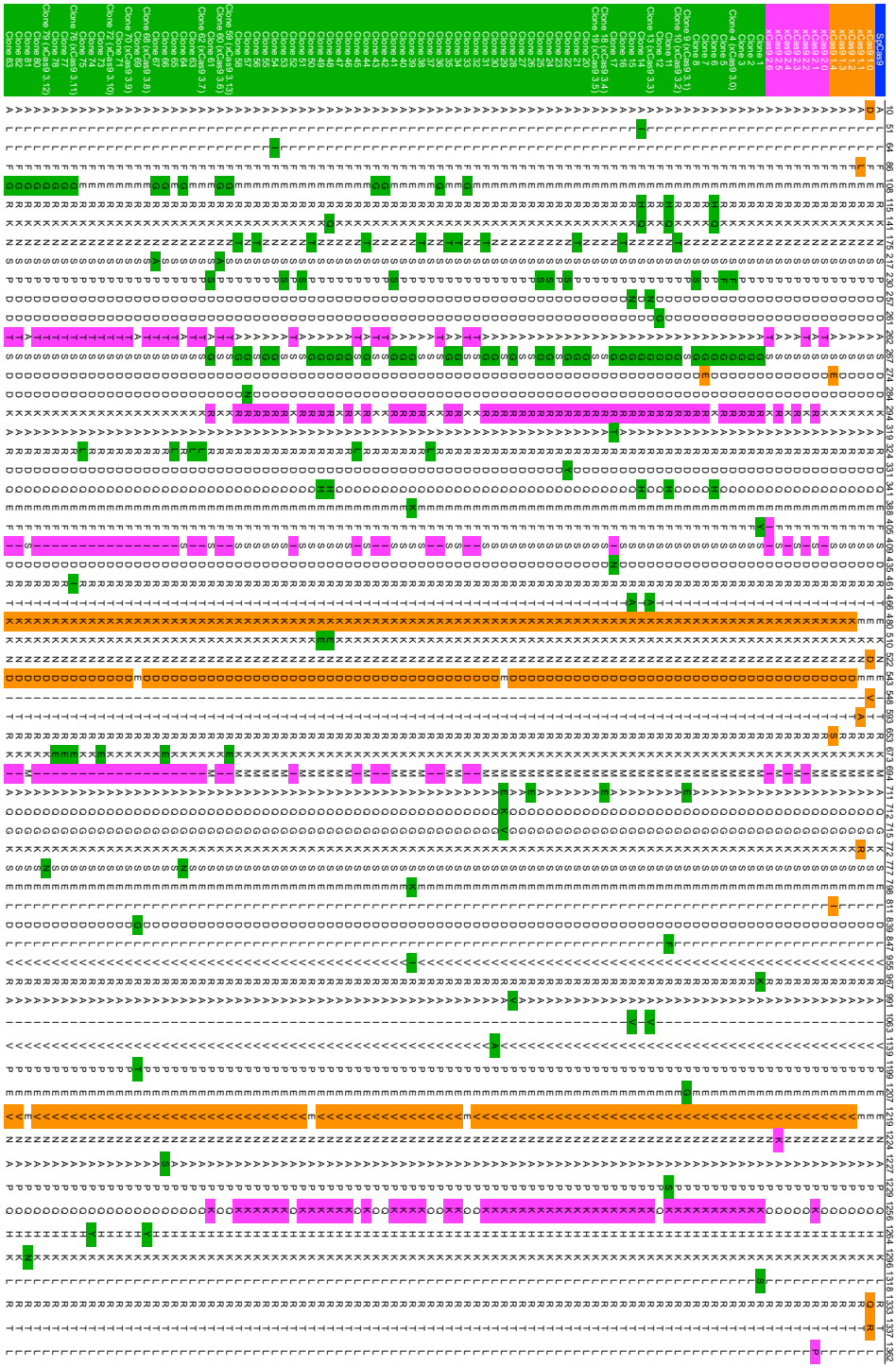
Supplementary Table 3. Mutations in ω–dCas9 variants from Extended Data Figure 1d.

ω/dCas9	Mutations relative to canonical ω–dCas9
wt/wt	I12N(ω)
PACE1/PACE1	I12N(ω), H112Y, L300F, Y449H, N802S, K1288E
PACE2/PACE2	F574L, L641I
PACE3/PACE3	Q30K(ω)
wt/PACE1	H112Y, L300F, Y449H, D706D, N802S, K1288E

Supplementary Table 4. Target sites used in xCas9 PACE.

Label	Sequence
G7’	AGTCTCCTCAGCAAAACGAA
CD13	GGAGCCCACCGAGTACCTGG
VEGFA	GACCCCCTCCACCCCGCCTC

Supplementary Table 5. Genotypes of xCas9 variants evolved in this work.



Supplementary Table 6. Plasmids used for Cas9 PACE and xCas9 characterization.

Plasmid name	Antibiotic resistance	Origin of replication	ORF1		ORF2		Comments
			Promoter	Gene	Promoter	Gene	
PhJH008	None	M13 f1	P_{gIII}	ω–dCas9			SP for PACE[73,90]
MP4	chlor^R	CloDF13	p_{BAD}	dnaQ926, dam, seqA	PC	araC	MP for PACE[88]
MP6	chlor^R	CloDF13	p_{BAD}	dnaQ926, dam, seqA, emrR, ugi, cda1	PC	araC	MP for PACE[88]
pJH414	carb^R	SC101	P_{lacZ}	gIII, luxAB	P_{lacZ} (no LacO)	sgRNA	AP (G7' NGG PAM)
pJH447	carb^R	SC101	P_{lacZ}	gIII, luxAB	P_{lacZ} (no LacO)	sgRNA	AP (G7' NNN PAM library)
pJH451	carb^R	SC101	P_{lacZ}	gIII, luxAB	P_{lacZ} (no LacO)	sgRNA	AP (CD13 NNN PAM library)
pJH453	carb^R	SC101	P_{lacZ}	gIII, luxAB	P_{lacZ} (no LacO)	sgRNA	AP (*VEGFA* NNN PAM library)
pJH958	spect^R	SC101	P_{tet}	xCas9 3.0	Plac (no LacO)	sgRNA	PAM depletion
pJH959	spect^R	SC101	P_{tet}	xCas9 3.1	Plac (no LacO)	sgRNA	PAM depletion
pJH960	spect^R	SC101	P_{tet}	xCas9 3.2	Plac (no LacO)	sgRNA	PAM depletion
pJH961	spect^R	SC101	P_{tet}	xCas9 3.3	Plac (no LacO)	sgRNA	PAM depletion
pJH962	spect^R	SC101	P_{tet}	xCas9 3.4	Plac (no LacO)	sgRNA	PAM depletion
pJH963	spect^R	SC101	P_{tet}	xCas9 3.5	Plac (no LacO)	sgRNA	PAM depletion
pJH964	spect^R	SC101	P_{tet}	xCas9 3.6	Plac (no LacO)	sgRNA	PAM depletion
pJH965	spect^R	SC101	P_{tet}	xCas9 3.7	Plac (no LacO)	sgRNA	PAM depletion
pJH966	spect^R	SC101	P_{tet}	xCas9 3.8	Plac (no LacO)	sgRNA	PAM depletion
pJH967	spect^R	SC101	P_{tet}	xCas9 3.9	Plac (no LacO)	sgRNA	PAM depletion
pJH968	spect^R	SC101	P_{tet}	xCas9 3.10	Plac (no LacO)	sgRNA	PAM depletion
pJH969	spect^R	SC101	P_{tet}	xCas9 3.11	Plac (no LacO)	sgRNA	PAM depletion
pJH970	spect^R	SC101	P_{tet}	xCas9 3.12	Plac (no LacO)	sgRNA	PAM depletion
pJH971	spect^R	SC101	P_{tet}	xCas9 3.13	Plac (no LacO)	sgRNA	PAM depletion
pJH306	carb^R	BR322	CMV	dSpCas9-VPR			GFP activation[100]
pJH599	carb^R	BR322	CMV	dxCas9(2.0)-VPR			GFP activation
pJH11172	carb^R	BR322	CMV	dxCas9(3.6)-VPR			GFP activation
pJH11173	carb^R	BR322	CMV	dxCas9(3.7)-VPR			GFP activation
pJH407	carb^R	BR322	CMV	SpCas9			DNA cutting
pJH1009	carb^R	BR322	CMV	xCas9 3.6			DNA cutting
pJH1010	carb^R	BR322	CMV	xCas9 3.7			DNA cutting
pJH1054	carb^R	BR322	CMV	SpCas9-BE3			C•G to T•A base editor[77]
pJH1056	carb^R	BR322	CMV	xCas9(3.6)-BE3			C•G to T•A base editor
pJH11171	carb^R	BR322	CMV	xCas9(3.7)-BE3			C•G to T•A base editor
pJH11174	carb^R	BR322	CMV	SpCas9-ABE			A•T to G•C base editor[78]
pJH11175	carb^R	BR322	CMV	xCas9(3.6)-ABE			A•T to G•C base editor
pJH11176	carb^R	BR322	CMV	xCas9(3.7)-ABE			A•T to G•C base editor

Supplementary Table 7. Primers used for PAM depletion assay HTS.

Primer name	sequence
PAM depletion fwd 1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTATATAGGCCTCCCACCGTAC
PAM depletion rev 1	TGGAGTTCAGACGTGTGCTCTTCCGATCTACAGGGTAATAGTGTCCCCTC
PAM depletion fwd 2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTTATATAGGCCTCCCACCG
PAM depletion rev 2	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGCTCGAGCGGCCGCC

Supplementary Table 8. Target sequences used for GFP transcriptional activation assays.

Site	NGG PAM	Protospacer sequence
R1	GGG	GTCCCCTCCACCCCACAGTG
R2	TGG	GGGGCCACTAGGGACAGGAT

Supplementary Table 9. Target sites used for endogenous gene transcriptional activation.

Gene	PAM	Protospacer sequence
NEUROD1	GGG	GTCTCTAACTGGCGACAGAT
ASCL1	GGG	GGGAGAAAGGAACGGGAGGG
MIAT	CGC	GGCAAGCGTTGCCATGACAG
NEUROD1	AGT	GGGTAAAAACAGGTCCGCGG
RHOXF2	GGC	GCTCTCCCTCATCCGGGCAA
MIAT	CTG	GCGGGTGGGGATGCGGGAGG

Supplementary Table 10. Primers used for endogenous gene transcriptional activation analysis.

Site	Primer sequence
NEUROD1 fwd	GGATGACGATCAAAAGCCCAA
NEUROD1 rev	GCGTCTTAGAATAGCAAGGCA
ASCL1 fwd	CGCGGCCAACAAGAAGATG
ASCL1 rev	CGACGAGTAGGATGAGACCG
MIAT fwd	TGGCTGGGGTTTGAACCTTT
MIAT rev	AGGAAGCTGTTCCAGACTGC
RHOXF2 fwd	GGCAAGAAGCATGAATGTGA
RHOXF2 rev	TGTCTCCTCCATTTGGCTCT

Supplementary Table 11. Target sites used in the GFP disruption assay.

Site	PAM	Protospacer sequence
GFP1	CGG	GACCAGGATGGGCACCACCC
GFP2	GGG	GTGGTCACGAGGGTGGGCCA
GFP3	GGA	GTGCCCATCCTGGTCGAGC
GFP4	CGT	GTGACCACCTTCACCTACGG
GFP5	GGT	GTGGTGCAGATGAACTTCAG
GFP6	GGT	GAAGTCGTGCTGCTTCATGT
GFP7	TGC	GGAGCTGTTCACCGGGGTGG
GFP8	GAT	GAGTTCGTGACCGCCGCCGG
GFP9	CAA	GTGAACTTCAAGACCCGCCA
GFP10	ACG	GGTAGTGGTCGGCGAGCTGC
GFP11	ATG	GCTGCACGCTGCCGTCCTCG

sgRNA plasmids were generated with reverse primer GGTGTTTCGTCCTTTCCACAAGATATATAAAG and forward primer $(N)_{20}$GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC, where $(N)_{20}$ is the protospacer sequence.

Supplementary Table 12. Genomic sites used for targeted genome editing.

Site	Genomic locus	PAM	Protospacer sequence
1	HEK site 4	CGG	GATGACAGGCAGGGGCACCG
2	*EMX1*	GGG	GGCCCCAGTGGCTGCTCTGG
3	*EMX1*	GGG	GGCCCCCAGAGCAGCCACTG
4	HEK site 4	GGG	GGCACTGCGGCTGGAGGTGG
5	*VEGFA*	AGT	GAGCGAGCAGCGTCTTCGAG
6	*FANCF*	CGT	GCGGTCTCAAGCACTACCTA
7	HEK site 3	TGT	GAGCTGCACATACTAGCCCC
8	HEK site 3	GGT	GCTTCTCCAGCCCTGGCCTG
9	HEK site 4	AGC	GGCCCCCACTGTAGTCACAC
10	HEK site 4	CGC	GACAGGCAGGGGCACCGCGG
11	HEK site 4	TGC	GTGCCACCGGGGCGCCGCGG
12	*VEGFA*	GGC	GCAGACGGCAGTCACTAGGG
13	HEK site 3	AGA	GTCGCAGGACAGCTTTTCCT
14	*VEGFA*	CGA	GGGAAGCTGGGTGAATGGAG
15	HEK site 4	TGA	GCGGAGACTCTGGTGCTGTG
16	*VEGFA*	GGA	GGTCAGAAATAGGGGGTCCA
17	*FANCF*	GAA	GACTCTCTGCGTACTGATTG
18	*FANCF*	GAT	GATCCAGGTGCTGCAGAAGG
19	*VEGFA*	GAT	GTCACTCCAGGATTCCAATA
20	*EMX1*	CAA	GTGGTTCCAGAACCGGAGGA
F1	*FANCF*	TGG	GGAATCCCTTCTGCAGC
F2	*FANCF*	AGT	GGCGGGCCAGGCTCTCTTGG
F3	*FANCF*	CGT	GGGCCCTCGGGGATGCAGCT
F4	*FANCF*	TGT	GCCCTCTTGCCTCCACTGGT
F5	*FANCF*	GGT	GCCACATCCATCGGCGCTTT
F6	*FANCF*	GGT	GCCGCCCTCTTGCCTCCACT
F7	*FANCF*	CGC	GTGACGTCCTGCTCTCTCTG
F8	*FANCF*	CGC	GCGCCGGGCCTTGCAGTGGG
F9	*FANCF*	CGC	GCGCCCACTGCAAGGCCCGG
F10	*FANCF*	TGC	GAGAACCGGGCCCTCGGGGA
F11	*FANCF*	GGC	GAGACACTCCAAGAGAGCCT
F12	*FANCF*	AGA	GCGCTTCAATGGCTATAGAG
F13	*FANCF*	TGA	GAAGGTCATAGTGCAAACGT
F14	*FANCF*	GGA	GAGAACCCAAATCTCCAGGA
F15	*FANCF*	GGA	GGCCCCATTCGCACGGCTCT

Supplementary Table 12 (Continued)

sgRNA plasmids were generated with reverse primer GGTGTTTCGTCCTTTCCACAAGATATATAAAG and forward primer $(N)_{20}$GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC, where $(N)_{20}$ is the protospacer sequence.

Supplementary Table 13. Primers for mammalian cell genomic DNA amplification.

Primer name	Sequence
VEGFA fwd 1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCCCATTCCCTCTTTAGCCA
VEGFA rev 1	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCAACCCCACACGCACA
VEGFA fwd 2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCACCACAGGGAAGCTGG
VEGFA rev 2	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCAACCCCACACGCACA
VEGFA fwd 3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCAGAGGGACACACTGTGG
VEGFA rev 3	TGGAGTTCAGACGTGTGCTCTTCCGATCTAATGGGCTTTGGAAAGGGGG
VEGFA fwd 4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGGACGTGTGTGTCTGTGT
VEGFA rev 4	TGGAGTTCAGACGTGTGCTCTTCCGATCTATATTGAAGGGGGCAGGGGA
VEGFA fwd 5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGGATCGCTTTTCCGAGCTT
VEGFA rev 5	TGGAGTTCAGACGTGTGCTCTTCCGATCTGACGTCACAGTGACCGAGG
VEGFA fwd 6	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACAACCAGTGGAGGCAAGA
VEGFA rev 6	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGGCTCTCTTGGAGTGTC
VEGFA fwd 7	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCTGGTGCAGCAACTCTTT
VEGFA rev 7	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTGGGTCTTCATCAGAGAG
VEGFA fwd 8	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGACGCCGATGAGGAGAC
VEGFA rev 8	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTGTTCTGAGGCAAGCGCTC
VEGFA fwd 9	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCTAGTCCACTGGCTTCTGG
VEGFA rev 9	TGGAGTTCAGACGTGTGCTCTTCCGATCTATTGTGCAACTCCTCCCAGG
EMX1 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGCTCAGCCTGAGTGTTGA
EMX1 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGGGTTTGTGGTTGC
FANCF fwd 1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCATTGCAGAGAGGCGTATCA
FANCF rev 1	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGTCCCAGGTGCTGAC
FANCF fwd 2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAATCCCTTCTGCAGCACCT
FANCF rev 2	TGGAGTTCAGACGTGTGCTCTTCCGATCTGATGGATGTGGCGCAGGTAG
HEK site 4 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAACCCAGGTAGCCAGAGAC
HEK site 4 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTTCAACCCGAACGGAG
HEK site 3 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTGGGCTGCCTAGAAAGG
HEK site 3 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC
EMX1 on-target fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGAGGACAAAGTACAAACGGC
EMX1 on-target rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTATCGATGTCCTCCCCATTGG
EMX1 off-target 1 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCGGCCTTTGCAAATAGAGCC
EMX1 off-target 1 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTGGCCTTTGTAGGAAAACACC
HEK site 1 on-target fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGCCAGGACCTGACATGCAG
HEK site 1 on-target rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTTCCTCTTTCCCCTGTAGT

Supplementary Table 13 (Continued)

HEK site 1 off-target 1 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTCAACTCTGGAGCTGAT
HEK site 1 off-target 1 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCTCCTGGCTAGAAATCAGC
HEK site 2 on-target fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCTGGCTGAGCTAACTGTGA
HEK site 2 on-target rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGCCCCATCTGTCAAACT
HEK site 2 off-target 1 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCTCTCAGCTCAGCCTGAGT
HEK site 2 off-target 1 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTGGAGGTGACATCGATGTC
HEK site 2 off-target 2 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGGGTTCAAAAACAAACAGAGAA
HEK site 2 off-target 2 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTTAATGCTCCCACACCATTTTT
HEK site 3 on-target fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGAAACGCCCATGCAATTAGT
HEK site 3 on-target rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAGCTGCACATACTAGCCCC
HEK site 3 off-target 1 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCTTGCTGAGATGTGGGCAG
HEK site 3 off-target 1 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCAGGAGTACACAGAGGTCC
HEK site 3 off-target 2 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATTAACTGCACCAGTGGGCA
HEK site 3 off-target 2 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGGTGTGCCTGTCACTGTAC
HEK site 3 xCas9 off-target 1 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGGAAATAGGCCCTCTTCCT
HEK site 3 xCas9 off-target 1 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGAATCCCATGGGCACCATG
HEK site 3 xCas9 off-target 2 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTCAGTCTGAGAAAGTGGCG
HEK site 3 xCas9 off-target 2 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTAAGAACAAAGGCTGATGCGT
HEK site 4 on-target fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGGTCCAAAGCAGGATGACA
HEK site 4 on-target rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACACAGCACCAGAGTCTCC
HEK site 4 off-target 1 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGCAGGTCCTTCATGGCAAG
HEK site 4 off-target 1 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCATATCCCTGCACCTC
HEK site 4 off-target 2 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGCTCTCCAGAGTGGTCCAG
HEK site 4 off-target 2 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTCCTCGGAGTCCTCAAG
HEK site 4 off-target 3 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGCAGAGCAAGTCTCACATTT
HEK site 4 off-target 3 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCTTGTGATCTGTCCGTCTC

Supplementary Table 13 (Continued)

HEK site 4 off-target 4 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCGGATGTTCTTTCTGGGCT
HEK site 4 off-target 4 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGGTTCCTCCACAACACA
HEK site 4 off-target 5 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCATTACTCACCGGACCTCC
HEK site 4 off-target 5 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACGGGAAGGACAGGAGAAG
HEK site 4 xCas9 off-target 1 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGTCCCCAGAGCTTAGTTCA
HEK site 4 xCas9 off-target 1 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCGGATGCAATGAGCCAGGA
HEK site 4 xCas9 off-target 2 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCACCCGGAATACCGGTCT
HEK site 4 xCas9 off-target 2 rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCGTCTCTGTCATCATGCTG

Supplementary Table 14. HTS sequencing results at 35 genomic sites treated with BE3[77] and ABE[78] variants. HEK293T cells were treated with SpCas9–BE3, xCas9(3.6)–BE3, xCas9(3.7)–BE3, SpCas9–ABE, xCas9(3.6)–ABE, or xCas9(3.7)–ABE and an sgRNA that targets each of the 35 sites below. One arbitrarily chosen replicate is shown; the data for all replicates is available from the NCBI sequencing read archive.

