Engineering and evolution of precision genome editing agents

Citation

Huang, Tony. 2022. Engineering and evolution of precision genome editing agents. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Permanent link

https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37371983

Terms of Use

Share Your Story

HARVARD UNIVERSITY
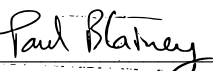## Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Chemical Biology

have examined a dissertation entitled

Engineering and Evolution of Precision Genome Editing Agents

presented by

Tony P. Huang

candidate for the degree of Doctor of Philosophy and hereby

certify that it is worthy of acceptance.

Signature _____

Typed name: Prof. Brian Liau

Signature _____

Typed name: Prof. Paul Blainey

Signature _____

Typed name: Prof. George Church

Signature _____

Typed name: Prof. David Liu

Signature _____

Typed name: Prof.

Date: 04/04/2022

Engineering and evolution of precision genome editing agents

A dissertation presented

by

Tony P. Huang

To

The Committee on Higher Degrees in Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemical Biology

Harvard University

Cambridge, Massachusetts

April 2022

Dissertation Advisor: David R. Liu   Tony P. Huang

Engineering and evolution of precision genome editing agents

Abstract

Genome editing agents have had a transformative effect on the study of biological systems, the engineering of cells and organisms with novel properties, and even the treatment of genetic diseases in the clinic. This revolution was first enabled through the discovery of programmable nucleases, including zinc-finger nucleases (ZFNs)[1-5], transcription activator-like effector nucleases (TALENs)[6-11], and CRISPR-associated (Cas) nucleases[12-15]. These nucleases induce targeted double-stranded DNA breaks (DSBs), which stimulate cellular repair processes that ultimately result in genome modification. While effective for some desired genome editing applications, such as targeted gene disruption, the highly stochastic nature of DSB repair results in a variety of outcomes, typically small insertions or deletions (indels), making targeted modification and correction of a gene difficult and imprecise.

To address this limitation, several precision genome editing technologies building upon CRISPR-Cas nucleases have been developed to enable precise installation of desired modifications without DSBs. These tools include base editing (BE), which enables the installation of any of the four nucleotide transitions (A-to-G, C-to-T, T-to-C, or G-to-A) and some transversions (C-to-G), prime editing (PE), which enables the programmable modification of a small region (generally < 50 bp) of DNA, and CRISPR-associated transposons or PE+recombinases, which have the potential to enable the installation of gene-sized fragments. While in theory these tools greatly broaden the scope of genome modifications available to researchers, in practice their initial application was partially constrained by three major types of limitations: 1) efficiency and generalizability, 2) specificity, and 3) deliverability. As such, recent

developments have focused on minimizing these limitations through a variety of natural discovery, rational engineering, or evolution approaches.

In this thesis, I will describe efforts to improve the efficiency, generalizability, and specificity of precision genome editing technologies using engineering and directed evolution, with a focus on base editing. In chapter two, I describe the successful incorporation of natural and engineered variants of the canonically used SpCas9 to expand the targeting range of adenine base editing. In chapter three, we show that a novel, sequence-agnostic functional selection is better able to evolve high-activity variants of non-SpCas9 orthologs capable of efficient base editing. Lastly, in chapter four, I describe early efforts to incorporate a negative counterselection into the sequence-agnostic selection, with the goal of developing a generalizable method for generating tailor-made, genome site-specific Cas variants for use in precision genome editing. Together, the work described herein include novel tools that enhance the generalizability of precision genome editing while also providing a set of methods that will enable the further development of improved variants.

Table of Contents

Table of Figures

List of Tables

Acknowledgements

This thesis would not have been possible without the support and mentorship of a great number of people throughout my life. First, I would like to thank my family – my mother, Shinan, my father, Guoyuan, and my sister, Sunny, who have always believed in me and gotten me to where I am today. My parents immigrated to the United States over 30 years ago with less than $5,000 to their name. Through their hard work and perseverance, they made sure that my sister and I had all that we needed and the opportunity to achieve the American Dream. My sister, being the older of us, has always served as a fantastic role model for me. Among many things, she was the one who introduced me to science, and she has always been my greatest supporter. My partner, Shannon, who has immensely brightened the latter half of my graduate career. Being able to share the joys and pains (in science and in life) of the past few years with her has been immeasurably wonderful.

I would also like to thank the mentors, scientific and otherwise, who have guided me throughout my academic career. First, I would like to thank my PhD advisor, Professor David R. Liu, for his support and mentorship throughout the past six years. David's enthusiasm and passion for pursuing the unknown bounds of science, particularly in areas relevant to human health, have and will continue to inspire me in my own scientific endeavors. His mentorship has been invaluable in making me a more critical and thorough researcher. Furthermore, I have greatly enjoyed my time in the laboratory environment that he has created for his students. Second, I would like to thank my dissertation committee, Professors Brian Liau, Paul Blainey, and George Church. Their advice over the past few years have greatly helped me clarify paths in my research and in my career. My undergraduate advisor, Dr. Jimmy Blair, who introduced me to biochemical research and pushed me to continue pursuing a career in research. His unwavering support and advice during my time at Williams were integral to the path I chose for my PhD. Finally, much of the work described in this thesis was made possible by the great

Chapter 1. Introduction

This chapter has been partially adapted from:

Contributions:

The adapted parts of this chapter were written by T.P.H., G.A.N. and D.R.L. of the referenced work. I would like to thank Dr. Gregory Newby for being a constant resource for key developments in the genome editing delivery field.

1.1 Introduction to precision genome editing

1.1.1 Overview

Programmable genome editing has been a longstanding goal of the scientific community, with widespread implications in addressing pressing challenges from human health to agricultural sustainability. The earliest class of widely used modern genome editing agents were the programmable nucleases, including meganucleases, zinc-finger nucleases (ZFNs)[1-5], transcription activator-like effector nucleases (TALENs)[6-11], and more recently CRISPR-associated (Cas) nucleases[12-15]. These agents induce targeted DSBs to stimulate cellular repair processes that ultimately result in genome modification. Cellular repair of DSBs, however, typically leads to a variety of outcomes through multiple competing pathways[16-19], including potential cell cycle arrest from p53-mediated DNA damage responses[20,21] or complex chromosomal rearrangements. DSBs are typically repaired through non-homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ), which resect then join DNA-ends together, sometimes utilizing microhomology (in the case of MMEJ) flanking the DNA break to align end-joining. While both NHEJ and MMEJ are predictable[22-24], they are uncontrollable and stochastic, resulting in a distribution of gene disruptions through insertions, deletions, duplications, or other DNA rearrangements[16,17], limiting the precision of these processes for precise gene modification.

To more controllably correct DSBs, researchers can include a donor DNA template with homology to the genomic region containing the DSB, which stimulates homology-directed repair (HDR). HDR utilizes the donor template to precisely replace the corresponding homologous region around the DSB[25,26]. However, this process is largely limited to dividing cells. Furthermore, in cells that support HDR, desired products are typically disfavored relative to indels from end-joining processes, resulting in a mixture of HDR-corrected products and NHEJ or MMEJ-mediated insertions or deletions[26-29].

The limitations of DSB-mediated genome editing, particularly the lack of precision and range of accessible modifications, inspired the development of further refined tools capable of efficiently and precisely installing or removing specific DNA modifications without the stochastic outcomes that accompany generation of DSBs[30]. These precision genome editing tools comprise of base editing (BE), which enables the installation of any of the four nucleotide transitions (A-to-G, C-to-T, T-to-C, or G-to-A)[31,32] and some transversions (C-to-G)[33-35], prime editing (PE)[36], which enables the programmable modification of a small region (generally < 50 bp) of DNA, and CRISPR-associated transposons[37-39] or PE+recombinase[40,41] combination approaches, which have the potential to enable the installation of gene-sized fragments. Together, these tools in theory enable highly flexible modification of any target genomic locus and the potential correction of many of the pathogenic genetic lesions that affect human health.

1.1.2   Development of base editing

Among the many pathogenic human genetic variants that have been characterized, the majority are single-nucleotide polymorphisms (SNPs)[42] – changes that occur at a single base pair within the genome. Despite the relative simplicity of the genomic modification, SNPs can have devastating effects, being the cause of several uncured diseases including sickle cell anemia[43], variants of cystic fibrosis[44], and progeria[45], among others. Existing treatments are both costly and primarily palliative. While partially addressable by DSB-mediated methods (MMEJ, NHEJ, or HDR), many of these SNPs occur in essential genomic regions (e.g. coding sequences) that are highly susceptible to disruption by the stochastic outcomes of DSB repair[16,17]. Methods to efficiently install targeted single-nucleotide mutations without inducing DSBs are therefore desirable.

To address this need, base editing, a genome editing method that directly converts targeted base pairs without requiring double-strand DNA breaks, HDR, or donor DNA template[31,32], was developed. Base editors contain two core components: a programmable DNA-

binding protein such as a catalytically impaired Cas nuclease that cannot make DSBs, and a DNA modifying enzyme that selectively targets single-stranded DNA (ssDNA)[31,32]. Cas domain binding to a programmed genomic locus exposes a small segment of ssDNA in an R-loop[12,46]. In base editing, this ssDNA R-loop then serves as the substrate for the fused ssDNA-modifying enzyme. The high effective concentration of target DNA to the ssDNA-modifying enzyme due to fusion with the Cas domain promotes efficient and selective nucleobase modification. Editing is canonically localized to a narrow window of nucleobases ("editing window") exposed by Cas binding, although strategies have been introduced to expand[47-51] or further restrict[52-54] this region. Following DNA modification, the resulting base pairing mismatch is subject to cellular DNA repair. Most base editors bias the outcome of DNA repair by using a nickase form of the Cas protein to nick the non-deaminated strand[31,32]. Nicking stimulates resynthesis of the non-edited strand using the deaminated strand as a template[55,56], resulting in conversion of both DNA strands at the target position to the desired base pair[32].

Three classes of DNA base editors have been developed: cytosine base editors (CBEs), which convert C•G base pairs to T•A base pairs, adenine base editors (ABEs), which convert A•T base pairs to G•C base pairs[31,32], and CGBEs[33-35], which convert C•G base pairs to G•C base pairs. CBEs typically have three components: a cytidine deaminase, a Cas nickase, and uracil glycosylase inhibitor (UGI)[32], which inhibits counterproductive DNA repair processes. ABEs are comprised of two components: an evolved adenosine deaminase capable of recognizing ssDNA and a Cas ortholog nickase. Finally, CGBEs comprise of three or more components: a cytidine deaminase, a Cas nickase, and one or more DNA repair proteins that facilitate formation and repair of abasic sites[33-35]. Together, ABEs, CBEs, and CGBEs can generate all four transition mutations and some transversion mutations, which in theory could be used to install or correct the majority of known human pathogenic SNPs[42,57,58].

1.1.3   Development of prime editing

Although base editing enables the potential correction of a significant proportion of

SNPs, genetic lesions of different nature (insertions, deletions, other transversion SNPs, etc.)

remained largely inaccessible. To further enable access to some of these classes of mutations,

prime editing was developed[36]. In prime editing (PE), a nickase Cas domain is fused to a

reverse transcriptase (RT) rather than a deaminase. The guide RNA is modified into a prime

editing guide RNA (pegRNA), in which the 3' end is extended with a template region containing

the desired modification and a primer binding site (PBS) which anneals to the nicked strand

generated by the Cas nickase. Following Cas nicking and pegRNA annealing, the 3' end of the

nick is extended by the RT using the pegRNA as a template. The newly synthesized 3' flap is

then preferentially retained over the existing 5' flap due to natural DNA replication and repair

mechanisms that typically remove 5' flaps. Finally, mismatch repair (MMR) either removes the

newly installed modification or corrects the unedited strand, leading to a mixture of unedited or

desired product. Notably, because of the mechanism of modification, reverse transcription, PE

in theory can enable any modification that can be encoded within one or more[40,59,60] pegRNAs

and subsequently incorporated by cellular repair.

1.1.4   Discovery of CRISPR-associated transposases

Together, BE and PE approaches enable the installation or removal of any small genetic

lesion. However, one remaining frontier is the precise, programmable installation of large, gene-

sized fragments. Although a few viral gene therapy approaches have been approved in the

clinic, many of these treatments provide expression of a corrected copy of a protein of interest

through episomal DNA, rather than endogenous replacement or insertion of the corrected

gene[61]. This episomal approach can lead to undesired outcomes such as incorrect dosing or

unintended and uncontrolled insertion of the exogenous DNA[61,62]. Other experimental

approaches, such as recombinases, have also been attempted. However, recombinases are

much more site specific, limiting their utility[63,64]. Recent approaches to combine PE and recombinases[40,41] have yielded some success in enabling programmable insertion of large genetic payloads. However, due to the current limitations of both PE and recombinase efficiencies as standalone technologies, their multiplicative efficiency will require substantial further optimization.

Alternatively, several groups have recently discovered transposases associated with CRISPR-Cas effectors[37-39,65]. These systems comprise of Tn7-like transposases which are directed to a genomic site by an associated CRISPR-Cas complex, enabling site-specific insertion of large genomic payloads. However, given the novelty of this technology, significant efficiency limitations remain, limiting its applicability for therapeutic precision genome editing. The CRISPR-associated transposase systems described so far have only been validated in bacterial systems. While insertion in bacterial system can be quite robust, including insertion into the *E. coli* native genome, these properties have not yet translated to mammalian cell transposition[37-39]. Subsequent engineering or evolution approaches may further enable this technology. Nevertheless, technologies like PE+recombinases and CRISPR-associated transposases provide starting platforms for the future delivery of large genetic payloads into native genomes without the formation of double stranded breaks or undesired viral integration.

1.2   Improvements to precision genome editing

The development of precision genome editing technologies have rapidly enabled the study and correction of previously inaccessible genetic lesions and have already entered the clinic for previously uncurable diseases. However, while these technologies in theory could install or remove any genetic modification of interest, their early implementations in practice were limited by several constraints. These constraints include but are not limited to: (1) on-target efficiency of the target edit, (2) targeting scope of the editing technology, (3) on-to-off target specificity of the target edit, and/or (4) delivery of the editing agent[30,66]. To address these limitations, researchers have used a variety of engineering or evolution approaches to great success. These successes were largely aided by the modular nature of most precision genome editing technologies, enabling the improvement of various components (e.g. the targeting domain or the effector domain) in isolation that could then be generalized across editing modalities[30,66]. While newer technologies like CRISPR-associated transposases require further optimization, more mature technologies like base editing have seen a large array of architectural and enzymatic improvements that have drastically expanded their applicability[30].

In this thesis, I describe efforts to apply engineering and directed evolution approaches to improve the efficiency, generalizability, and specificity of precision genome editing technologies, with a focus on base editing. In chapter two, successful incorporation of natural and engineered Cas variants enables an expansion of the targeting range of base editing. In chapter three, we show that a novel, sequence-agnostic functional selection generalizable to any Cas protein is better able to evolve high-activity variants of non-SpCas9 orthologs capable of efficient base editing. Lastly, in chapter four, I describe early efforts to incorporate a negative counterselection into the sequence-agnostic selection, with the goal of developing a generalizable platform for facile development of tailor-made, genome site-specific Cas variants capable of precision genome editing. Collectively, the work described provide novel insights that enhance the

7

generalizability of precision genome editing while potentially providing a platform for the future

development of tools optimized for applications ranging from agriculture to therapeutics.

Chapter 2. Broadening the targeting scope of base editing using circularly permuted SpCas9 and Cas9 PAM variants

This chapter was adapted from:

Contributions:

This chapter was written by T.P.H., K.T.Z., and D.R.L. with input and review from all other coauthors. The work described in this chapter reflect experiments designed and conducted by T.P.H. and K.T.Z. with help from N.M.G. and S.M.M. Circularly permuted SpCas9 variants were provided by the Savage laboratory, and insights on usage were provided by B.L.F., C.F., and D.F.S. I want to especially thank Prof. Dave Savage for being a fantastic collaborator and a wonderful resource over my graduate career.

2.1 Introduction

Base editing has enabled the programmable conversion of one base pair to another without making double-stranded DNA breaks (DSBs),[31,32,67-69] and has already been widely used to install or correct point mutations in both research and therapeutic settings. Two classes of base editors: cytosine base editors (CBEs), which convert C•G to T•A[52,70,71], and adenine base editors (ABEs), which convert A•T to G•C[72,73], collectively enable the installation or correction of all four transition mutations, accounting for the majority of pathogenic human point mutations[74,75]. Unlike nuclease-based genome editing agents, base editors do not make DSBs, thereby minimizing the formation of detrimental editing byproducts or the activation of p53-mediated stress responses.

Base editors use a catalytically impaired, programmable DNA binding domain (typically a Cas nickase) to recognize the target DNA site. Cas protein binding generates a single-stranded DNA (ssDNA) R-loop, in which the exposed nucleobases can then be deaminated by a fused effector domain. Because Cas proteins have a protospacer adjacent motif (PAM) requirement to bind target DNA, base editing typically can only occur ~15±2 nucleotides upstream of an accessible PAM. To minimize this targeting limitation, CBE variants that use Cas9 homologs with different PAM requirements[52,69,76,77] were generated to increase the likelihood that a target site supports cytosine base editing. In contrast, far fewer ABE variants with distinct PAMs had been described[78-80].

Additionally, both CBEs and ABEs, in the originally described architecture comprising nickase SpCas9, mutate target base pairs within a small (typically ~4- to 5-nucleotide) window. The width of this editing window, together with PAM availability, define the accessible scope of base editing. In some cases, the target base is located outside the base editing window relative to an available PAM. Moreover, for applications such as broad-spectrum mutagenesis, gene

disruption via introduction of premature stop codons, or splice site disruption, a wider editing window is desirable.

In this chapter, we describe the development and characterization ABEmax[81] variants optimized for mammalian cell use at target sites with non-NGG PAMs. We also develop CBEs and ABEs that use circularly permuted SpCas9 (CP-Cas9) variants, which exhibit expanded base editing windows that span up to ~8-9 nucleotides of the protospacer. The resulting CP-CBEmax variants exhibit higher product purities, in addition to expanded editing windows, while CP-ABEmax variants maintain the high product purities typical of ABEs. Together, these new PAM variant ABEs and circularly permuted base editors expand the targeting scope of base editing.

2.2 Results

2.2.1 Development of PAM-variant ABEs using Cas9 orthologs

We and others had previously reported the compatibility of CBEs with CRISPR proteins that recognize PAMs other than NGG, thereby expanding their targeting scope. These variants included evolved *S. pyogenes* Cas9 variants SpCas9-VQR/SpCas9-VRQR (PAM: NGA)[82,83], SpCas9-VRER (PAM: NGCG)[82], xCas9 (PAM: NGN)[77], SpCas9-NG (PAM: NG)[84], *S. aureus* Cas9 (PAM: NNGRRT)[85,86] and its modified variant KKH (PAM: NNNRRT)[87], and *L. bacterium* Cas12a (Cpf1, PAM: TTTV where V = A, C, or G)[76,88]. Furthermore, we previously showed that optimization of both codon usage and nuclear localization in both cytosine (BE4max, referred to hereafter as CBEmax) and adenine base editors (ABEmax), greatly enhances base editing activity in mammalian cells[81,89]. We hypothesized that the evolved TadA deoxyadenosine deaminase domain might be similarly compatible with other CRISPR proteins, yielding high activity PAM variant ABEs when constructed in the optimized architecture.

To evaluate the compatibility of evolved TadA with alternate Cas variants, we created ABEmax variants replacing the SpCas9 nickase component with two engineered SpCas9 variants with altered PAM specificities: VRQR-SpCas9 (PAM: NGA) and VRER-SpCas9 (PAM: NGCG) (**Figure 2.1**), yielding VRQR-ABEmax and VRER-ABEmax, respectively. We then evaluated base editing activity at six endogenous human genomic loci for each PAM in human HEK293T cells.
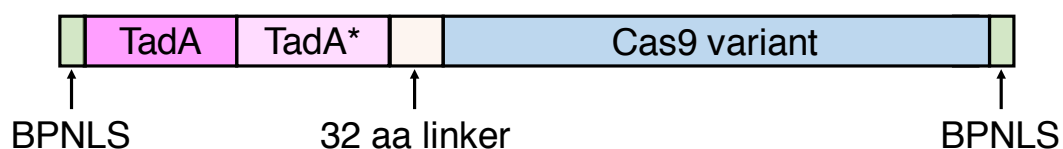
Figure 2.1. Optimized ABEmax architecture used for PAM variant construction.

Across six endogenous NGA PAM-containing sites, we observed low editing efficiency with ABEmax, averaging 11±2.1% A•T-to-G•C conversion (mean and s.d. of three biological replicates at six genomic sites, reporting the target A with the highest conversion frequency). In contrast, VRQR-ABEmax resulted in 35±4.6%A•T-to-G•C conversion across the same six genomic sites, a 3.2-fold average improvement (**Figure 2.2**). We also compared VRQR-ABEmax activity to that of ABEmax variants that use either xCas9[77] or the recently reported SpCas9-NG[84], both of which are active on some NGN PAM sites (xABEmax and NG-ABEmax, respectively). The average editing activity of xABEmax was 2.7-fold lower than that of VRQR-ABEmax (**Figure 2.2**). NG-ABEmax exhibited comparable activity to VRQR-ABEmax at some sites, but overall lower activity than VRQR-ABEmax with an average of 24±3.9% A•T-to-G•C conversion at these six genomic sites. Thus, VRQR-SpCas9, engineered specifically to recognize NGA PAM sites[28], supports more efficient editing at these sites than other evolved Cas9 variants.
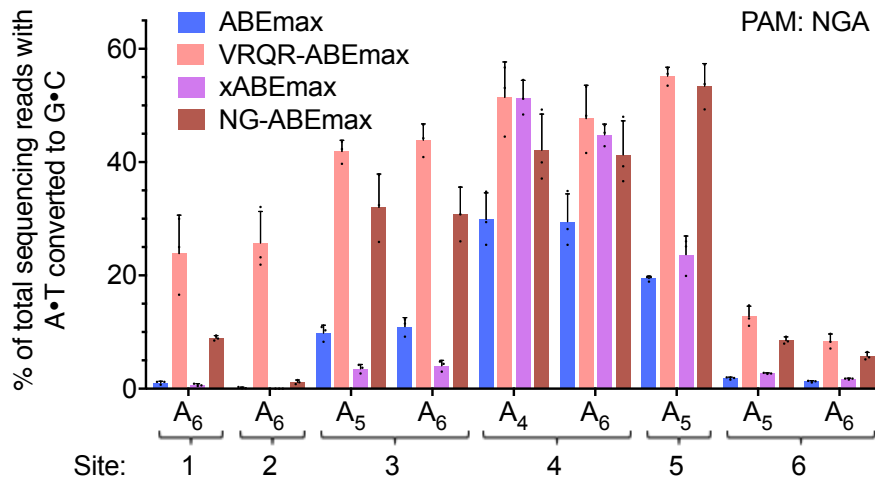
Figure 2.2. Activity of SpCas9-variant ABEs at genomic sites with NGA PAMs.
Base editing in HEK293T cells by ABEmax, VRQR-ABEmax, xABEmax, and NG-ABEmax at six genomic sites containing an NGA PAM. Subscripted numbers indicate protospacer positions, counting the first base of the PAM as position 21. Values and error bars reflect the mean±s.d. of three independent biological replicates performed by different researchers on different days.