No sgRNA control

Site 1	G_1	A_2	T_3	G_4	A_5	C_6	A_7	G_8	G_9	C_{10}	A_{11}	G_{12}	G_{13}	G_{14}	G_{15}	C_{16}	A_{17}	C_{18}	C_{19}	G_{20}	C	G	G	indel%
A	0.01	100	0	0.01	100	0	100	0.01	0.01	0.01	100	0.01	0.01	0.01	0.01	0.01	100	0	0	0.01	0.01	0.02	0.01	0.03
C	0	0.03	0.01	0	0.02	100	0.03	0	0	100	0.01	0	0	0	0.01	100	0.01	100	100	0	100	0.01	0	
G	100	0.01	0.01	100	0.01	0	0.02	100	100	0	0.02	100	100	100	100	0	0	0	0	100	0	100	100	
T	0.02	0	100	0.02	0.01	0.02	0	0.01	0.01	0.01	0.02	0.03	0.02	0.02	0.01	0	0	0.01	0.02	0.02	0.02	0.01	0.01	

SpCas9–BE3

Site 1	G_1	A_2	T_3	G_4	A_5	C_6	A_7	G_8	G_9	C_{10}	A_{11}	G_{12}	G_{13}	G_{14}	G_{15}	C_{16}	A_{17}	C_{18}	C_{19}	G_{20}	C	G	G	indel%
A	0.5	100.0	0.0	0.2	100.0	0.7	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.08
C	0.1	0.0	0.0	0.1	0.0	55.4	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	99.8	100.0	0.0	100.0	0.0	0.0	
G	99.3	0.0	0.0	99.7	0.0	3.6	0.0	100.0	100.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	
T	0.1	0.0	100.0	0.1	0.0	40.3	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	

xCas9(3.6)–BE3

Site 1	G_1	A_2	T_3	G_4	A_5	C_6	A_7	G_8	G_9	C_{10}	A_{11}	G_{12}	G_{13}	G_{14}	G_{15}	C_{16}	A_{17}	C_{18}	C_{19}	G_{20}	C	G	G	indel%
A	0.0	100.0	0.0	0.0	100.0	0.6	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.83
C	0.0	0.0	0.0	0.0	0.0	52.7	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	
G	99.9	0.0	0.0	99.9	0.0	3.3	0.0	100.0	100.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	
T	0.0	0.0	100.0	0.0	0.0	43.4	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

Site 1	G_1	A_2	T_3	G_4	A_5	C_6	A_7	G_8	G_9	C_{10}	A_{11}	G_{12}	G_{13}	G_{14}	G_{15}	C_{16}	A_{17}	C_{18}	C_{19}	G_{20}	C	G	G	indel%
A	0.1	100.0	0.0	0.1	100.0	1.1	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	1.05
C	0.0	0.0	0.0	0.0	0.0	32.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	99.9	100.0	0.0	100.0	0.0	0.0	
G	99.9	0.0	0.0	99.9	0.0	5.4	0.0	100.0	100.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	
T	0.0	0.0	100.0	0.0	0.0	61.5	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	

SpCas9–ABE

Site 1	G_1	A_2	T_3	G_4	A_5	C_6	A_7	G_8	G_9	C_{10}	A_{11}	G_{12}	G_{13}	G_{14}	G_{15}	C_{16}	A_{17}	C_{18}	C_{19}	G_{20}	C	G	G	indel%
A	0.0	99.9	0.0	0.0	49.8	0.0	83.5	0.0	0.0	0.0	99.6	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.03
C	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	
G	100.0	0.1	0.0	100.0	50.1	0.0	16.5	100.0	100.0	0.0	0.4	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	
T	0.0	0.0	100.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.6)–ABE

Site 1	G_1	A_2	T_3	G_4	A_5	C_6	A_7	G_8	G_9	C_{10}	A_{11}	G_{12}	G_{13}	G_{14}	G_{15}	C_{16}	A_{17}	C_{18}	C_{19}	G_{20}	C	G	G	indel%
A	0.0	100.0	0.0	0.0	48.4	0.0	89.7	0.0	0.0	0.0	99.7	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.66
C	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	99.9	99.9	0.0	100.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	51.6	0.0	10.3	100.0	100.0	0.0	0.3	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	
T	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	

xCas9(3.7)–ABE

Site 1	G_1	A_2	T_3	G_4	A_5	C_6	A_7	G_8	G_9	C_{10}	A_{11}	G_{12}	G_{13}	G_{14}	G_{15}	C_{16}	A_{17}	C_{18}	C_{19}	G_{20}	C	G	G	indel%
A	0.0	99.9	0.0	0.0	35.8	0.0	82.7	0.0	0.0	0.0	99.6	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.84
C	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	
G	100.0	0.1	0.0	100.0	64.2	0.0	17.3	100.0	100.0	0.0	0.4	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	99.9	0.0	100.0	100.0	
T	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

No sgRNA control

Site 2	G_1	G_2	C_3	C_4	C_5	C_6	A_7	G_8	T_9	G_{10}	G_{11}	C_{12}	T_{13}	G_{14}	C_{15}	T_{16}	C_{17}	T_{18}	G_{19}	G_{20}	G	G	G	indel%
A	0.1	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.32
C	0.0	0.0	99.9	100.0	99.9	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	
G	99.8	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.9	99.9	100.0	100.0	100.0	
T	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0	

SpCas9–BE3

Site 2	G_1	G_2	C_3	C_4	C_5	C_6	A_7	G_8	T_9	G_{10}	G_{11}	C_{12}	T_{13}	G_{14}	C_{15}	T_{16}	C_{17}	T_{18}	G_{19}	G_{20}	G	G	G	indel%
A	0.3	0.0	0.0	0.0	0.1	0.2	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.44
C	0.0	0.0	99.2	93.2	81.4	79.7	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	99.6	0.0	0.0	0.0	0.0	0.0	0.0	
G	99.6	99.9	0.0	0.0	0.1	0.0	0.0	100.0	0.0	99.9	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	99.9	99.9	99.9	99.9	100.0	
T	0.1	0.1	0.8	6.8	18.4	20.1	0.0	0.0	100.0	0.1	0.1	0.0	100.0	0.0	0.0	100.0	0.3	100.0	0.1	0.1	0.0	0.0	0.0	

xCas9(3.6)–BE3

Site 2	G_1	G_2	C_3	C_4	C_5	C_6	A_7	G_8	T_9	G_{10}	G_{11}	C_{12}	T_{13}	G_{14}	C_{15}	T_{16}	C_{17}	T_{18}	G_{19}	G_{20}	G	G	G	indel%
A	0.2	0.0	0.0	0.0	0.1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.52
C	0.0	0.0	99.5	90.8	83.3	83.5	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	100.0	0.0	99.0	0.0	0.0	0.0	0.0	0.0	0.0	
G	99.8	99.9	0.0	0.0	0.1	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	99.9	99.9	100.0	99.9	100.0	
T	0.0	0.0	0.5	9.1	16.5	16.4	0.0	0.0	100.0	0.0	0.0	0.1	100.0	0.0	0.0	100.0	1.0	100.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

Site 2	G_1	G_2	C_3	C_4	C_5	C_6	A_7	G_8	T_9	G_{10}	G_{11}	C_{12}	T_{13}	G_{14}	C_{15}	T_{16}	C_{17}	T_{18}	G_{19}	G_{20}	G	G	G	indel%
A	0.1	0.1	0.0	0.0	0.1	0.1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.82
C	0.0	0.0	99.5	86.9	71.4	70.4	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	100.0	0.0	98.3	0.0	0.0	0.0	0.0	0.0	0.0	
G	99.9	99.9	0.0	0.0	0.1	0.0	0.0	100.0	0.0	100.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.9	99.9	99.9	99.9	99.9	
T	0.0	0.0	0.5	13.1	28.3	29.5	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	1.7	100.0	0.1	0.1	0.1	0.0	0.0	

No sgRNA control

Site 3	C	C	C	C_{20}	A_{19}	G_{18}	T_{17}	G_{16}	G_{15}	C_{14}	T_{13}	G_{12}	C_{11}	T_{10}	C_9	T_8	G_7	G_6	G_5	G_4	G_3	C_2	C_1	indel%
A	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.35
C	99.9	100.0	99.9	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	
G	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.9	99.9	100.0	100.0	100.0	0.0	0.0	
T	0.1	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	

Supplementary Table 14 (Continued)

SpCas9–BE3

Site 3	C	C	C	C_{20}	A_{19}	G_{18}	T_{17}	G_{16}	G_{15}	C_{14}	T_{13}	G_{12}	C_{11}	T_{10}	C_9	T_8	G_7	G_6	G_5	G_4	G_3	C_2	C_1	indel%
A	0.1	0.1	0.3	0.0	99.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	37.4	38.4	38.1	21.3	2.0	0.0	0.0	1.02
C	99.8	99.9	99.6	99.8	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	99.6	0.0	0.0	0.4	0.1	0.0	0.0	100.0	100.0	
G	0.0	0.0	0.0	0.1	0.0	99.9	0.0	99.9	99.8	0.0	0.0	99.9	0.0	0.0	0.1	0.0	62.4	60.8	61.5	78.6	98.0	0.0	0.0	
T	0.1	0.0	0.1	0.1	0.0	0.0	100.0	0.1	0.2	0.0	100.0	0.0	0.0	100.0	0.3	100.0	0.2	0.4	0.3	0.1	0.0	0.0	0.0	

xCas9(3.6)–BE3

Site 3	C	C	C	C_{20}	A_{19}	G_{18}	T_{17}	G_{16}	G_{15}	C_{14}	T_{13}	G_{12}	C_{11}	T_{10}	C_9	T_8	G_7	G_6	G_5	G_4	G_3	C_2	C_1	indel%
A	0.0	0.0	0.1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	30.4	31.1	30.7	14.7	0.9	0.0	0.0	0.91
C	99.9	100.0	99.9	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.2	0.0	0.0	0.0	100.0	100.0	
G	0.0	0.0	0.0	0.0	0.0	99.9	0.0	99.9	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	69.4	68.6	69.1	85.3	99.1	0.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.2	0.1	0.2	0.1	0.0	0.0	0.0	

xCas9(3.7)–BE3

Site 3	C	C	C	C_{20}	A_{19}	G_{18}	T_{17}	G_{16}	G_{15}	C_{14}	T_{13}	G_{12}	C_{11}	T_{10}	C_9	T_8	G_7	G_6	G_5	G_4	G_3	C_2	C_1	indel%
A	0.0	0.0	0.2	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	50.8	51.9	49.7	20.3	1.1	0.0	0.0	1.29
C	99.9	99.9	99.7	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	99.9	0.0	0.1	0.3	0.1	0.0	0.0	100.0	100.0	
G	0.0	0.0	0.0	0.0	0.0	99.9	0.0	99.9	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	48.9	47.4	50.0	79.6	98.9	0.0	0.0	
T	0.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	99.9	0.2	0.4	0.2	0.1	0.1	0.0	0.0	

No sgRNA control

Site 4	T_1	G_2	G_3	C_4	A_5	C_6	T_7	G_8	C_9	G_{10}	G_{11}	C_{12}	T_{13}	G_{14}	G_{15}	A_{16}	G_{17}	G_{18}	T_{19}	G_{20}	G	G	G	indel%
A	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.03
C	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
G	0.0	100.0	99.9	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	
T	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	

SpCas9–BE3

Site 4	T_1	G_2	G_3	C_4	A_5	C_6	T_7	G_8	C_9	G_{10}	G_{11}	C_{12}	T_{13}	G_{14}	G_{15}	A_{16}	G_{17}	G_{18}	T_{19}	G_{20}	G	G	G	indel%
A	0.0	0.0	0.0	0.0	100.0	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.16
C	0.0	0.0	0.0	98.6	0.0	55.5	0.0	0.0	99.1	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
G	0.0	100.0	100.0	0.0	0.0	3.8	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	99.9	0.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	
T	100.0	0.0	0.0	1.4	0.0	39.2	100.0	0.0	0.8	0.0	0.0	0.1	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	

xCas9(3.6)–BE3

Site 4	T_1	G_2	G_3	C_4	A_5	C_6	T_7	G_8	C_9	G_{10}	G_{11}	C_{12}	T_{13}	G_{14}	G_{15}	A_{16}	G_{17}	G_{18}	T_{19}	G_{20}	G	G	G	indel%
A	0.0	0.0	0.0	0.0	100.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.79
C	0.0	0.0	0.0	98.5	0.0	49.1	0.0	0.0	99.5	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
G	0.0	100.0	100.0	0.0	0.0	2.8	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	
T	100.0	0.0	0.0	1.5	0.0	46.9	100.0	0.0	0.5	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

Site 4	T_1	G_2	G_3	C_4	A_5	C_6	T_7	G_8	C_9	G_{10}	G_{11}	C_{12}	T_{13}	G_{14}	G_{15}	A_{16}	G_{17}	G_{18}	T_{19}	G_{20}	G	G	G	indel%
A	0.0	0.0	0.0	0.0	100.0	1.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.80
C	0.0	0.0	0.0	98.7	0.0	38.1	0.0	0.0	98.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
G	0.0	100.0	100.0	0.0	0.0	4.1	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	
T	100.0	0.0	0.0	1.2	0.0	56.2	100.0	0.0	1.1	0.0	0.0	0.1	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	

No sgRNA control

Site 5	G_1	A_2	G_3	C_4	G_5	A_6	G_7	C_8	A_9	G_{10}	C_{11}	G_{12}	T_{13}	C_{14}	T_{15}	T_{16}	C_{17}	G_{18}	A_{19}	G_{20}	A	G	T	indel%
A	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	100.0	0.0	0.0	0.50
C	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.2	0.0	0.0	0.0	0.0	
G	100.0	0.0	100.0	0.0	99.9	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	99.9	0.0	100.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.1	0.0	0.0	100.0	

SpCas9–BE3

Site 5	G_1	A_2	G_3	C_4	G_5	A_6	G_7	C_8	A_9	G_{10}	C_{11}	G_{12}	T_{13}	C_{14}	T_{15}	T_{16}	C_{17}	G_{18}	A_{19}	G_{20}	A	G	T	indel%
A	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	99.9	0.0	0.0	0.38
C	0.0	0.0	0.0	99.9	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.2	0.0	0.0	0.0	0.0	
G	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	99.9	0.0	100.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	

xCas9(3.6)–BE3

Site 5	G_1	A_2	G_3	C_4	G_5	A_6	G_7	C_8	A_9	G_{10}	C_{11}	G_{12}	T_{13}	C_{14}	T_{15}	T_{16}	C_{17}	G_{18}	A_{19}	G_{20}	A	G	T	indel%
A	0.0	99.9	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	100.0	0.0	0.0	0.59
C	0.0	0.0	0.0	99.7	0.0	0.0	0.0	99.5	0.0	0.0	99.9	0.0	0.0	93.8	0.0	0.0	93.6	0.0	0.2	0.0	0.0	0.0	0.0	
G	99.9	0.0	100.0	0.0	99.9	0.0	100.0	0.0	0.0	99.9	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	99.9	0.0	100.0	0.0	
T	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.5	0.0	0.0	0.1	0.0	100.0	6.2	100.0	100.0	6.4	0.1	0.0	0.1	0.0	0.0	100.0	

xCas9(3.7)–BE3

Site 5	G_1	A_2	G_3	C_4	G_5	A_6	G_7	C_8	A_9	G_{10}	C_{11}	G_{12}	T_{13}	C_{14}	T_{15}	T_{16}	C_{17}	G_{18}	A_{19}	G_{20}	A	G	T	indel%
A	0.0	100.0	0.0	0.0	0.0	99.9	0.0	0.1	100.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	99.7	0.0	99.9	0.0	0.0	0.81
C	0.0	0.0	0.0	99.6	0.0	0.0	0.0	99.3	0.0	0.0	99.8	0.0	0.0	91.9	0.0	0.0	91.6	0.0	0.2	0.0	0.0	0.0	0.0	
G	99.9	0.0	99.9	0.0	99.9	0.0	100.0	0.0	0.0	100.0	0.0	99.9	0.1	0.1	0.0	0.0	0.0	100.0	0.0	99.9	0.0	100.0	0.0	
T	0.0	0.0	0.1	0.3	0.1	0.0	0.0	0.7	0.0	0.0	0.2	0.0	99.9	8.0	100.0	100.0	8.3	0.0	0.0	0.1	0.0	0.0	100.0	

SpCas9–ABE

Site 5	G_1	A_2	G_3	C_4	G_5	A_6	G_7	C_8	A_9	G_{10}	C_{11}	G_{12}	T_{13}	C_{14}	T_{15}	T_{16}	C_{17}	G_{18}	A_{19}	G_{20}	A	G	T	indel%
A	0.0	100.0	0.0	0.0	0.0	99.2	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.81
C	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	
G	99.9	0.0	100.0	0.0	100.0	0.8	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	

Supplementary Table 14 (Continued)

xCas9(3.6)–ABE

Site 5	G_1	A_2	G_3	C_4	G_5	A_6	G_7	C_8	A_9	G_{10}	C_{11}	G_{12}	T_{13}	C_{14}	T_{15}	T_{16}	C_{17}	G_{18}	A_{19}	G_{20}	A	G	T	indel%
A	0.0	100.0	0.0	0.1	0.0	97.0	0.0	0.0	99.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	1.79
C	0.0	0.0	0.0	99.9	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	99.9	0.0	99.9	3.0	99.9	0.0	0.3	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	100.0	0.0	100.0	0.0	
T	0.0	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	

xCas9(3.7)–ABE

Site 5	G_1	A_2	G_3	C_4	G_5	A_6	G_7	C_8	A_9	G_{10}	C_{11}	G_{12}	T_{13}	C_{14}	T_{15}	T_{16}	C_{17}	G_{18}	A_{19}	G_{20}	A	G	T	indel%
A	0.0	99.9	0.0	0.0	0.0	94.1	0.0	0.1	99.2	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	99.9	0.0	0.0	2.50
C	0.0	0.0	0.0	99.9	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	
G	99.9	0.0	99.9	0.0	100.0	5.9	99.9	0.0	0.8	99.9	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	99.9	0.0	100.0	0.0	
T	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	

No sgRNA control

Site 6	G_1	C_2	G_3	G_4	T_5	C_6	T_7	C_8	A_9	A_{10}	G_{11}	C_{12}	A_{13}	C_{14}	T_{15}	A_{16}	C_{17}	C_{18}	T_{19}	A_{20}	C	G	T	indel%
A	0.02	0.01	0.02	0.02	0	0.02	0.01	0.01	99.9	87.7	0.02	0.01	99.9	0.02	0	99.9	0.01	0.01	0	99.9	0.01	0	0	0.25
C	0	99.9	0	0	0.01	100	0.03	100	0.05	12.3	0	100	0.04	100	0.01	0.06	100	100	0.01	0.05	100	0	0.01	
G	99.9	0	99.9	99.9	0	0	0	0	0.03	0.01	99.9	0.01	0	0	0	0.01	0	0	0.01	0.01	0	99.9	0.02	
T	0.05	0.12	0.06	0.05	100	0.01	100	0.02	0	0.01	0.05	0.03	0.01	0.01	100	0.01	0.02	0.01	100	0.02	0.01	0.07	100	

SpCas9–BE3

Site 6	G_1	C_2	G_3	G_4	T_5	C_6	T_7	C_8	A_9	A_{10}	G_{11}	C_{12}	A_{13}	C_{14}	T_{15}	A_{16}	C_{17}	C_{18}	T_{19}	A_{20}	C	G	T	indel%
A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	87.8	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.53
C	0.0	99.9	0.0	0.0	0.0	98.9	0.0	98.9	0.0	12.2	0.0	100.0	0.0	99.9	0.0	0.1	100.0	100.0	0.0	0.0	100.0	0.0	0.0	
G	99.9	0.0	99.9	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	
T	0.0	0.1	0.1	0.0	100.0	1.1	100.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.1	100.0	

xCas9(3.6)–BE3

Site 6	G_1	C_2	G_3	G_4	T_5	C_6	T_7	C_8	A_9	A_{10}	G_{11}	C_{12}	A_{13}	C_{14}	T_{15}	A_{16}	C_{17}	C_{18}	T_{19}	A_{20}	C	G	T	indel%
A	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	99.9	88.3	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	99.9	0.0	0.0	0.0	1.61
C	0.0	99.9	0.0	0.0	0.0	67.7	0.0	69.3	0.1	11.6	0.0	99.9	0.1	99.2	0.0	0.1	99.8	100.0	0.0	0.0	100.0	0.0	0.0	
G	99.9	0.0	99.9	100.0	0.0	0.1	0.0	0.5	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	
T	0.0	0.1	0.1	0.0	100.0	32.1	100.0	30.0	0.0	0.0	0.0	0.0	0.0	0.8	100.0	0.0	0.2	0.0	100.0	0.0	0.0	0.1	100.0	

xCas9(3.7)–BE3

Site 6	G_1	C_2	G_3	G_4	T_5	C_6	T_7	C_8	A_9	A_{10}	G_{11}	C_{12}	A_{13}	C_{14}	T_{15}	A_{16}	C_{17}	C_{18}	T_{19}	A_{20}	C	G	T	indel%
A	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.4	99.9	89.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0	99.9	0.0	0.0	0.0	2.95
C	0.0	99.9	0.0	0.0	0.0	44.8	0.0	47.8	0.0	11.0	0.0	99.9	0.0	99.3	0.0	0.1	99.7	99.9	0.0	0.0	100.0	0.0	0.0	
G	100.0	0.0	99.9	99.9	0.0	0.3	0.0	0.8	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	
T	0.0	0.1	0.0	0.0	100.0	54.7	100.0	51.0	0.0	0.0	0.1	0.0	0.0	0.7	100.0	0.0	0.3	0.1	100.0	0.0	0.0	0.1	100.0	

No sgRNA control

Site 7	A	C	A	G_{20}	G_{19}	G_{18}	G_{17}	C_{16}	T_{15}	A_{14}	G_{13}	T_{12}	A_{11}	T_{10}	G_9	T_8	G_7	C_6	A_5	G_4	C_3	T_2	C_1	indel%
A	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.04
C	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	
G	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	

SpCas9–BE3

Site 7	A	C	A	G_{20}	G_{19}	G_{18}	G_{17}	C_{16}	T_{15}	A_{14}	G_{13}	T_{12}	A_{11}	T_{10}	G_9	T_8	G_7	C_6	A_5	G_4	C_3	T_2	C_1	indel%
A	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.1	0.0	0.3	0.0	100.0	0.1	0.0	0.0	0.0	0.03
C	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	
G	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	99.9	0.0	99.7	0.0	0.0	99.9	0.0	0.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	

xCas9(3.6)–BE3

Site 7	A	C	A	G_{20}	G_{19}	G_{18}	G_{17}	C_{16}	T_{15}	A_{14}	G_{13}	T_{12}	A_{11}	T_{10}	G_9	T_8	G_7	C_6	A_5	G_4	C_3	T_2	C_1	indel%
A	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.7	0.0	100.0	0.0	0.7	0.0	7.3	0.0	100.0	3.7	0.0	0.0	0.0	0.04
C	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	
G	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	99.3	0.0	0.0	0.0	99.2	0.0	92.6	0.0	0.0	96.3	0.0	0.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	

xCas9(3.7)–BE3

Site 7	A	C	A	G_{20}	G_{19}	G_{18}	G_{17}	C_{16}	T_{15}	A_{14}	G_{13}	T_{12}	A_{11}	T_{10}	G_9	T_8	G_7	C_6	A_5	G_4	C_3	T_2	C_1	indel%
A	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.8	0.0	100.0	0.0	0.9	0.0	10.2	0.0	100.0	4.9	0.0	0.0	0.0	0.04
C	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	99.9	
G	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	99.2	0.0	0.0	0.0	99.1	0.0	89.7	0.0	0.0	95.0	0.0	0.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	99.9	0.1	0.0	0.0	0.1	0.0	100.0	0.0	

No sgRNA control

Site 8	G_1	C_2	T_3	T_4	C_5	T_6	C_7	C_8	A_9	G_{10}	C_{11}	C_{12}	C_{13}	T_{14}	G_{15}	G_{16}	C_{17}	C_{18}	T_{19}	G_{20}	G	G	T	indel%
A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.03
C	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	0.0	100.0	100.0	0.1	0.0	0.0	0.0	0.0	
G	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	
T	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	100.0	

SpCas9–BE3

Site 8	G_1	C_2	T_3	T_4	C_5	T_6	C_7	C_8	A_9	G_{10}	C_{11}	C_{12}	C_{13}	T_{14}	G_{15}	G_{16}	C_{17}	C_{18}	T_{19}	G_{20}	G	G	T	indel%
A	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.29
C	0.0	99.9	0.0	0.0	90.5	0.0	90.4	92.8	0.0	0.0	100.0	100.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	
G	99.9	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	
T	0.0	0.1	100.0	100.0	9.5	100.0	9.4	7.2	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	

Supplementary Table 14 (Continued)

xCas9(3.6)–BE3

Site 8	G_1	C_2	T_3	T_4	C_5	T_6	C_7	C_8	A_9	G_{10}	C_{11}	C_{12}	C_{13}	T_{14}	G_{15}	G_{16}	C_{17}	C_{18}	T_{19}	G_{20}	G	G	T	indel%
A	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.1	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
C	0.0	99.8	0.0	0.0	59.8	0.0	59.5	68.6	0.0	0.0	100.0	99.9	99.9	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	
G	99.9	0.0	0.0	0.0	0.0	0.0	1.5	0.0	0.0	99.9	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	
T	0.1	0.2	100.0	100.0	40.2	100.0	38.7	31.3	0.0	0.1	0.0	0.0	0.1	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	4.44

xCas9(3.7)–BE3

Site 8	G_1	C_2	T_3	T_4	C_5	T_6	C_7	C_8	A_9	G_{10}	C_{11}	C_{12}	C_{13}	T_{14}	G_{15}	G_{16}	C_{17}	C_{18}	T_{19}	G_{20}	G	G	T	indel%
A	0.0	0.0	0.0	0.0	0.1	0.0	0.3	0.1	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
C	0.0	99.8	0.0	0.0	53.8	0.0	53.6	62.8	0.0	0.0	99.9	99.9	99.9	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	
G	99.9	0.0	0.0	0.0	0.0	0.0	1.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	
T	0.1	0.2	100.0	100.0	46.1	100.0	44.2	37.1	0.0	0.0	0.1	0.1	0.1	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	4.20

No sgRNA control

Site 9	G	C	T	G_{20}	T_{19}	G_{18}	T_{17}	G_{16}	A_{15}	C_{14}	T_{13}	A_{12}	C_{11}	A_{10}	G_9	T_8	G_7	G_6	G_5	G_4	G_3	C_2	C_1	indel%
A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
C	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	
G	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	
T	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.03

SpCas9–BE3

Site 9	G	C	T	G_{20}	T_{19}	G_{18}	T_{17}	G_{16}	A_{15}	C_{14}	T_{13}	A_{12}	C_{11}	A_{10}	G_9	T_8	G_7	G_6	G_5	G_4	G_3	C_2	C_1	indel%
A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.5	0.5	0.5	0.3	0.0	0.0	0.0	
C	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	
G	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	99.4	99.5	99.5	99.7	99.9	0.0	0.0	
T	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.07

xCas9(3.6)–BE3

Site 9	G	C	T	G_{20}	T_{19}	G_{18}	T_{17}	G_{16}	A_{15}	C_{14}	T_{13}	A_{12}	C_{11}	A_{10}	G_9	T_8	G_7	G_6	G_5	G_4	G_3	C_2	C_1	indel%
A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	100.0	0.0	0.0	100.0	0.0	100.0	1.1	0.0	4.8	4.7	4.3	2.0	0.2	0.0	0.0	
C	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	
G	100.0	0.0	0.0	100.0	0.0	100.0	0.0	99.1	0.0	0.0	0.0	0.0	0.0	0.0	98.9	0.0	95.2	95.3	95.6	98.0	99.8	0.0	0.0	
T	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.17

xCas9(3.7)–BE3

Site 9	G	C	T	G_{20}	T_{19}	G_{18}	T_{17}	G_{16}	A_{15}	C_{14}	T_{13}	A_{12}	C_{11}	A_{10}	G_9	T_8	G_7	G_6	G_5	G_4	G_3	C_2	C_1	indel%
A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.4	100.0	0.0	0.0	100.0	0.0	100.0	2.4	0.0	10.5	10.4	9.5	4.1	0.4	0.0	0.0	
C	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	
G	100.0	0.0	0.0	100.0	0.0	99.9	0.0	97.6	0.0	0.0	0.0	0.0	0.0	0.0	97.5	0.0	89.5	89.6	90.5	95.9	99.6	0.0	0.0	
T	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.27

No sgRNA control

Site 10	G_1	A_2	C_3	A_4	G_5	G_6	C_7	A_8	G_9	G_{10}	G_{11}	G_{12}	C_{13}	A_{14}	C_{15}	C_{16}	G_{17}	C_{18}	G_{19}	G_{20}	C	G	C	indel%
A	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
C	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	
G	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.04

SpCas9–BE3

Site 10	G_1	A_2	C_3	A_4	G_5	G_6	C_7	A_8	G_9	G_{10}	G_{11}	G_{12}	C_{13}	A_{14}	C_{15}	C_{16}	G_{17}	C_{18}	G_{19}	G_{20}	C	G	C	indel%
A	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
C	0.0	0.0	99.9	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	
G	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	99.9	100.0	0.0	100.0	0.0	
T	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.04

xCas9(3.6)–BE3

Site 10	G_1	A_2	C_3	A_4	G_5	G_6	C_7	A_8	G_9	G_{10}	G_{11}	G_{12}	C_{13}	A_{14}	C_{15}	C_{16}	G_{17}	C_{18}	G_{19}	G_{20}	C	G	C	indel%
A	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
C	0.0	0.0	98.2	0.0	0.0	0.0	95.7	0.0	0.0	0.0	0.0	0.0	100.0	0.0	99.7	99.9	0.0	100.0	0.0	0.0	100.0	0.0	100.0	
G	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	
T	0.0	0.0	1.8	0.0	0.0	0.0	4.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.09

xCas9(3.7)–BE3

Site 10	G_1	A_2	C_3	A_4	G_5	G_6	C_7	A_8	G_9	G_{10}	G_{11}	G_{12}	C_{13}	A_{14}	C_{15}	C_{16}	G_{17}	C_{18}	G_{19}	G_{20}	C	G	C	indel%
A	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
C	0.0	0.0	97.4	0.0	0.0	0.0	94.2	0.0	0.0	0.0	0.0	0.0	100.0	0.0	99.6	99.9	0.0	100.0	0.0	0.0	100.0	0.0	100.0	
G	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	
T	0.0	0.0	2.6	0.0	0.0	0.0	5.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.13

No sgRNA control

Site 11	G	C	A	C_{20}	C_{19}	G_{18}	C_{17}	G_{16}	G_{15}	C_{14}	G_{13}	C_{12}	C_{11}	C_{10}	C_9	G_8	G_7	T_6	G_5	G_4	C_3	A_2	C_1	indel%
A	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	
C	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	
G	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	99.9	0.0	0.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.04

SpCas9–BE3

Site 11	G	C	A	C_{20}	C_{19}	G_{18}	C_{17}	G_{16}	G_{15}	C_{14}	G_{13}	C_{12}	C_{11}	C_{10}	C_9	G_8	G_7	T_6	G_5	G_4	C_3	A_2	C_1	indel%
A	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.0	0.3	0.2	0.0	100.0	0.0	
C	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	
G	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	99.8	99.7	0.0	99.7	99.8	0.0	0.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.06

Supplementary Table 14 (Continued)

xCas9(3.6)–BE3

Site 11	G	C	A	C_{20}	C_{19}	G_{18}	C_{17}	G_{16}	G_{15}	C_{14}	G_{13}	C_{12}	C_{11}	C_{10}	C_9	G_8	G_7	T_6	G_5	G_4	C_3	A_2	C_1	indel%
A	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.5	1.7	0.0	1.8	0.4	0.0	100.0	0.0	0.26
C	0.0	100.0	0.0	99.9	99.9	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	
G	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	98.5	98.3	0.0	98.2	99.6	0.0	0.0	0.0	
T	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

Site 11	G	C	A	C_{20}	C_{19}	G_{18}	C_{17}	G_{16}	G_{15}	C_{14}	G_{13}	C_{12}	C_{11}	C_{10}	C_9	G_8	G_7	T_6	G_5	G_4	C_3	A_2	C_1	indel%
A	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.3	5.3	0.0	5.4	1.4	0.0	100.0	0.0	0.40
C	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	
G	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	95.7	94.6	0.0	94.6	98.6	0.0	0.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	100.0	0.0	0.0	0.0	0.0	0.0	

SpCas9–ABE

Site 11	G	C	A	C_{20}	C_{19}	G_{18}	C_{17}	G_{16}	G_{15}	C_{14}	G_{13}	C_{12}	C_{11}	C_{10}	C_9	G_8	G_7	T_6	G_5	G_4	C_3	A_2	C_1	indel%
A	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.07
C	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	2.2	0.0	0.0	100.0	0.0	100.0	
G	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	97.8	0.0	0.0	0.0	0.0	0.0	

xCas9(3.6)–ABE

Site 11	G	C	A	C_{20}	C_{19}	G_{18}	C_{17}	G_{16}	G_{15}	C_{14}	G_{13}	C_{12}	C_{11}	C_{10}	C_9	G_8	G_7	T_6	G_5	G_4	C_3	A_2	C_1	indel%
A	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.37
C	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	17.4	0.0	0.0	100.0	0.0	100.0	
G	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	82.6	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–ABE

Site 11	G	C	A	C_{20}	C_{19}	G_{18}	C_{17}	G_{16}	G_{15}	C_{14}	G_{13}	C_{12}	C_{11}	C_{10}	C_9	G_8	G_7	T_6	G_5	G_4	C_3	A_2	C_1	indel%
A	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.27
C	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	12.6	0.0	0.0	100.0	0.0	100.0	
G	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	87.4	0.0	0.0	0.0	0.0	0.0	

No sgRNA control

Site 12	G_1	C_2	A_3	G_4	A_5	C_6	G_7	G_8	C_9	A_{10}	G_{11}	T_{12}	C_{13}	A_{14}	C_{15}	T_{16}	A_{17}	G_{18}	G_{19}	G_{20}	G	G	C	indel%
A	0.1	0.0	100.0	0.0	100.0	0.0	0.0	0.1	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.22
C	0.0	99.9	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	
G	99.9	0.0	0.0	100.0	0.0	0.0	100.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0	0.0	
T	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

SpCas9–BE3

Site 12	G_1	C_2	A_3	G_4	A_5	C_6	G_7	G_8	C_9	A_{10}	G_{11}	T_{12}	C_{13}	A_{14}	C_{15}	T_{16}	A_{17}	G_{18}	G_{19}	G_{20}	G	G	C	indel%
A	0.1	0.0	100.0	0.0	100.0	0.0	0.0	0.1	0.0	100.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.27
C	0.0	99.8	0.0	0.0	0.0	99.0	0.0	0.0	100.0	0.0	0.0	0.0	99.7	0.1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	
G	99.9	0.0	0.0	100.0	0.0	0.0	100.0	99.8	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.1	100.0	100.0	99.9	100.0	100.0	0.0	
T	0.0	0.1	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	99.9	0.2	0.0	0.0	99.9	0.0	0.0	0.0	0.1	0.0	0.0	0.1	

xCas9(3.6)–BE3

Site 12	G_1	C_2	A_3	G_4	A_5	C_6	G_7	G_8	C_9	A_{10}	G_{11}	T_{12}	C_{13}	A_{14}	C_{15}	T_{16}	A_{17}	G_{18}	G_{19}	G_{20}	G	G	C	indel%
A	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.18
C	0.0	99.9	0.0	0.0	0.0	99.0	0.0	0.0	100.0	0.0	0.0	0.0	99.8	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	
G	99.9	0.0	0.0	100.0	0.0	0.0	100.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0	0.0	
T	0.0	0.1	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	100.0	0.2	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