At the six tested endogenous genomic sites containing NGCG PAMs, we observed minimal activity from ABEmax and xABEmax in HEK293T cells (**Figure 2.3**). VRER-ABEmax, however, greatly improved A•T-to-G•C conversion efficiencies at all tested sites, averaging 40±3.6% conversion, a 7.0-fold improvement over ABEmax. Because the VRQR variant differs from the VQR variant (which functions on both NGA and NGCG PAMs[82]) only by the addition of the G1218R mutation, which is also present in VRER-SpCas9, we suspected that VRQR retains activity on NGCG PAMs. Indeed, VRQR-ABEmax exhibited a further 1.3-fold improvement in editing efficiencies (averaging 50±3.6% A•T-to-G•C conversion) at the same six NGCG PAM sites compared to VRER-ABEmax (**Figure 2.3**). NG-ABEmax performed equally well on the six NGCG PAM-containing genomic sites as VRQR-ABEmax, averaging 51±5.9% A•T-to-G•C conversion. VRER-ABEmax, VRQR-ABEmax, and NG-ABEmax did not exhibit significant indel formation or an apparent shift in base editing window (**Figure 2.4, Figure 2.5**).
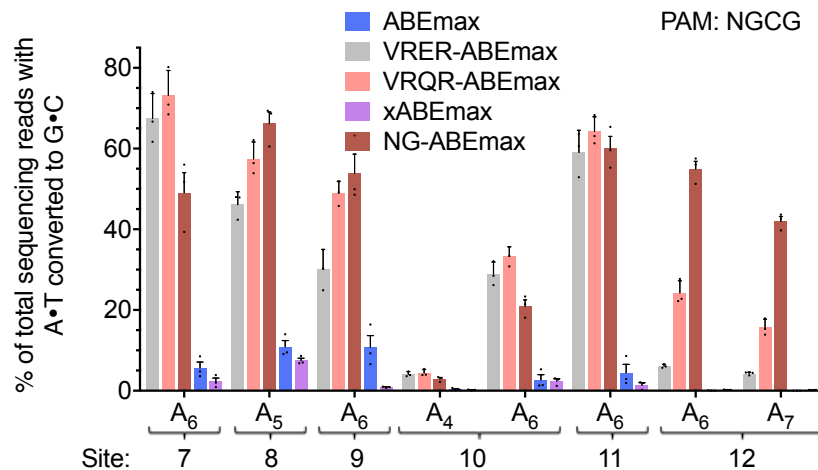
Figure 2.3. Activity of SpCas9-variant ABEs at genomic sites with NGCG PAMs.
Base editing in HEK293T cells by ABEmax, VRQR-ABEmax, xABEmax, and NG-ABEmax at six genomic sites containing an NGCG PAM. Subscripted numbers indicate protospacer positions, counting the first base of the PAM as position 21. Values and error bars reflect the mean±s.d. of three independent biological replicates performed by different researchers on different days.

a

b

a

b

PAM: NGCG
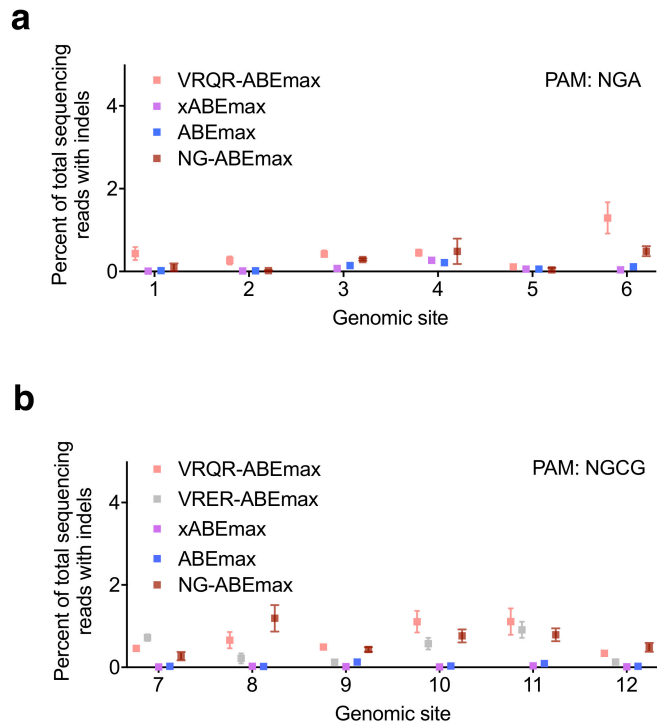
Figure 2.4. Indel frequencies for
(a) Percent of all sequencing reads
ABEmax, xABEmax, ABEmax, or N
HEK293T cells. **(b)** Percent of all s
VRQR-ABEmax, VRER-ABEmax,
containing an NGCG PAM in HEK2
three independent biological replic

Figure 2.5. Average editing window of SpCas9-variant ABEs.
Heat maps showing average editing efficiency by SpCas9-derived ABE variants at each protospacer position across sites containing each PAM listed (n = 6). Spaces crossed out indicate a position for which no target base was present among all the genomic sites tested. Values and error bars reflect the pooled mean of three independent biological replicates at sites containing an adenine at the target protospacer position.

Next, we evaluated the base editing activity of VRQR-ABEmax and NG-ABEmax on three genomic sites that we previously showed were edited by xABE[77] containing PAMs other than NGA or NGCG. Of the three sites tested (with GAT and two NGCC PAMs), VRQR-ABEmax exhibited an average of 2.3-fold greater A•T-to-G•C conversion activity on both NGCC PAM sites, but 2.6-fold lower activity on a GAT PAM site, compared to xABEmax (**Figure 2.6**). NG-ABEmax exhibited a 1.5-fold greater A•T-to-G•C conversion activity on both NGCC PAM sites and a 1.7-fold greater activity on the GAT PAM site. These data indicate that the VRER-, VRQR-, and SpCas9-NG variants are compatible with the ABEmax architecture and retain base editing activity at sites containing their cognate non-NGG PAMs. In most cases, VRQR-ABEmax tends to outperform other reported SpCas9 variants on sites containing either NGCG or NGA PAMs, and at some sites with NGCC PAMs. For alternative PAMs, such as the GAT PAM site examined in this study, xABEmax or base editors derived from Cas9-NG[84] are preferred. Taken together, these results indicate that targets lacking NGG PAMs can be accessed by multiple

16

evolved or engineered Cas9 variants, and that the optimal base editor may vary on a site-by-site basis.
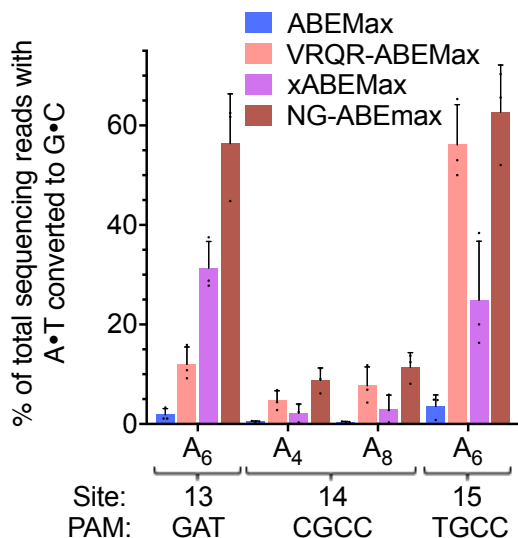
Figure 2.6. Activity of SpCas9-variant ABEs at three non-NGA, NGCG PAM sites.
Base editing in HEK293T cells by ABEmax, VRQR-ABEmax, xABEmax, and NG-ABEmax at three genomic sites (PAMs: GAT, CGCC, TGCC). Values and error bars reflect the mean±s.d. of three independent biological replicates performed by different researchers on different days.

To further expand the targeting scope of ABE, we examined whether the Cas9 ortholog from *S. aureus* (SaCas9) is compatible with the ABEmax architecture. SaCas9 naturally targets NNGRRT PAMs[85], and an evolved variant, SaKKH, recognizes NNNRRT PAMs[87]. We generated both SaCas9 and SaKKH ABEmax variants and tested them on six endogenous NNGRRT PAM sites and six endogenous NNHRRT PAM sites in HEK293T cells.

Observed A•T-to-G•C conversion activity varied substantially from site-to-site, but averaged 22±2.3% and 26±5.7% A•T-to-G•C conversion for SaABEmax on six NNGRRT PAM sites and SaKKH-ABEmax on six NNHRRT sites, respectively, with minimal indels (**Figure 2.7, Figure 2.8**). SaABEmax was unable to efficiently target the six NNHRRT PAM sites (**Figure 2.7**). However, the editing efficiency of SaABEmax was slightly higher than that of SaKKH-ABEmax on NNGRRT sites (22±2.3% vs. 15±2.8%). The moderate editing efficiencies of

SaABEmax and SaKKH-ABEmax contrast with the high activities of SaCas9-derived CBEs[52], which generally edit more efficiently than the corresponding SpCas9 CBE. These results suggest further engineering or evolution may benefit targeting ABE with SaCas9 derivatives.
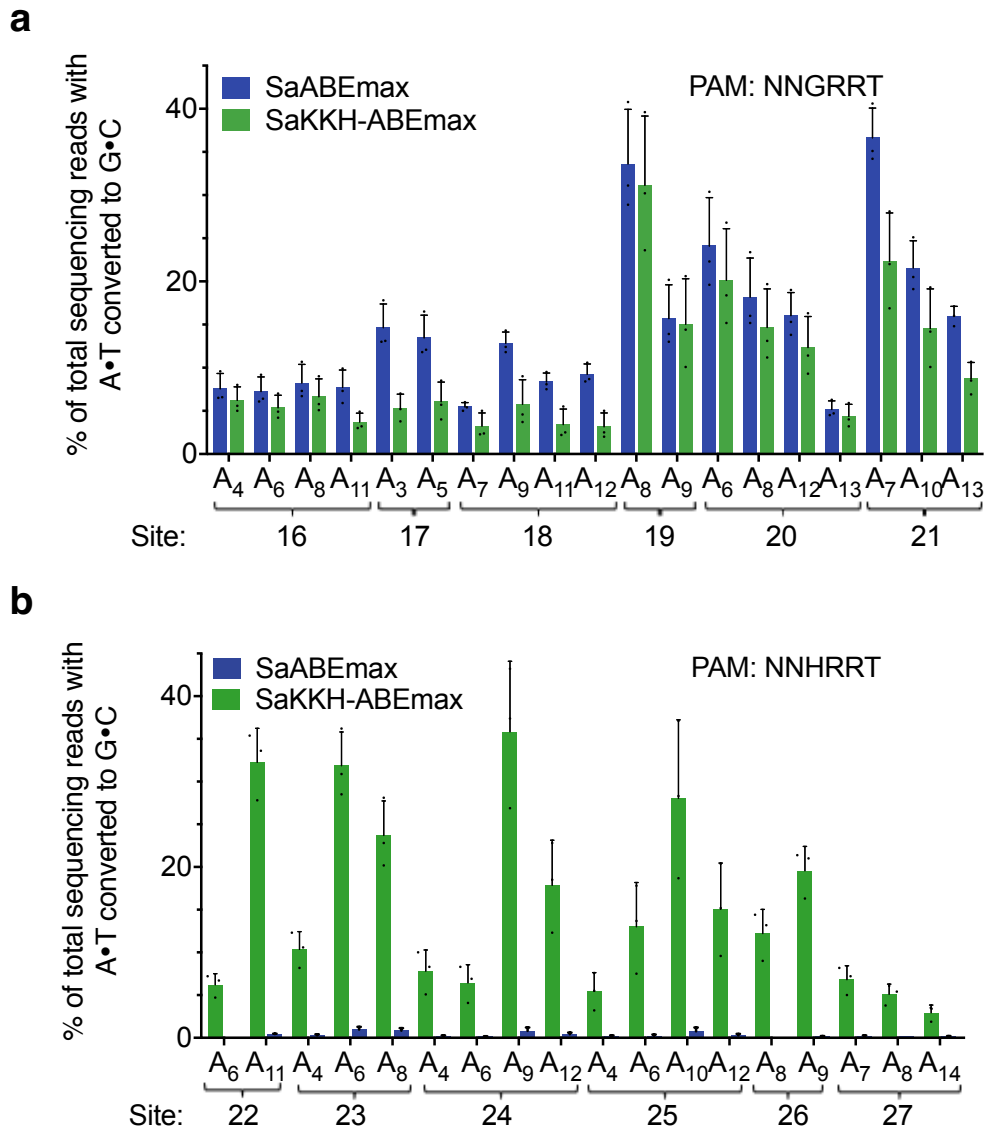
Figure 2.7. Activity of SaABEmax and SaKKH-ABEmax at NNGRRT and NNHRRT PAM sites.
(a) Base editing in HEK293T cells by SaABEmax and SaKKH-ABEmax at six genomic sites containing an NNGRRT PAM. (**b**) Base editing in HEK293T cells by SaABEmax and SaKKH-ABEmax at six genomic sites containing an NNHRRT PAM. Values and error bars reflect the mean±s.d. of three independent biological replicates performed by different researchers on different days.
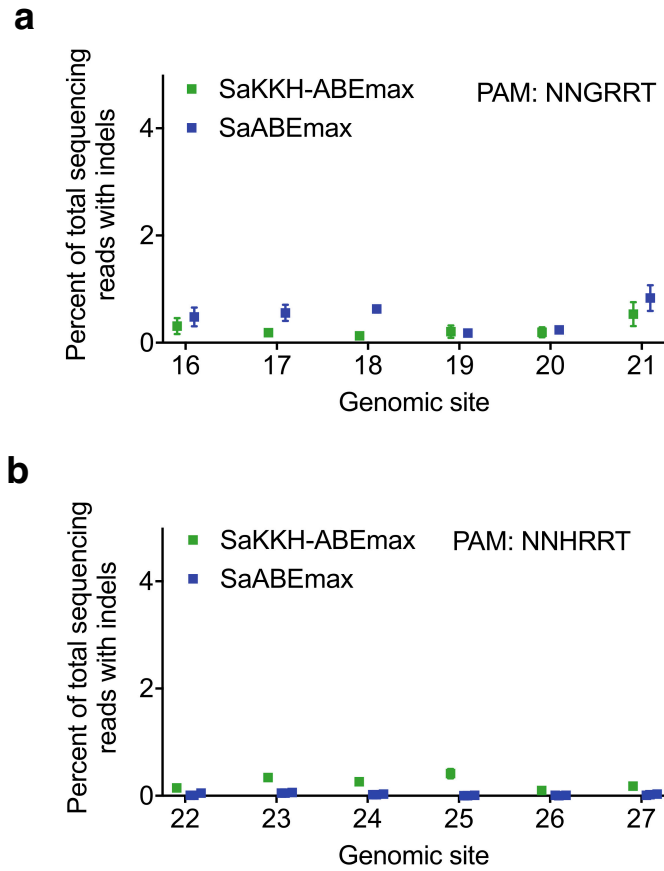
Figure 2.8. Indel frequencies for SaABEmax and SaKKH-ABEmax at NNGRRT and NNHRRT PAM sites.
(a) Percent of all sequencing reads containing an indel following modification by SaABEmax or SaKKH-ABEmax at six genomic sites containing an NNGRRT PAM. **(b)** Percent of all sequencing reads containing an indel following modification by SaABEmax or SaKKH-ABEmax at six genomic sites containing an NNHRRT PAM (where H = A, C, or T). Values and error bars reflect the mean±s.d. of three independent biological replicates performed by different researchers on different days.

Consistent with our previous observations of SaCas9-derived CBEs[52], SaABEmax and SaKKH-ABEmax exhibited an expanded base editing activity window from protospacer positions 4-14 (numbering the PAM as positions 21-26). Maximum editing typically occurred around positions 7-11, with the most frequent outcome being a single A•T-to-G•C edit within this window (**Figure 2.9**). Although an expanded window increases the likelihood of bystander base editing (the editing of non-target adenines within the activity window), a larger window is useful when target adenines would otherwise be inaccessible due to the lack of a PAM, when bystander editing is not consequential, when undesired genotypes can be removed by screening, or when broad mutagenesis is desired.
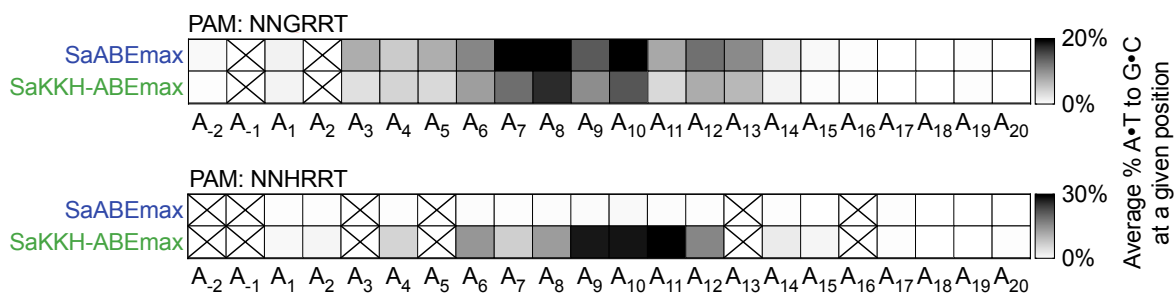
Figure 2.9. Average editing window of SaABEmax and SaKKH-ABEmax.
Heat maps showing average editing efficiency by SaCas9-derived ABE variants at each protospacer position across sites containing each PAM listed (n = 6). Subscripted numbers indicate protospacer positions, counting the first base of the PAM as position 21. Values and error bars reflect the pooled mean of three independent biological replicates at sites containing an adenine at the target protospacer position.

2.2.2 Circularly-permuted SpCas9 base editor variants enable alternate base editing properties

Given the potential utility of base editors with shifted or expanded activity windows[90], we next sought to engineer new base editor architectures that enable editing at different protospacer positions. The activity window of base editors in mammalian cells has proven surprisingly difficult to broaden, with multimeric deaminase assembly for CBEs[48] and extended guide RNAs for ABEs[91] representing the only window-broadening strategies reported to date.

Oakes and coworkers recently generated circularly permuted SpCas9 variants that retain both binding and DNA cleavage activity[92]. For several active SpCas9 circular permutants, the new termini are predicted to lie closer to the ssDNA loop that is the substrate for base editing than the original SpCas9 termini (**Figure 2.10**)[93]. We hypothesized that these circular permutants might provide the deaminase domains in CBEs and ABEs greater access to the ssDNA loop, resulting in expanded or otherwise altered activity windows.
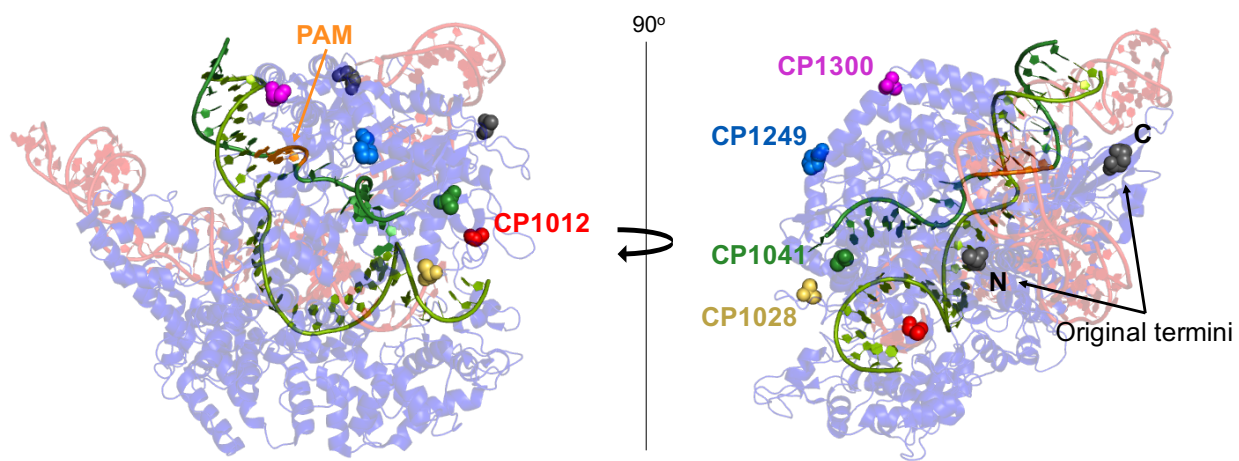
Figure 2.10. Circularly-permuted SpCas9 variants likely have shifted termini positioning. Two views of the SpCas9:sgRNA:DNA crystal structure (PDB: 5F9R[36]) showing the location of the N and C termini in the wild-type protein (black) and in circularly permuted variants tested in this work (red, yellow, green, blue, and magenta). The DNA strand paired with the guide RNA is shown in light green and the other strand, targeted for base editing and partially disordered in the structure, is shown in dark green.

We chose five SpCas9 circular permutants (CP1012, CP1028, CP1041, CP1249, and CP1300, in which the number identifies the amino acid that serves as the new N-terminus) based on both retention of DNA binding activity and predicted proximity to the ssDNA loop[92]. We generated five CP-CBEmax and five CP-ABEmax variants by fusing the circularly permuted Cas9 nickase variants in bis-bpNLS and codon-optimized forms (**Figure 2.10, Figure 2.11**).
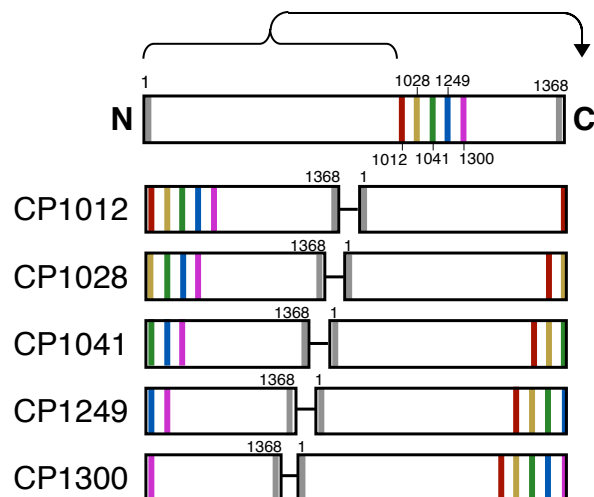
Figure 2.11. Architecture of CP-Cas9 variants used to construct ABE variants.

The resulting CP-CBEmax and CP-ABEmax variants were transfected into HEK293T cells and tested for base editing activity at five endogenous genomic sites containing adenines and cytosines throughout the target 20-nt protospacer (**Figure 2.12, Figure 2.13**). We chose genomic loci for which target bases were mostly located outside of the canonical editing window of positions 4-8 to more precisely define the editing window of these new base editors (**Figure 2.14**). Four of the five CP-CBE variants were capable of base editing at all five sites without substantial indel formation (generally < 2%) (**Figure 2.12, Figure 2.15**), while CP1300-CBEmax demonstrated highly site-dependent base editing activity. Three of the remaining four CP-CBEmax variants exhibited efficient editing activity, averaging 47±3.6%, 46±4.9%, 18±5.6% and 42±4.6% C•G-to-T•A conversion for CP1012-CBEmax, CP1028-CBEmax, CP1041-CBEmax, and CP1249-CBEmax, respectively, compared to 66±5.9% C•G-to-T•A conversion for CBEmax at the same genomic sites. Two of the variants, CP1012-CBEmax and CP1028-CBEmax, showed broadening of the editing window from the canonical positions 4-8 to positions 4-11 of the protospacer, averaging 12±1.6% and 15±3.9% C•G-to-T•A conversion at positions 9-11, respectively, compared to 5.8±2.2% for CBEmax at these positions (**Figure 2.14**). These results

together establish that circularly permuted CBEmax variants indeed exhibit broadened editing