Site 12	G_1	C_2	A_3	G_4	A_5	C_6	G_7	G_8	C_9	A_{10}	G_{11}	T_{12}	C_{13}	A_{14}	C_{15}	T_{16}	A_{17}	G_{18}	G_{19}	G_{20}	G	G	C	indel%
A	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.1	0.0	100.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.26
C	0.0	99.9	0.0	0.0	0.0	96.0	0.0	0.0	100.0	0.0	0.0	0.0	99.2	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	
G	99.9	0.0	0.0	100.0	0.0	0.0	100.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0	0.0	
T	0.0	0.1	0.0	0.0	0.0	3.9	0.0	0.0	0.0	0.0	0.0	100.0	0.8	0.0	0.1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

SpCas9–ABE

Site 12	G_1	C_2	A_3	G_4	A_5	C_6	G_7	G_8	C_9	A_{10}	G_{11}	T_{12}	C_{13}	A_{14}	C_{15}	T_{16}	A_{17}	G_{18}	G_{19}	G_{20}	G	G	C	indel%
A	0.0	0.0	99.7	0.0	89.0	0.0	0.0	0.0	0.0	99.7	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.60
C	0.0	99.9	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	
G	100.0	0.0	0.3	100.0	11.0	0.0	100.0	99.9	0.0	0.3	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0	0.0	
T	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	

xCas9(3.6)–ABE

Site 12	G_1	C_2	A_3	G_4	A_5	C_6	G_7	G_8	C_9	A_{10}	G_{11}	T_{12}	C_{13}	A_{14}	C_{15}	T_{16}	A_{17}	G_{18}	G_{19}	G_{20}	G	G	C	indel%
A	0.0	0.0	99.6	0.0	83.7	0.0	0.0	0.1	0.0	99.4	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.44
C	0.0	99.9	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	
G	99.9	0.0	0.4	100.0	16.3	0.0	100.0	99.9	0.0	0.6	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0	0.0	
T	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	

xCas9(3.7)–ABE

Site 12	G_1	C_2	A_3	G_4	A_5	C_6	G_7	G_8	C_9	A_{10}	G_{11}	T_{12}	C_{13}	A_{14}	C_{15}	T_{16}	A_{17}	G_{18}	G_{19}	G_{20}	G	G	C	indel%
A	0.0	0.0	99.5	0.0	77.8	0.0	0.0	0.0	0.0	99.2	0.0	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	1.17
C	0.0	99.9	0.0	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	99.9	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	
G	100.0	0.0	0.5	100.0	22.2	0.0	100.0	100.0	0.0	0.8	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0	0.0	
T	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	

Supplementary Table 14 (Continued)

No sgRNA control

Site 13	G_1	T_2	C_3	G_4	C_5	A_6	G_7	G_8	A_9	C_{10}	A_{11}	G_{12}	C_{13}	T_{14}	T_{15}	T_{16}	T_{17}	C_{18}	C_{19}	T_{20}	A	G	A	indel%
A	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.04
C	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	
T	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	

SpCas9–BE3

Site 13	G_1	T_2	C_3	G_4	C_5	A_6	G_7	G_8	A_9	C_{10}	A_{11}	G_{12}	C_{13}	T_{14}	T_{15}	T_{16}	T_{17}	C_{18}	C_{19}	T_{20}	A	G	A	indel%
A	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.05
C	0.0	0.0	98.7	0.0	98.7	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.9	99.9	0.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	0.0	99.9	99.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	
T	0.0	100.0	1.3	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.1	0.1	100.0	0.0	0.0	0.0	

xCas9(3.6)–BE3

Site 13	G_1	T_2	C_3	G_4	C_5	A_6	G_7	G_8	A_9	C_{10}	A_{11}	G_{12}	C_{13}	T_{14}	T_{15}	T_{16}	T_{17}	C_{18}	C_{19}	T_{20}	A	G	A	indel%
A	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.08
C	0.0	0.0	95.2	0.0	94.4	0.0	0.0	0.0	0.0	99.8	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.8	100.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	
T	0.0	100.0	4.7	0.0	5.6	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.2	0.0	100.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

Site 13	G_1	T_2	C_3	G_4	C_5	A_6	G_7	G_8	A_9	C_{10}	A_{11}	G_{12}	C_{13}	T_{14}	T_{15}	T_{16}	T_{17}	C_{18}	C_{19}	T_{20}	A	G	A	indel%
A	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.22
C	0.0	0.0	87.6	0.0	85.0	0.0	0.0	0.0	0.0	99.7	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.7	100.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	
T	0.0	100.0	12.4	0.0	15.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.3	0.0	100.0	0.0	0.0	0.0	

SpCas9–ABE

Site 13	G_1	T_2	C_3	G_4	C_5	A_6	G_7	G_8	A_9	C_{10}	A_{11}	G_{12}	C_{13}	T_{14}	T_{15}	T_{16}	T_{17}	C_{18}	C_{19}	T_{20}	A	G	A	indel%
A	0.0	0.0	0.0	0.0	0.0	94.2	0.0	0.0	99.8	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.05
C	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	99.9	0.0	5.8	100.0	100.0	0.2	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	100.0	0.0	
T	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	

xCas9(3.6)–ABE

Site 13	G_1	T_2	C_3	G_4	C_5	A_6	G_7	G_8	A_9	C_{10}	A_{11}	G_{12}	C_{13}	T_{14}	T_{15}	T_{16}	T_{17}	C_{18}	C_{19}	T_{20}	A	G	A	indel%
A	0.0	0.0	0.0	0.0	0.0	69.1	0.0	0.0	99.1	0.0	99.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	100.0	0.06
C	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	30.9	100.0	100.0	0.9	0.0	0.6	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	
T	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	

xCas9(3.7)–ABE

Site 13	G_1	T_2	C_3	G_4	C_5	A_6	G_7	G_8	A_9	C_{10}	A_{11}	G_{12}	C_{13}	T_{14}	T_{15}	T_{16}	T_{17}	C_{18}	C_{19}	T_{20}	A	G	A	indel%
A	0.0	0.0	0.0	0.0	0.0	53.0	0.0	0.0	98.3	0.0	99.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.08
C	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	46.9	100.0	100.0	1.7	0.0	0.4	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	
T	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	

No sgRNA control

Site 14	G_1	G_2	G_3	A_4	A_5	G_6	C_7	T_8	G_9	G_{10}	G_{11}	T_{12}	G_{13}	A_{14}	A_{15}	T_{16}	G_{17}	G_{18}	A_{19}	G_{20}	C	G	A	indel%
A	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	100.0	0.19
C	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	
G	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.1	100.0	0.0	100.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

SpCas9–BE3

Site 14	G_1	G_2	G_3	A_4	A_5	G_6	C_7	T_8	G_9	G_{10}	G_{11}	T_{12}	G_{13}	A_{14}	A_{15}	T_{16}	G_{17}	G_{18}	A_{19}	G_{20}	C	G	A	indel%
A	0.0	0.0	0.1	100.0	100.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	99.9	0.0	0.0	0.1	100.0	0.16
C	0.0	0.0	0.0	0.0	0.0	0.0	86.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
G	99.9	99.9	99.8	0.0	0.0	100.0	0.8	0.0	100.0	100.0	100.0	0.0	99.9	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	99.9	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	13.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.6)–BE3

Site 14	G_1	G_2	G_3	A_4	A_5	G_6	C_7	T_8	G_9	G_{10}	G_{11}	T_{12}	G_{13}	A_{14}	A_{15}	T_{16}	G_{17}	G_{18}	A_{19}	G_{20}	C	G	A	indel%
A	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	100.0	1.15
C	0.0	0.0	0.0	0.0	0.0	0.0	85.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
G	100.0	100.0	100.0	0.0	0.0	100.0	0.4	0.0	100.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	99.9	0.0	100.0	0.0	100.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	13.7	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

Site 14	G_1	G_2	G_3	A_4	A_5	G_6	C_7	T_8	G_9	G_{10}	G_{11}	T_{12}	G_{13}	A_{14}	A_{15}	T_{16}	G_{17}	G_{18}	A_{19}	G_{20}	C	G	A	indel%
A	0.0	0.0	0.1	100.0	100.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	100.0	3.08
C	0.0	0.0	0.0	0.0	0.0	0.0	57.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
G	100.0	100.0	99.9	0.0	0.0	100.0	2.7	0.0	100.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	100.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	39.4	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

SpCas9–ABE

Site 14	G_1	G_2	G_3	A_4	A_5	G_6	C_7	T_8	G_9	G_{10}	G_{11}	T_{12}	G_{13}	A_{14}	A_{15}	T_{16}	G_{17}	G_{18}	A_{19}	G_{20}	C	G	A	indel%
A	0.0	0.0	0.0	73.0	54.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.48
C	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
G	100.0	100.0	100.0	26.9	45.5	100.0	0.0	0.0	100.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	100.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Supplementary Table 14 (Continued)

xCas9(3.6)–ABE

Site 14	G_1	G_2	G_3	A_4	A_5	G_6	C_7	T_8	G_9	G_{10}	G_{11}	T_{12}	G_{13}	A_{14}	A_{15}	T_{16}	G_{17}	G_{18}	A_{19}	G_{20}	C	G	A	indel%
A	0.0	0.0	0.0	93.6	67.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	1.25
C	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
G	100.0	100.0	100.0	6.4	32.1	100.0	0.0	0.0	100.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	100.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

Site 14	G_1	G_2	G_3	A_4	A_5	G_6	C_7	T_8	G_9	G_{10}	G_{11}	T_{12}	G_{13}	A_{14}	A_{15}	T_{16}	G_{17}	G_{18}	A_{19}	G_{20}	C	G	A	indel%
A	0.0	0.0	0.0	84.6	50.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	99.9	1.68
C	0.0	0.0	0.0	0.1	0.0	0.1	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	
G	100.0	100.0	100.0	15.4	49.7	99.9	0.1	0.1	100.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.1	100.0	0.1	100.0	0.0	
T	0.0	0.0	0.0	0.0	0.1	0.0	0.0	99.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

No sgRNA control

Site 15	G_1	C_2	G_3	G_4	A_5	G_6	A_7	C_8	T_9	C_{10}	T_{11}	G_{12}	G_{13}	T_{14}	G_{15}	C_{16}	T_{17}	G_{18}	T_{19}	G_{20}	T	G	A	indel%
A	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.03
C	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	

SpCas9–BE3

Site 15	G_1	C_2	G_3	G_4	A_5	G_6	A_7	C_8	T_9	C_{10}	T_{11}	G_{12}	G_{13}	T_{14}	G_{15}	C_{16}	T_{17}	G_{18}	T_{19}	G_{20}	T	G	A	indel%
A	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.06
C	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.6	0.0	99.8	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	99.9	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	100.0	0.2	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	99.9	0.0	100.0	0.0	0.0	

xCas9(3.6)–BE3

Site 15	G_1	C_2	G_3	G_4	A_5	G_6	A_7	C_8	T_9	C_{10}	T_{11}	G_{12}	G_{13}	T_{14}	G_{15}	C_{16}	T_{17}	G_{18}	T_{19}	G_{20}	T	G	A	indel%
A	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.19
C	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.4	0.0	97.3	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	100.0	2.6	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	

xCas9(3.7)–BE3

Site 15	G_1	C_2	G_3	G_4	A_5	G_6	A_7	C_8	T_9	C_{10}	T_{11}	G_{12}	G_{13}	T_{14}	G_{15}	C_{16}	T_{17}	G_{18}	T_{19}	G_{20}	T	G	A	indel%
A	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	1.00
C	0.0	100.0	0.0	0.0	0.0	0.0	0.0	94.1	0.0	86.8	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	100.0	100.0	0.0	99.9	0.0	0.0	0.0	0.1	0.0	100.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.8	100.0	13.1	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	

No sgRNA control

Site 16	G_1	G_2	T_3	C_4	A_5	G_6	A_7	A_8	A_9	T_{10}	A_{11}	G_{12}	G_{13}	G_{14}	G_{15}	G_{16}	T_{17}	C_{18}	C_{19}	A_{20}	G	G	A	indel%
A	0.0	0.0	0.0	0.0	99.9	0.0	99.9	99.9	99.9	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	99.9	0.11
C	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.2	0.0	0.0	0.0	
G	99.9	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	99.9	0.0	
T	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

SpCas9–BE3

Site 16	G_1	G_2	T_3	C_4	A_5	G_6	A_7	A_8	A_9	T_{10}	A_{11}	G_{12}	G_{13}	G_{14}	G_{15}	G_{16}	T_{17}	C_{18}	C_{19}	A_{20}	G	G	A	indel%
A	0.0	0.0	0.0	0.0	99.9	0.0	100.0	100.0	100.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	99.9	0.10
C	0.0	0.0	0.0	97.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.2	0.0	0.0	0.0	
G	99.9	100.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	99.9	0.0	
T	0.0	0.0	100.0	2.4	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.6)–BE3

Site 16	G_1	G_2	T_3	C_4	A_5	G_6	A_7	A_8	A_9	T_{10}	A_{11}	G_{12}	G_{13}	G_{14}	G_{15}	G_{16}	T_{17}	C_{18}	C_{19}	A_{20}	G	G	A	indel%
A	0.0	0.0	0.0	0.0	99.9	0.0	99.9	99.9	99.9	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.7	0.0	0.0	99.9	0.15
C	0.0	0.0	0.0	92.4	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0	0.2	0.0	0.0	0.1	
G	99.9	99.9	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	99.9	99.9	0.0	
T	0.1	0.0	100.0	7.6	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

Site 16	G_1	G_2	T_3	C_4	A_5	G_6	A_7	A_8	A_9	T_{10}	A_{11}	G_{12}	G_{13}	G_{14}	G_{15}	G_{16}	T_{17}	C_{18}	C_{19}	A_{20}	G	G	A	indel%
A	0.0	0.0	0.0	0.0	99.9	0.0	99.9	99.9	99.9	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	99.9	0.21
C	0.0	0.0	0.0	86.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0	0.2	0.0	0.0	0.1	
G	99.9	100.0	0.0	0.1	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	99.9	99.9	0.0	
T	0.1	0.0	100.0	13.8	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0	

SpCas9–ABE

Site 16	G_1	G_2	T_3	C_4	A_5	G_6	A_7	A_8	A_9	T_{10}	A_{11}	G_{12}	G_{13}	G_{14}	G_{15}	G_{16}	T_{17}	C_{18}	C_{19}	A_{20}	G	G	A	indel%
A	0.0	0.0	0.0	0.0	86.4	0.0	94.2	99.1	99.5	0.0	99.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.13
C	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	
G	100.0	100.0	0.0	0.0	13.6	100.0	5.8	0.9	0.5	0.0	0.5	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	
T	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.6)–ABE

Site 16	G_1	G_2	T_3	C_4	A_5	G_6	A_7	A_8	A_9	T_{10}	A_{11}	G_{12}	G_{13}	G_{14}	G_{15}	G_{16}	T_{17}	C_{18}	C_{19}	A_{20}	G	G	A	indel%
A	0.0	0.0	0.0	0.0	64.5	0.0	88.4	98.6	98.9	0.0	98.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.53
C	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	
G	100.0	100.0	0.0	0.0	35.4	100.0	11.5	1.4	1.0	0.0	1.1	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	
T	0.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

Supplementary Table 14 (Continued)

xCas9(3.7)–ABE

Site 16	G_1	G_2	T_3	C_4	A_5	G_6	A_7	A_8	A_9	T_{10}	A_{11}	G_{12}	G_{13}	G_{14}	G_{15}	G_{16}	T_{17}	C_{18}	C_{19}	A_{20}	G	G	A	indel%
A	0.0	0.0	0.0	0.0	69.2	0.0	84.2	97.2	98.2	0.0	98.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.95
C	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	
G	100.0	100.0	0.0	0.0	30.7	100.0	15.8	2.8	1.8	0.0	2.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	
T	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

No sgRNA control

Site 17	T	T	C	C_{20}	A_{19}	A_{18}	T_{17}	C_{16}	A_{15}	G_{14}	T_{13}	A_{12}	C_{11}	G_{10}	C_9	A_8	G_7	A_6	G_5	A_4	G_3	T_2	C_1	indel%
A	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.01
C	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	99.9	0.0	0.0	100.0	0.0	99.9	0.0	100.0	0.0	0.0	
T	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	

SpCas9–BE3

Site 17	T	T	C	C_{20}	A_{19}	A_{18}	T_{17}	C_{16}	A_{15}	G_{14}	T_{13}	A_{12}	C_{11}	G_{10}	C_9	A_8	G_7	A_6	G_5	A_4	G_3	T_2	C_1	indel%
A	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	99.9	0.1	100.0	0.1	100.0	0.0	0.0	0.0	0.01
C	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	99.9	0.0	0.0	99.8	0.0	99.8	0.0	99.9	0.0	0.0	
T	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	100.0	0.0	

xCas9(3.6)–BE3

Site 17	T	T	C	C_{20}	A_{19}	A_{18}	T_{17}	C_{16}	A_{15}	G_{14}	T_{13}	A_{12}	C_{11}	G_{10}	C_9	A_8	G_7	A_6	G_5	A_4	G_3	T_2	C_1	indel%
A	0.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0	100.0	0.3	0.0	100.0	0.0	0.1	0.0	100.0	4.7	100.0	4.8	100.0	0.5	0.0	0.0	0.04
C	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.7	0.0	0.0	0.0	99.9	0.0	0.0	95.3	0.0	95.2	0.0	99.5	0.0	0.0	
T	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1	

xCas9(3.7)–BE3

Site 17	T	T	C	C_{20}	A_{19}	A_{18}	T_{17}	C_{16}	A_{15}	G_{14}	T_{13}	A_{12}	C_{11}	G_{10}	C_9	A_8	G_7	A_6	G_5	A_4	G_3	T_2	C_1	indel%
A	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	0.4	0.0	100.0	0.0	0.1	0.0	99.9	10.5	100.0	10.7	100.0	0.6	0.0	0.0	0.06
C	0.0	0.0	100.0	100.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	100.0	
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.6	0.0	0.0	0.0	99.9	0.0	0.0	89.4	0.0	89.3	0.0	99.3	0.0	0.0	
T	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	100.0	0.0	

No sgRNA control

Site 18	A	T	C	C_{20}	C_{19}	T_{18}	T_{17}	C_{16}	T_{15}	G_{14}	C_{13}	A_{12}	G_{11}	C_{10}	A_9	C_8	C_7	T_6	G_5	G_4	A_3	T_2	C_1	indel%
A	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.02
C	0.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.1	0.0	100.0	0.1	100.0	100.0	0.0	0.0	0.0	0.1	0.0	100.0	
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
T	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	