windows.
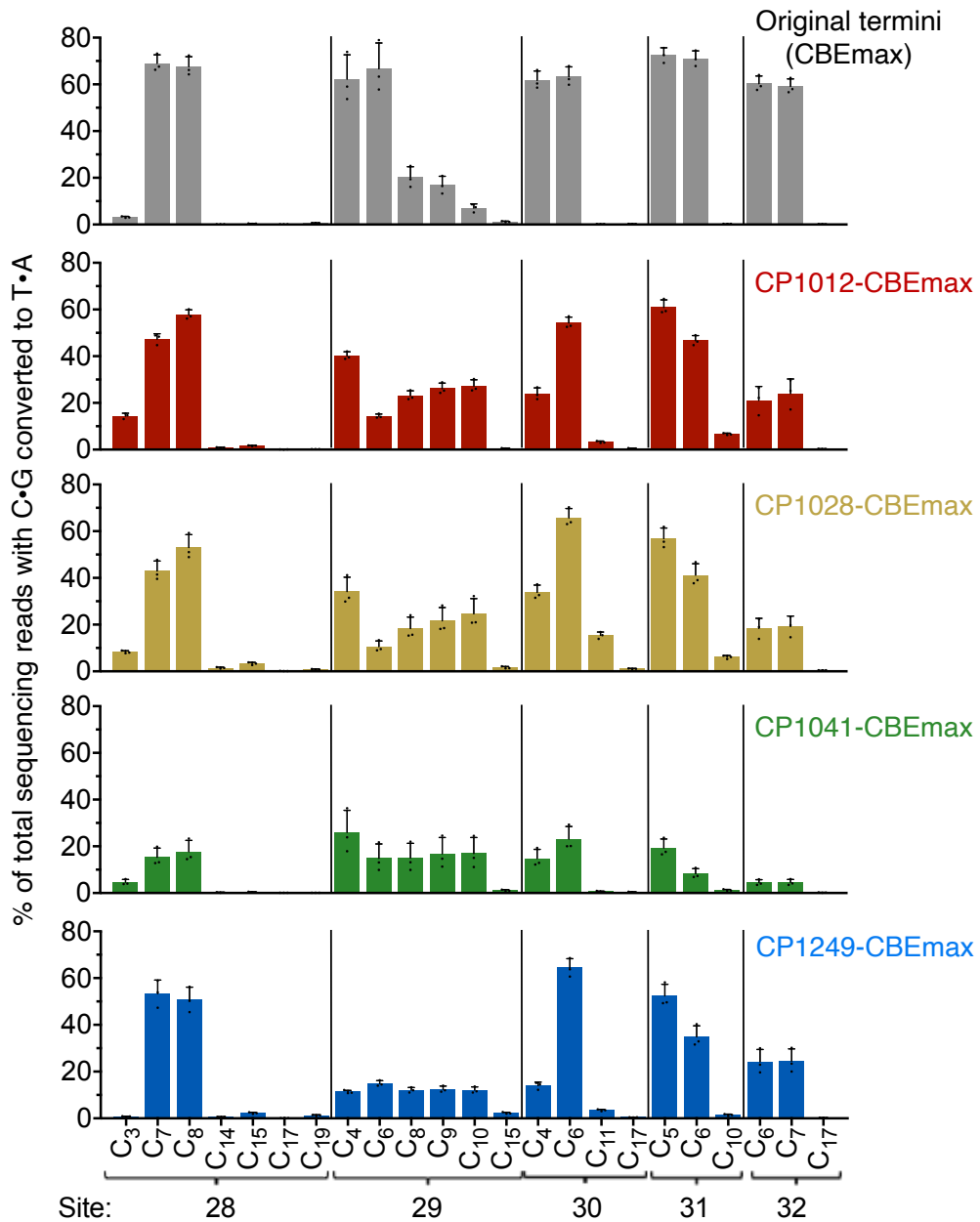
Figure 2.12. Cytosine base editing activity of CP-CBEmax variants in mammalian cells.
For clarity, protospacer positions with editing lower than 0.5% across all editors are not shown. Subscripted numbers indicate protospacer positions, counting the first base of the PAM as position 21. Values and error bars reflect the mean±s.d. of three biological replicates performed on different days at each site.
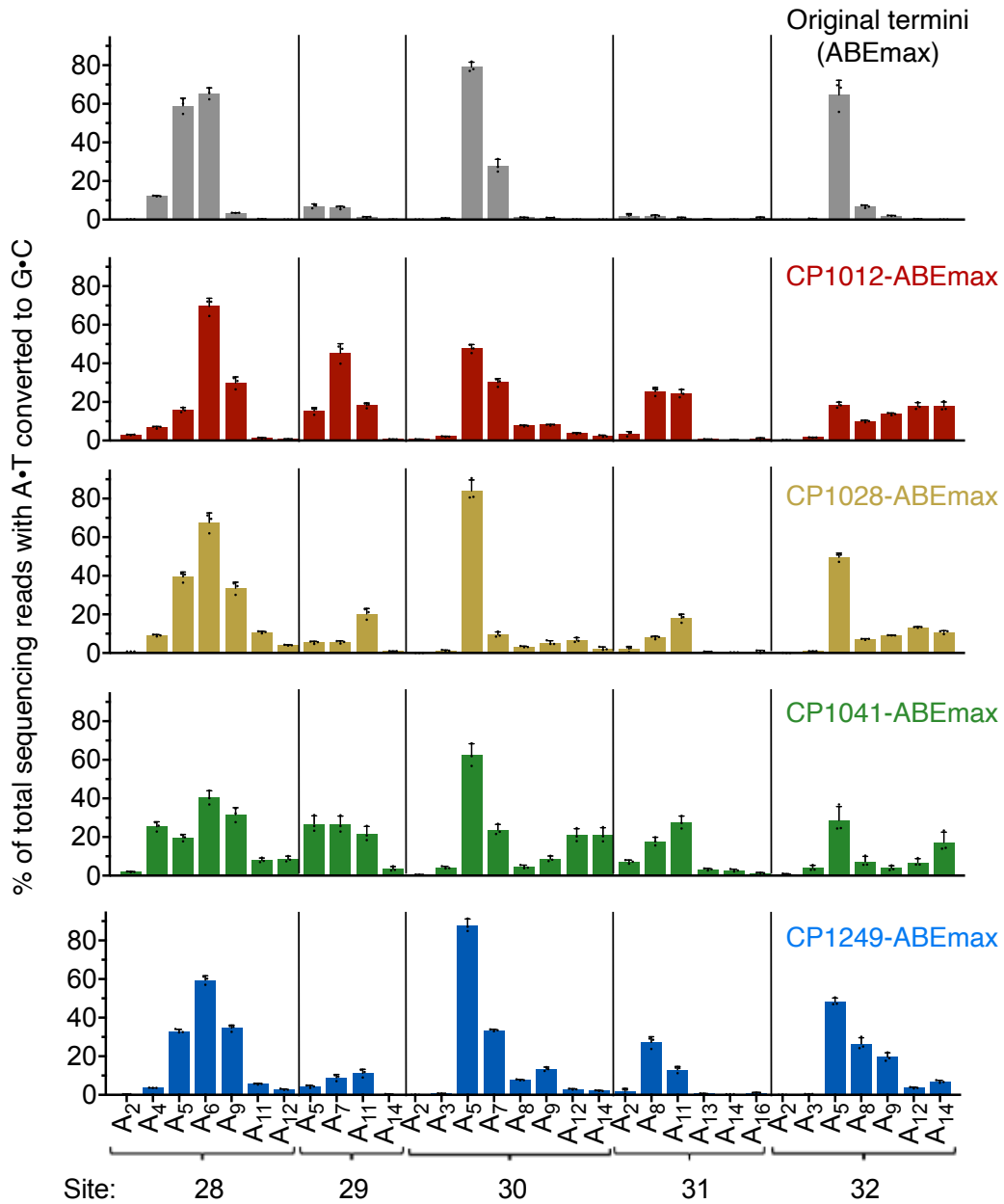
Figure 2.13. Adenine base editing activity of CP-ABEmax variants in mammalian cells. For clarity, protospacer positions with editing lower than 0.5% across all editors are not shown. Subscripted numbers indicate protospacer positions, counting the first base of the PAM as position 21. Values and error bars reflect the mean±s.d. of three biological replicates performed on different days at each site.
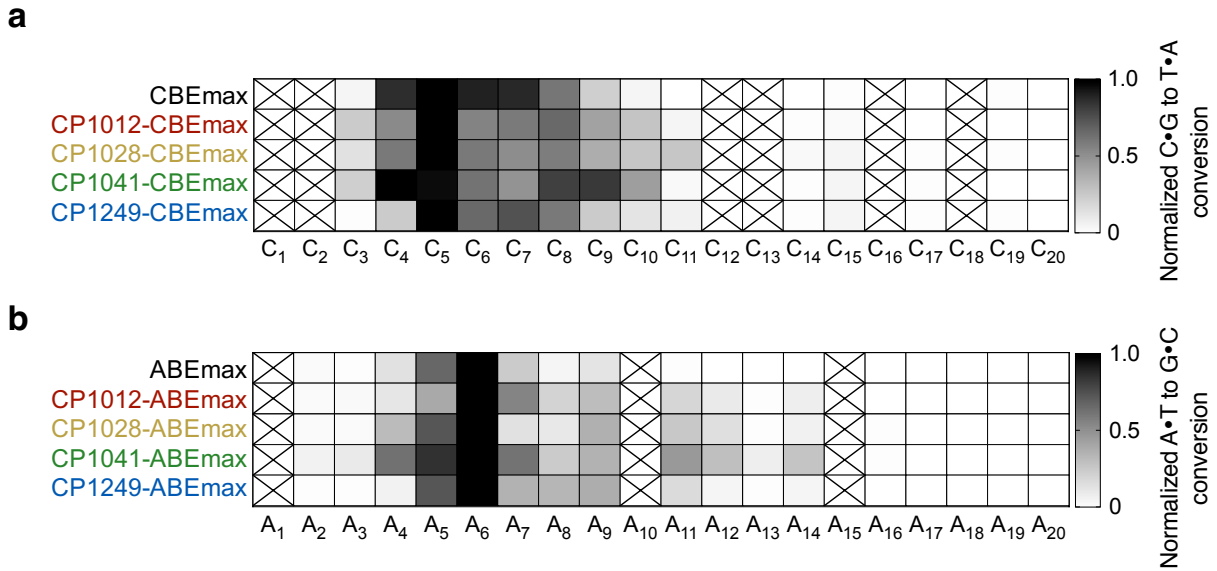
Figure 2.14. Average editing window of CP-CBEmax and CP-ABEmax variants.
Heat maps showing average base editing efficiency at each position within the protospacer across five sites tested with (**e**) CP-CBEmax variants or (**f**) CP-ABEmax variants, normalized to the maximum observed editing within the protospacer (1.0). Boxes crossed out indicate positions for which no target base was present among all genomic sites tested.
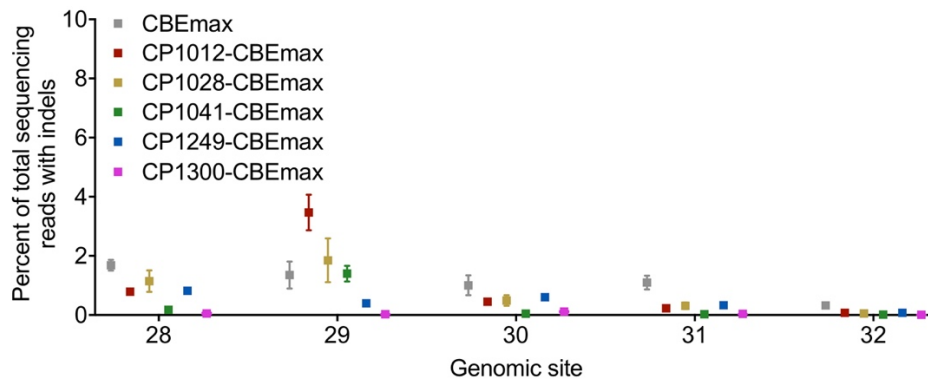
Figure 2.15. Indel frequencies for CP-CBEmax variants.
Percent of all sequencing reads containing an indel following modification by a CP-CBEmax variant at five genomic sites in HEK293T cells. Values and error bars reflect the mean±s.d. of three independent biological replicates performed by different researchers on different days.

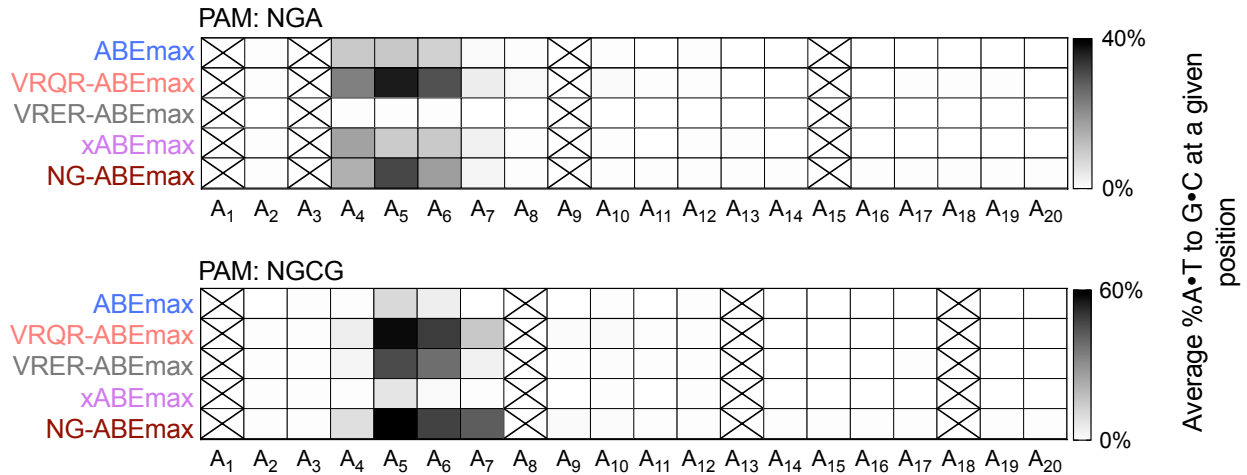Surprisingly, at three of the five genomic sites tested with CP-CBEmax variants, CP1012-CBEmax, CP1028-CBEmax, and CP1041-CBEmax also edited bases upstream of the protospacer on both the target strand (the strand normally targeted for nucleobase deamination)

25

and the non-target strand. This out-of-protospacer editing was particularly evident for CP1012-CBEmax, with editing observed as far upstream as the -13 position of the target strand (**Figure 2.16**). These upstream editing events may arise from extended R-loop formation by some circular permutants at sites prone to unwinding, creating larger accessible ssDNA regions.
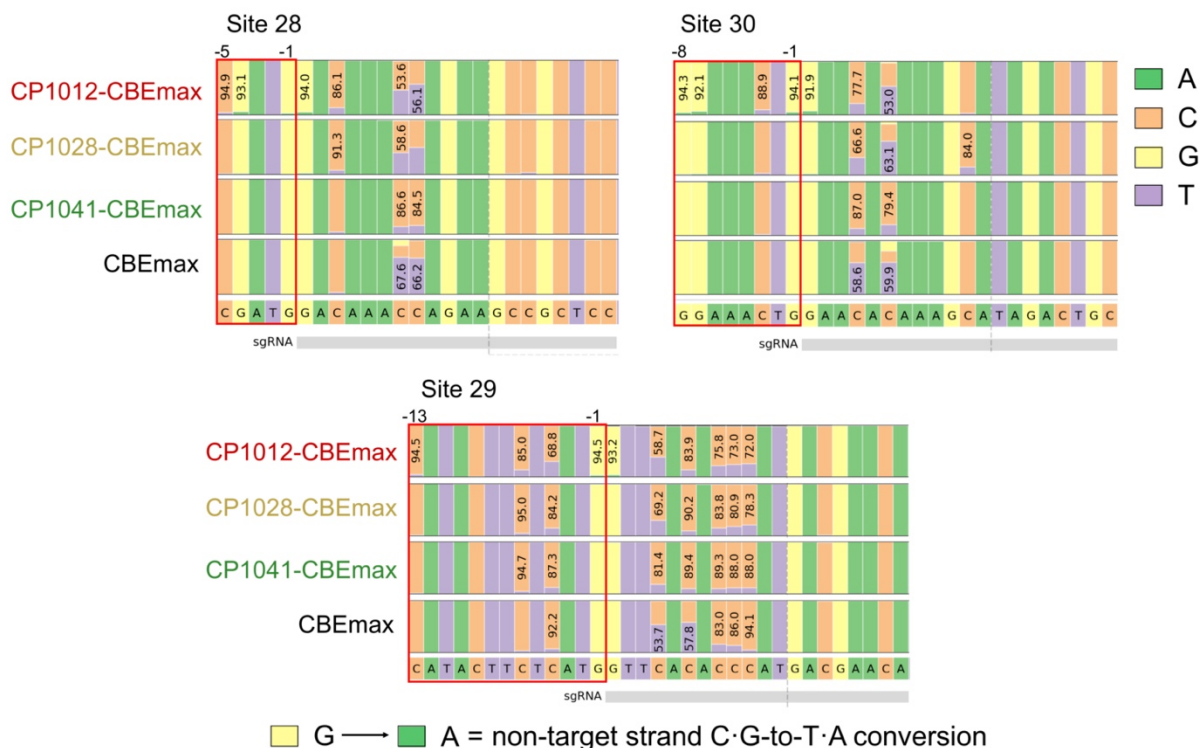
Figure 2.16. Out-of-protospacer editing by CP-CBEmax variants.
Three of the five genomic sites treated with CP-CBEmax variants exhibited both nontarget strand editing and out-of-protospacer editing. Representative samples of three sites are shown, with the protospacer designated by the grey box, and out-of-protospacer C·G-to-T·A conversion highlighted in the red box. CP1012-CBEmax exhibited the most frequent out-of-protospacer editing, with CP1028-CBEmax and CP1041-CBEmax exhibiting this property on only one of the sites.

As we previously reported[32,70], CBEs can generate both desired C-to-T edits and unanticipated C-to-G and C-to-A mutations resulting from error-prone base excision repair of the uracil intermediate. Among the five genomic sites tested, three sites when treated with CBEmax resulted in < 1% non-C-to-T byproducts, but two sites unusually prone to unanticipated editing byproducts showed an average of 19±3.3% non-C-to-T byproducts among CBEmax-edited

products. Surprisingly, CP1012-CBEmax, CP1028-CBEmax, CP1041-CBEmax, and CP1249-

CBEmax demonstrated greatly reduced (2.1- to 19-fold lower than CBEmax) byproduct

formation at these two problematic sites (**Figure 2.17**). The improved product purity of CP-

CBEmax variants might result from the newly positioned termini of Cas9 allowing the C-terminal

UGI (or a UGI•UNG complex) to better impede uracil excision by UNG. Consistent with this

model, the minimum linear distance between the predicted location of the C-termini and the

ssDNA target in two different SpCas9 crystal structures[93,94] is inversely related to observed

product purity (decreasing distance and increasing product purities: CBEmax, CP1249-CBEmax

~ CP1012-CBEmax, CP1028-CBEmax, CP1041-CBEmax) (**Figure 2.18**).
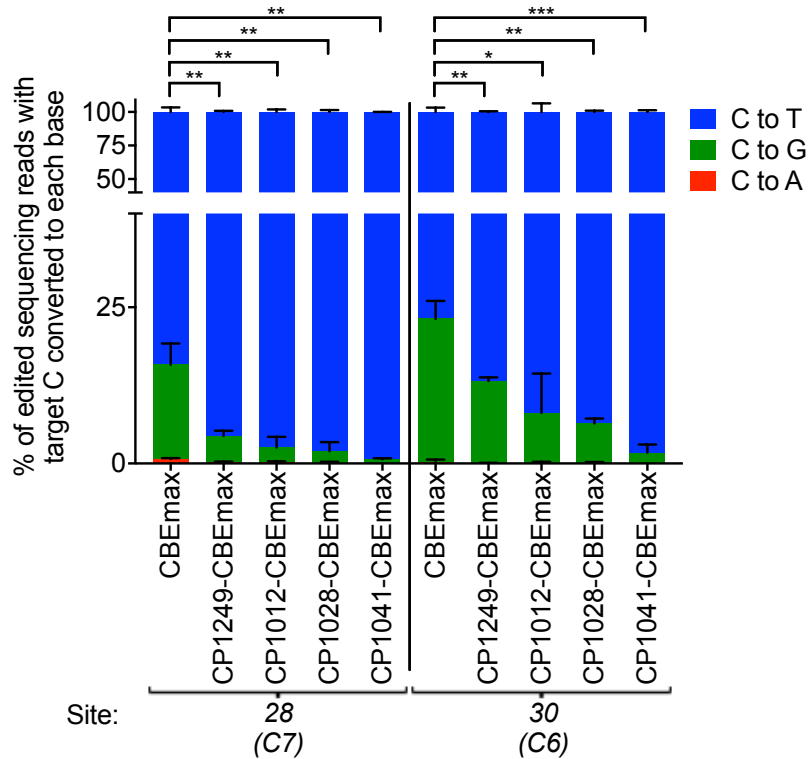
Figure 2.17. Editing product purity of CP-CBEmax variants.
The product distribution among edited DNA sequencing reads (reads in which the target C is base edited) is shown for each CP-CBEmax variant tested at two different genomic sites that are especially prone to non-C-to-T byproduct formation. Values and error bars reflect the mean±s.d. of three biological replicates performed on different days at each site. ns, $P>0.05$; *$P<0.05$; **$P<0.01$; ***$P<0.001$, by two tailed Student's t-test.
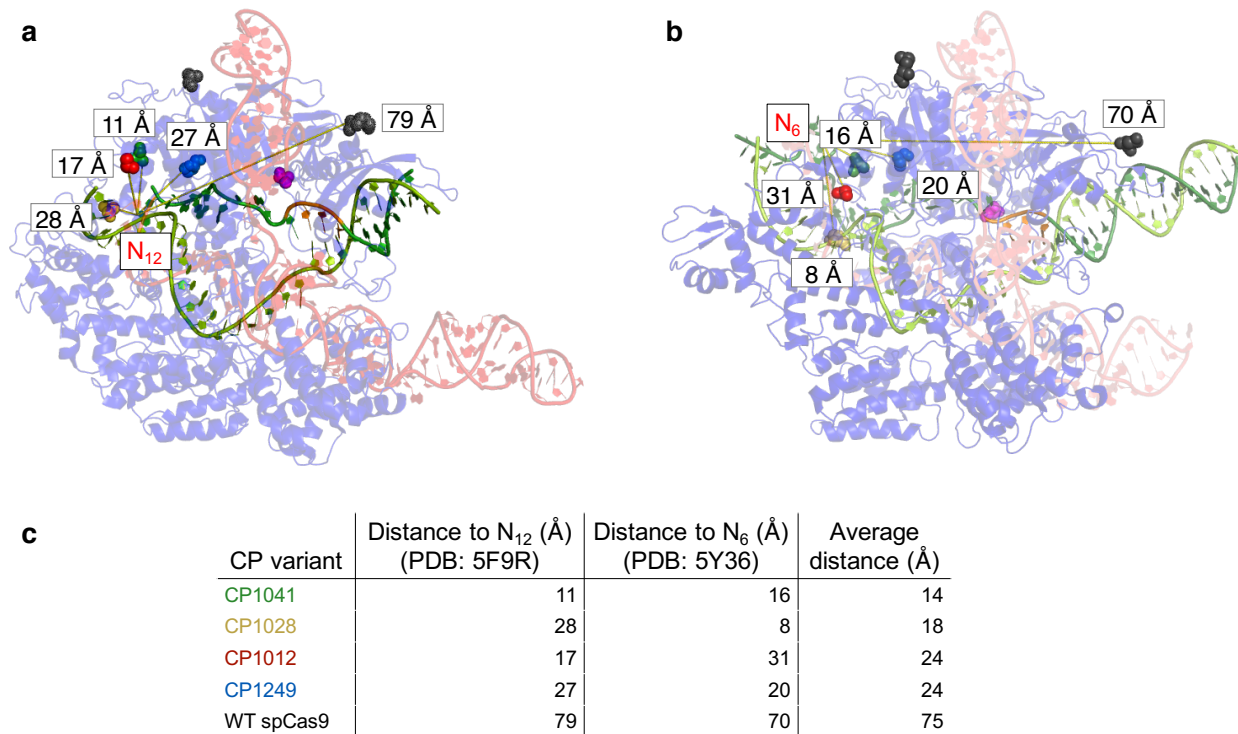
Figure 2.18. Predicted average minimum distance between original or circularly-permuted C-termini and the ssDNA substrate for base editing.
(a) Crystal structure of the SpCas9:gRNA:DNA ternary complex with the ssDNA bubble partially resolved (PDB: 5F9R)[36]. Novel CP termini are represented as spheres (original N- and C-termini in dark grey). The minimal linear distance between the predicted position of the novel CP termini (and WT C-terminus) and the furthest resolved nucleobase in the ssDNA bubble (corresponding to protospacer position 12, counting the PAM as positions 21-23) is depicted. **(b)** Cryo-EM structure of the SpCas9:gRNA:DNA ternary complex with the ssDNA bubble fully resolved (PDB: 5Y36)[38]. Novel CP termini are represented as spheres (original N- and C-termini in dark grey). The minimal linear distance between the predicted position of the novel CP termini (and WT C-terminus) and a protospacer position typically targeted for base editing (corresponding to protospacer position 6) is depicted. **(c)** Average of the distances to two different target positions on the ssDNA substrate measured for the novel CP termini (or WT C-terminus) from (a) and (b), listed in ascending order.

To probe the relationship between product purity and UGI positioning, we generated CP-CBEmax base editors without UGI, denoted CP-CBEmax-B variants. At one of the genomic sites prone to product mixtures, CP-CBEmax-B variants no longer showed a correlation between linear distance and product purity (**Figure 2.19**). This correlation was still evident, however, at the other site, suggesting that factors other than UGI positioning can also govern

product purity[12] on a site-dependent basis. Together, these results reveal that circularly

permuted CBEmax variants generate undesired byproducts less frequently than CBEmax,

possibly by enhancing access of the UGI domain to the edited site.
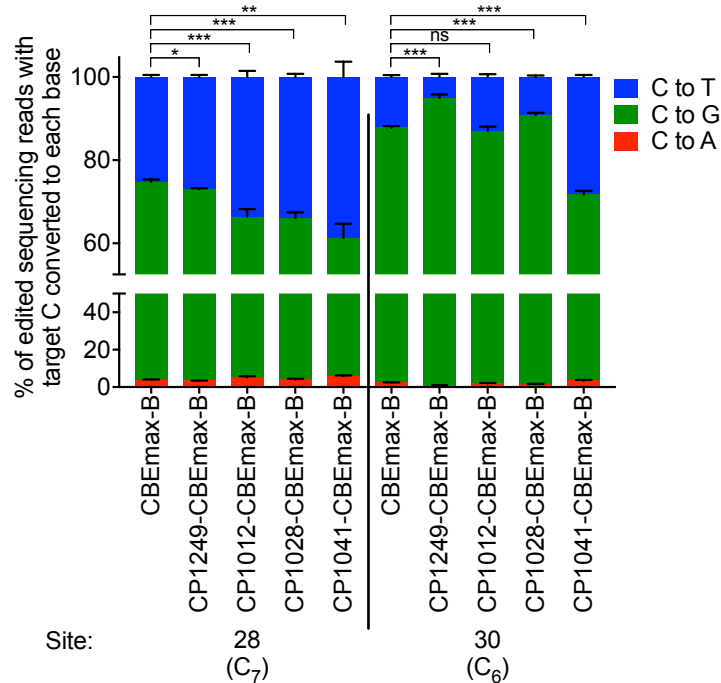
Figure 2.19. Editing product purity of CP-CBEmax variants without UGI.
The product distribution among edited DNA sequencing reads (reads in which the target C is
base edited) is shown for each CBEmax variant with no UGI ("CBEmax-B" variants) tested at
the same two sites as in **Figure 2.17**. Subscripted numbers indicate protospacer positions,
counting the first base of the PAM as position 21. Values and error bars reflect the mean±s.d. of
three biological replicates performed on different days at each site. ns, $P>0.05$; *$P<0.05$;
**$P<0.01$; ***$P<0.001$, by two tailed Student's t-test.