SpCas9–BE3

Site 18	A	T	C	C_{20}	C_{19}	T_{18}	T_{17}	C_{16}	T_{15}	G_{14}	C_{13}	A_{12}	G_{11}	C_{10}	A_9	C_8	C_7	T_6	G_5	G_4	A_3	T_2	C_1	indel%
A	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.02
C	0.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.9	99.9	0.0	0.0	0.0	
T	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	

xCas9(3.6)–BE3

Site 18	A	T	C	C_{20}	C_{19}	T_{18}	T_{17}	C_{16}	T_{15}	G_{14}	C_{13}	A_{12}	G_{11}	C_{10}	A_9	C_8	C_7	T_6	G_5	G_4	A_3	T_2	C_1	indel%
A	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	5.6	5.5	100.0	0.0	0.0	0.07
C	0.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	94.4	94.5	0.0	0.0	0.0	
T	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	

xCas9(3.7)–BE3

Site 18	A	T	C	C_{20}	C_{19}	T_{18}	T_{17}	C_{16}	T_{15}	G_{14}	C_{13}	A_{12}	G_{11}	C_{10}	A_9	C_8	C_7	T_6	G_5	G_4	A_3	T_2	C_1	indel%
A	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	12.0	11.9	99.9	0.0	0.0	0.23
C	0.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.1	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	87.9	88.0	0.0	0.0	0.0	
T	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1	0.0	100.0	0.0	

SpCas9–ABE

Site 18	A	T	C	C_{20}	C_{19}	T_{18}	T_{17}	C_{16}	T_{15}	G_{14}	C_{13}	A_{12}	G_{11}	C_{10}	A_9	C_8	C_7	T_6	G_5	G_4	A_3	T_2	C_1	indel%
A	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.01
C	0.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	0.2	0.0	0.0	0.0	0.0	100.0	
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
T	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	100.0	0.0	

xCas9(3.6)–ABE

Site 18	A	T	C	C_{20}	C_{19}	T_{18}	T_{17}	C_{16}	T_{15}	G_{14}	C_{13}	A_{12}	G_{11}	C_{10}	A_9	C_8	C_7	T_6	G_5	G_4	A_3	T_2	C_1	indel%
A	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.30
C	0.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	16.2	0.0	0.0	0.0	0.0	100.0	
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0	0.0	0.0	0.0	
T	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	83.8	0.0	0.0	0.0	100.0	0.0	

xCas9(3.7)–ABE

Site 18	A	T	C	C_{20}	C_{19}	T_{18}	T_{17}	C_{16}	T_{15}	G_{14}	C_{13}	A_{12}	G_{11}	C_{10}	A_9	C_8	C_7	T_6	G_5	G_4	A_3	T_2	C_1	indel%
A	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.46
C	0.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	18.5	0.0	0.0	0.0	0.0	100.0	
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
T	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	81.5	0.0	0.0	0.0	100.0	0.0	

Supplementary Table 14 (Continued)

No sgRNA control

Site 19	G_1	T_2	C_3	A_4	C_5	T_6	C_7	C_8	A_9	G_{10}	G_{11}	A_{12}	T_{13}	T_{14}	C_{15}	C_{16}	A_{17}	A_{18}	T_{19}	A_{20}	G	A	T	indel%
A	0.0	0.0	0.0	95.2	0.0	0.0	0.0	0.0	95.1	0.0	0.0	99.9	0.0	0.0	0.0	0.0	99.9	99.9	0.0	99.9	0.0	99.9	0.0	0.03
C	0.0	0.0	100.0	4.8	100.0	0.0	100.0	100.0	4.9	0.0	0.0	0.1	0.0	0.0	100.0	100.0	0.0	0.1	0.0	0.1	0.0	0.1	0.0	
G	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
T	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.1	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	

SpCas9–BE3

Site 19	G_1	T_2	C_3	A_4	C_5	T_6	C_7	C_8	A_9	G_{10}	G_{11}	A_{12}	T_{13}	T_{14}	C_{15}	C_{16}	A_{17}	A_{18}	T_{19}	A_{20}	G	A	T	indel%
A	0.0	0.0	0.0	95.3	0.0	0.0	0.0	0.0	95.3	0.0	0.0	99.9	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	100.0	0.0	0.23
C	0.0	0.0	100.0	4.7	100.0	0.0	100.0	100.0	4.7	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	
T	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.1	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1	0.0	100.0	

xCas9(3.6)–BE3

Site 19	G_1	T_2	C_3	A_4	C_5	T_6	C_7	C_8	A_9	G_{10}	G_{11}	A_{12}	T_{13}	T_{14}	C_{15}	C_{16}	A_{17}	A_{18}	T_{19}	A_{20}	G	A	T	indel%
A	0.0	0.0	0.0	95.3	0.0	0.0	0.0	0.0	95.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	100.0	0.0	0.18
C	0.0	0.0	98.9	4.7	92.8	0.0	90.0	92.0	5.0	0.0	0.0	0.0	0.0	0.0	97.7	98.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
T	0.0	100.0	1.1	0.0	7.1	100.0	10.0	8.0	0.0	0.0	0.0	0.0	100.0	100.0	2.3	1.1	0.0	0.0	100.0	0.0	0.0	0.0	100.0	

xCas9(3.7)–BE3

Site 19	G_1	T_2	C_3	A_4	C_5	T_6	C_7	C_8	A_9	G_{10}	G_{11}	A_{12}	T_{13}	T_{14}	C_{15}	C_{16}	A_{17}	A_{18}	T_{19}	A_{20}	G	A	T	indel%
A	0.0	0.0	0.0	95.3	0.0	0.0	0.0	0.0	95.2	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.9	99.9	0.0	99.9	0.0	100.0	0.0	0.33
C	0.0	0.0	97.8	4.7	87.2	0.0	83.1	86.8	4.7	0.0	0.0	0.0	0.0	0.0	95.8	97.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
T	0.0	100.0	2.2	0.0	12.8	100.0	16.8	13.2	0.0	0.0	0.0	0.0	100.0	100.0	4.2	2.1	0.0	0.0	100.0	0.0	0.0	0.0	100.0	

No sgRNA control

Site 20	G_1	T_2	G_3	G_4	T_5	T_6	C_7	C_8	A_9	G_{10}	A_{11}	A_{12}	C_{13}	C_{14}	G_{15}	G_{16}	A_{17}	G_{18}	G_{19}	A_{20}	C	A	A	indel%
A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	0.02
C	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
G	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	
T	0.0	100.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

SpCas9–BE3

Site 20	G_1	T_2	G_3	G_4	T_5	T_6	C_7	C_8	A_9	G_{10}	A_{11}	A_{12}	C_{13}	C_{14}	G_{15}	G_{16}	A_{17}	G_{18}	G_{19}	A_{20}	C	A	A	indel%
A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	100.0	100.0	0.03
C	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
G	100.0	0.0	99.9	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.9	99.9	0.0	100.0	100.0	0.0	0.0	0.0	0.0	
T	0.0	100.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.6)–BE3

Site 20	G_1	T_2	G_3	G_4	T_5	T_6	C_7	C_8	A_9	G_{10}	A_{11}	A_{12}	C_{13}	C_{14}	G_{15}	G_{16}	A_{17}	G_{18}	G_{19}	A_{20}	C	A	A	indel%
A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	100.0	100.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	99.9	99.9	0.02
C	0.0	0.0	0.0	0.0	0.0	0.0	99.9	99.9	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
G	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.9	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	
T	0.0	100.0	0.0	0.0	100.0	100.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

Site 20	G_1	T_2	G_3	G_4	T_5	T_6	C_7	C_8	A_9	G_{10}	A_{11}	A_{12}	C_{13}	C_{14}	G_{15}	G_{16}	A_{17}	G_{18}	G_{19}	A_{20}	C	A	A	indel%
A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	100.0	99.9	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	99.9	100.0	0.05
C	0.0	0.0	0.0	0.0	0.0	0.0	99.7	99.7	0.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
G	99.9	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.9	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	
T	0.0	100.0	0.0	0.0	100.0	100.0	0.3	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

SpCas9–BE3

FANCF site 1	G_1	G_2	A_3	A_4	T_5	C_6	C_7	C_8	T_9	T_{10}	C_{11}	T_{12}	G_{13}	C_{14}	A_{15}	G_{16}	C_{17}	A_{18}	C_{19}	C_{20}	T	G	G	indel%
	0.1	0.3	100.0	100.0	0.0	0.2	0.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.1	0.29
	0.0	0.0	0.0	0.0	0.0	79.1	80.4	80.9	0.0	0.0	97.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	
	99.9	99.6	0.0	0.0	0.0	0.4	0.3	0.1	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	99.8	
	0.0	0.0	0.0	0.0	100.0	20.3	19.0	18.7	100.0	100.0	2.9	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	

xCas9(3.7)–BE3

FANCF site 1	G_1	G_2	A_3	A_4	T_5	C_6	C_7	C_8	T_9	T_{10}	C_{11}	T_{12}	G_{13}	C_{14}	A_{15}	G_{16}	C_{17}	A_{18}	C_{19}	C_{20}	T	G	G	indel%
	0.0	0.0	100.0	100.0	0.0	0.1	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	1.77
	0.0	0.0	0.0	0.0	0.0	70.9	72.7	73.2	0.0	0.0	78.2	0.0	0.0	99.5	0.0	0.0	99.9	0.0	100.0	100.0	0.0	0.0	0.0	
	100.0	100.0	0.0	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.1	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	
	0.0	0.0	0.0	0.0	100.0	28.9	27.1	26.5	100.0	100.0	21.6	100.0	0.0	0.5	0.0	0.0	0.1	0.0	0.0	0.0	100.0	0.0	0.0	

SpCas9–BE3

FANCF site 2	A	C	T	C_{20}	C_{19}	A_{18}	A_{17}	G_{16}	A_{15}	G_{14}	A_{13}	G_{12}	C_{11}	C_{10}	T_9	G_8	G_7	C_6	C_5	C_4	G_3	C_2	C_1	indel%
	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.01
	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	100.0	100.0	
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	
	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

FANCF site 2	A	C	T	C_{20}	C_{19}	A_{18}	A_{17}	G_{16}	A_{15}	G_{14}	A_{13}	G_{12}	C_{11}	C_{10}	T_9	G_8	G_7	C_6	C_5	C_4	G_3	C_2	C_1	indel%
	100.0	0.0	0.0	0.0	0.0	100.0	100.0	3.1	100.0	1.9	100.0	0.0	0.0	0.0	0.0	1.1	1.3	0.0	0.0	0.0	0.1	0.0	0.0	0.01
	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	100.0	100.0	
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	96.9	0.0	98.1	0.0	100.0	0.0	0.0	0.0	98.9	98.7	0.0	0.0	0.0	99.9	0.0	0.0	
	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Supplementary Table 14 (Continued)

SpCas9–BE3

FANCF site 3	G_1	G_2	G_3	C_4	C_5	C_6	T_7	C_8	G_9	G_{10}	G_{11}	G_{12}	A_{13}	T_{14}	G_{15}	C_{16}	A_{17}	G_{18}	C_{19}	T_{20}	C	G	T	indel%
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.04
	0.0	0.0	0.0	100.0	99.9	99.9	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	
	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	
	0.0	0.0	0.0	0.0	0.1	0.1	100.0	0.1	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	

xCas9(3.7)–BE3

FANCF site 3	G_1	G_2	G_3	C_4	C_5	C_6	T_7	C_8	G_9	G_{10}	G_{11}	G_{12}	A_{13}	T_{14}	G_{15}	C_{16}	A_{17}	G_{18}	C_{19}	T_{20}	C	G	T	indel%
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.18
	0.0	0.0	0.0	99.6	95.1	94.0	0.0	94.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	
	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	
	0.0	0.0	0.0	0.4	4.9	6.0	100.0	5.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	

SpCas9–BE3

FANCF site 4	A	C	A	A_{20}	C_{19}	C_{18}	A_{17}	G_{16}	T_{15}	G_{14}	G_{13}	A_{12}	G_{11}	G_{10}	C_9	A_8	A_7	G_6	A_5	G_4	G_3	C_2	C_1	indel%
	100.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.02
	0.0	100.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	
	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0	0.0	
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

FANCF site 4	A	C	A	A_{20}	C_{19}	C_{18}	A_{17}	G_{16}	T_{15}	G_{14}	G_{13}	A_{12}	G_{11}	G_{10}	C_9	A_8	A_7	G_6	A_5	G_4	G_3	C_2	C_1	indel%
	100.0	0.0	100.0	100.0	0.0	0.0	100.0	0.3	0.0	1.1	2.4	100.0	0.8	0.1	0.0	100.0	100.0	7.4	100.0	3.1	0.8	0.0	0.0	0.02
	0.0	100.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	100.0	
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.7	0.0	98.9	97.6	0.0	99.2	99.9	0.0	0.0	0.0	92.5	0.0	96.9	99.2	100.0	0.0	
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

SpCas9–BE3

FANCF site 5	G_1	C_2	C_3	A_4	C_5	A_6	T_7	C_8	C_9	A_{10}	T_{11}	C_{12}	G_{13}	G_{14}	C_{15}	G_{16}	C_{17}	T_{18}	T_{19}	T_{20}	G	G	T	indel%
	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.01
	0.0	100.0	100.0	0.0	100.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	
	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	
	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	

xCas9(3.7)–BE3

FANCF site5	G_1	C_2	C_3	A_4	C_5	A_6	T_7	C_8	C_9	A_{10}	T_{11}	C_{12}	G_{13}	G_{14}	C_{15}	G_{16}	C_{17}	T_{18}	T_{19}	T_{20}	G	G	T	indel%
	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.05
	0.0	100.0	99.9	0.0	99.7	0.0	0.0	99.7	99.7	0.0	0.0	99.8	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	
	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	
	0.0	0.0	0.1	0.0	0.3	0.0	100.0	0.3	0.3	0.0	100.0	0.2	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	

SpCas9–BE3

FANCF site 6	A	A	C	C_{20}	A_{19}	G_{18}	T_{17}	G_{16}	G_{15}	A_{14}	G_{13}	G_{12}	C_{11}	A_{10}	A_9	G_8	A_7	G_6	G_5	C_4	C_3	G_2	G_1	indel%
	100.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.1	100.0	0.1	0.0	0.0	0.0	0.0	0.0	0.02
	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	99.9	0.0	99.9	100.0	100.0	0.0	100.0	100.0	
	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

FANCF site 6	A	A	C	C_{20}	A_{19}	G_{18}	T_{17}	G_{16}	G_{15}	A_{14}	G_{13}	G_{12}	C_{11}	A_{10}	A_9	G_8	A_7	G_6	G_5	C_4	C_3	G_2	G_1	indel%
	100.0	100.0	0.0	0.0	100.0	0.0	0.0	0.5	1.6	100.0	0.2	0.1	0.0	100.0	100.0	4.1	100.0	3.4	2.9	1.6	0.0	0.0	0.0	0.03
	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
	0.0	0.0	0.0	0.0	0.0	100.0	0.0	99.5	98.4	0.0	99.8	99.9	0.0	0.0	0.0	95.8	0.0	96.6	97.1	98.4	0.0	100.0	100.0	
	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

SpCas9–BE3

FANCF site 7	G_1	T_2	G_3	A_4	C_5	G_6	T_7	C_8	C_9	T_{10}	G_{11}	C_{12}	T_{13}	C_{14}	T_{15}	C_{16}	T_{17}	C_{18}	T_{19}	G_{20}	C	G	C	indel%
	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.01
	0.0	0.0	0.0	0.0	99.5	0.0	0.0	99.5	99.7	0.0	0.0	100.0	0.0	100.0	0.0	99.9	0.0	100.0	0.0	0.0	99.9	0.0	100.0	
	100.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	
	0.0	100.0	0.0	0.0	0.5	0.0	100.0	0.5	0.3	100.0	0.0	0.0	100.0	0.0	100.0	0.1	100.0	0.0	100.0	0.0	0.1	0.0	0.0	

xCas9(3.7)–BE3

FANCF site 7	G_1	T_2	G_3	A_4	C_5	G_6	T_7	C_8	C_9	T_{10}	G_{11}	C_{12}	T_{13}	C_{14}	T_{15}	C_{16}	T_{17}	C_{18}	T_{19}	G_{20}	C	G	C	indel%
	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.72
	0.0	0.0	0.0	0.0	86.3	0.0	0.0	83.3	92.6	0.0	0.0	99.9	0.0	95.5	0.0	92.9	0.0	98.2	0.0	0.0	100.0	0.0	100.0	
	100.0	0.0	100.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	
	0.0	100.0	0.0	0.0	13.7	0.0	100.0	16.6	7.4	100.0	0.0	0.1	100.0	4.5	100.0	7.1	100.0	1.8	100.0	0.0	0.0	0.0	0.0	

SpCas9–BE3

FANCF site 8	G_1	C_2	G_3	C_4	C_5	G_6	G_7	G_8	C_9	C_{10}	T_{11}	T_{12}	G_{13}	C_{14}	A_{15}	G_{16}	T_{17}	G_{18}	G_{19}	G_{20}	C	G	A	indel%
	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.02
	0.0	100.0	0.0	99.9	99.9	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	
	100.0	0.0	99.9	0.0	0.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0	0.0	100.0	0.0	
	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

FANCF site 8	G_1	C_2	G_3	C_4	C_5	G_6	G_7	G_8	C_9	C_{10}	T_{11}	T_{12}	G_{13}	C_{14}	A_{15}	G_{16}	T_{17}	G_{18}	G_{19}	G_{20}	C	G	A	indel%
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.01
	0.0	100.0	0.0	99.1	98.5	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	
	100.0	0.0	100.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0	0.0	100.0	0.0	
	0.0	0.0	0.0	0.9	1.5	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

Supplementary Table 14 (Continued)

SpCas9–BE3

FANCF site 9	G	C	G	C_{20}	C_{19}	G_{18}	G_{17}	G_{16}	C_{15}	C_{14}	T_{13}	T_{12}	G_{11}	C_{10}	A_9	G_8	T_7	G_6	G_5	G_4	C_3	G_2	C_1	indel%
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.3	0.0	0.5	0.4	0.1	0.0	0.0	0.0	0.01
	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	
	100.0	0.0	100.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.7	0.0	99.5	99.5	99.9	0.0	100.0	0.0	
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

FANCF site 9	G	C	G	C_{20}	C_{19}	G_{18}	G_{17}	G_{16}	C_{15}	C_{14}	T_{13}	T_{12}	G_{11}	C_{10}	A_9	G_8	T_7	G_6	G_5	G_4	C_3	G_2	C_1	indel%
	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	100.0	1.9	0.0	4.1	3.9	0.5	0.0	0.0	0.0	0.01
	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0	0.0	100.0	0.0	100.0	
	100.0	0.0	100.0	0.0	0.0	100.0	99.9	100.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	98.1	0.0	95.9	96.1	99.5	0.0	100.0	0.0	
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

SpCas9–BE3

FANCF site 10	G_1	A_2	G_3	A_4	A_5	C_6	C_7	G_8	G_9	G_{10}	C_{11}	C_{12}	C_{13}	T_{14}	C_{15}	G_{16}	G_{17}	G_{18}	G_{19}	A_{20}	T	G	C	indel%
	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.03
	0.0	0.0	0.0	0.0	0.0	99.9	99.8	0.0	0.0	0.0	100.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	
	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	
	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	

xCas9(3.7)–BE3

FANCF site 10	G_1	A_2	G_3	A_4	A_5	C_6	C_7	G_8	G_9	G_{10}	C_{11}	C_{12}	C_{13}	T_{14}	C_{15}	G_{16}	G_{17}	G_{18}	G_{19}	A_{20}	T	G	C	indel%
	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.12
	0.0	0.0	0.0	0.0	0.0	96.7	97.0	0.0	0.0	0.0	100.0	99.9	99.9	0.0	98.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	
	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	
	0.0	0.0	0.0	0.0	0.0	3.3	3.0	0.0	0.0	0.0	0.0	0.1	0.1	100.0	1.5	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	

SpCas9–BE3

FANCF site 11	G_1	A_2	G_3	A_4	C_5	A_6	C_7	T_8	C_9	C_{10}	A_{11}	A_{12}	G_{13}	A_{14}	G_{15}	A_{16}	G_{17}	C_{18}	C_{19}	T_{20}	G	G	C	indel%
	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.01
	0.0	0.0	0.0	0.0	99.9	0.0	99.9	0.0	99.9	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	
	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	
	0.0	0.0	0.0	0.0	0.1	0.0	0.1	100.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

FANCF site 11	G_1	A_2	G_3	A_4	C_5	A_6	C_7	T_8	C_9	C_{10}	A_{11}	A_{12}	G_{13}	A_{14}	G_{15}	A_{16}	G_{17}	C_{18}	C_{19}	T_{20}	G	G	C	indel%
	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.04
	0.0	0.0	0.0	0.0	98.7	0.0	98.7	0.0	99.1	99.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	
	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	
	0.0	0.0	0.0	0.0	1.3	0.0	1.3	100.0	0.8	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	

SpCas9–BE3

FANCF site 12	G_1	G_2	G_3	C_4	T_5	T_6	C_7	A_8	A_9	T_{10}	G_{11}	G_{12}	C_{13}	T_{14}	A_{15}	T_{16}	A_{17}	G_{18}	A_{19}	G_{20}	A	G	A	indel%
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.01
	0.0	100.0	0.0	99.7	0.0	0.0	99.2	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	
	0.0	0.0	0.0	0.3	100.0	100.0	0.8	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

FANCF site 12	G_1	G_2	G_3	C_4	T_5	T_6	C_7	A_8	A_9	T_{10}	G_{11}	G_{12}	C_{13}	T_{14}	A_{15}	T_{16}	A_{17}	G_{18}	A_{19}	G_{20}	A	G	A	indel%
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.62
	0.0	99.9	0.0	98.8	0.0	0.0	90.4	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	
	0.0	0.1	0.0	1.2	100.0	100.0	9.5	0.0	0.0	100.0	0.0	0.0	0.1	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

SpCas9–BE3

FANCF site 13	T	C	A	A_{20}	C_{19}	G_{18}	T_{17}	T_{16}	T_{15}	G_{14}	C_{13}	A_{12}	C_{11}	T_{10}	A_9	T_8	G_7	A_6	C_5	C_4	T_3	T_2	C_1	indel%
	0.0	0.0	100.0	97.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.02
	0.0	100.0	0.0	2.4	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	
	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	
	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	

xCas9(3.7)–BE3

FANCF site 13	T	C	A	A_{20}	C_{19}	G_{18}	T_{17}	T_{16}	T_{15}	G_{14}	C_{13}	A_{12}	C_{11}	T_{10}	A_9	T_8	G_7	A_6	C_5	C_4	T_3	T_2	C_1	indel%
	0.0	0.0	100.0	97.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	6.3	100.0	0.0	0.0	0.0	0.0	0.0	0.02
	0.0	100.0	0.0	2.5	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	
	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	93.6	0.0	0.0	0.0	0.0	0.0	0.0	
	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	

SpCas9–BE3

FANCF site 14	G_1	A_2	G_3	A_4	A_5	C_6	C_7	C_8	A_9	A_{10}	A_{11}	T_{12}	C_{13}	T_{14}	C_{15}	C_{16}	A_{17}	G_{18}	G_{19}	A_{20}	G	G	A	indel%
	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	99.9	99.8	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.01
	0.0	0.0	0.0	0.0	0.0	99.2	99.2	99.2	0.0	0.1	0.2	0.0	99.9	0.0	99.9	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	
	0.0	0.0	0.0	0.0	0.0	0.8	0.8	0.7	0.0	0.0	0.0	100.0	0.1	100.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

FANCF site 14	G_1	A_2	G_3	A_4	A_5	C_6	C_7	C_8	A_9	A_{10}	A_{11}	T_{12}	C_{13}	T_{14}	C_{15}	C_{16}	A_{17}	G_{18}	G_{19}	A_{20}	G	G	A	indel%
	0.1	100.0	0.0	100.0	100.0	0.6	0.3	0.1	100.0	99.9	99.7	0.0	0.1	0.0	0.3	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	1.00
	0.0	0.0	0.0	0.0	0.0	72.6	73.1	78.8	0.0	0.1	0.3	0.0	89.2	0.0	83.8	94.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	99.9	0.0	100.0	0.0	0.0	0.1	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	
	0.0	0.0	0.0	0.0	0.0	26.7	26.6	20.8	0.0	0.0	0.0	100.0	10.7	100.0	15.8	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Supplementary Table 14 (Continued)

SpCas9–BE3

FANCF site 15	G_1	C_2	C_3	C_4	C_5	A_6	T_7	T_8	C_9	G_{10}	C_{11}	A_{12}	C_{13}	G_{14}	G_{15}	C_{16}	T_{17}	C_{18}	T_{19}	G_{20}	C	A	G	indel%
	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.02
	0.0	99.9	99.3	99.3	99.2	0.0	0.0	0.0	99.8	0.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	
	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	
	0.0	0.1	0.7	0.7	0.8	0.0	100.0	100.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	

xCas9(3.7)–BE3

FANCF site 15	G_1	C_2	C_3	C_4	C_5	A_6	T_7	T_8	C_9	G_{10}	C_{11}	A_{12}	C_{13}	G_{14}	G_{15}	C_{16}	T_{17}	C_{18}	T_{19}	G_{20}	C	A	G	indel%
	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.17
	0.0	99.7	97.8	96.7	96.4	0.0	0.0	0.0	98.4	0.0	100.0	0.0	99.8	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	
	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	
	0.0	0.3	2.2	3.3	3.6	0.0	100.0	100.0	1.5	0.0	0.0	0.0	0.2	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	

Supplementary Table 15. Genomic sites used for GUIDE-seq analysis.

Site	PAM	Protospacer sequence
EMX1 site	GGG	GAGTCCGAGCAGAAGAAGAA
HEK site 1	GGG	GGGAAAGACCCAGCATCCGT
HEK site 2	GGG	GAACACAAAGCATAGACTGC
HEK site 3	GGG	GGCCCAGACTGAGCACGTGA
HEK site 4	GGG	GGCACTGCGGCTGGAGGTGG
VEGFA site 1	TGG	GGGTGGGGGGAGTTTGCTCC
FANCF GAA PAM site	GAA	GACTCTCTGCGTACTGATTG
FANCF CGT PAM site	CGT	GCGGTCTCAAGCACTACCTA

sgRNA plasmids were generated with reverse primer GGTGTTTCGTCCTTTCCACAAGATATATAAAG and forward primer $(N)_{20}$GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC, where $(N)_{20}$ is the protospacer sequence.

Supplementary Table 16. List of GUIDE-seq off target sites, with corresponding chromosomal location, PAM, and protospacer sequences.

Site	Chromosome	PAM	Protospacer sequence
EMX1 off-target 1	15	GAG	GAGTCTAAGCAGAAGAAGAA
HEK site 1 off-target 1	1	TGG	GGGAAAGTCCCAGCATCCTT
HEK site 2 off-target 1	2	AGA	GAGGACAAAGTACAAACGGC
HEK site 2 off-target 2	4	CGG	GAACACAATGCATAGATTGC
HEK site 3 off-target 1	7	GGG	GACACAGACTGGGCACGTGA
HEK site 3 off-target 2	15	TGG	CACCCAGACTGAGCACGTGC
HEK site 3 xCas9 off-target 1	15	CGG	AGGCAAAACGCACCACGTGA
HEK site 3 xCas9 off-target 2	12	AGT	GGGCCAGACTGAGCACGTGA
HEK site 4 off-target 1	10	GGG	GGCACGACGGCTGGAGGTGG
HEK site 4 off-target 2	20	GGG	GGCACTGCGGCTGGAGGTGG
HEK site 4 off-target 3	7	GGG	GGCACTGCGGCTGGAGGTGG
HEK site 4 off-target 4	10	GGG	GGCACTGCGGCTGGAGGTGG
HEK site 4 off-target 5	15	GGG	GGCACTGCGGCTGGAGGTGG
HEK site 4 xCas9 off-target 1	5	GGG	GGCACTGCGGCTGGAGGTGG
HEK site 4 xCas9 off-target 2	13	GGG	GGCACTGCGGCTGGAGGTGG

Supplementary Table 17. GUIDE-seq HTS files contained in NCBI sequencing read archive, and the corresponding GUIDE-seq data within each file.

File name	Data
Guideseq 1	*EMX1* SpCas9 in HEK293T cells
Guideseq 3	*EMX1* xCas9 3.6 in HEK293T cells
Guideseq 1	*EMX1* xCas9 3.7 in HEK293T cells
Guideseq 5	*EMX1* SpCas9 in U2OS cells
Guideseq 5	*EMX1* xCas9 3.6 in U2OS cells
Guideseq 5	*EMX1* xCas9 3.7 in U2OS cells
Guideseq 3	HEK site 1 SpCas9 in HEK293T cells
Guideseq 3	HEK site 1 xCas9 3.6 in HEK293T cells
Guideseq 3	HEK site 1 xCas9 3.7 in HEK293T cells
Guideseq 3	HEK site 2 SpCas9 in HEK293T cells
Guideseq 3	HEK site 2 xCas9 3.6 in HEK293T cells
Guideseq 3	HEK site 2 xCas9 3.7 in HEK293T cells
Guideseq 3	HEK site 3 SpCas9 in HEK293T cells
Guideseq 3	HEK site 3 xCas9 3.6 in HEK293T cells
Guideseq 3	HEK site 3 xCas9 3.7 in HEK293T cells
Guideseq 1	HEK site 4 SpCas9 in HEK293T cells
Guideseq 1	HEK site 4 xCas9 3.7 in HEK293T cells
Guideseq 3	HEK site 4 xCas9 3.6 in HEK293T cells
Guideseq 2	*FANCF* GAA xCas9 3.7 in HEK293T cells
Guideseq 4	*FANCF* NGT xCas9 3.7 in HEK293T cells
Guideseq 6	*VEGFA* site 1 SpCas9 in U2OS cells
Guideseq 6	*VEGFA* site 1 xCas9 3.6 in U2OS cells
Guideseq 6	*VEGFA* site 1 xCas9 3.7 in U2OS cells

Supplementary Sequences 1. DNA sequences of PACE selection components. Color coding is as follows:

green = protospacer sequence

orange = linker

black = dCas9

Magenta = sd8

purple = NLS

light blue = wild-type *E. coli* TadA

blue = ω(I12N) subunit of bacterial RNA polymerase

AP construct (ω(I12N)-linker(2aa)-dCas9):

ATGGCACGCGTAACTGTTCAGGACGCTGTAGAGAAAAATGGTAACCGTTTTGACCTGGTACTGGTCG
CCGCGCGTCGCGCTCGTCAGATGCAGGTAGGCGGAAAGGATCCGCTGGTACCGGAAGAAAACGAT
AAAACCACTGTAATCGCGCTGCGCGAAATCGAAGAAGGTCTGATCAACAACCAGATCCTCGACGTTC
GCGAACGCCAGGAACAGCAAGAGCAGGAAGCCGCTGAATTACAAGCCGTTACCGCTATTGCTGAAG
GTCGTCGTGCTGCAGACAAGAAGTACTCCATTGGGCTCGCTATCGGCACAAACAGCGTCGGCTGGG
CCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCCA
CAGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCACGC
GGCTCAAAAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAGGAGA
TCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGGTG
GAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCAT
GAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGC
GGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTCGGGGACACTTCCTCATCGAGGGGGACCT
GAACCCAGACAACAGCGATGTCGATAAACTCTTTATCCAACTGGTTCAGACTTACAATCAGCTTTTCG
AAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAAAT
CCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACGGCCTGTTTGGTAATC
TTATCGCCCTGTCACTCGGGCTGACCCCCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAA
GCTTCAACTGAGCAAAGACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCA
GTACGCAGACCTTTTTTTGGCGGCAAAGAACCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGA
GTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCAC
CAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAAGTACAAGGAAATTTTCT
TCGATCAGTCTAAAAATGGCTACGCCGGATACATTGACGGCGGAGCAAGCCAGGAGGAATTTTACAA
ATTTATTAAGCCCATCTTGGAAAAAATGGACGGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAA
GATCTGTTGCGCAAACAGCGCACTTTCGACAATGGAAGCATCCCCCACCAGATTCACCTGGGCGAA
CTGCACGCTATCCTCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTG
AGAAAATCCTCACATTTCGGATACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGC
GTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACTTCGAGGAAGTCGTGGATAAGGG
GGCCTCTGCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACGAAAAGGTG
CTTCCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGT
CACAGAAGGGATGAGAAAGCCAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCT
CTTCAAGACGAACCGGAAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTT
TCGACTCTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCT
CCTGAAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATT
GTCCTCACCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACTTACGCTCATCT
CTTCGACGACAAAGTCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGTCAAG
AAAACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCTGGATTTTCTTAAGTCCGAT
GGATTTGCCAACCGGAACTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCC
AGAAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCC
CAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGTCAAAGTAATGGGAAG

Supplementary Sequence 1 (Continued)

GCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACTACCCAGAAGGGACAGAA
GAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAGAACTGGGGTCCCAAATCCTTAA
GGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGC
AGGGACATGTACGTGGATCAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGACGCTATC
GTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTTGACAAGATCCGATAAAAACAG
AGGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTG
CTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGT
CTGAGTTGGATAAAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCACGT
GGCCCAAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAAACTGATTCGAGAGGTG
AAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAG
AGAGATCAACAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATC
AAAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAAT
GATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATG
AATTTTTTCAAGACCGAGATTACACTGGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAACAA
ACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTGT
CCATGCCGCAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTA
TCCTCCCGAAAAGGAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACG
GCGGATTCGATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTC
TAAAAAACTCAAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAA
AACCCCATCGACTTTCTCGAGGCGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTTC
CCAAGTACTCTCTCTTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCGGGCGAGCTGC
AGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAA
AAGCTCAAAGGGTCTCCCGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTACC
TTGATGAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGA
TAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAAACATTATC
CACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACCATAGACA
GAAAGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCAATTACGGGGC
TCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGAC

SP:

CCGTTCGTTTTGCTGAGGAGACTTAGGGACCCTACAACACGCACGGTGTTACATTAGGCACCCCGG
GCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTCGACCGACCTGCAGGTGCAGTAAAGGAAAAA
AAAA

Supplementary Sequences 2. Amino acid sequence of VPR activator dxCas9 3.7 construct

green = VP64

orange = linker

black = dxCas9 3.7, with mutations highlighted in red

Magenta = p65 AD

purple = NLS

blue = Rta AD

dCas9–VPR (dxCas9(3.7)–NLS–linker(22aa)–VP64–linker(4aa)–NLS–p65AD–linker(6aa)-RtaAD):

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV
DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPN
EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVM
KQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDS
LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEL
GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA
QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL
ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGV
LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL
SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGDSRADPKKKRKVSPGIRRLDALISTSLYKKAGYKEASGSGRADALDDFDLDMLGSDALDDFDLDMLG
SDALDDFDLDMLGSDALDDFDLDMLINSRSGGSPKKKRKVGSQYLPDTDDRHRIEEKRKRTYETFKSIMK
KSPFSGPTDPRPPPRRIAVPSRSSASVPKPAPQPYPFTSSLSTINYDEFPTMVFPSGQISQASALAPAPPQ
VLPQAPAPAPAPAMVSALAQAPAPVPVLAPGPPQAVAPPAPKPTQAGEGTLSEALLQLQFDDEDLGALL
GNSTDPAVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLP
NGLLSGDEDFSSIADMDFSALLGSGSGSRDSREGMFLPKPEAGSAISDVFEGREVCQPKRIRPFHPPGS
PWANRPLPASLAPTPTGPVHEPVGSLTPAPVPQPLDPAPAVTPEASHLLEDPDEETSQAVKALREMADT
VIPQKEEAAICGQMDLSHPPPRGHLDELTTTLESMTEDLNLDSPLTPELNEILDTFLNDECLLHAMHISTGL
SIFDTSLF

Supplementary Sequences 3. Amino acid sequences of xCas9 3.7 constructs developed in this study. Color coding is as follows:

green = UGI

orange = linker

black + red = evolved xCas9, with mutations highlighted in red

Magenta = APOBEC

purple = NLS

light blue = wild-type *E. coli* TadA

blue = evolved 7.10 *E. coli* TadA

xCas9 3.7 (xCas9 3.7–linker(4aa)–NLS):

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV
DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPN
EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVM
KQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDS
LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEL
GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA
QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL
ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGV
LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL
SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGDSRADPKKKRKV

xCas9(3.7)–BE3 (APOBEC–linker(16aa)–xCas9(3.7)–linker(4aa)–UGI–linker(4aa)–NLS):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQ
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF
IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQD
LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR
TFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR
NFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM

Supplementary Sequence 3 (Continued)

ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA
ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK
LIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
VKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKP
ESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

xCas9(3.7)–ABE (ecTadA(wt)–linker(32 aa)–ecTadA*(7.10)–linker(32 aa)–nxCas9(3.7)–NLS):

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSH
EYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID
ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCY
FFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVG
WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHM
IKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILR
VNTEITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPIL
EKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVG
PLARGNSRFAWMTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLI
NGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK
LYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK
NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI
AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN
IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEF
ESPKKKRKV

xCas9(3.7)–BE4 (APOBEC–linker(32aa)–xCas9(3.7)–linker(10aa)–UGI– linker(10aa)–UGI–linker(4aa)–
NLS):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQ
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVIT
DEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV
DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR
GHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGL
FGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK

Supplementary Sequence 3 (Continued)

MDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLING
IRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN
YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA
KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV
KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYL
ASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSD
IIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGEN
KIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLL
TSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