Most CP-ABEmax variants similarly exhibited a broadening of the editing window

(**Figure 2.14**). CP-ABEmax variants retained efficient editing activity similar to that of ABEmax,

averaging 45±3.5%, 52±3.9%, 40±5.4%, and 51±2.2% A•T-to-G•C conversion for CP1012-

ABEmax, CP1028-ABEmax, CP1041-ABEmax, and CP1249-ABEmax, respectively compared

to an average of 54±4.3% A•T-to-G•C conversion for ABEmax at the same four genomic sites.

Both ABEmax and the circularly permuted variants generated minimal indels (generally < 2%)

(**Figure 2.20**). While no out-of-protospacer editing was observed among circularly permutated

ABEmax variants, the window-broadening effect was pronounced, generally resulting in an

expansion from the canonical window of protospacer positions 4-7 for ABEmax to a window

spanning positions 4-12 (**Figure 2.14**). Base editing efficiencies at protospacer positions 8-12

averaged 20±1.9%, 18±2.2%, 24±4.0%, and 23±2.3% A•T-to-G•C conversion for CP1012-

ABEmax, CP1028-ABEmax, CP1041-ABEmax, and CP1249-ABEmax, respectively, a 6.4- to

8.6-fold increase over that of ABEmax at these distal protospacer positions. Intriguingly,

CP1012-ABEmax also exhibited a shifted (rather than broadened) base editing window, with

maximal editing taking place at A6 or A7, compared to A5 or A6 for ABEmax. CP1041-ABEmax

showed the broadest editing window, with up to 11±3.3% average editing at position 14 of the

protospacer. CP-ABEmax variants thus are able to edit target As that lie outside the canonical

ABE editing window.
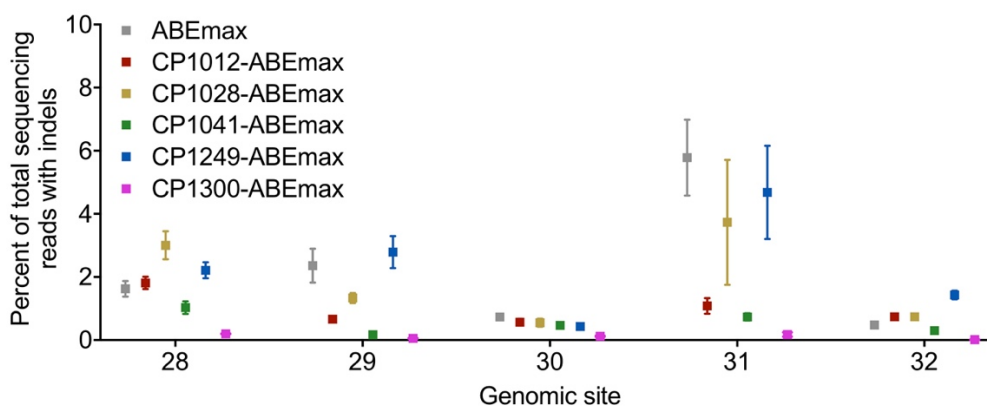
Figure 2.20. Indel frequencies for CP-ABEmax variants.
Percent of all sequencing reads containing an indel following modification by a CP-ABEmax
variant at five genomic sites in HEK293T cells. Values and error bars reflect the mean±s.d. of
three independent biological replicates performed by different researchers on different days.

To assess possible effects of circular permutation on off-target base editing, we

measured off-target editing of all ten CP-CBEmax and CP-ABEmax variants at nine genomic

off-target sites previously identified by GUIDE-Seq as the most highly edited off-target

substrates of SpCas9 nuclease for three target loci[95]. Off-target base editing efficiency of

circularly permuted base editors was similar to or less than that of CBEmax or ABEmax for C or

A nucleotides within the canonical editing window. As expected, for C or A nucleotides outside

the canonical editing window, the expanded editing windows of circularly permuted base editors

in some cases allowed higher off-target editing than CBEmax or ABEmax. Nevertheless, the

alternate targeting capabilities of CP-CBEmax and CP-ABEmax variants should enable access

to novel target sites and editing outcomes.

2.3 Conclusions and outlook

Together, the results described in this chapter demonstrate that base editing is a highly modular technology compatible with a variety of Cas orthologs, natural or engineered. By using Cas variants with alternate PAM compatibilities, we developed six new ABEmax variants, VRQR-ABEmax, VRER-ABEmax, xABEmax, NG-ABEmax, SaABEmax, and SaKKH-ABEmax, capable of accessing new target sites with robust adenine base editing efficiencies. We further demonstrated that circularly permuting the Cas9 nickase domain of SpCas9 base editors results in CBEmax and ABEmax variants with broadened or shifted editing windows. These altered targeting properties enable efforts to perform base editing at currently inaccessible target nucleotides and can also substantially improve product purity. Indeed, an analysis of human pathogenic SNPs in ClinVar[17,18] reflects a substantial improvement in the fraction of targetable SNPs when considering the expanded CP-CBEmax or CP-ABEmax editing windows (51% of SNPs correctable by A•T-to-G•C conversion or 51% of SNPs correctable by C•G-to-T•A conversion, respectively) compared to their unpermuted CBEmax and ABEmax counterparts (27% and 31%, respectively) (**Figure 2.21**). Alternative-PAM ABEmax variants and circularly permuted CBEmax and ABEmax variants thus expand the capabilities of base editors, now widely used in the biomedical research community[69].

Pathogenic SNPs corrected by C•G-to-T•A conversion
(4,052 total SNPs analyzed)

27%
73%

49% 51%

Targetable
Not targetable
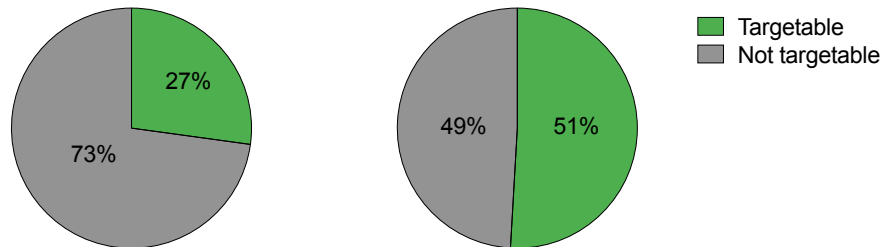
Pathogenic SNPs corrected by A•T-to-G•C conversion
(12,120 total SNPs analyzed)

31%
69%

49% 51%

Targetable
Not targetable
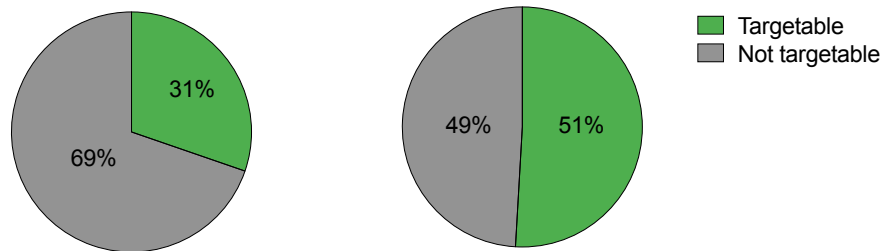
Figure 2.21 ClinVar analysis of targetable human pathogenic SNPs with expanded editing window CP-CBEmax and CP-ABEmax variants.
Fraction of pathogenic T•A-to-C•G to SNPs in ClinVar[17,18] that could, in principle, be corrected by (**a**) CBEmax with an editing window of positions 4-8 (left) versus the SNPs correctable by CP-CBEmax with an editing window of positions 4-14 (right). Fraction of G•C-to-A•T pathogenic SNPs in ClinVar that could, in principle, be corrected by (**b**) ABEmax with an editing window of positions 4-8 (left) versus the SNPs correctable by CP-ABEmax with an editing window of positions 4-14 (right).

2.4 Methods

2.4.1 General methods

PCR was performed using Phusion U Green Multiplex PCR Master Mix (ThermoFisher Scientific). All plasmids were assembled by the USER cloning method as previously described[31]. Guide RNA plasmids for SpCas9, SaCas9, and all engineered variants were assembled as previously described[52]. Plasmids for mammalian cell transfections were prepared using the ZymoPURE Plasmid Midiprep kit (Zymo Research Corporation).

2.4.2 Cell culture

HEK293T cells (ATCC CRL-3216) were cultured in Dulbecco's Modified Eagle's Medium (DMEM, Corning) supplemented with 10% fetal bovine serum (FBS, ThermoFisher Scientific) and maintained at 37 °C with 5% CO_2.

2.4.3 Transfections

HEK293T cells were seeded on 48-well Poly-D-Lysine plates (Corning) in the same culture medium. Cells were transfected 12-16 hours after plating with 1.5 μL Lipofectamine 2000 (ThermoFisher Scientific) using 750 ng of base editor plasmid, 250 ng of guide RNA plasmid, and 10 ng of green fluorescent protein as a transfection control. Cells were cultured for 3 d with media exchanged following the first day, then washed with 1x phosphate-buffered saline (PBS, ThermoFisher Scientific), followed by genomic DNA extraction by addition of 100 μL of freshly prepared lysis buffer (10 mM Tris-HCl, pH 7.5, 0.05 % SDS, 25 μg/mL proteinase K (ThermoFisher Scientific)) directly into each transfected well. The mixture was incubated at 37 °C for 1 h then heat inactivated at 80 °C for 30 min. Genomic DNA lysate was subsequently used immediately for HTS sequencing.

2.4.4 HTS of genomic DNA samples

HTS of genomic DNA from HEK293T cells was performed as previously described[31]. Following Illumina barcoding, PCR products were pooled and purified by electrophoresis with a 2% agarose gel using a Monarch DNA Gel Extraction Kit (NEB), eluting with 30 μL H_2O. DNA concentration was quantified with Qubit dsDNA High Sensitivity Assay Kit (ThermoFisher Scientific) and sequenced on an Illumina MiSeq instrument (paired-end read – R1: 250-280 cycles, R2: 0 cycles) according to the manufacturer's protocols.

2.4.5 HTS data analysis

Sequencing reads were demultiplexed using the MiSeq Reporter (Illumina) and fastq files were analyzed using Crispresso2. Base editing values are representative of n=3 independent biological replicates collected over different days by different researchers with the mean$\pm$s.d. shown. Base editing values are reported as a percentage of the number of reads with cytosine or adenine mutagenesis over the total aligned reads.

2.4.6 Statistics and reproducibility

All statistical analyses were performed on n = 3 biologically independent experiments using the unpaired two tailed Student's *t*-test. Biologically independent experiments reported here were performed by different researchers using independent splits of the mammalian cell type used.

2.4.7 Code availability

Access to and usage for Crispresso2 script can be found here[96]. ClinVar analysis of pathogenic human SNPs targetable by the base editors described in this study was executed using a custom Matlab script described previously[1,2].

2.4.8   Data availability

Plasmids encoding modified PAM adenine base editors and circularly permuted cytidine and adenine base editors have been deposited to Addgene. High-throughput sequencing data are deposited in the NCBI Sequence Read Archive (PRJNA498804).

Chapter 3. High-throughput directed evolution of Cas9 variants using a sequence-agnostic functional selection

This chapter was adapted from:

Huang, T. P., Heins, Z.J., Miller, S.M., Wong, B.G., Balivada, P.A., Wang, T., Khalil, A.S., & Liu, D. R. High-throughput continuous evolution of compact Cas9 variants targeting single-nucleotide-pyrimidine PAMs. *Under review* (2022).

Contributions:

This chapter was written by T.P.H., Z.J.H, A.S.K., and D.R.L. with input from all coauthors of the referenced work. The work described in this chapter reflect experiments designed and performed by T.P.H. and Z.J.H., with help from all coauthors. Specifically, T.P.H. conceptualized and validated the SAC-PACE selection and the PAM profiling assay, developed and cloned plasmids and phage, designed evolution schemes and mammalian cell experiments, conducted characterization of evolved variants in bacteria and mammalian cells, and analyzed data. Z.J.H. developed and characterized the ePACE platform, conceptualized and characterized millifluidic devices and pressure regulation unit, wrote ePACE software, carried out ePACE evolution experiments, performed mutational analysis of evolved variants, and analyzed data. S.M.M. performed and assisted with analysis of GUIDE-Seq. B.G.W. aided with conceptualization and design of millifluidic devices and assisted with validation of the ePACE platform. T.W. provided materials and assisted with analysis of the MBP evolution during ePACE validation. P.A.B. aided in fabrication and characterization of millifluidic devices.

I thank Prof. Mo Khalil and Zack Heins for being enthusiastic and helpful collaborators. It has been a pleasure working with you on the project described in this chapter.

3.1 Introduction

Target sites engaged by Cas9 must contain a protospacer adjacent motif (PAM) that is recognized through a protein:DNA interaction prior to single guide RNA (sgRNA) binding[12]. While not prohibitive for some gene editing applications, such as target gene disruption, this PAM requirement limits the applicability of precision gene editing methods, including base editing, prime editing, or site-specific DNA integration[30,66]. For these technologies, the target modification must occur either at a specific distance or within a certain range of the PAM[66]. Thus, the availability of a PAM sequence compatible with a Cas protein that retains robust activity in mammalian cells strongly determines the application scope of precision gene editing. Indeed, recent *ex vivo* and *in vivo* therapeutic base editing to rescue sickle-cell disease[43] and progeria[45] in mice used evolved or engineered Cas9 variants to precisely position the base editor at CACC or NGA PAMs, respectively.

The limitations imposed by PAM restrictions have motivated efforts to engineer or evolve Cas protein variants with broadened or altered PAM compatibility. These approaches have generated variants of the most widely used Cas9 from *Streptococcus pyogenes* (SpCas9)[77,82,84,97,98], which offers robust mammalian cell activity and engages sites with NGG PAMs[12], where N = A, C, G, or T. The wild-type and evolved or engineered variants of SpCas9 described to date can collectively access essentially all purine-containing PAMs and a subset of pyrimidine-containing PAMs[77,84,97,98].

Researchers have also parsed the genomes of other bacterial species or bacteriophage to identify Cas variants with different PAM requirements[66,99]. These Cas variants vary dramatically in size, PAM compatibility, and enzymatic activity[30,66,100]. Unfortunately, most of these natural homologs are less well characterized, less active in mammalian cells, or have highly restrictive PAM requirements compared to SpCas9[100], limiting their utility for precision gene editing applications and the ease with which they can be modified. As such, engineering or

38

evolution of non-SpCas9 orthologs has been uncommon, with only a few reported examples[87,101,102].

Novel engineering or evolution methods to address the limitations of reprogramming non-SpCas9 orthologs could provide new precision gene editing capabilities that expand upon and complement the suite of commonly used SpCas9-derived variants. Nme2Cas9, a Cas9 variant from *Neisseria meningitidis*, is an attractive Cas ortholog for evolving PAM compatibility[103]. The wild-type enzyme is active on N_4CC PAMs, and thus may serve as a promising starting point to all pyrimidine PAMs previously inaccessible by SpCas9 variants. In addition, Nme2Cas9 has a smaller size than SpCas9 (1,082 aa vs 1,368 aa), making it attractive for future delivery applications. Nme2Cas9 has also shown robust activity in mammalian cells as both a nuclease and a base editor[103,104].

In this chapter, we report the directed evolution of Nme2Cas9[103], expanding its PAM scope from the N_4CC requirement of the wild-type protein to include most N_4YN sequences, where Y = C or T. To enable the evolution of this non-SpCas9 ortholog, we developed and integrated three technologies. First, we established a new, generalizable selection strategy requiring both PAM recognition and functional editing activity. We carried out selections in parallel across single PAM sequences using phage-assisted non-continuous evolution (PANCE)[105] and a novel, high-throughput eVOLVER-enabled[106] phage-assisted continuous evolution (ePACE) platform. Lastly, we developed a high-throughput base editing-dependent PAM profiling assay (BE-PPA) to rapidly and thoroughly characterize evolving Nme2Cas9 variants and to guide evolutionary trajectories. With these developments, we evolved four Nme2Cas9 variants that enable robust precision genome editing at PAMs with a single specified pyrimidine nucleotide: eNme2-C, eNme2-C.NR, eNme2-T.1, and eNme2-T.2. The evolved Nme2 variants exhibit comparable (eNme2-T.1 and eNme2-T.2) or more robust (eNme2-C) base editing and lower off-target editing than SpRY, the only other engineered variant capable

of accessing similar PAMs for a subset of target sites[98]. Together, these new variants offer broad PAM accessibility that is complementary to the suite of PAMs previously targetable by SpCas9-derived variants. Moreover, the selection strategy developed in this study is highly scalable and general. Because of the lack of target site requirements, this selection could in principle be applied to evolve functional activities in any Cas ortholog or to optimize editing at a specific PAM or target site.

3.2 Results

3.2.1   Limitations of existing PACE selections for evolving Cas9 orthologs

We hypothesized that our continuous evolution system, PACE[107], in which the propagation of M13 bacteriophage is coupled to the desired activity of a protein of interest (POI), could be used to evolve Nme2Cas9 variants with expanded pyrimidine-rich PAM scope. Previously, we broadened the PAM scope of SpCas9 variants using a one-hybrid, DNA-binding PACE circuit[77,97]. In those efforts, SpCas9 variants encoded on selection phage (SP) capable of simply binding the target PAM(s) successfully produce gene III (gIII), a gene essential for phage propagation. The resulting SpCas9 variants could access most NR PAM sequences (where R = A or G), but efforts to apply the DNA-binding selection to evolve pyrimidine PAM recognition were less successful[77,97].

While this binding selection could be adapted to evolve Nme2Cas9, fundamental differences between the activities of SpCas9 and Nme2Cas9 could impede efforts to evolve the PAM scope of the latter. Nme2Cas9, and more broadly Type II-C Cas variants, may have slower nuclease kinetics relative to SpCas9[100]. This weaker nuclease activity is attributed to slower Cas9 helicase activity, as artificially introduced bulges mimicking partially unwound DNA in the PAM proximal region increase the cleavage rate of Type II-C Cas variants but not of SpCas9[100]. This theory is supported by observations that miniaturized SpCas9 variants with partially deleted domains have reduced DNA binding affinity that can also be rescued by the introduction of PAM-proximal bulges in target DNA[108]. Because a primary motivation for broadening PAM compatibility is to improve the applicability of precision gene editing technologies that require DNA unwinding[66], it is critical that a selection preserves or improves R-loop formation, maintenance, and nuclease activation. Notably, these Cas properties are dependent on domains outside of the PAM-interacting domain (PID), which has been the focus of rational engineering approaches[82,87,98,101]. Together, this analysis suggests that while DNA-

binding selections or PID engineering can yield robust SpCas9 variants with altered PAM

compatibilities, the same type of binding-only selection applied to the evolution of Nme2Cas9 or

similar Cas orthologs may not yield both desired PAM recognition and efficient downstream

activity (**Figure 3.1**). This hypothesis motivated us to envision a new, functional selection in

PACE for evolving PAM compatibility.
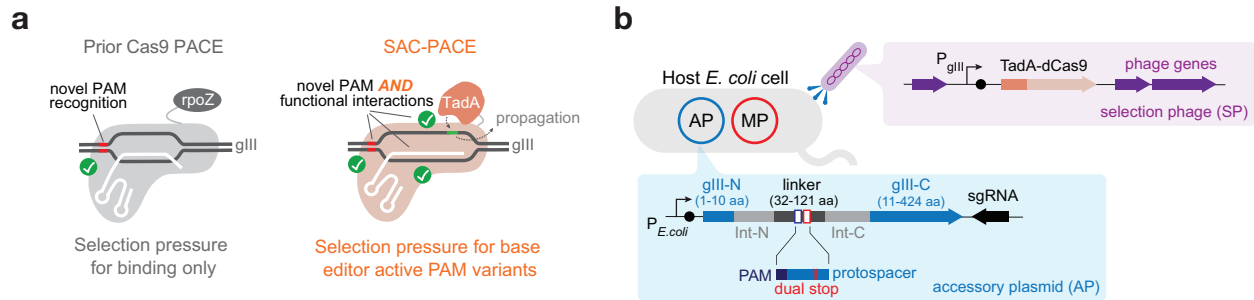
Figure 3.1. Development of a general functional selection for Cas PAM evolution.
(a) Overview of prior Cas9 PACE (left) requiring only PAM binding upstream of a promoter
controlling expression of gIII, compared to the sequence-agnostic Cas PACE selection (SAC-
PACE) developed in this study, which requires both PAM binding and subsequent base editing.
(b) The selection circuit in SAC-PACE. The selection phage (SP) encodes an adenine base
editor in place of gIII. In the host cells, an accessory plasmid (AP) contains a *cis* intein-split gIII,
with a linker (31–121 aa) containing stop codons. Correction of the stop codons through
recognition of a novel PAM and subsequent base editing results in excision of the *cis*-intein,
production of functional gIII, and phage propagation.

3.2.2 Development of a general functional selection for evolving PAM compatibility in PACE

To develop a functional selection for Cas9-based genome editing agents with altered

PAM compatibilities, we combined elements of a DNA-binding selection[77,97] with a base editing

(BE) selection[109,110], such that both novel PAM recognition and subsequent BE within the

protospacer are required to pass the selection. Although we previously developed BE selections

to evolve high-activity adenine and cytidine deaminases[109,110], these selections place targeted

nucleotides within the coding sequence of T7 RNA polymerase (T7 RNAP). This selection

strategy is not broadly applicable to evolve altered PAM compatibility since changing the target

PAM and protospacer likely requires changing the coding sequence of T7 RNAP. Furthermore,

evolved variants with high activity that edit over large activity windows may inadvertently alter the activity of T7 RNAP through bystander editing.

To address these limitations, we designed a new selection strategy in which the target protospacer and PAM can be fully specified without impacting the coding sequence of the gene responsible for selection survival (**Figure 3.1**)**.** To achieve this programmability, we used the splicing capabilities of inteins, protein elements that insert and remove themselves from other proteins *in cis*, leaving only a small (~3- to 10-aa) extein scar[111,112]. We hypothesized that *trans* split-inteins could function effectively as *cis* splicing elements when the N- and C-inteins are fused together with a linker containing a programmed PAM and protospacer. We used the split-intein pair from *N. punciforme* (Npu)[113] since we previously showed that gIII split after Leu 10 with the Npu intein supports robust phage propagation after *trans* splicing[114].