Supplementary Note 1. Matlab script for base calling.

```
function basecall()
%cycle through fastq files for different samples
WTnuc='GAATCCCTTCTGCAGCACCTGGATCGCTTTTCCGAGCTTCTGGCGGTCTCAAGCACTACCT
ACGTCAGCACCTGGGACCCCGCCACCGTGCGCCGGGCCTTGCAGTGGGCGCGCTACCTGCGCCAC
ATCCA';
files=dir('*.fastq');
for d=1:9
    filename=files(d).name;
    %read fastq file
    [header,seqs,qscore] = fastqread(filename);
    seqsLength = length(seqs);          % number of sequences
    seqsFile = strrep(filename,'.fastq','');      % trims off .fastq
    %create a directory with the same name as fastq file
    if exist(seqsFile,'dir');
        error('Directory already exists. Please rename or move it before moving on.');
    end
    mkdir(seqsFile);                    % make directory
    wtLength = length(WTnuc);               % length of wildtype sequence
    %% aligning back to the wildtype nucleotide sequence
    %
    % AlN is a matrix of the nucleotide alignment
    window=1:wtLength;
    sBLength = length(seqs);         % number of sequences
    % counts number of skips
    nSkips = 0;
    ALN=repmat(' ',[sBLength wtLength]);
    % iterate through each sequencing read
    for i = 1:sBLength
        %If you only have forward read fastq files leave as is
        %If you have R1 foward and R2 is reverse fastq files uncomment the
        %next four lines of code and the subsequent end statement
%         if mod(d,2)==0;
%             reverse = seqrcomplement(seqs{i});
%             [score,alignment,start] = swalign(reverse,WTnuc,'Alphabet','NT');
%         else

        [score,alignment,start] = swalign(seqs{i},WTnuc,'Alphabet','NT');
%         end

        % length of the sequencing read
        len = length(alignment(3,:));
        % if there is a gap in the alignment , skip = 1 and we will
        % throw away the entire read
        skip = 0;
        for j = 1:len
            if (alignment(3,j) == '-' || alignment(1,j) == '-')
                skip = 1;
                break;
            end
            %in addition if the qscore for any given base in the read is
            %below 31 the nucleotide is turned into an N (fastq qscores that are not letters)
            if isletter(qscore{i}(start(1)+j-1))
```

Supplementary Note 1 (Continued)

```
            else
                alignment(1,j) = 'N';
            end

        end
        if skip == 0 && len>10
            ALN(i, start(2):(start(2)+length(alignment)-1))=alignment(1,:);
        end
    end
    % with the alignment matrices we can simply tally up the occurrences of
    % each nucleotide at each column in the alignment these
    % tallies ignore bases annotated as N
    % due to low qscores
    TallyNTD=zeros(5,wtLength);
    FreqNTD=zeros(4,wtLength);
    SUM=zeros(1,wtLength);
    for i=1:wtLength
        TallyNTD(:,i)=[sum(ALN(:,i)=='A'),sum(ALN(:,i)=='C'),sum(ALN(:,i)=='G'),sum(ALN(:,i)=='T'),sum(ALN
(:,i)=='N')];
    end

    for i=1:wtLength
        FreqNTD(:,i)=100*TallyNTD(1:4,i)/sum(TallyNTD(1:4,i));
    end
    for i=1:wtLength
        SUM(:,i)=sum(TallyNTD(1:4,i));
    end

    % we then save these tally matrices in the respective folder for
    % further processing

    save(strcat(seqsFile, '/TallyNTD'), 'TallyNTD');
    dlmwrite(strcat(seqsFile, '/TallyNTD.csv'), TallyNTD, 'precision', '%.3f', 'newline', 'pc');
    save(strcat(seqsFile, '/FreqNTD'), 'FreqNTD');
    dlmwrite(strcat(seqsFile, '/FreqNTD.csv'), FreqNTD, 'precision', '%.3f', 'newline', 'pc');
    fid = fopen('FrequencySummary.csv', 'a');
    fprintf(fid, '\n \n');
    fprintf(fid, filename);
    fprintf(fid, '\n \n');
    dlmwrite('FrequencySummary.csv', FreqNTD, 'precision', '%.3f', 'newline', 'pc', '-append');
    dlmwrite('FrequencySummary.csv', SUM, 'precision', '%.3f', 'newline', 'pc', '-append');
end
```

Supplementary Note 2. Matlab script for indel analysis.

```
function indelcall()

%cycle through fastq files for different samples

s11048='GCCCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGTCACTAGGGGGCGCTC
GGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGCAGCGTCTTCGAGAGTGAGGACGTGTGTGTC
TGTGTGGGTGAGTGAGTGTGTGCGTGTGGGGTTGAG';

s11150='GCCCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGTCACTAGGGGGCGCTC
GGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGCAGCGTCTTCGAGAGTGAGGACGTGTGTGTC
TGTGTGGGTGAGTGAGTGTGTGCGTGTGGGGTTGAG';


s11048is=34

s11150is=71


WTnuc=s11048;

%WTnuc='CGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGAGATC
CGCGGCCGCTAATACGACTCACCCTAGGGAGAGCCGCCACCGTGGTGAGCAAGGGCGAGGAGCTG
TTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGT
GTCCGGCGAG';

%cycle through fastq files for different samples

files=dir('*.fastq');

indelstart=s11048is;

width=30;

flank=10;

for d=1:9

    filename=files(d).name;

    %read fastq file

    [header,seqs,qscore] = fastqread(filename);

    seqsLength = length(seqs);          % number of sequences

    seqsFile = strcat(strrep(filename,'.fastq',''),'_INDELS');      % trims off .fastq

    %create a directory with the same name as fastq file+_INDELS

    if exist(seqsFile,'dir');

        error('Directory already exists. Please rename or move it before moving on.');

    end

    mkdir(seqsFile);                    % make directory

    wtLength = length(WTnuc);           % length of wildtype sequence
```

Supplementary Note 2 (Continued)

```
    sBLength = length(seqs);          % number of sequences

    % initialize counters and cell arrays
    nSkips = 0;
    notINDEL=0;
    ins={};
    dels={};
    NumIns=0;
    NumDels=0;
    % iterate through each sequencing read
    for i = 1:sBLength
    %search for 10BP sequences that should flank both sides of the "INDEL WINDOW"
        windowstart=strfind(seqs{i},WTnuc(indelstart-flank:indelstart));
        windowend=strfind(seqs{i},WTnuc(indelstart+width:indelstart+width+flank));
        %if these flanks are found and more than half of base calls
        %are above Q31 THEN proceed OTHERWISE save as a skip
        if length(windowstart)==1 && length(windowend)==1 &&
(sum(isletter(qscore{i}))/length(qscore{i}))>=0.5
            %if the sequence length matches the INDEL window length save as
            %not INDEL
            if windowend-windowstart==width+flank
                notINDEL=notINDEL+1;
            %if the sequence is ONE or more baseslonger than the INDEL
            %window length save as an Insertion
            elseif windowend-windowstart>=width+flank+1
                NumIns=NumIns+1;
                ins{NumIns}=seqs{i};
            %if the sequence is ONE or more bases shorter than the INDEL
            %window length save as a Deletion
            elseif windowend-windowstart<=width+flank-1
                NumDels=NumDels+1;
                dels{NumDels}=seqs{i};
```

Supplementary Note 2 (Continued)

```
            end
        %keep track of skipped sequences that do not posess matching flank
        %sequences and do not pass quality cutoff
        else
            nSkips=nSkips+1;


        end
    end
    INDELrate=(NumIns+NumDels)/(NumIns+NumDels+notINDEL)*100.;
    FID = fopen('INDELSummary.csv', 'a');
    fprintf(FID, '\n \n');
    fprintf(FID, filename);
    fprintf(FID, '\n');
    fprintf(FID, num2str(INDELrate));


    fid=fopen(strcat(seqsFile, '/summary.txt'), 'wt');
    fprintf(fid, 'Skipped reads %i\n not INDEL %i\n Insertions %i\n Deletions %i\n INDEL percent %e\n',
[nSkips, notINDEL, NumIns, NumDels,INDELrate]);
    fclose(fid);
    save(strcat(seqsFile, '/nSkips'), 'nSkips');
    save(strcat(seqsFile, '/notINDEL'), 'notINDEL');
    save(strcat(seqsFile, '/NumIns'), 'NumIns');
    save(strcat(seqsFile, '/NumDels'), 'NumDels');
    save(strcat(seqsFile, '/INDELrate'), 'INDELrate');
    save(strcat(seqsFile, '/dels'), 'dels');
    C = dels;
    fid = fopen(strcat(seqsFile, '/dels.txt'), 'wt');
    fprintf(fid, '"%s"\n', C{:});
    fclose(fid);
    save(strcat(seqsFile, '/ins'), 'ins');
    C = ins;
    fid = fopen(strcat(seqsFile, '/ins.txt'), 'wt');
    fprintf(fid, '"%s"\n', C{:});
```

Supplementary Note 2 (Continued)

```
    fclose(fid);


end
```

Supplementary Note 3. Script for Clinvar SNP analysis for C•G to T•A base conversion and T•A to G•C base conversion

```
import numpy as np
import pandas as pd
import regex
import re
from Bio import SeqIO
import Bio

###CHANGE PAM AND WINDOW INFO HERE###
PAM = 'GAT'
windowstart = 4
windowend = 8
###CHANGE PAM AND WINDOW INFO HERE###

def RC(seq):
	encoder = {'A':'T','T':'A','C':'G','G':'C','N':'N','R':'Y','Y':'R', 'M':'K', 'K':'M', 'S':'S', 'W':'W', 'H':'D', 'B':'V', 'V':'B', 'D':'H'}
	rc = ''
	for n in reversed(seq):
		rc += encoder[n]
	return rc

def create_PAM(pam):
	encoder = {'A':'A','T':'T','G':'G','C':'C','R':'[A|G]','Y':'[C|T]','N':'[A|T|C|G]','M':'[A|C]','K':'[G|T]','S':'[C|G]','W':'[A|T]','H':'[A|C|T]','B':'[C|G|T]','V':'[A|C|G]','D':'[A|G|T]'}
	enc_pam = {'f':'','r':''}
	rc_pam = RC(pam)
	for n,m in zip(pam, rc_pam):
		enc_pam['f'] += encoder[n]
		enc_pam['r'] += encoder[m]
	return enc_pam
enc_pam = create_PAM(PAM)
```

Supplementary Note 3 (Continued)

```
windowlen=windowend-windowstart+1
lenpam=len(PAM)

################################################################################################################################################################################

Clinvar=pd.read_csv('Clinvar_DB.csv', encoding = "ISO-8859-1")
Clinvar_size = len(Clinvar.index)
Clinvar_dedup = Clinvar.drop_duplicates(subset='RCVaccession', keep='first')
Clinvar_dedup_size = len(Clinvar_dedup.index)
print "size of the dedupped Clinvar Database:", Clinvar_dedup_size

pathogenic = Clinvar_dedup['ClinicalSignificance'] == "Pathogenic"
Clinvar_Patho_df = Clinvar_dedup[pathogenic]
Clinvar_Patho_size = len(Clinvar_Patho_df.index)
print "Pathogenic entities in Clinvar:", Clinvar_Patho_size

SNV = Clinvar_Patho_df['Type'] == "single nucleotide variant"
Clinvar_Patho_SNV = Clinvar_Patho_df[SNV]
Clinvar_Patho_SNV_size = len(Clinvar_Patho_SNV)
print "Pathogenic SNV's in Clinvar:", Clinvar_Patho_SNV_size

BE_CtoT_editable = Clinvar_Patho_SNV[Clinvar_Patho_SNV["Name"].str.contains("T>C")]
BE_GtoA_editable = Clinvar_Patho_SNV[Clinvar_Patho_SNV["Name"].str.contains("A>G")]
ABE_TtoC_editable = Clinvar_Patho_SNV[Clinvar_Patho_SNV["Name"].str.contains("C>T")]
ABE_AtoG_editable = Clinvar_Patho_SNV[Clinvar_Patho_SNV["Name"].str.contains("G>A")]

AT_GC = Clinvar_Patho_SNV[Clinvar_Patho_SNV["Name"].str.contains("G>A|C>T")] # C>T and G>A in clinvar but A>G , T>C base editing #ABE
GC_AT = Clinvar_Patho_SNV[Clinvar_Patho_SNV["Name"].str.contains("A>G|T>C")]  # BE
```

Supplementary Note 3 (Continued)

```
# editing opportunities for transitions that can not yet be edited by base editing
CG_GC = Clinvar_Patho_SNV[Clinvar_Patho_SNV["Name"].str.contains("C>G|G>C")]
AT_TA = Clinvar_Patho_SNV[Clinvar_Patho_SNV["Name"].str.contains("A>T|T>A")]
GC_TA = Clinvar_Patho_SNV[Clinvar_Patho_SNV["Name"].str.contains("A>C|T>G")]
AT_CG = Clinvar_Patho_SNV[Clinvar_Patho_SNV["Name"].str.contains("C>A|G>T")]

GC_AT_size= len(GC_AT.index)
print "BE editable bases(GC_AT):", GC_AT_size
AT_GC_size = len(AT_GC.index)
print "ABE editable bases(AT_GC):", AT_GC_size
CG_GC_size = len(CG_GC.index)
print "CG_GC:", CG_GC_size
AT_TA_size = len(AT_TA.index)
print "AT_TA:", AT_TA_size
GC_TA_size = len(GC_TA.index)
print "GC_TA:", GC_TA_size
AT_CG_size = len(AT_CG.index)
print "AT_CG:", AT_CG_size



#open flanking sequence fasta files for all pathogenic human SNPs
#downloaded as fasta file from:
handle = open("SNP_FASTA.txt", "rU")
flanks={}
#save as a dictionary keyed on rsID as an Integer with values being 25nt of flanking sequence on each side of the SNP
for record in SeqIO.parse(handle, "fasta") :
    flanks[int(record.id.split("|")[-1].strip('rs'))]=regex.findall('.{25}[^A,T,C,G].{25}', record.seq.tostring())
handle.close()
#merge flanking sequences to the CtoT frame on rsID
F=pd.DataFrame({'RS# (dbSNP)': list(flanks.keys()), 'Flanks': [x for x in flanks.values()]})
```

Supplementary Note 3 (Continued)

```
BE_CtoT_editable = F.merge(BE_CtoT_editable, left_on='RS# (dbSNP)', right_on='RS# (dbSNP)', how='right')

BE_GtoA_editable = F.merge(BE_GtoA_editable, left_on='RS# (dbSNP)', right_on='RS# (dbSNP)', how='right')

ABE_TtoC_editable = F.merge(ABE_TtoC_editable, left_on='RS# (dbSNP)', right_on='RS# (dbSNP)', how='right')

ABE_AtoG_editable = F.merge(ABE_AtoG_editable, left_on='RS# (dbSNP)', right_on='RS# (dbSNP)', how='right')

# clinvar may refer to the opposite strand that was used in dbSNP;

# we want to allow clinvar reference alleles A and T with alternate alleles G and C respectively

#BE_CtoT_editable.to_csv('BE_CtoT_editable.csv', sep='\t', encoding = 'utf-8')

#BE_GtoA_editable.to_csv('BE_GtoA_editable.csv', sep='\t', encoding = 'utf-8')


BE_CtoT_editable['gRNAs']=None


BE_CtoT_editable['gRNAall']=None

for i in range(len(BE_CtoT_editable)):

    if type(BE_CtoT_editable.iloc[i].Flanks)==list and BE_CtoT_editable.iloc[i].Flanks!=[]:

        test=BE_CtoT_editable.iloc[i].Flanks[0]

        # define a potential gRNA spacer for each window positioning

        gRNAoptions=[test[(26-windowstart-j):(26-windowstart-j+lenpam+20)] for j in range(windowlen)]

        #if there is an appropriate PAM placed for a given gRNA spacer

        #save tuple of gRNA spacer, and the position of off-target Cs in the window

        gRNA=[(gRNAoptions[k],[x.start()+1 for x in re.finditer('C',gRNAoptions[k]) if windowstart-1<x.start()+1<windowend+1]) for k in range(len(gRNAoptions)) if regex.match(enc_pam['f'], gRNAoptions[k][-lenpam:])]

        gRNAsingleC=[]

        for g,c in gRNA:

            #if the target C is the only C in the window save this as a single C site

            if g[windowstart-1:windowend].count('C')==0:

                gRNAsingleC.append(g)

            #OPTIONAL uncomment the ELIF statement if you are interest in filtered based upon position of off-target C
```

Supplementary Note 3 (Continued)

```
        #if the target C is expected to be editted more efficiently than the off-target Cs, also save as a single C Site
        #elif all([p<priority[x] for x in c]):
            #gRNAsingleC.append(g)

    BE_CtoT_editable.gRNAs.iloc[i]=gRNAsingleC
    BE_CtoT_editable.gRNAall.iloc[i]=[g for g,c in gRNA]

#merge in phenotypes based upon MedGen IDs; remove redundant columns
BE_CtoT_editable=BE_CtoT_editable[['RS# (dbSNP)','GeneSymbol', 'gRNAs', 'gRNAall','Name', 'PhenotypeIDS', 'Origin', 'ReviewStatus', 'NumberSubmitters', 'LastEvaluated']]
BE_CtoT_editable.drop('PhenotypeIDS', inplace=True, axis=1)



BE_GtoA_editable['gRNAs']=None
BE_GtoA_editable['gRNAall']=None
for i in range(len(BE_GtoA_editable)):
    if type(BE_GtoA_editable.iloc[i].Flanks)==list and BE_GtoA_editable.iloc[i].Flanks!=[]:
        test=BE_GtoA_editable.iloc[i].Flanks[0]
        gRNAoptions=[test[(25+windowstart+j-20-lenpam):(25+windowstart+j)] for j in range(windowlen)]
        gRNA=[(gRNAoptions[k],[20+lenpam-x.start() for x in re.finditer('G',gRNAoptions[k]) if windowstart-1<20+lenpam-x.start()<windowend+1]) for k in range(len(gRNAoptions)) if regex.match(enc_pam['r'], gRNAoptions[k][:lenpam])]
        gRNAsingleG=[]
        for g,c in gRNA:
            if g[20+lenpam-windowstart-windowlen+1:20+lenpam-windowstart+1].count('G')==0:
                gRNAsingleG.append(g)
            #elif all([p<priority[x] for x in c]):
                #gRNAsingleC.append(g)
        BE_GtoA_editable.gRNAs.iloc[i]=gRNAsingleG
        BE_GtoA_editable.gRNAall.iloc[i]=[g for g,c in gRNA]
```