To test whether the reconfigured *cis*-splicing Npu intein supports phage propagation, we constructed an accessory plasmid (AP) with the N- and C-terminal halves of the Npu intein fused together with a flexible 32-aa linker and inserted into the coding sequence of gIII after Leu 10 under the control of the phage shock promoter (psp)[115] (**Figure 3.2**). When infected with ∆gIII-phage, host cells containing this AP supported robust phage propagation in a splicing-dependent manner similar to cells containing psp-driven wild-type gIII. Importantly, installation of stop codons within the linker sequence reduced phage propagation by >10^5-fold relative to the unmutated construct (**Figure 3.2**), indicating that this selection, which we term sequence-agnostic Cas PACE (SAC-PACE), should enable robust selection of variants capable of correcting targeted stop codons.
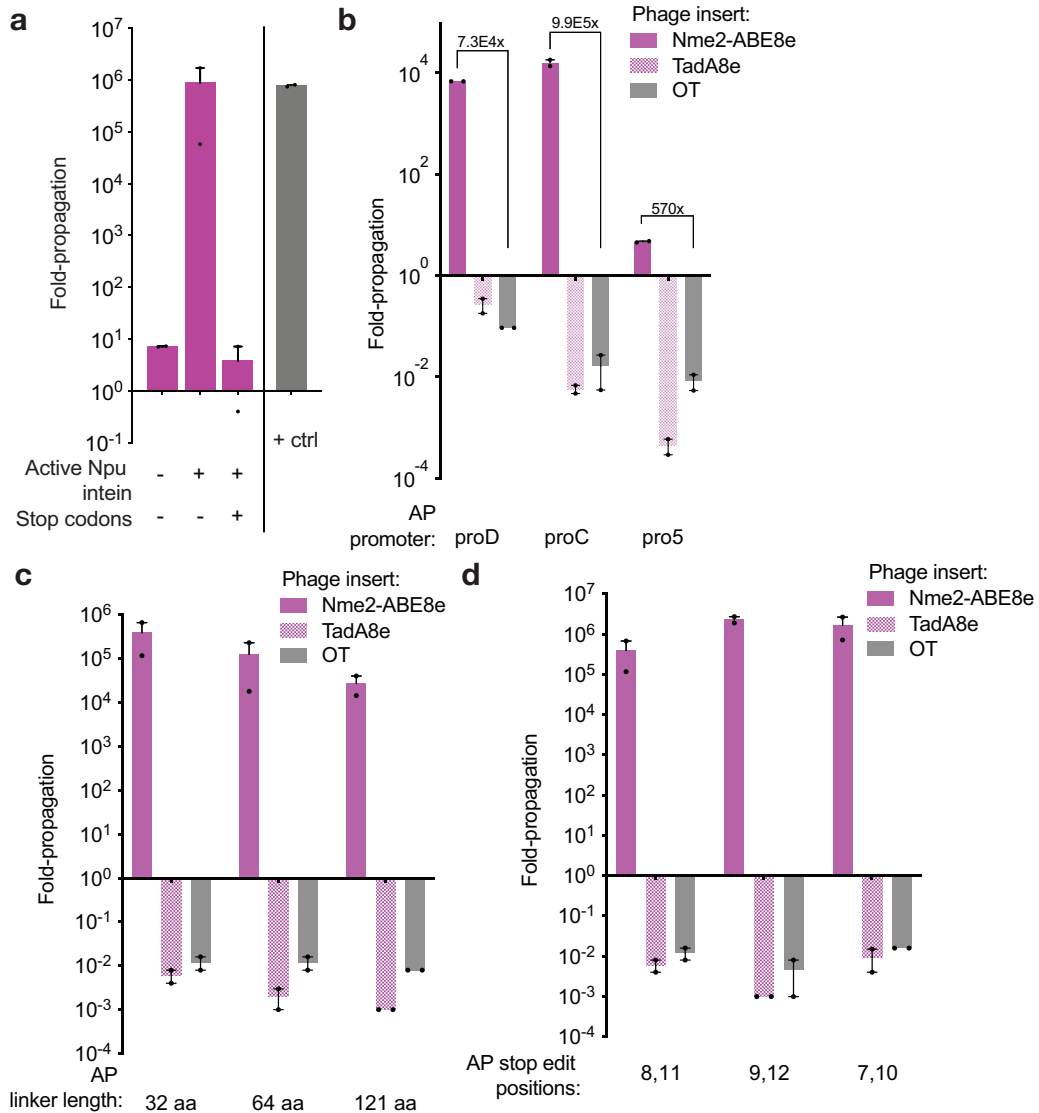
Figure 3.2. Validation of the SAC-PACE selection.
(a) Overnight propagation assay to test the requirements of active intein splicing and stop codons to turn on or off, respectively, the SAC-PACE circuit. **(b)** Overnight phage propagation assays to test the selection stringency of SAC-PACE with various AP promoter strengths. **(c)** Overnight propagation assay to test the linker length limitations of SAC-PACE, OT phage did not contain Nme2-ABE8e or TadA8e. **(d)** Overnight propagation assay to test the relative activity of Nme2-ABE8e phage when the target adenines within the stop codons are placed at different locations in the 23 nucleotide Nme2Cas9 protospacer (counting the PAM as positions 24-29). For (**a-d**), Mean±SEM is shown and are representative of *n = 2* independent biological replicates. Fold-propagation is calculated as the ratio of titer after overnight propagation over inoculating titer.

Next, we tested whether adenine base editing could support phage propagation in SAC-PACE. Indeed, on host cells harboring an AP containing gIII with two stop codons flanked by a cognate Nme2Cas9 N_4CC PAM, phage encoding dead Nme2Cas9 fused to the adenosine deaminase TadA8e[109] (Nme2-ABE8e) enriched 10^2- to 10^6-fold after overnight propagation, depending on the expression level of the gIII-construct (**Figure 3.2**). In contrast, phage containing only TadA8e or a non-targeting gene de-enriched in these host cells below the limit of detection at any tested expression level, indicating a large base-editing dependent dynamic range for this selection.

To test the generality of the selection circuit, we generated a series of APs containing linkers between 32 and 121 aa or with stop codons placed at different positions within the protospacer (**Figure 3.2**). Although propagation decreased with increasing linker length, the maximum tested linker length of 121 aa still supported strong overnight propagation sufficient to support phage survival during PACE ($> 10^4$-fold)[105]. This linker length can encode up to 10 simultaneous protospacer/PAM combinations (23 to 30 nt in length) with at least 7 nt between targets, a spacing shown to be compatible for multiple Cas protein binding events[116]. Together, these results suggest that the SAC-PACE selection is a highly flexible system that could be used to evolve the PAM scope of Cas variants.

3.2.3   A high-throughput platform for phage-assisted continuous evolution (ePACE)

Previous efforts to evolve SpCas9 on specific PAM sequences (NAG, NAC, NAT, etc.) yielded variants with both higher activity and specificity compared to variants evolved on a broad set of pooled PAMs[97]. Evolving on specific PAM sequences using traditional PACE methodology, however, is limited by throughput, since PACE is inherently challenging to parallelize due to cost, space, and design complexity, requiring temperature-controlled rooms and fluid-handling equipment[117]. This constraint limits the number of conditions that can be

explored in a PACE campaign, a drawback given the difficulty of predicting the set of conditions that will evolve molecules with desired properties.

To address this throughput challenge and enable large-scale parallel PACE of Nme2Cas9 towards specific PAMs, we developed ePACE (**Figure 3.3**). The ePACE system combines the continuous mutagenesis and selection of PACE with the highly scalable, customizable, and automated eVOLVER continuous culture platform, which has already proven effective for directed evolution[118]. Three key design features of eVOLVER make it an ideal choice for facilitating parallel PACE selections. First, eVOLVER enables individual programmatic control of continuous culture conditions, allowing the platform to simultaneously operate PACE chemostat cell reservoirs and lagoons on a standard lab benchtop. Second, eVOLVER can scale in a cost-effective manner to arbitrary throughput, enabling large-scale parallelization of miniature PACE reactors. Lastly, the do-it-yourself and open-source nature of eVOLVER allow it to be rapidly adapted and reconfigured for novel actuation elements, making it amenable to the customization necessary to run PACE. Integrating PACE and eVOLVER enables the simultaneous execution of PACE experiments across eight different PAMs (or other selection conditions) in parallel. Given that PACE experiments typically require 1-2 weeks each, this 8-fold increase in throughput represents a 2- to 4-month reduction in experimental time compared to traditional single-lagoon PACE at a 10-fold reduction in cost.
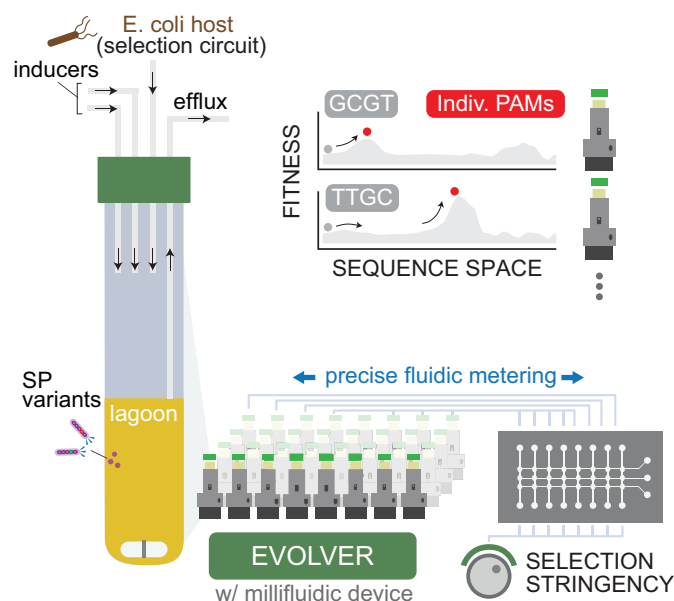
Figure 3.3. Schematic of ePACE.
A microfluidic control panel enables parallelization of up to 16 independent lagoons for PACE, greatly increasing the number of conditions that can be tested in parallel during continuous evolution.

To test the evolutionary capabilities of ePACE, we evolved a folding-defective

(G32D/I33S) maltose-binding protein (MBP) variant validated in traditional PACE[114]. Previously,

this folding defective MBP was evolved using a two-hybrid selection scheme to optimize both

soluble expression of the MBP variant and binding to an anti-MBP monobody[114]. We replicated

this evolution using ePACE, yielding evolved MBP variants with mutations at residues clustered

around the monobody-MBP interaction interface (D32G, A63T, R66L) that we previously

observed in PACE[114]. These results demonstrate that eVOLVER equipped with IPP devices can

successfully support and automate PACE, validating the ePACE platform for high-throughput

continuous directed evolution.

3.2.4   Development of a high-throughput base editing-dependent PAM profiling method

Next, we developed a method to rapidly profile the PAM scope of Nme2Cas9 variants

that emerge during evolution. Assessing PAM compatibility by testing individual sites in

mammalian cells is throughput-limited. Although many library-based PAM-profiling methods have been described, these methods rely on nuclease activity (PAM depletion[82], PAMDA[98,101], TXTL PAM profiling[119], CHAMP[120], etc.) or Cas protein binding activity (PAM-SCANR[121], CHAMP[120], etc.), which may not fully reflect PAM compatibility in precision gene editing applications such as base editing. We previously reported a mammalian cell base editing profiling assay[97,122]; however, this method is both slower and costlier than cell-free[119,120] or *E. coli*-based[82,98,101,121] methods, making it better suited for the characterization of late-stage variants.

      To address the need to rapidly assess the PAM specificities of newly evolved Cas9 variants in base editor form, we developed a base editing-dependent PAM profiling assay (BE-PPA). In BE-PPA, a protospacer or library of protospacers containing target adenines (ABE-PPA) or cytosines (CBE-PPA) is installed upstream of a library of PAM sequences (**Figure 3.4**). This library is transformed into *E. coli* along with a plasmid expressing a base editor of interest. Since base editing at each PAM is measured independently of other PAMs, BE-PPA offers greater sensitivity compared to nuclease-based assays. The PAM profile we observed for BE2 (rAPOBEC1-dSpCas9-UGI) using CBE-PPA closely matched (R^2 = 0.97) the PAM profile we previously observed for the related CBE, BE4, in mammalian HEK293T cells[97] (**Figure 3.4**), validating BE-PPA as a rapid base editor PAM profiling method.
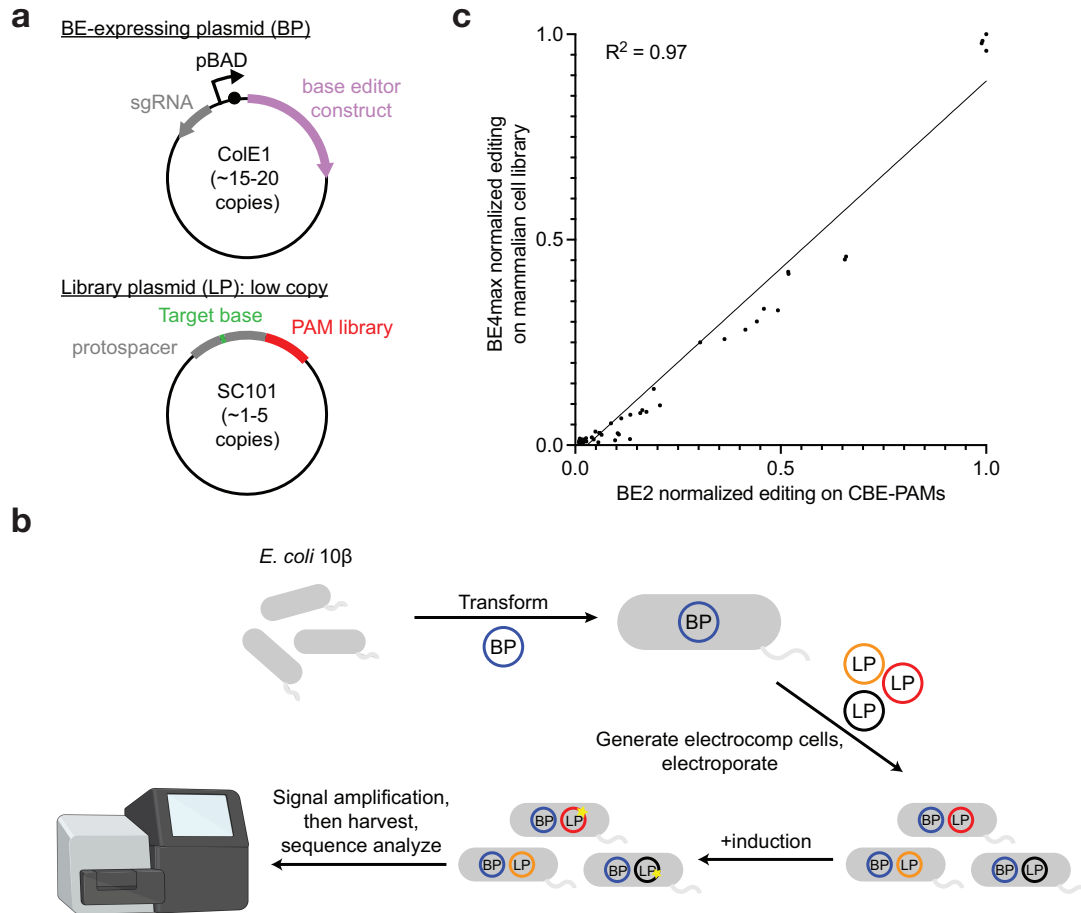
Figure 3.4. Development of a base editing-dependent PAM profiling assay (BE-PPA).
(a) Schematic of BE-PPA constructs. A BE-expressing plasmid (BP) containing the base editor to be evaluated is cloned along with a library plasmid (LP) containing a target protospacer and target base (adenine or cytosine for ABE-PPA or CBE-PPA, respectively) flanked by a library of PAMs of interest. **(b)** BE-PPA workflow. A cell line containing the BP is first generated, then the LP is electroporated into that cell line before base editor expression is induced. Induced cells are grown for 22-36 hours (with dilution after 24 hours if necessary), before plasmid DNA is harvested and sequenced by high-throughput sequencing. **(c)** Comparison of the BE-PPA assay against existing mammalian cell base editing PAM profiling[97]. Each point represents 1 of 64 NNN PAMs, normalized to the activity of the highest PAM for BE2 (rAPOBEC1-dSpCas) along the x-axis in BE-PPA or for BE4max along the y-axis for the previously assessed mammalian library. All points reflect the average normalized activity of *n* = 2 independent biological replicates. The line reflects a simple OLS regression, with the R-squared value shown.

3.2.5   Strategy for evolving the PAM scope of Nme2Cas9

Having validated the SAC-PACE selection, the ePACE system for high-throughput

continuous evolution, and the BE-PPA method for profiling PAM compatibility of base editors,

we next identified desirable target PAMs for evolving Nme2Cas9. In overnight propagation

assays, phage containing Nme2-ABE8e exhibited modest to strong propagation ($N_3NCG <$

$N_3NCA < N_3NCT < N_3NCC$) on the set of 16 $N_3\underline{NCN}$ PAMs, and strong propagation on $N_3\underline{NTC}$

PAMs if the base immediately downstream of the canonical six base pair PAM was a C (PAM

position 7, NNNNNN$\underline{N}$, counting the canonical PAM as positions 1-6), likely due to PAM

slippage (**Figure 3.5**)[123]. This initial activity suggested an overall evolution campaign along two

trajectories: a more difficult trajectory towards activity on N_4TN PAMs that could require several

selection stringencies, and a simpler trajectory towards N_4CN-active variants. If successful,

these variants could together enable targeting of PAM sequences largely complementary to the

PAM scope of existing, high-activity SpCas9 variants.
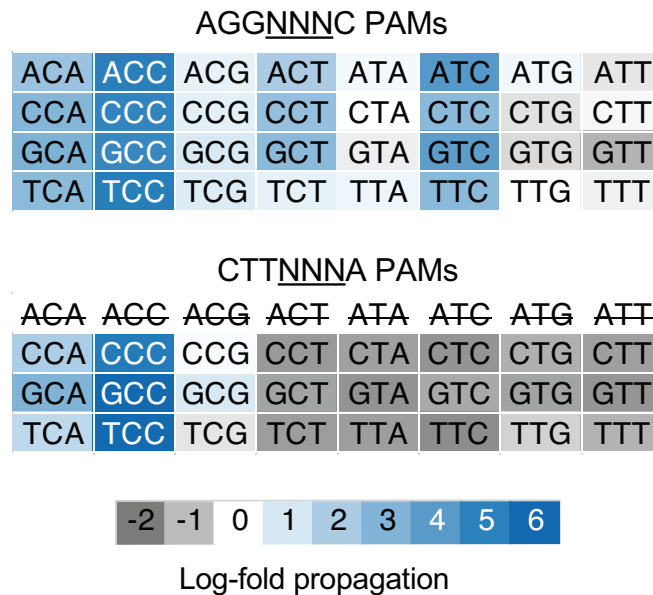
Figure 3.5. Overnight propagation of Nme2-ABE8e on N_3NYN PAMs.
Overnight propagation assays of wild-type Nme2-ABE8e on two sets of 32 $N_3\underline{NYN}$ PAMs. Fold-propagation was measured by qPCR and is reflective of the average of two independent biological replicates. The eight CTT$\underline{AYN}$A PAMs are excluded as they introduce an additional stop codon in the AP, preventing Cas-dependent propagation.

3.2.6 Low stringency evolution of Nme2Cas9 towards N$_4$TN PAM sequences

We first used our evolution platform to perform parallel SAC-PACE selections to evolve

Nme2Cas9 variants towards specific N$_4$TN PAM sequences (**Figure 3.6, left side**). We

envisioned using the initial activity of wild-type Nme2Cas9 on some N$_4$TC PAMs (**Figure 3.5**) as

an evolutionary stepping-stone to access other N$_4$TN PAMs. Using the original (low stringency)

SAC-PACE selection featuring one protospacer, two stop codons, and one target PAM (**Figure

3.7, left panel**), we evolved wild-type Nme2-ABE8e on host cells containing APs with each of

the eight possible N$_3$YTN APs and the mutagenesis plasmid (MP6)[124] (**ePACE1**, **Figure 3.6**). As

expected, all APs aside from those containing a N$_3$TTC or N$_3$CTC PAM washed out rapidly.

However, those two PAM-containing lagoons persisted at up to 2 volumes/hr and yielded

Nme2Cas9 variants with PAM-dependent mutational convergence (**Table 3.1**). Consensus

mutations occurred both inside (I1025S, R1033K, S1043R for CTC PAM variants, Y1035C/H for

TTC PAM variants) and outside of the PID (Y441C, K581R, D844V/G for CTC PAM variants;

I462V, N616S, D844V for TTC PAM variants), suggesting potential PAM-specific and PAM-

independent improvements to Nme2Cas9. Indeed, early evolved variants (e.g. E1-2-ABE8e)

supported base editing activity on non-canonical PAMs and improved activity on wild-type N$_4$CC

PAMs in human cells (**Figure 3.9**). Surprisingly, expanded PAM activity appeared strongest on

N$_4$CN PAMs and was minimal on N$_4$TN PAMs.
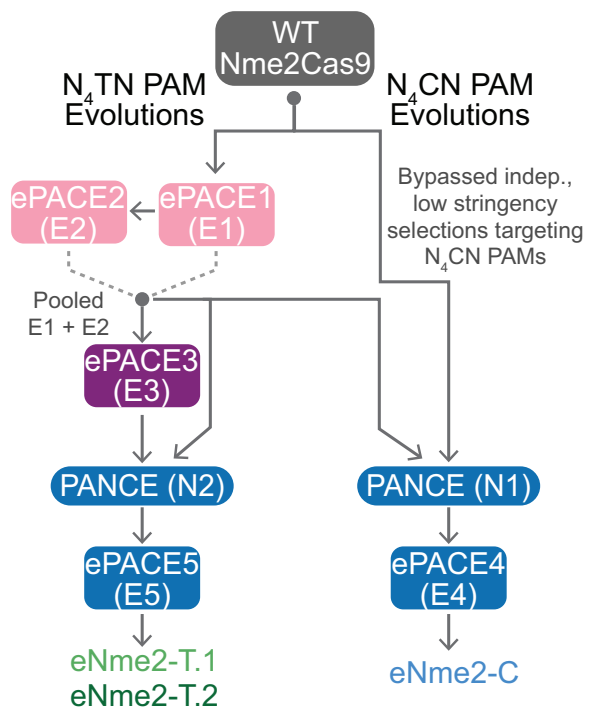
Figure 3.6. Overview of Nme2Cas9 evolution trajectories towards N$_4$TN and N$_4$CN PAM compatibility.

SAC-PACE | Split SAC-PACE | Dual PAM split SAC-PACE

SAC-PACE
gIII-N Int-N Int-C gIII-C [AP]
P_E.coli
PAM   STOP
NYNNNCTGGCTCTGAGTAGTGAGGAACA
G S E * * G T
E. coli host
Ribosome — gIII-N / gIII-C — ✗STOP codon — Inactive phage ✗
BE correction — Ribosome → gIII / pIII — gIII-N / gIII-C — Phage propagation ✓
Low stringency

Split SAC-PACE
gp41-8C-dCas9 [SP]
P_gIII
TadA8e-gp41-8N [CP]
P_psp
E. coli host
Limited TadA8e conc.
Cas9
TadA — BE conc.
Translation level set by synthetic RBS in host cell

Dual PAM split SAC-PACE
gIII-N Int-N Int-C gIII-C [AP]
P_E.coli
PAM1 PAM2
gp41-8C-dCas9 [SP]
P_gIII
TadA8e-gp41-8N [CP]
P_psp
E. coli host
gIII-N / Cas9 — Two edits required — gIII-C / Cas9 — Limited BE conc.
High stringency

Figure 3.7. SAC-PACE selection schemes.
Overview of SAC-PACE modifications increasing selection stringency. (Left) original selection scheme; (middle) split SAC-PACE selection in which the expression of TadA8e is placed on a complementary plasmid (CP) in the host cell, enabling tunable control of active enzyme concentration; (right) dual PAM split SAC-PACE selection in which limited active enzyme concentration is coupled with a requirement to edit an additional protospacer and PAM sequence containing a stop codon. In the evolutions described in this work, the protospacer was kept constant for multi-site edits.

	TadA8e	dNme2Cas9 (N-terminal: REC/HNH/RuvC)																					
Residue number	26	47	60	67	105	212	296	335	336	337	441	451	462	560	581	616	647	657	768	777	805	808	811
wild-type	R	E	M	S	S	L	P	L	R	Y	Y	D	I	E	K	N	S	F	Q	E	K	T	A
CTC.E1-1	G					P						V					N						
CTC.E1-2	G								C						R								I
CTC.E1-3	G								C						R						G		
CTC.E1-4	G								C						R								
TTC.E1-5						G		P	K				V						R				
TTC.E1-6	G			R			L				C				S								T
TTC.E1-7		K	V								E	V	G						V		G	R	
TTC.E1-8	G			R											S								

	dNme2Cas9 (C-terminal: WED, PID)									
Residue number	844	904	908	1025	1033	1035	1043	1045	1051	1075
wild-type	D	K	D	I	R	Y	S	E	S	L
CTC.E1-1	V			S	K		R			
CTC.E1-2	G			S	K		R			M
CTC.E1-3	G			S	K		R			M
CTC.E1-4	G			S	K		R			M
TTC.E1-5	V						C	A	P	
TTC.E1-6	V	N					H		A	
TTC.E1-7	V		G				H			
TTC.E1-8	V	N					H			

Table 3.1. Mutation table of ePACE1 evolved variants.
Genotypes of individually sequenced plaques following ePACE1, with positions varying from wild-type displayed. Clones evolved on different PAMs are delineated by a bold line.
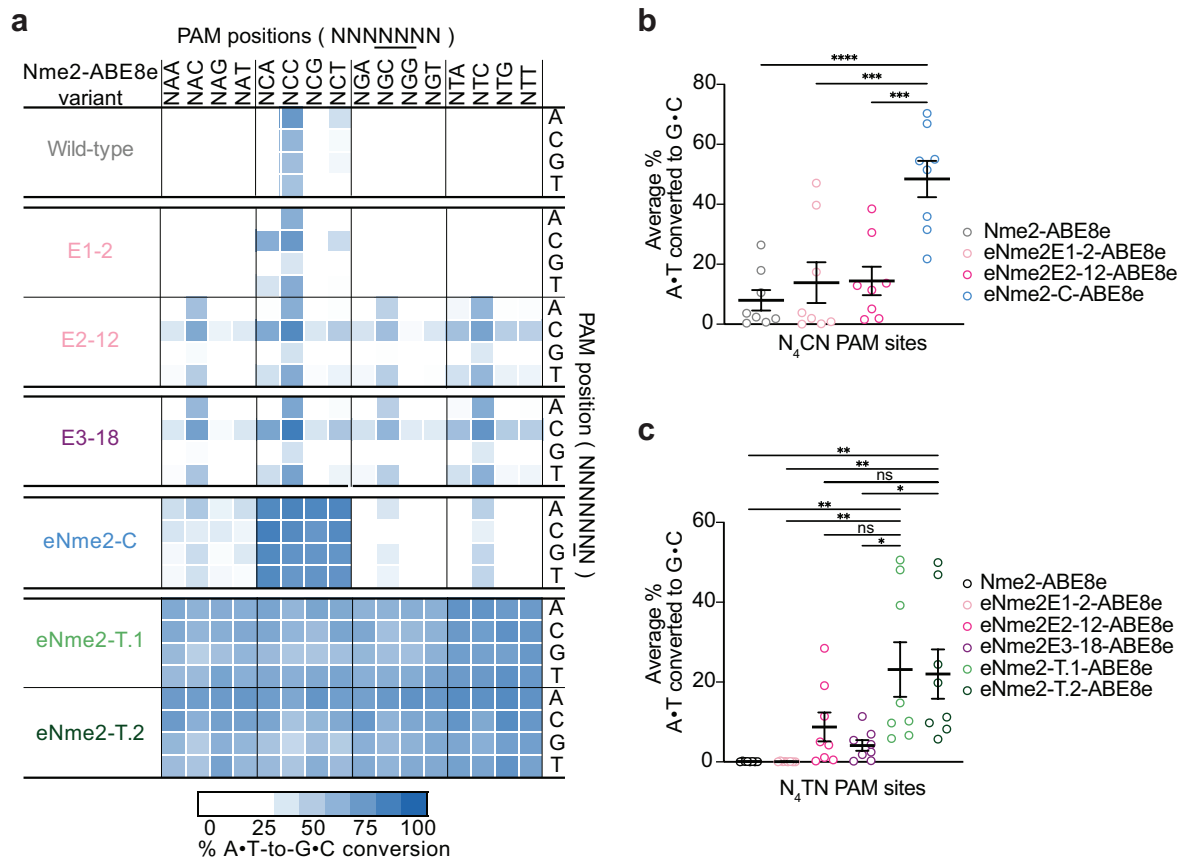
Figure 3.8. Improved activity of SAC-PACE evolved Nme2Cas9 variants on N4YN PAM sequences in ABE-PPE and mammalian cells.
(a) Summary heat map showing ABE-PPA activity for representative variants across both evolutionary trajectories. Values plotted are raw observed % A•T-to-G•C conversion for one replicate of each base editor. **(b)** Summary dot-plots showing the progression of mammalian cell adenine base editing activity at eight $N_4\underline{CN}$ PAM-containing sites for representative variants from the $N_4\underline{CN}$ evolution trajectory. **(c)** Summary dot-plots showing the progression of mammalian cell adenine base editing activity at eight $N_4\underline{TN}$ PAM-containing sites for representative variants from the $N_4\underline{TN}$ evolution trajectory. For (**b,c**), each point represents the average editing of $n = 3$ independent biological replicates measured at the maximally edited position within each given genomic site. Mean±SEM is shown and reflects the average activity and standard error of the pooled genomic site averages. ns, $p > 0.05$; *, $p \leq 0.05$; **, $p \leq 0.01$, ***, $p \leq 0.001$, ****, $p \leq 0.0001$. p-values determined by Sidak's multiple comparisons test following ordinary one-way ANOVA.
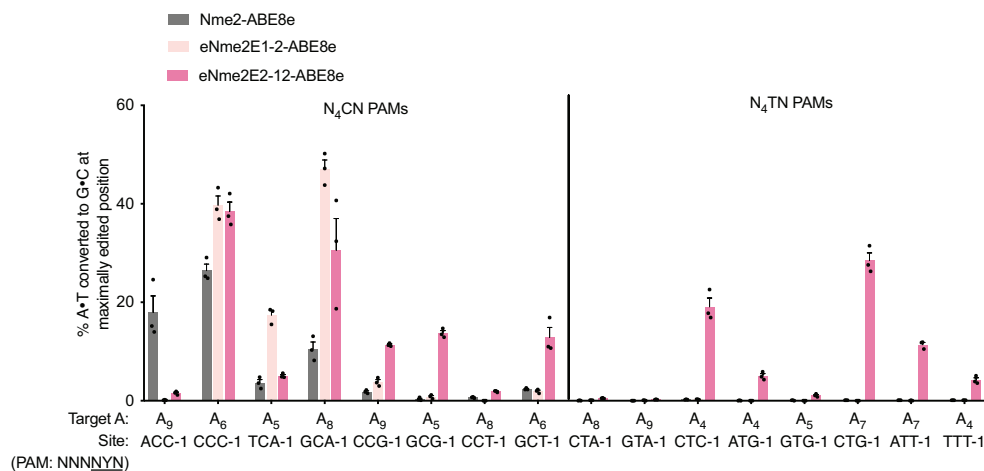
Figure 3.9. Mammalian cell adenine base editing activity of early evolved Nme2Cas9 variants.
Adenine base editing activity of a representative ePACE1 and ePACE2 clone (E1-2-ABE8e and E2-12-ABE8e) at eight N_3NCN PAM-containing sites and eight N_4TN PAM-containing sites in HEK293T cells. Mean±SEM are shown and are representative of *n* = 3 independent biological replicates.

We reseeded all PAM lagoons with pooled phage from the two surviving PAMs

(**ePACE2**) (**Figure 3.6**). All lagoons now exhibited strong propagation at up to 2.5 volumes/hr,

but surviving phage appeared to lose the Nme2-ABE8e cassette, indicating recombination to

bypass the selection. We sequenced clones that did not show recombination and found novel

mutations that again appeared to cluster by PAM/lagoon both in and outside of the PID (**Table**

3.2). In mammalian cells, while expanded PAM compatibility did extend to some N_4TN PAMs,

activity appeared to be site-dependent while moderate activity on N_4CN PAMs was retained

(**Figure 3.8, Figure 3.9**). These ePACE1 and ePACE2 outcomes suggested that the low

stringency SAC-PACE selection may be insufficient to generate highly active Nme2Cas9 PAM

variants.

Table 3.2. Mutation table of ePACE2 evolved variants (dNme2Cas9 N-terminal: REC/HNH/RuvC and TadA8e)

Residue number	26	49	53	71	102	124	156	163	2	47	53	68	94	119	123	154	186	323	340	361	396	409	424	431	441	451	452	462	508	520	532	536	545	581	624	633	634	659	701	719	720	762	767	768	769	770	771	792	813
wild-type	R	A	H	M	R	Y	V	A	A	E	K	V	A	D	T	E	E	L	D	E	T	E	S	I	Y	D	H	I	E	E	K	K	V	K	N	Q	E	E	Q	N	D	T	H	Q	K	T	H	G	K
CTC-L1.E2-1	G						A		S				E						N								C		V																R	R				R
CTC-L1.E2-2	G																										C		V																	R			R
CTC-L1.E2-3	G																										C		V																	R			R
CTC-L1.E2-4	G																										C		V									T								R			R
CTC-L2.E2-5	G							T				R															C		V												D					R		R	R
CTC-L2.E2-6	G																										C		V																	R			R
CTC-L2.E2-7	G																										C		V																				R
CTC-L2.E2-8	G																										C		V																R			R	R
CTG-L2.E2-9	G																								F		C							R											A			A	
CTG-L2.E2-10	G													D											F		C			R									S					A	R		A	V	R
CTG-L2.E2-11	G													D											F		C					D			K									A	R		A	V	R
CTT-L2.E2-12	G					H			K			M			A	K					A	A				R																	A	Y	R		L	R
CTT-L2.E2-13	G	P	I			H			K			M		G	A	K						A				R							R			D							A					
CTT-L2.E2-14	G					H			K			M			A	K						A				R																	A					
CTT-L2.E2-15	G								K			M			A	K						A				R																	A			R		
TTG-L2.E2-16		V							K			M				K	R					A				R					M								R				A		R			
TTG-L2.E2-17	G								K			M				K										V											R											
TTT-L1.E2-18	G								K							K	A								N	R					K																	
TTT-L1.E2-19	G								K							K	A								N	R					K																	
TTT-L1.E2-20	G								K							K						K			N	R					K																	

dNme2Cas9 (C-terminal: WED, PID)

Residue number	844	858	869	911	929	932	933	951	968	981	986	991	1005	1025	1028	1033	1043	1045	1049	1075
wild-type	D	K	I	D	K	E	S	M	Y	N	I	Y	K	I	D	R	S	E	R	L
CTC-L1.E2-1	V	T												S		K	R			M
CTC-L1.E2-2	V													S		K	R			M
CTC-L1.E2-3	V													S		K	R			
CTC-L1.E2-4	V	R						H						S		K	R			
CTC-L2.E2-5	V													S		K	R			
CTC-L2.E2-6	V													S		K	R			
CTC-L2.E2-7	V													S		K	R			
CTC-L2.E2-8	V													S		K	R			
CTG-L2.E2-9	V													S	N	K	R		C	
CTG-L2.E2-10	V		A	R			R							S	N	K	R		C	
CTG-L2.E2-11	V		A	R			R							S	N	K	R		C	
CTT-L2.E2-12	A							R								K				
CTT-L2.E2-13	A				G			R			H					K		A		
CTT-L2.E2-14	A							R								K				
CTT-L2.E2-15	A							R								K				M
TTG-L2.E2-16	A							R	S							K				
TTG-L2.E2-17	A							R								K	R			
TTT-L1.E2-18	A							R		T		R				K	R			
TTT-L1.E2-19	A							R				R				K	R			
TTT-L1.E2-20	A							R				R				K	R		D	

Table 3.2. Mutation table of ePACE2 evolved variants.
Genotypes of individually sequenced plaques following ePACE2, with positions varying from wild-type displayed. Clones evolved on different PAMs are delineated by a bold line. Mutations that had previously appeared in ePACE1 are shown in light pink, while novel mutations are shown in magenta.

We used ABE-PPA to profile the PAM compatibility of wild-type Nme2-ABE8e and a representative ABE variant from both ePACE1 (E1-2-ABE8e) and ePACE2 (E2-12-ABE8e) that had exhibited improved mammalian cell base editing activity on N_4YN PAMs (**Figure 3.8, Figure 3.10**). While both evolved variants exhibited improved activity on N_4CD (where D = A, G, or T) PAMs over Nme2-ABE8e (17%, 23%, and 32% average A•T-to-G•C conversion for Nme2-ABE8e, E1-2-ABE8e, and E2-12-ABE8e, respectively), only the more evolved variant, E2-12-ABE8e, exhibited improved N_4TN PAM activity (2%, 2%, and 39% average A•T-to-G•C conversion for Nme2-ABE8e, E1-2-ABE8e, and E2-12-ABE8e, respectively). This result suggests a model in which broadened activity on N_4CN PAMs precedes activity on N_4TN PAMs.
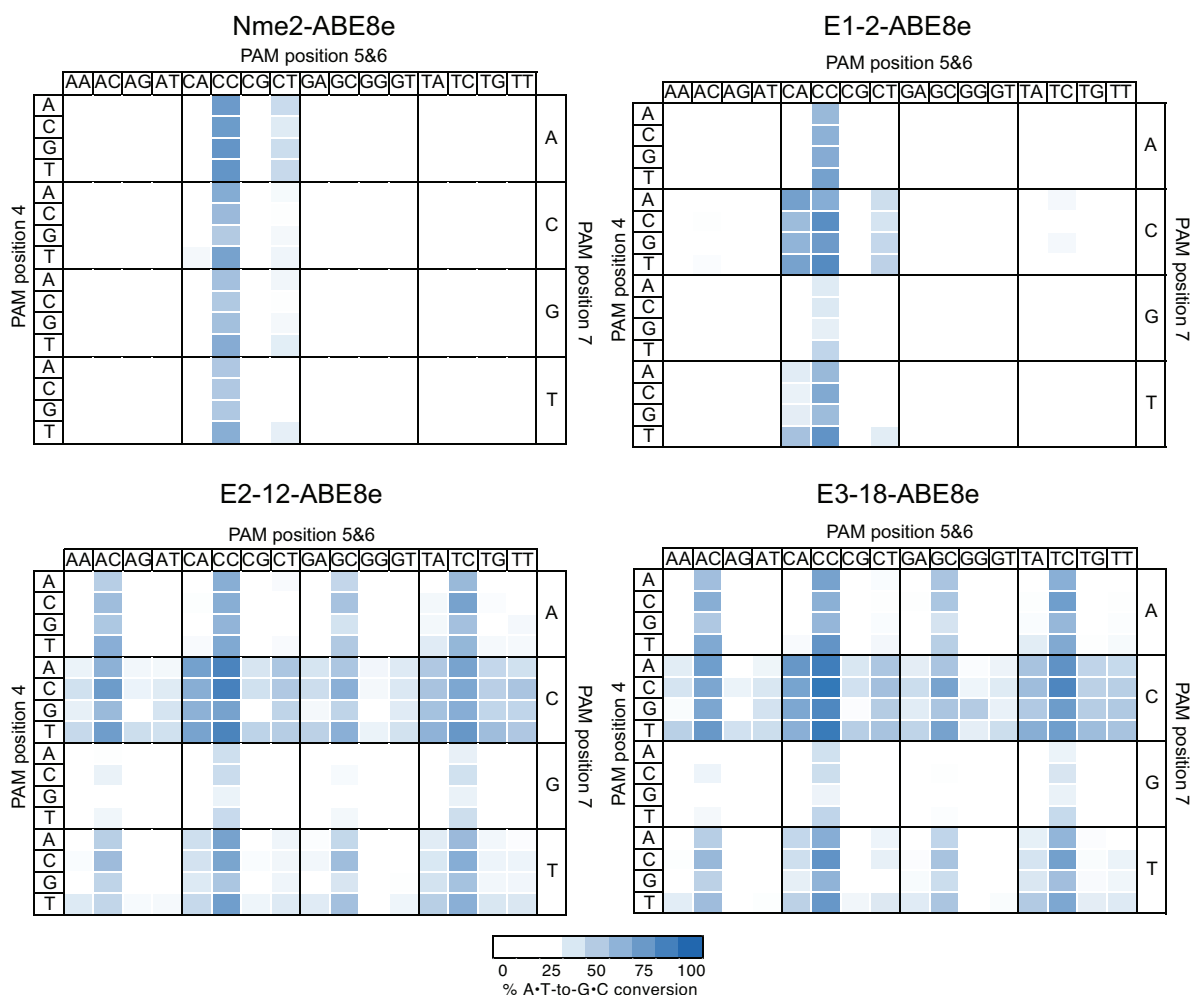
Figure 3.10. ABE-PPA activity of wild-type Nme2-ABE8e and early evolved Nme2Cas9 variants.
Heat maps showing ABE-PPA activity of wild-type Nme2-ABE8e (top-left) and representative clones from ePACE1 (top-right), ePACE2 (bottom-left), and ePACE3 (bottom-right) on the set of 256 N₃NNNN PAMs (PAM positions 1-3 fixed). Values are raw % A•T-to-G•C conversion observed for one replicate of each editor.

Further examination of the ABE-PPA data indicated that broadened PAM activity of early evolved Nme2Cas9 variants was primarily driven by an acquired C preference at the undesired PAM position 7, a position not recognized by the wild-type enzyme[125]. While E1-2-ABE8e and E2-12-ABE8e progressively improve base editing activity compared to wild-type Nme2-ABE8e on N₄YNC PAM sites (18%, 29%, and 58% average A•T-to-G•C conversion for Nme2-ABE8e, E1-2-ABE8e, and E2-12-ABE8e, respectively), base editing activity was improved to a lesser

extent at $N_4\underline{YND}$ PAM sites (14%, 14%, and 33% average A•T-to-G•C conversion for Nme2WT

ABE8e, E1-2-ABE8e, and E2-12-ABE8e, respectively). This discrepancy suggested the need

for higher selection stringency to restrict the survival of Cas variants that acquire expanded

PAM recognition at undesired positions.

3.2.7   Increasing SAC-PACE selection stringency to evolve high-activity Nme2Cas9 variants

In previous efforts evolving SpCas9, restricting the amount of active enzyme and

requiring additional PAM recognition via a multi-PAM system increased selection stringency and

enabled evolution of higher activity variants[97]. We hypothesized similar strategies could be

implemented in SAC-PACE to evolve high-activity Nme2Cas9 variants while preventing

selectivity at undesired PAM positions (**Figure 3.7**). To limit the amount of active base editor,

we used a split-intein strategy with the base editor split at the linker between TadA8e and

dNme2Cas9, which we hypothesized could tolerate the insertion of an extein scar (split SAC-

PACE) (**Figure 3.7, middle panel**). We selected the fast-splicing gp41-8 intein pair[126,127] as the

Npu intein pair was already in use in the AP. In overnight propagation assays, only host cells

containing a psp-driven TadA8e-gp41-8N construct on a complementary plasmid (CP) enabled

survival of SP expressing gp41-8C-dNme2Cas9 (**Figure 3.11**). Since we can control the

expression level of the TadA8e construct on the CP, this result validated the ability of the split

SAC-PACE selection to limit base editor concentrations while continuing to select for evolving

Cas9-containing SP.
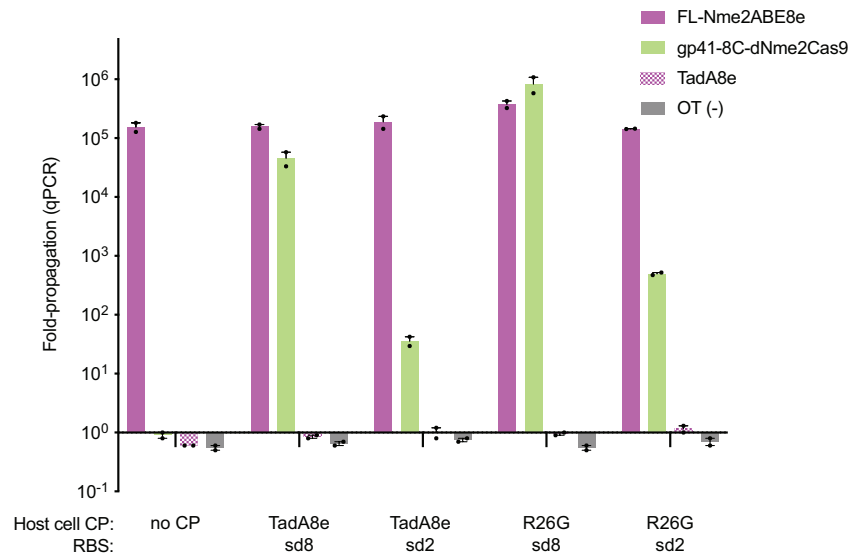
Figure 3.11. Validation of the split-SAC PACE selection.
Overnight propagation assay to test the activity of the split-SAC-PACE selection with different
TadA8e variants. Each TadA8e variant was fused to the N-terminal half of an intein (gp41-8N)
and placed on a complementary plasmid (CP) in host cells. FL-Nme2ABE8e phage contained
full-length, active Nme2ABE8e, and OT phage did not contain Nme2Cas9, intein, or TadA8e.
Mean±SEM are shown and are representative of n = 2 independent biological replicates. Fold-
propagation is calculated as the ratio of phage titer after overnight propagation over inoculating
titer.

Using the intermediate-stringency split SAC-PACE selection, we further evolved

Nme2Cas9 variants that had emerged from low-stringency selections. We pooled endpoint

phage from ePACE1 and ePACE2 and cloned them into the split SP architecture, then seeded

those SP into the split SAC-PACE selection (**ePACE3**) (**Figure 3.6**). All targeted PAMs

exhibited moderate phage persistence (>10^5 titers) within at least one lagoon at or above 2

vol/hr. Sequenced clones from lagoons other than the one targeting an N_3CTG PAM showed

very strong mutational convergence across lagoons and PAMs, suggesting that the resulting

Nme2Cas9 variants likely were not acquiring PAM specificity at the positions defined in our

evolutions (PAM positions 4 and 6) (**Table 3.3**). ABE-PPE profiling of a representative variant

from ePACE3 (E3-18-ABE8e) showed comparable activity (31% and 39% average A•T-to-G•C

conversion on N_4CD and N_4TN PAM sites, respectively) to the earlier evolved E2-12-ABE8e

59

variant. However, this broadened PAM compatibility was again accompanied by a PAM position 7 C preference (61% vs. 33% average A•T-to-G•C conversion on N$_4$<u>Y</u>NC and N$_4$<u>Y</u>ND PAM sites, respectively) (**Figure 3.8, Figure 3.10**), indicating that restricting enzyme concentration alone is insufficient to evolve higher activity variants with desired PAM preferences.

Table 3.3. Mutation table of ePACE3 evolved variants.
Genotypes of individually sequenced plaques following ePACE3, with positions varying from wild-type displayed. Clones evolved on different PAMs are delineated by a bold line. Mutations that had previously appeared in ePACE1 and ePACE2 are shown in light pink and magenta, respectively, while novel mutations are shown in purple.

Thus, we added another layer of stringency control to increase the likelihood of evolving higher activity variants. We implemented a multiplexed-PAM selection requiring correction of a stop codon in two protospacers flanked by PAM sequences with alternate sequence identity at PAM positions 1-3 and 7 (<u>NNN</u>NNN<u>N</u>), thereby forcing evolving Nme2Cas9 variants to recognize multiple nucleotides at undesired PAM positions. We coupled this selection with split SAC-PACE to produce a third (high stringency) scheme that we term dual-PAM split SAC-PACE (**Figure 3.7, right panel**). With these developments, we could now pursue high-stringency evolutions along *both* trajectories (N$_4$CN and N$_4$TN PAM sequences).

3.2.8 High stringency evolution of Nme2Cas9 towards N_4CN PAM sequences

The outcomes of ePACE1 and ePACE2 revealed that improved activity on $N_4\underline{TN}$ PAMs was accompanied by broadened activity on N_4CN PAMs. We hypothesized that the mutational diversity from these evolutions might provide useful starting points for the evolution of N_4CN PAM compatibility. We thus pursued this trajectory with *both* wild-type Nme2Cas9 and pooled ePACE1 and ePACE2 (E1+E2) phage, subjecting these starting points to high stringency evolutions in parallel via dual PAM split SAC-PACE (**Figure 3.6**).