Supplementary Note 3 (Continued)

```
BE_GtoA_editable=BE_GtoA_editable[['RS# (dbSNP)','GeneSymbol', 'gRNAs', 'gRNAall','Name', 'PhenotypeIDS', 'Origin', 'ReviewStatus', 'NumberSubmitters', 'LastEvaluated']]

BE_GtoA_editable.drop('PhenotypeIDS', inplace=True, axis=1)


ABE_AtoG_editable['gRNAs']=None

ABE_AtoG_editable['gRNAall']=None

for i in range(len(ABE_AtoG_editable)):

    if type(ABE_AtoG_editable.iloc[i].Flanks)==list and ABE_AtoG_editable.iloc[i].Flanks!=[]:

        test=ABE_AtoG_editable.iloc[i].Flanks[0]

        # define a potential gRNA spacer for each window positioning

        gRNAoptions=[test[(26-windowstart-j):(26-windowstart-j+lenpam+20)] for j in range(windowlen)]

        #if there is an appropriate PAM placed for a given gRNA spacer

        #save tuple of gRNA spacer, and the position of off-target Cs in the window

        gRNA=[(gRNAoptions[k],[x.start()+1 for x in re.finditer('A',gRNAoptions[k]) if windowstart-1<x.start()+1<windowend+1]) for k in range(len(gRNAoptions)) if regex.match(enc_pam['f'], gRNAoptions[k][-lenpam:])]

        gRNAsingleA=[]

        for g,c in gRNA:

            #if the target C is the only C in the window save this as a single C site

            if g[windowstart-1:windowend].count('A')==0:

                gRNAsingleA.append(g)

            #OPTIONAL uncomment the ELIF statement if you are interest in filtered based upon position of off-target C

            #if the target C is expected to be editted more efficiently than the off-target Cs, also save as a single C Site

            #elif all([p<priority[x] for x in c]):

                #gRNAsingleC.append(g)


        ABE_AtoG_editable.gRNAs.iloc[i]=gRNAsingleA

        ABE_AtoG_editable.gRNAall.iloc[i]=[g for g,c in gRNA]


#merge in phenotypes based upon MedGen IDs; remove redundant columns

ABE_AtoG_editable=ABE_AtoG_editable[['RS# (dbSNP)','GeneSymbol', 'gRNAs', 'gRNAall','Name', 'PhenotypeIDS', 'Origin', 'ReviewStatus', 'NumberSubmitters', 'LastEvaluated']]
```

Supplementary Note 3 (Continued)

```
ABE_AtoG_editable.drop('PhenotypeIDS', inplace=True, axis=1)
ABE_AtoG_editable.to_csv('ABE_AtoG_editable.csv')


ABE_TtoC_editable['gRNAs']=None
ABE_TtoC_editable['gRNAall']=None
for i in range(len(ABE_TtoC_editable)):
    if type(ABE_TtoC_editable.iloc[i].Flanks)==list and ABE_TtoC_editable.iloc[i].Flanks!=[]:
        test=ABE_TtoC_editable.iloc[i].Flanks[0]
        gRNAoptions=[test[(25+windowstart+j-20-lenpam):(25+windowstart+j)] for j in range(windowlen)]
        gRNA=[(gRNAoptions[k],[20+lenpam-x.start() for x in re.finditer('T',gRNAoptions[k]) if windowstart-
1<20+lenpam-x.start()<windowend+1]) for k in range(len(gRNAoptions)) if regex.match(enc_pam['r'],
gRNAoptions[k][:lenpam])]
        gRNAsingleT=[]
        for g,c in gRNA:
            if g[20+lenpam-windowstart-windowlen+1:20+lenpam-windowstart+1].count('T')==0:
                gRNAsingleT.append(g)
            #elif all([p<priority[x] for x in c]):
                #gRNAsingleC.append(g)
        ABE_TtoC_editable.gRNAs.iloc[i]=gRNAsingleT
        ABE_TtoC_editable.gRNAall.iloc[i]=[g for g,c in gRNA]


ABE_TtoC_editable=ABE_TtoC_editable[['RS# (dbSNP)','GeneSymbol', 'gRNAs', 'gRNAall','Name',
'PhenotypeIDS', 'Origin', 'ReviewStatus', 'NumberSubmitters', 'LastEvaluated']]
ABE_TtoC_editable.drop('PhenotypeIDS', inplace=True, axis=1)



BE_CtoT_hasPAM=BE_CtoT_editable[[type(x)==list and x!=[] for x in BE_CtoT_editable.gRNAall]]
BE_CtoT_hasPAM.to_csv('BE_CtoT_hasPAM.csv')


BE_CtoT_SingleC=BE_CtoT_editable[[type(x)==list and x!=[] for x in BE_CtoT_editable.gRNAs]]
BE_CtoT_SingleC.to_csv('BE_CtoT_SingleC.csv')
```

Supplementary Note 3 (Continued)

```
BE_GtoA_hasPAM=BE_GtoA_editable[[type(x)==list and x!=[] for x in BE_GtoA_editable.gRNAall]]
BE_GtoA_hasPAM.to_csv('BE_GtoA_hasPAM.csv')
BE_GtoA_SingleG=BE_GtoA_editable[[type(x)==list and x!=[] for x in BE_GtoA_editable.gRNAs]]
BE_GtoA_SingleG.to_csv('BE_GtoA_SingleG.csv')

ABE_AtoG_hasPAM=ABE_AtoG_editable[[type(x)==list and x!=[] for x in ABE_AtoG_editable.gRNAall]]
ABE_AtoG_hasPAM.to_csv('ABE_AtoG_hasPAM.csv')
ABE_AtoG_SingleA=ABE_AtoG_editable[[type(x)==list and x!=[] for x in ABE_AtoG_editable.gRNAs]]
ABE_AtoG_SingleA.to_csv('ABE_AtoG_SingleA.csv')

ABE_TtoC_hasPAM=ABE_TtoC_editable[[type(x)==list and x!=[] for x in ABE_TtoC_editable.gRNAall]]
ABE_TtoC_hasPAM.to_csv('ABE_TtoC_hasPAM.csv')
ABE_TtoC_SingleT=ABE_TtoC_editable[[type(x)==list and x!=[] for x in ABE_TtoC_editable.gRNAs]]
ABE_TtoC_SingleT.to_csv('ABE_TtoC_SingleT.csv')

with open("Summary.txt", "w") as text_file:
    text_file.write("BE3/4 \n")
    text_file.write("single base %s \n" % (len(BE_CtoT_SingleC)+len(BE_GtoA_SingleG)))
    text_file.write("hasPAM %s \n" % (len(BE_CtoT_hasPAM)+len(BE_GtoA_hasPAM)))
    text_file.write("Pathogenic SNPs that can be targeted with BE %s \n" % (GC_AT_size))
    text_file.write("ABE \n")
    text_file.write("single base %s \n" % (len(ABE_AtoG_SingleA)+len(ABE_TtoC_SingleT)))
    text_file.write("hasPAM %s \n" % (len(ABE_AtoG_hasPAM)+len(ABE_TtoC_hasPAM)))
text_file.write("Pathogenic SNPs that can be targeted with ABE %s" % (AT_GC_size)
```

References

1 Friedmann, T. & Roblin, R. Gene therapy for human genetic disease? *Science* **175**, 949-955 (1972).

2 Cohen, S. N., Chang, A. C., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci U S A* **70**, 3240-3244 (1973).

3 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467 (1977).

4 Wu, R., Guo, L. H., Georges, F. & Narang, S. A. Rapid methods for sequencing DNA. *Trans N Y Acad Sci* **41**, 253-267 (1983).

5 Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A* **74**, 560-564 (1977).

6 Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-throughput sequencing technologies. *Mol Cell* **58**, 586-597, doi:10.1016/j.molcel.2015.05.004 (2015).

7 Muir, P. *et al.* The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biol* **17**, 53, doi:10.1186/s13059-016-0917-0 (2016).

8 Elliott, B. & Jasin, M. Repair of double-strand breaks by homologous recombination in mismatch repair-defective mammalian cells. *Mol Cell Biol* **21**, 2671-2682, doi:10.1128/MCB.21.8.2671-2682.2001 (2001).

9 Osman, F. & Subramani, S. Double-strand break-induced recombination in eukaryotes. *Prog Nucleic Acid Res Mol Biol* **58**, 263-299 (1998).

10 Scherer, S. & Davis, R. W. Replacement of chromosome segments with altered DNA sequences constructed in vitro. *Proc Natl Acad Sci U S A* **76**, 4951-4955 (1979).

11 Capecchi, M. R. Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet* **6**, 507-512, doi:10.1038/nrg1619 (2005).

12 Carroll, D. Genome engineering with zinc-finger nucleases. *Genetics* **188**, 773-782, doi:10.1534/genetics.111.131433 (2011).

13 Chevalier, B. S. *et al.* Design, activity, and structure of a highly specific artificial endonuclease. *Mol Cell* **10**, 895-905 (2002).

14 Miller, J., McLachlan, A. D. & Klug, A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J* **4**, 1609-1614 (1985).

15 Pavletich, N. P. & Pabo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* **252**, 809-817 (1991).

16 Li, L., Wu, L. P. & Chandrasegaran, S. Functional domains in Fok I restriction endonuclease. *Proc Natl Acad Sci U S A* **89**, 4275-4279 (1992).

17 Tebas, P. *et al.* Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med* **370**, 901-910, doi:10.1056/NEJMoa1300662 (2014).

18 Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* **14**, 49-55, doi:10.1038/nrm3486 (2013).

19 Liu, J. *et al.* Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. *J Genet Genomics* **39**, 209-215, doi:10.1016/j.jgg.2012.04.003 (2012).

20 Wood, A. J. *et al.* Targeted genome editing across species using ZFNs and TALENs. *Science* **333**, 307, doi:10.1126/science.1207773 (2011).

21 Sander, J. D. *et al.* Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol* **29**, 697-698, doi:10.1038/nbt.1934 (2011).

22 Lei, Y. *et al.* Efficient targeted gene disruption in Xenopus embryos using engineered transcription activator-like effector nucleases (TALENs). *Proc Natl Acad Sci U S A* **109**, 17484-17489, doi:10.1073/pnas.1215421109 (2012).

23 Tesson, L. *et al.* Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol* **29**, 695-696, doi:10.1038/nbt.1940 (2011).

24 Carlson, D. F. *et al.* Efficient TALEN-mediated gene knockout in livestock. *Proc Natl Acad Sci U S A* **109**, 17382-17387, doi:10.1073/pnas.1211446109 (2012).

25 Mussolino, C. *et al.* TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic Acids Res* **42**, 6762-6773, doi:10.1093/nar/gku305 (2014).

26 Gupta, R. M. & Musunuru, K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J Clin Invest* **124**, 4154-4161, doi:10.1172/JCI72992 (2014).

27 Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* **169**, 5429-5433 (1987).

28 Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551-2561, doi:10.1099/mic.0.28048-0 (2005).

29 Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712, doi:10.1126/science.1138140 (2007).

30 Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964, doi:10.1126/science.1159689 (2008).

31 Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607, doi:10.1038/nature09886 (2011).

32 Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096, doi:10.1126/science.1258096 (2014).

33 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

34 Jiang, F. & Doudna, J. A. CRISPR-Cas9 Structures and Mechanisms. *Annu Rev Biophys* **46**, 505-529, doi:10.1146/annurev-biophys-062215-010822 (2017).

35 Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477-1481, doi:10.1126/science.aab1452 (2015).

36 Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573, doi:10.1038/nature13579 (2014).

37 Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62-67, doi:10.1038/nature13011 (2014).

38 Fazio, T., Visnapuu, M. L., Wind, S. & Greene, E. C. DNA curtains and nanoscale curtain rods: high-throughput tools for single molecule imaging. *Langmuir* **24**, 10524-10531, doi:10.1021/la801762h (2008).

39 Visnapuu, M. L. & Greene, E. C. Single-molecule imaging of DNA curtains reveals intrinsic energy landscapes for nucleosome deposition. *Nat Struct Mol Biol* **16**, 1056-1062, doi:10.1038/nsmb.1655 (2009).

40 Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* **32**, 677-683, doi:10.1038/nbt.2916 (2014).

41 O'Geen, H., Henry, I. M., Bhakta, M. S., Meckler, J. F. & Segal, D. J. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res* **43**, 3389-3404, doi:10.1093/nar/gkv137 (2015).

42 Wu, X. *et al*. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* **32**, 670-676, doi:10.1038/nbt.2889 (2014).

43 Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110-113, doi:10.1038/nature15544 (2015).

44 Josephs, E. A. *et al.* Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res* **43**, 8924-8941, doi:10.1093/nar/gkv892 (2015).

45 Guilinger, J. P. *et al.* Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nat Methods* **11**, 429-435, doi:10.1038/nmeth.2845 (2014).

46 Frock, R. L. *et al.* Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol* **33**, 179-186, doi:10.1038/nbt.3101 (2015).

47 Rogers, J. M. *et al.* Context influences on TALE-DNA binding revealed by quantitative profiling. *Nat Commun* **6**, 7440, doi:10.1038/ncomms8440 (2015).

48 Meckler, J. F. *et al.* Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res* **41**, 4118-4128, doi:10.1093/nar/gkt085 (2013).

49 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).

50 Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **31**, 233-239, doi:10.1038/nbt.2508 (2013).

51 Wiedenheft, B. *et al.* RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A* **108**, 10092-10097, doi:10.1073/pnas.1102716108 (2011).

52 Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* **108**, 10098-10103, doi:10.1073/pnas.1104144108 (2011).

53 Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31**, 839-843, doi:10.1038/nbt.2673 (2013).

54 Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822-826, doi:10.1038/nbt.2623 (2013).

55 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832, doi:10.1038/nbt.2647 (2013).

56 Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**, 833-838, doi:10.1038/nbt.2675 (2013).

57 Lin, Y. *et al.* CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res* **42**, 7473-7485, doi:10.1093/nar/gku402 (2014).

58 Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473-1475, doi:10.1093/bioinformatics/btu048 (2014).

59 Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186-191, doi:10.1038/nature14299 (2015).

60 Duan, J. *et al.* Genome-wide identification of CRISPR/Cas9 off-targets in human genome. *Cell Res* **24**, 1009-1012, doi:10.1038/cr.2014.87 (2014).

61 Veres, A. *et al.* Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell* **15**, 27-30, doi:10.1016/j.stem.2014.04.020 (2014).

62 Smith, C. *et al.* Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell* **15**, 12-13, doi:10.1016/j.stem.2014.06.011 (2014).

63 Yang, L. *et al.* Targeted and genome-wide sequencing reveal single nucleotide variations impacting specificity of Cas9 in human stem cells. *Nat Commun* **5**, 5507, doi:10.1038/ncomms6507 (2014).

64 Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187-197, doi:10.1038/nbt.3117 (2015).

65 Kim, D. *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods* **12**, 237-243, 231 p following 243, doi:10.1038/nmeth.3284 (2015).

66 Crosetto, N. *et al.* Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat Methods* **10**, 361-365, doi:10.1038/nmeth.2408 (2013).

67 Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* **32**, 577-582, doi:10.1038/nbt.2909 (2014).

68 Tsai, S. Q. *et al.* Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol* **32**, 569-576, doi:10.1038/nbt.2908 (2014).

69 Aouida, M. *et al.* Efficient fdCas9 Synthetic Endonuclease with Improved Specificity for Precise Genome Engineering. *PLoS One* **10**, e0133373, doi:10.1371/journal.pone.0133373 (2015).

70 Pattanayak, V., Ramirez, C. L., Joung, J. K. & Liu, D. R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods* **8**, 765-770, doi:10.1038/nmeth.1670 (2011).

71 Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**, 279-284, doi:10.1038/nbt.2808 (2014).

72 Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84-88, doi:10.1126/science.aad5227 (2016).

73 Hubbard, B. P. *et al.* Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat Methods* **12**, 939-942, doi:10.1038/nmeth.3515 (2015).

74 Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-485, doi:10.1038/nature14592 (2015).

75 Kiani, S. *et al.* Cas9 gRNA engineering for genome editing, activation and repression. *Nat Methods* **12**, 1051-1054, doi:10.1038/nmeth.3580 (2015).

76 Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* **24**, 132-141, doi:10.1101/gr.162339.113 (2014).

77 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, doi:10.1038/nature17946 (2016).

78 Gaudelli, N. M. *et al.* Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471, doi:10.1038/nature24644 (2017).

79 Findlay, G. M., Boyle, E. A., Hause, R. J., Klein, J. C. & Shendure, J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513**, 120-123, doi:10.1038/nature13695 (2014).

80 Yang, L. *et al.* Optimization of scarless human stem cell genome editing. *Nucleic Acids Res* **41**, 9049-9061, doi:10.1093/nar/gkt555 (2013).

81 Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771, doi:10.1016/j.cell.2015.09.038 (2015).

82 Kim, E. *et al.* In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat Commun* **8**, 14500, doi:10.1038/ncomms14500 (2017).

83 Muller, M. *et al.* Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. *Mol Ther* **24**, 636-644, doi:10.1038/mt.2015.218 (2016).

84 Lee, C. M., Cradick, T. J. & Bao, G. The Neisseria meningitidis CRISPR-Cas9 System Enables Specific Genome Editing in Mammalian Cells. *Mol Ther* **24**, 645-654, doi:10.1038/mt.2016.8 (2016).

85 Kleinstiver, B. P. *et al.* Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol* **33**, 1293-1298, doi:10.1038/nbt.3404 (2015).

86 Mitsunobu, H., Teramoto, J., Nishida, K. & Kondo, A. Beyond Native Cas9: Manipulating Genomic Information and Function. *Trends Biotechnol* **35**, 983-996, doi:10.1016/j.tibtech.2017.06.004 (2017).

87 Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168**, 20-36, doi:10.1016/j.cell.2016.10.044 (2017).

88 Badran, A. H. & Liu, D. R. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat Commun* **6**, 8425, doi:10.1038/ncomms9425 (2015).

89 Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503, doi:10.1038/nature09929 (2011).

90 Badran, A. H. *et al.* Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. *Nature* **533**, 58-63, doi:10.1038/nature17938 (2016).

91 Bryson, D. I. *et al.* Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat Chem Biol*, doi:10.1038/nchembio.2474 (2017).

92 Packer, M. S., Rees, H. A. & Liu, D. R. Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nat Commun* **8**, 956, doi:10.1038/s41467-017-01055-9 (2017).

93 Meng, X. & Wolfe, S. A. Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid system. *Nat Protoc* **1**, 30-45, doi:10.1038/nprot.2006.6 (2006).

94 Dove, S. L., Joung, J. K. & Hochschild, A. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature* **386**, 627-630, doi:10.1038/386627a0 (1997).

95 Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407-410, doi:10.1038/nature24268 (2017).

96 Gao, L. *et al.* Engineered Cpf1 variants with altered PAM specificities. *Nat Biotechnol* **35**, 789-792, doi:10.1038/nbt.3900 (2017).

97 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190, doi:10.1101/gr.849004 (2004).

98 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, doi:10.1038/nchembio.1453 (2014).

99 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, doi:10.1007/s00335-017-9697-4 (2017).

100 Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* **12**, 326-328, doi:10.1038/nmeth.3312 (2015).

101 Komor, A. C. *et al.* Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv* **3**, eaao4774, doi:10.1126/sciadv.aao4774 (2017).

102 Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* **32**, 1262-1267, doi:10.1038/nbt.3026 (2014).

103 Zhang, Y. *et al.* Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci Rep* **4**, 5405, doi:10.1038/srep05405 (2014).

104 Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol* **35**, 371-376, doi:10.1038/nbt.3803 (2017).

105 Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980-985, doi:10.1093/nar/gkt1113 (2014).

106 Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495, doi:10.1038/nature16526 (2016).

107 Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949, doi:10.1016/j.cell.2014.02.001 (2014).