SP containing either wild-type or E1+E2 phage propagated insufficiently for PACE on N_4CN-containing APs requiring dual edits. As such, we started evolution with PANCE, a non-continuous version of PACE in which phage are discretely passaged following an incubation period (typically overnight)[105]. Using PANCE (**N1**), we evolved either wild-type gp41-8C-dNme2Cas9 or pooled E1+E2 endpoint phage on the set of six $N_3\underline{WCD}$ (where W = A or T) PAMs (**Figure 3.6**). Following 20 passages in PANCE, only some of the lagoons targeting $N_3\underline{TCD}$ PAMs appeared to consistently propagate. Phage from these lagoons were then seeded into ePACE (**ePACE4**) (**Figure 3.6**). Interestingly, few mutations from E1+E2 were retained in ePACE4, both within and outside the PID, suggesting evolution of a distinct mode of PAM recognition among ePACE4 clones (**Table 3.4**).
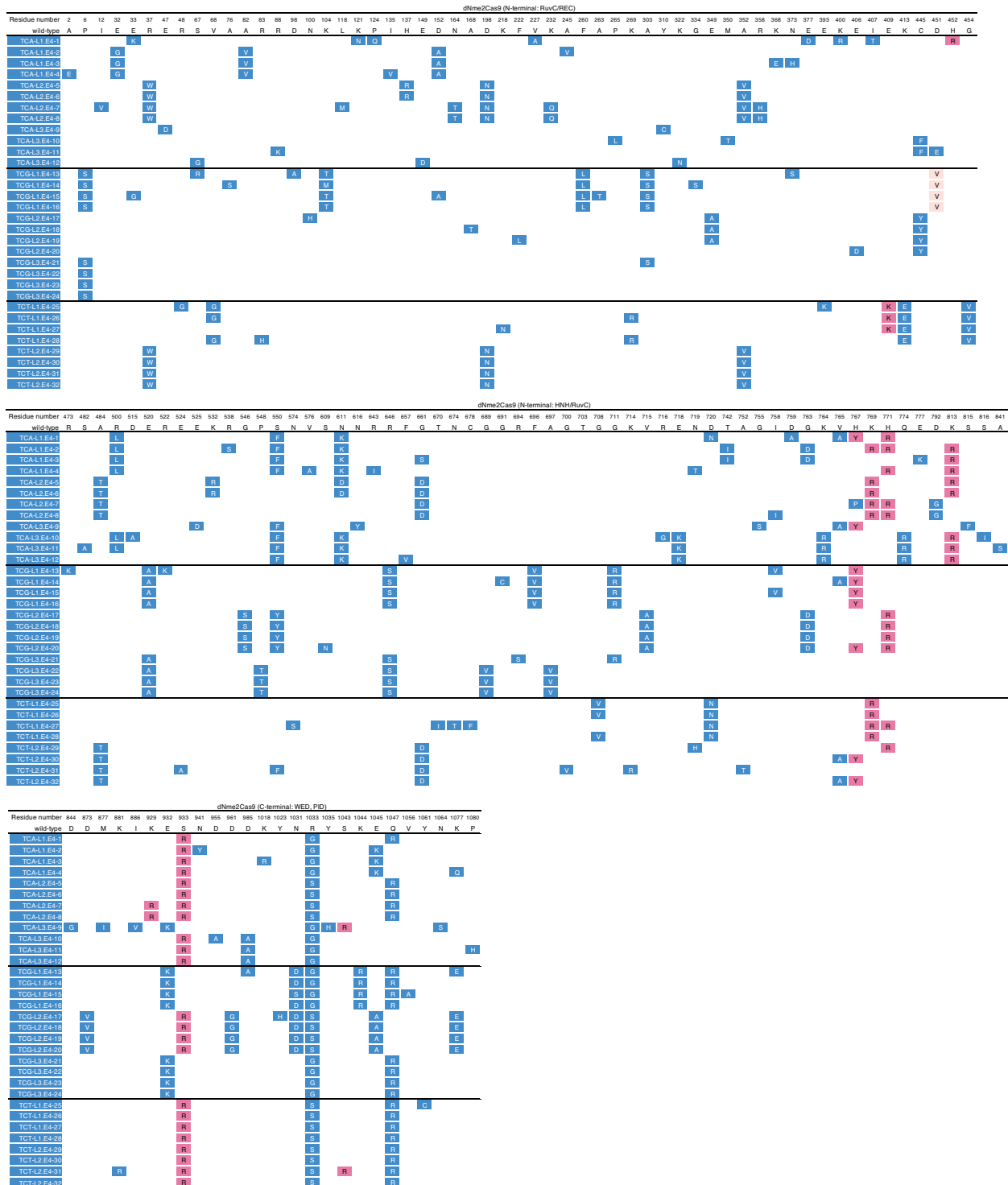
Table 3.4. Mutation table of ePACE4 evolved variants.
Genotypes of individually sequenced plaques following ePACE4, with positions varying from wild-type displayed. Clones evolved on different PAMs are delineated by a bold line. Mutations that had previously appeared in ePACE1 and ePACE2 are shown in light pink and magenta, respectively, while novel mutations are shown in blue.

Sixteen ePACE4 clones assayed using ABE-PPA exhibited strong and general ABE activity, averaging 66% editing across all N_4CN PAMs (**Figure 3.12**). The E4-15 variant in particular, which we denote as eNme2-C (Nme2Cas9 P6S, E33G, K104T, D152A, F260L, A263T, A303S, D451V, E520A, R646S, F696V, G711R, I758V, H767Y, E932K, N1031S, R1033G, K1044R, Q1047R, V1056A), achieved ≥80% A•T-to-G•C editing at all N_4CN PAM sites as an ABE8e, corresponding to a 4.8-fold average improvement in activity on N_4CD PAM sites over Nme2-ABE8e, and a 1.3-fold average improvement in activity even on N_4CC PAM sites natively recognized by wild-type Nme2Cas9 (**Figure 3.8, Figure 3.13**). Notably, activity improvements of ePACE4 variants on specific N_4CN PAMs appeared to be largely agnostic of the specific PAM offered during evolution, with most variants preferring N_4<u>CA</u> > N_4<u>CC</u> > N_4<u>CT</u> > N_4<u>CG</u> (**Figure 3.12**). The exceptions were variants evolved on the N_3<u>TCG</u> PAM, which exhibited N_4<u>CG</u> activity comparable to or better than activity on the other three groups of N_4<u>CD</u> PAMs. This result would suggest that binding of the position 6 G is distinct from binding to the other three nucleobases. In line with this hypothesis, the mutation profiles in the PID are relatively conserved between variants evolved on the N_3<u>TCA</u> and N_3<u>TCT</u> APs (S933R, R1033S/G, Q1047R). However, additional mutations outside of the three seen in those variants converged in the N_3<u>TCG</u> trajectory (D873V, E932K, D961G, N1031D/S, K1044R, E1045A, K1077E) (**Table 3.4**). Some of these additional mutations appear to contribute to improved N_4<u>CG</u> activity observed in ABE-PPA (**Figure 3.14**). Importantly, ePACE4 variants (e.g. eNme2-C, **Figure 3.8**) no longer exhibited the preference for a C at PAM position 7 exhibited in earlier evolved variants. Collectively, these findings establish that by requiring multiple PAM engagements, the dual PAM split SAC-PACE selection can successfully generate high-activity Cas9 variants with broadened PAM scope.
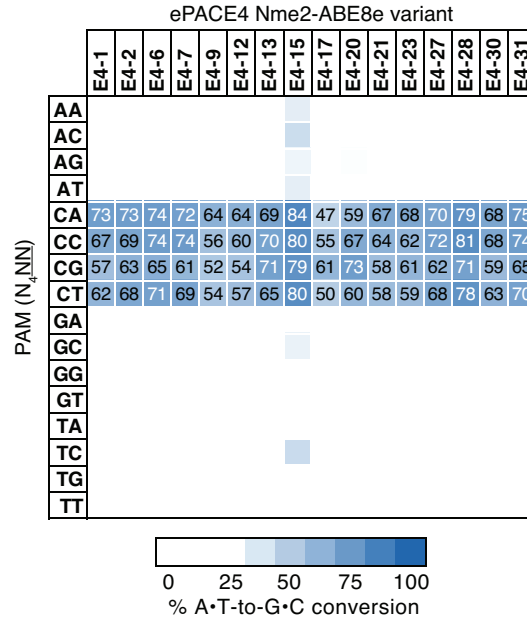
Figure 3.12. ABE-PPE activity of ePACE4 evolved variants.
Heat map showing ABE-PPA activity of representative clones from ePACE4 on the 16 combinations of PAM positions 5 and 6 (N4NN) Values are raw % A•T-to-G•C conversion observed for one replicate of each editor and are listed in each cell for the N4CN PAMs, with values above 70% A•T-to-G•C conversion colored white.
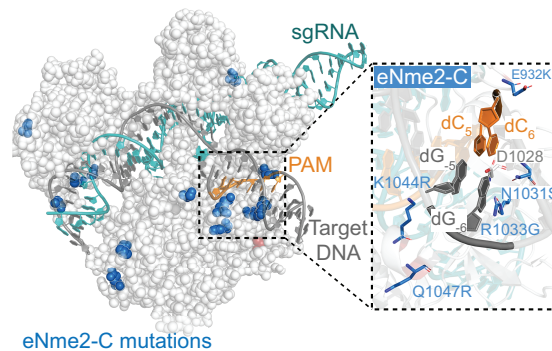
Figure 3.13. Crystal structure of Nme2Cas9 with mapped eNme2-C mutations.
Mutation overview of the eNme2-C variant, mapped onto the crystal structure of wild-type Nme2Cas9 (PDB: 6JE3), mutated positions are shown in blue. The inset shows the wild-type PAM and PAM-interacting residues (D1028, R1033), with evolved mutations listed.
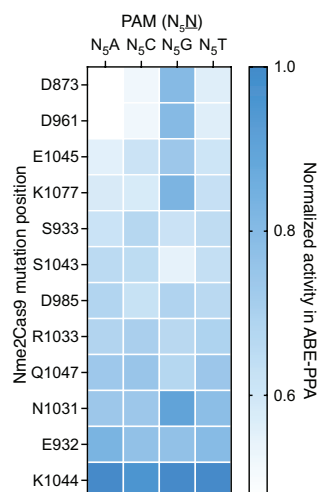
Figure 3.14. Mutational analysis of ABE-PPE activity of ePACE4 evolved variants.
ABE-PPA activity in **Figure 3.12** pooled and segregated by mutation position. Each column depicts the impact of a given position, when mutated, on ABE-PPA activity at each of the four PAM groups (N_5A, N_5C, N_5G, N_5T). Values are normalized against the highest activity within each set of PAMs. Only positions that were observed to be mutated more than once (**Table 3.4**) were included in this analysis.

Encouraged by the PAM profile of ePACE4 variants, we next tested whether the activity observed in bacterial cells successfully translated to mammalian cells. In HEK293T cells we observed robust ABE activity for eNme2-C-ABE8e across all eight endogenous human genomic N_4CN sites previously tested (**Figure 3.8**). Notably, eNme2-C-ABE8e showed 2.0-fold higher average editing efficiency on N_4CC PAM sites and 15-fold higher editing efficiency on N_4CD PAM sites than Nme2-ABE8e, and 2.3 to 3.3-fold improved editing at all sites compared to earlier evolved variants eNme2-E1-2-ABE8e and eNme2-E2-12-ABE8e, respectively. To further test the N_4CN PAM generality of eNme2-C-ABE8e, we evaluated activity at an additional 25 genomic sites flanked by N_4CN PAMs (for a total of 33 endogenous genomic sites tested) and observed an average of 34% A•T-to-G•C conversion at the tested sites exhibiting base editing above 1% (32 of 33 sites), a 1.8- and 30-fold average improvement at N_4CC and N_4CD PAM sites, respectively, over Nme2-ABE8e (**Figure 3.15**). The editing window of eNme2-C-ABE8e is approximately between protospacer positions 9 and 16 (counting the PAM as positions 24-29)

and retains a protospacer preference centered around 23 base pairs in length **(Figure** 3.15**).**

Together, the ABE-PPA data and this mammalian cell data suggest that eNme2-C-ABE8e is a

robust adenine base editor that provides general access to N_4CN PAMs.
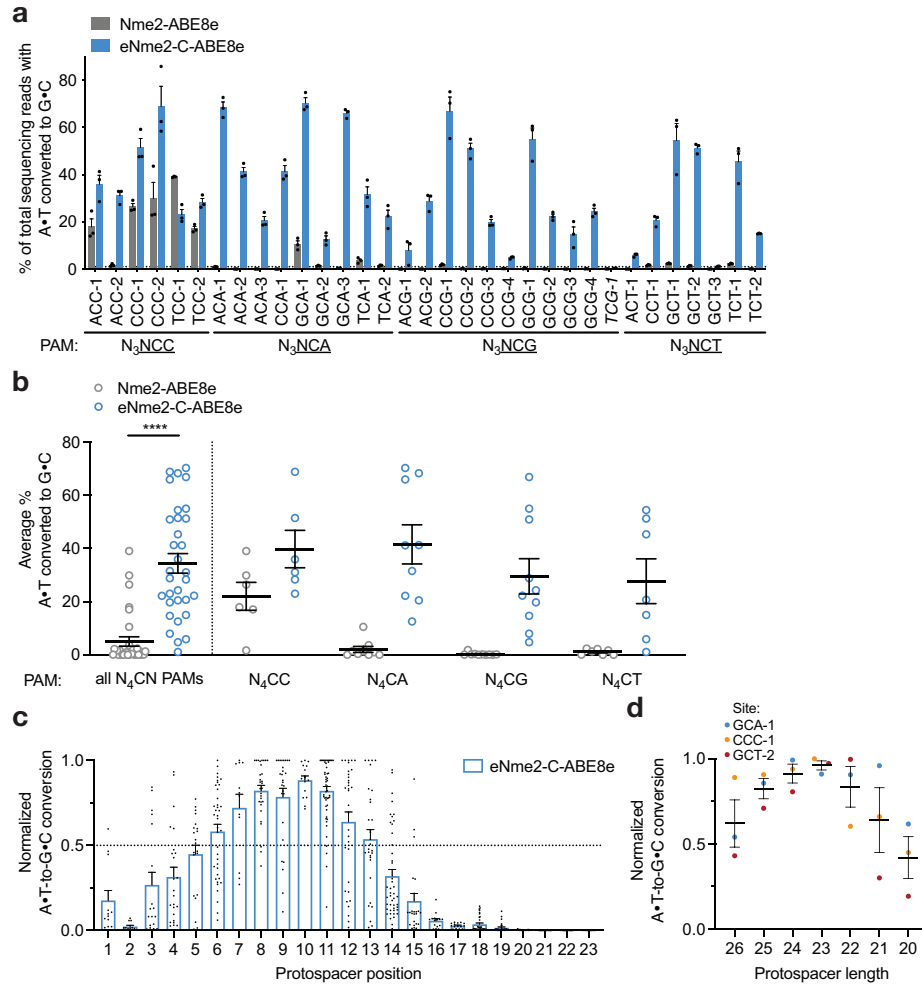
Figure 3.15. Activity characterization of eNme2-C-ABE8e in HEK293T cells.
(a) Adenine base editing activity of eNme2-C-ABE8e at 33 N$_3$NCN PAM-containing sites in HEK293T cells. Mean±SEM is shown and reflects the average activity and standard error of n = 3 replicates at the maximally edited position within each genomic site. The site that exhibited <1% base editing activity (line shown) that was excluded in subsequent analyses is italicized.
(b) Pooled adenine base editing activity of eNme2-C-ABE8e from (a). Each point represents the average editing of n = 3 independent biological replicates measured at a given genomic site. Mean±SEM is shown and reflects the average activity and standard error of the pooled genomic site averages. **(c)** Editing window of eNme2-C-ABE8e reflective of pooled adenine base editing activity at all 23 protospacer positions (PAM counted as positions 21-26) of the 32 sites shown in (a). Each point represents the % A•T-to-G•C conversion observed for an adenine that was present in one of the 32 protospacers, normalized to the highest editing observed within that protospacer. Mean±SEM is shown and reflects the average normalized activity and standard error at all observed adenines at that position. **(d)** Adenine base editing activity of eNme2-C-ABE8e as a function of protospacer length (between 26-20 nt) at three different genomic sites in HEK293T cells. Each point represents the average of n = 3 independent biological replicates observed for a given protospacer length at one genomic site, normalized to the protospacer length with the highest base editing activity for that site. Mean±SEM is shown and reflects the average normalized activity and standard error of the pooled averages at the observed sites. For (b), ****, $p \leq 0.0001$. p-value determined by unpaired Student's t-test.

3.2.9 *High stringency evolution of Nme2Cas9 towards N₄TN PAM sequences*

Following the success of the N₄CN trajectory using a high-stringency selection, we revisited the N₄TN trajectory using a similar approach. Starting with PANCE (**N2**), we attempted to evolve three different pools of MP6-diversified phage on each of the eight N₃YTN PAMs (**Figure 3.6**). Across eight PANCE passages, only lagoons seeded with ePACE3 endpoint phage propagated. These phage pools were subsequently seeded into ePACE (**ePACE5**). Under continuous evolution, these phage pools struggled to propagate, with phage washing out of many lagoons and only persisting with low titers ($\sim 10^5$ pfu/mL) at low flow rates (<1.5 vol/hr) among surviving lagoons. Phage clones were sequenced from each lagoon at a timepoint during which titers exceeded 10^5 pfu/mL. Most sequenced clones retained many of the strongly converged mutations from ePACE3, particularly in the non-PID region. However, in the PID, we observed intra-lagoon convergence at residue 1033 (which mediates the wild-type interaction with the PAM position 6 cytosine and previously converged to lysine in ePACE3) and residue 1049 (positioned proximal to the PAM) for lagoons evolved on the same PAM, but divergence across PAMs (R1033Y/E/N/H/T; R1049S/L/C), suggesting novel PAM-specific interactions at positions 4 or 6 made possible by the higher stringency selection (**Table 3.5**).

Table 3.5. Mutation table of ePACE5 evolved variants.
Genotypes of individually sequenced plaques following ePACE5, with positions varying from wild-type displayed. Clones evolved on different PAMs are delineated by a bold line. Mutations that had previously appeared in ePACE1, ePACE2, or ePACE3 are shown in light pink, magenta, or purple, respectively, while novel mutations are shown in green. Positions that were unable to be called due to low sequencing quality are denoted by a "-".

Using ABE-PPA, we observed that ePACE5 variants exhibited broad PAM compatibility (**Figure 3.16**), in contrast to ePACE4 variants which exhibited strong N_4CN-specific activity. While N_4TN activity was the most enriched, substantial adenine base editing activity was observed at all other PAMs, which could increase downstream Cas-dependent off-target editing. Two clones, E5-1, which we denote eNme2-T.1 (Nme2Cas9 E47K, V68M, T123A, D152G, E154K, T396A, H413N, A427S, H452R, E460A, A484T, S629P, N674S, D720A, V765A, H767Y, H771R, V821A, D844A, I859V, W865L, M951R, K1005R, D1028N, S1029A, R1033Y, R1049S, N1064S), and E5-40, which we denote eNme2-T.2 (Nme2Cas9 E47K, R63K, V68M, A116T, T123A, D152N, E154K, E221D, T396A, H452R, E460K, N674S, D720A, A724S, K769R, S816I, D844A, E932K, K940R, M951R, K1005R, D1028N, S1029A, R1033N, R1049C, L1075M), showed >70% average A•T-to-G•C editing across all N_4TN PAMs as ABE8e variants (**Figure 3.8**, **Figure 3.16**). As with the ePACE4 variants, many ePACE5 variants no longer exhibited a preference at PAM position 7 (e.g. eNme2-T.1, eNme2-T.2, **Figure 3.8**), further highlighting the benefit provided by the multiplexed-PAM selection scheme.
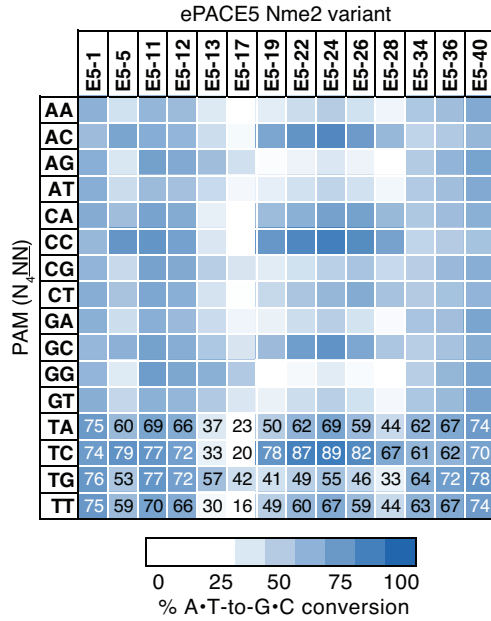
Figure 3.16. ABE-PPA activity of ePACE5 evolved variants.
Heat map showing ABE-PPA activity of representative clones from ePACE5 on the 16 combinations of PAM positions 5 and 6 (N4NN) Values are raw % A•T-to-G•C conversion observed for one replicate of each editor and are listed in each cell for the N4TN PAMs, with values above 70% A•T-to-G•C conversion colored white.
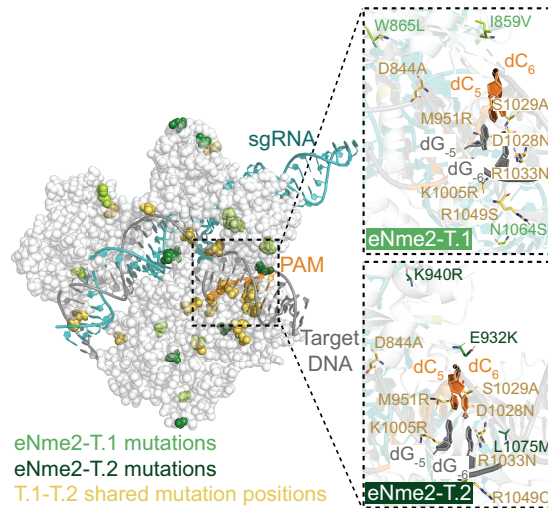
Figure 3.17. Crystal structure of Nme2Cas9 with mapped eNme2-T.1 and eNme2-T.2 mutations.
Mutation overview of the eNme2-T.1 and eNme2-T.2 variants, mapped onto the crystal structure of wild-type Nme2Cas9 (PDB: 6JE3), positions mutated in both variants are shown in yellow, while mutations unique to eNme2-T.1 are shown in light green and mutations unique to eNme2-T.2 are shown in dark green. The insets show the wild-type PAM and PAM-interacting residues (D1028, R1033), along with novel mutations listed.

71

We tested the eNme2-T.1 and eNme2-T.2 variants in HEK293T cells at the eight

endogenous human genomic N_4TN sites previously tested. At these eight sites, eNme2-T.1-

ABE8e and eNme2-T.2-ABE8e averaged 23% and 22% A•T-to-G•C editing, respectively,

representing a 278- and 264-fold improvement in activity over wild-type Nme2-ABE8e (**Figure**

3.18). After including eight additional genomic N_4TN sites, eNme2-T.1-ABE8e and eNme2-T.2-

ABE8e exhibited base editing efficiencies above 1% at 69% or 63% of the 16 total sites,

respectively. Within the sites showing >1% base editing, efficiencies ranged from 1.4-51% for

eNme2-T.1-ABE8e and from 1.4-50% for eNme2-T.2-ABE8e. Both variants appeared to have a

slightly 5' shifted base editing window compared to eNme2-C-ABE8e, between positions 7 and

12 of the protospacer (counting the PAM as positions 24-29), but showed similar protospacer

length preferences of 23 base pairs (**Figure 3.18**). This mammalian cell editing data suggests

that while capable of accessing many N_4TN PAMs, editing efficiencies supported by eNme2-

T.1-ABE8e and eNme2-T.1-ABE8e remain somewhat site-dependent. Nevertheless, together,

these evolved variants from both trajectories (eNme2-C, eNme2-T.1, and eNme2-T.2) enable

access to a large suite of pyrimidine-rich PAMs largely inaccessible to SpCas9-derived variants

while representing the first reported evolution of a non-*S. pyogenes* Cas protein towards single-

nucleotide PAM recognition.
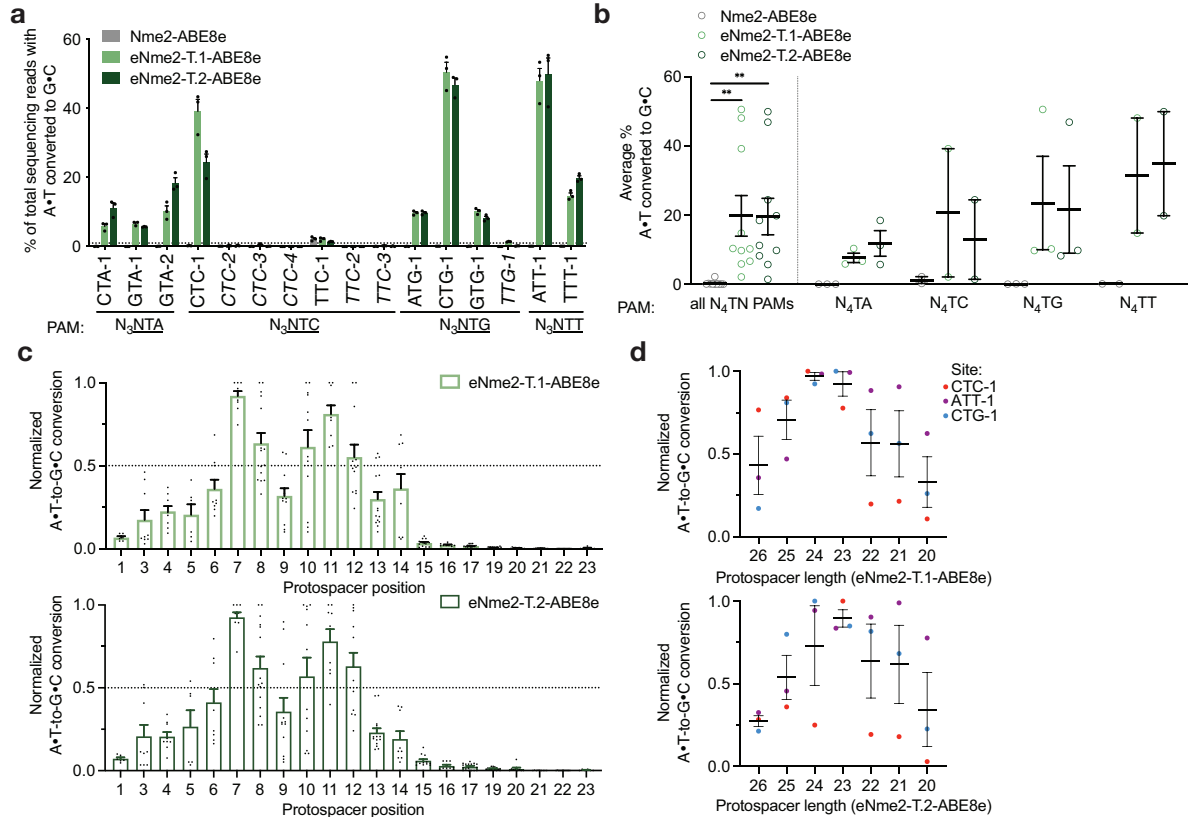
Figure 3.18. Activity characterization of eNme2-T.1-ABE8e and eNme2-T.2-ABE8e in HEK293T cells.
(a) Adenine base editing activity of eNme2-T.1-ABE8e and eNme2-T.2-ABE8e at 16 N$_3$<u>NTN</u> PAM-containing sites in HEK293T cells. Mean±SEM is shown and reflects the average activity and standard error of *n* = 3 independent biological replicates at the maximally edited position within each genomic site. The six sites that exhibited <1% base editing activity for either variant (line shown) that were excluded in subsequent analyses are italicized. **(b)** Pooled adenine base editing activity of eNme2-T.1-ABE8e and eNme2-T.2-ABE8e from (a). Each point represents the average of *n* = 3 independent biological replicates measured at the maximally edited position within each given genomic site. Mean±SEM is shown and reflects the average activity and standard error of the pooled genomic site averages. **(c)** Editing window of eNme2-T.1-ABE8e (top) or eNme2-T.2-ABE8e (bottom) reflective of pooled adenine base editing activity at all 23 protospacer positions (PAM counted as positions 24-29) of the 10 sites shown in (a). Each point represents the % A•T-to-G•C conversion observed for an adenine that was present in one of the 10 protospacers, normalized to the highest editing observed within that protospacer. Mean±SEM is shown and reflects the average normalized activity and standard error at all observed adenines at that position. **(d)** Adenine base editing activity of eNme2-T.1-ABE8e (top) or eNme2-T.2-ABE8e (bottom) as a function of protospacer length (between 26-20 nt) at three different genomic sites in HEK293T cells. Each point represents the average of *n* = 3 independent biological replicates observed for a given protospacer length at one genomic site, normalized to the protospacer length with the highest base editing activity for that site. Mean±SEM is shown and reflects the average normalized activity and standard error of the pooled averages at the observed sites. For (b), **, p ≤ 0.01. p-values determined by individual

73

"Figure 3.18 (Continued)" unpaired Student's t-tests comparing Nme2-ABE8e to either eNme2-T.1-ABE8e or eNme2-T.2-ABE8e.

3.2.10 Comparison of eNme2 and SpRY base editors and nucleases

Next, we compared the editing performance of evolved eNme2 variants with that of alternative Cas variants. No natural Cas variants capable of targeting single pyrimidine PAMs have been reported[66]. Among engineered Cas variants, only SpRY has shown activity on some NCN and NTN PAMs[98]. We selected PAM-matched genomic sites to directly compare the base editing activities of SpRY and eNme2 variants (**Figure 3.19**). At 14 matched C-containing PAM sites in HEK293T cells, eNme2-C-ABE8e showed a marked improvement in adenine base editing over SpRY, averaging 47% vs. 23% A•T-to-G•C editing. This difference is more pronounced (47% vs. 15% A•T-to-G•C editing) when compared to the ABE8e version of high-fidelity SpRY, SpRY-HF1-ABE8e (**Figure 3.19**). In contrast, at eight matched T-containing PAM sites in HEK293T cells, eNme2-T.1-ABE8e and eNme2-T.2-ABE8e are less active than either SpRY-ABE8e or SpRY-HF1-ABE8e (23% and 22% for eNme2-T.1-ABE8e and eNme2-T.2-ABE8e versus 35% and 38% for SpRY-ABE8e or SpRY-HF1-ABE8e, respectively) (**Figure 3.19**). These data indicate that eNme2-C offers a best-in-class option for modifying C-containing PAM sites, while eNme2-T.1 and eNme2-T.2 provide new options for targeting some T-containing PAMs together with the existing SpRY variants.
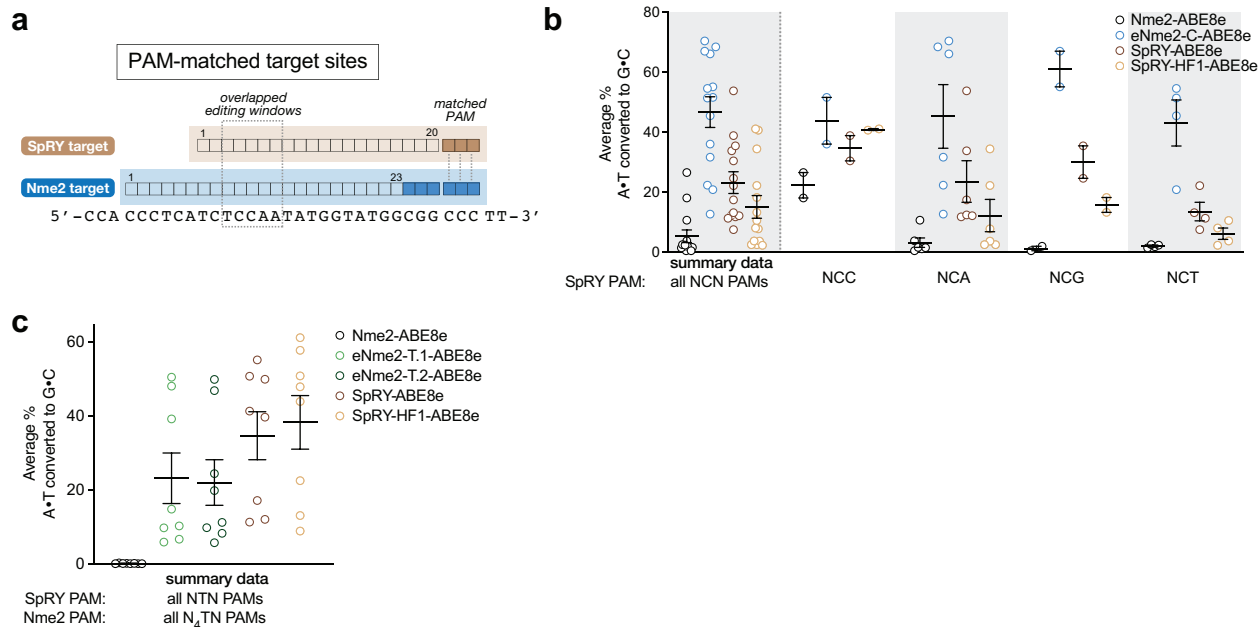
Figure 3.19. Comparison of eNme2-C-ABE8e, eNme2-T.1-ABE8e, and eNme2-T.1-ABE8e to SpRY-ABE8e and SpRY-HF1-ABE8e at PAM-matched sites.
(a) Overview of PAM-matched sites used to compare eNme2Cas9 variants to SpRY and SpRY-HF1. **(b)** Summary dot plots showing the activity of eNme2-C-ABE8e compared to SpRY-ABE8e and SpRY-HF1-ABE8e at 14 PAM-matched NCN/N4CN sites in HEK293T cells. Left-most data represent a summary of all 14 sites, and subsequent columns represent a subdivision into specific PAMs. **(c)** Summary dot plots showing the activity of eNme2-T.1-ABE8e and eNme2-T.2-ABE8e compared to SpRY-ABE8e and SpRY-HF1-ABE8e at eight PAM-matched NTN/N4TN sites in HEK293T cells. For (**b,c**), each point represents the average editing of $n = 3$ independent biological replicates measured at the maximally edited position within each given genomic site. Mean±SEM is shown and reflects the average activity and standard error of the pooled genomic site averages.

We then tested if the improvements to Nme2Cas9 were generalizable to other Cas9-dependent editing modalities. At six PAM-matched target sites in HEK293T cells, eNme2-C-BE4 exhibited an average of 28% C•G-to-T•A editing, a 3.2- and 4.8-fold improvement over SpRY-BE4 and SpRY-HF1-BE4, respectively (**Fig. 3d, Extended Data Fig. 7c**). Although less efficient than eNme2-C-ABE8e, eNme2-C-BE4 is capable of C•G-to-T•A editing at levels comparable to (within 2-fold of) those reported for SpCas9 or SpCas9-derived CBE variants at their canonical purine-containing PAMs[32,52,84,97,98].
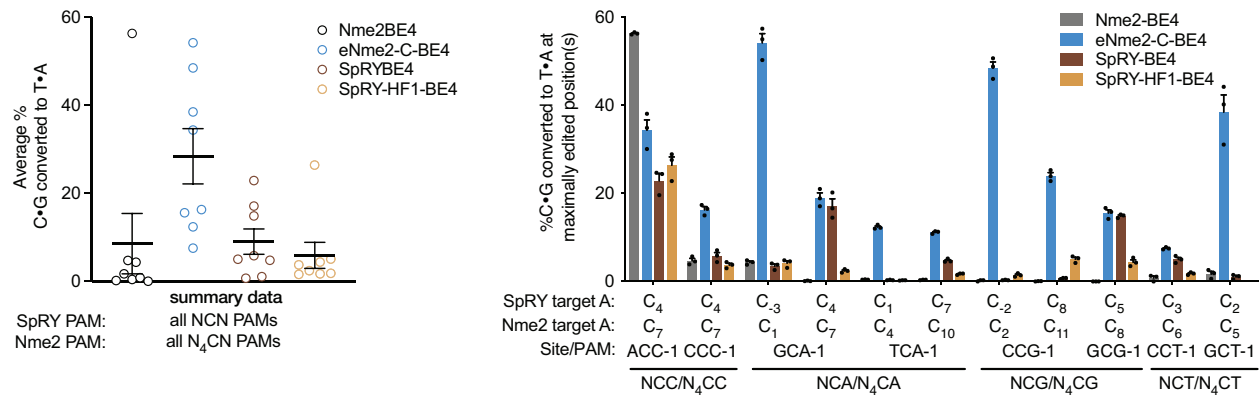
Figure 3.20. Comparison of eNme2-C-BE4 to SpRY-BE4 and SpRY-BE4 at PAM-matched sites.
(a) Summary dot plots showing the activity of eNme2-C-BE4 compared to SpRY-BE4 and SpRY-HF1-BE4 at eight PAM-matched NCN/N$_4$CN sites in HEK293T cells. Each point represents the average editing of n = 3 independent biological replicates measured at the maximally edited position within each given genomic site. Mean±SEM is shown and reflects the average activity and standard error of the pooled genomic site averages. **(b)** Cytosine base editing activity of eNme2-C-BE4 compared to SpRY-BE4 and SpRY-HF1-BE4 at eight NCN/N$_4$CN PAM-matched sites in HEK293T cells. Mean±SEM is shown and reflects the average activity and standard error of n = 3 independent biological replicates measured at the maximally edited position (if applicable) within each given genomic site.

Surprisingly, when the RuvC-inactivating mutation D16A[103] was reverted, eNme2-C nuclease was inefficient at generating indels in mammalian cell culture, averaging only 2.1% indels at eight N$_4$CN PAM sites (**Figure 3.21**). We hypothesized that this was due to the large number of mutations in the RuvC and HNH domains of eNme2-C, some of which could be nuclease-inactivating. Indeed, when we reverted all mutations in the nuclease and associated linker domains, the resulting variant, eNme2-C.NR (eNme2-C S6P, G33E, A520E, S646R, V696F, R711G, V758I, Y767H) had restored nuclease activity while retaining novel N$_4$CN PAM activity (average 34% indels across the same eight sites). However, reversion of these mutations had a negative impact on ABE activity, with eNme2-C.NR-ABE8e exhibiting 1.8-fold reduced A•T-to-G•C conversion compared to eNme2-C-ABE8e (**Figure 3.21**). These results suggest that some or all the mutations in the RuvC/HNH domains are important for robust base

editing of the eNme2-C variant, but the same mutations, if present, are detrimental to the subsequent activation or catalytic activity of eNme2-C.NR nuclease.

To further explore this idea and to potentially find an optimal dual base editor/nuclease variant, we generated the set of eight single-point reversion variants of mutations in the RuvC/HNH domains of eNme2-C and evaluated them as nucleases and ABEs (**Figure 3.21**). Only 2 of the 8 single-point reversion variants, eNme2-C V696F and eNme2-C R711G, showed significant rescue of nuclease activity (12.5- and 4.4-fold improvement over eNme2-C, respectively). Conversely, most of the reversions reduced ABE efficiency relative to eNme2-C-A8e. Notably, none of the eight variants outperformed eNme2-C as an ABE or eNme2-C.NR as a nuclease, highlighting the importance of the amino acid identities at these RuvC/HNH positions in differentiating between base editor and nuclease activities of evolved Nme2Cas9.
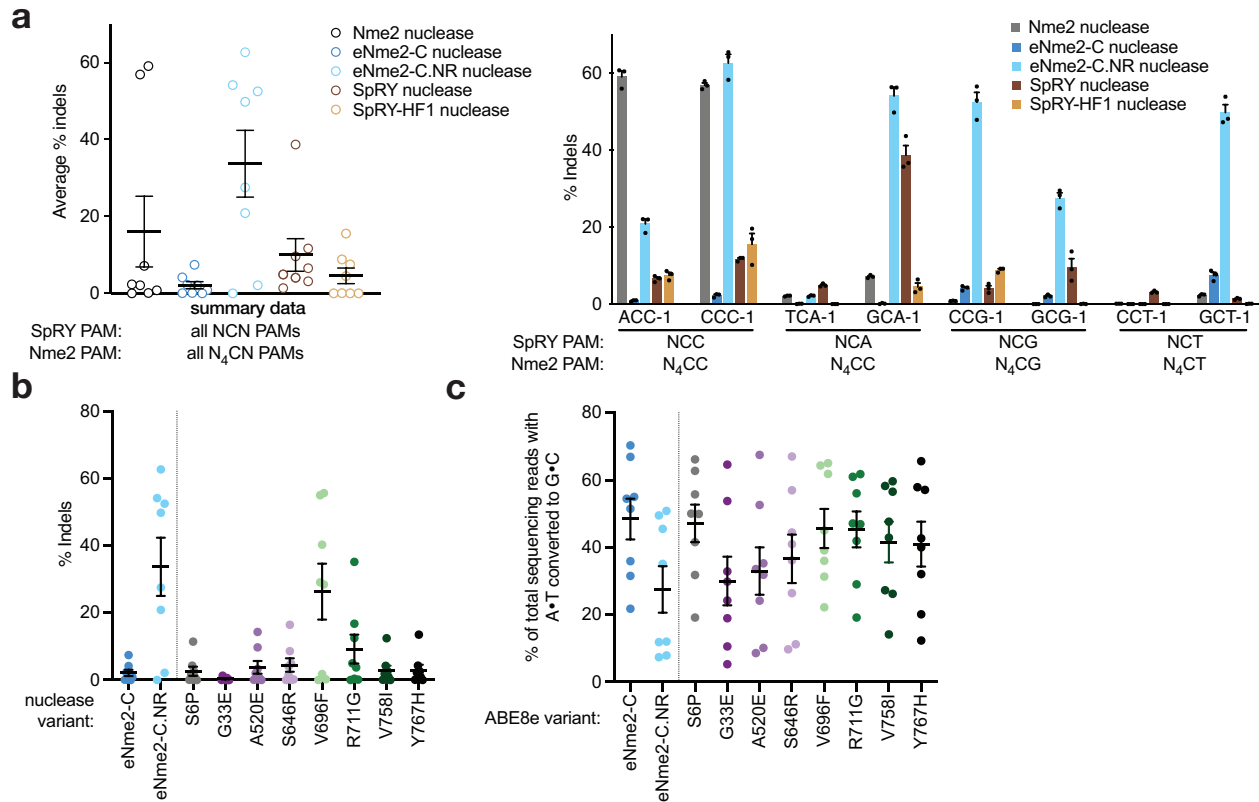
Figure 3.21. Development of a nuclease-active eNme2-C variant, eNme2-C.NR, and comparison to SpRY and SpRY-HF1 nucleases.
(a) Summary dot plot (left) or site-level bar graphs (right) showing the activity of eNme2-C nuclease and eNme2-C.NR nuclease compared to SpRY nuclease and SpRY-HF1 nuclease at eight PAM-matched NCN/N$_4$<u>CN</u> sites in HEK293T cells. (Left) each point represents the average editing of n = 3 independent biological replicates measured at the maximally edited position within each given genomic site. Mean±SEM is shown and reflects the average activity and standard error of the pooled genomic site averages. (Right) Mean±SEM is shown and reflects the average activity and standard error of n = 3 independent biological replicates measured at the maximally edited position (if applicable) within each given genomic site. **(b)** Pooled adenine base editing activity of eNme2-C-ABE8e compared to eNme2-C.NR-ABE8e or adenine base editors generated from reversion mutations at each of the eight RuvC/HNH domain mutations in eNme2-C at eight genomic sites in HEK293T cells. **(c)** Pooled nuclease activity of eNme2-C nuclease compared to eNme2-C.NR nuclease or nuclease-active variants generated from reversion mutations at each of the eight RuvC/HNH domain mutations in eNme2-C at eight genomic sites in HEK293T cells. For (**b,c**), each point represents the average of n = 3 independent biological replicates measured at the maximally edited position within each given genomic site in HEK293T cells. Mean±SEM is shown and reflects the average activity and standard error of the pooled genomic site averages.

Having established two distinct sub-variants of eNme2-C for either base editing or DNA cleavage, we next compared eNme2-C.NR nuclease to SpRY and SpRY-HF1 nucleases. Surprisingly, both SpRY and SpRY-HF1 nucleases were relatively inefficient at the NCN PAM-matched sites tested, being significantly outperformed by eNme2-C.NR nuclease (3.4- and 7.3-fold more efficient editing by eNme2-C.NR nuclease, respectively) (**Figure 3.21**). Given this data, we speculate that perhaps some mutations in SpRY, like with eNme2-C, may asymmetrically affect base editing versus nuclease activities (for instance sufficient R-loop formation for base editing but slow conformational shift for nuclease activation[128,129]). This hypothesis would also potentially explain why the activity observed for SpRY-ABE8e appears to be much more generalizable at NYN PAMs than what would be expected given the limited NYN PAM scope initially described for SpRY nuclease[98]. Together, these data highlight eNme2-C base editors and eNme2-C.NR nucleases as highly effective variants for genome editing, offering promising alternatives to SpRY and SpRY-HF1 in applications requiring access to C-containing PAMs.

3.2.11 Off-target analysis reveals high genome-wide specificity of eNme2-C variants

PAM-broadened Cas variants have been shown to increase off-target activity due to the increased number of sequences recognized as a PAM[84,97,98]. While this off-target activity can be compensated for by introducing high-fidelity mutations that increase protospacer-target binding fidelity[83,98], these mutations can sometimes result in a reduction in overall Cas activity (**Figure 3.19, Figure 3.20, Figure 3.21,** comparing SpRY to SpRY-HF1 variants). Nme2Cas9 has been shown to be highly accurate, exhibiting very few if any off-targets compared to SpCas9 at protospacer-matched sites[103]. We hypothesized that eNme2-C would be more specific than PAM-broadened SpCas9 variants. This higher specificity is potentially due to the longer protospacer requirement of Nme2Cas9 (22-23 nt[103] versus 20 nt), which naturally increases the

79

total possible sequence space and decreases the likelihood of finding perfectly or near-perfectly

($\leq$ 3 mismatches) matched sites (**Figure 3.22**).

a

b

Site	20 nt protospacer (SpCas9)				23 nt protospacer (Nme2Cas9)			
	0 MM*	1 MM	2 MM	3 MM	0 MM*	1 MM	2 MM	3 MM
ACC-1	0	0	3	111	0	0	0	0
GCT-1	0	0	14	271	0	0	0	6
CCC-1	4	11	149	1612	0	6	6	70
CCG-1	0	1	13	168	0	0	0	12
GCA-1	0	1	25	426	0	0	0	10
GCG-1	0	0	3	70	0	0	0	1
Average	0.7	2.2	34.5	443.0	0.0	1.0	1.0	16.5

*0 MM tally excludes the on-target site
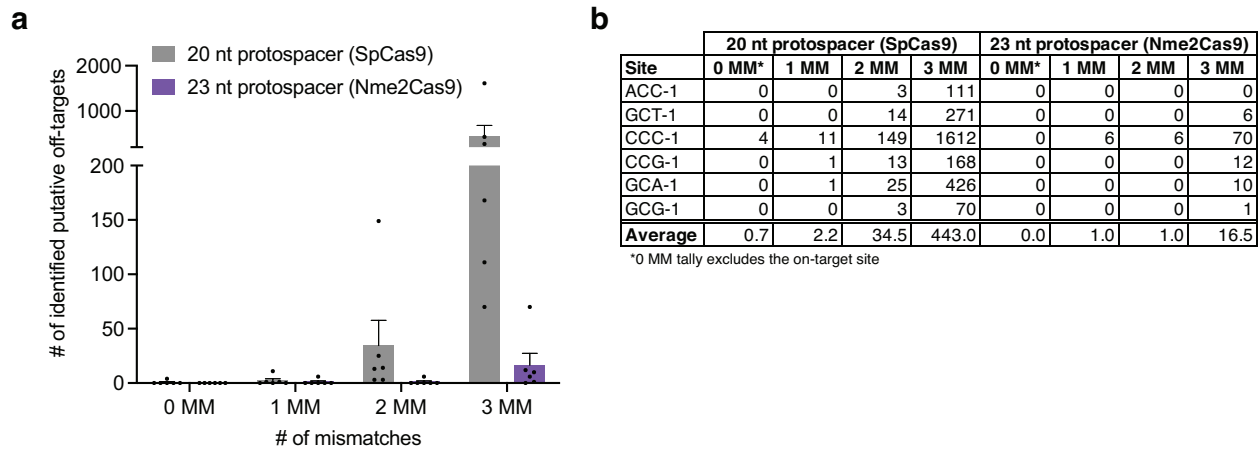
Figure 3.22. *In silico* predicted number of off-targets for a 23- vs. 20-nt protospacer.
(a) Count of genome-wide (GRCh38) sites with 0, 1, 2, or 3 mismatches to a 20-nt (SpCas9) or 23-nt (Nme2Cas9) protospacer identified with CHOPCHOPv3[130]. Mean±SEM representing identified off-targets at six randomly selected 20-nt or 23-nt protospacers are shown. **(b)** Table listing the number of identified sites with the corresponding number of mismatches to a 20-nt or 23-nt protospacer at six randomly selected genomic sites

To evaluate off-target activity, we first selected two protospacer-matched sites (Site 1 and Site 2) with validated nuclease and ABE activities for eNme2-C/eNme2-C.NR and SpRY variants (**Figure 3.23**). Using CHOPCHOPv3[130], we used *in silico* prediction to identify the set of potential off-target sites with $\leq$ 2 mismatches and no more than one PAM proximal (within 10 bp of the PAM) mismatch to at least one of the two protospacers (23 nt for Nme2Cas9, 20 nt for SpRY). We then evaluated off-target nuclease and ABE8e activities at all identified off-target sites (seven for Site 1, twelve for Site 2) using targeted amplicon sequencing.
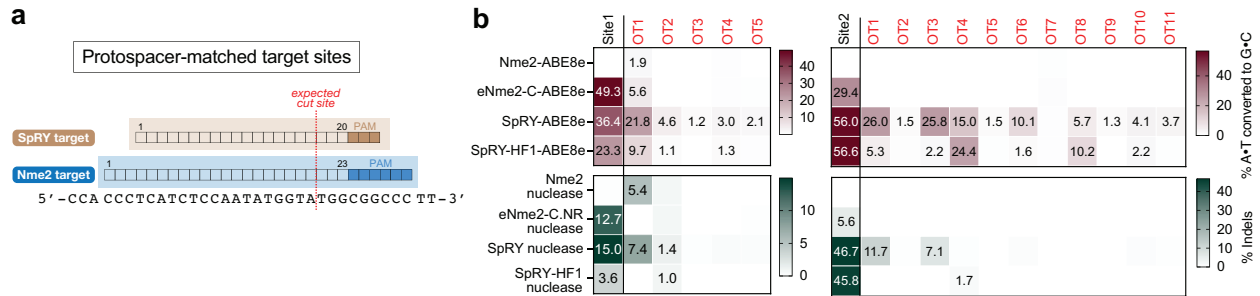
Figure 3.23. Off-target activities of eNme2-C, SpRY, and SpRY-HF1 at *in silico* predicted off-targets of protospacer-matched sites.
(a) Overview of protospacer-matched sites used to compare the DNA specificity of eNme2Cas9 variants against SpRY and SpRY-HF1. **(b)** Heat maps showing off-target adenine base editing activity (brown) or off-target indel formation (dark green) at computationally-determined off-targets for two sites in HEK293T cells for eNme2-C-ABE8e and eNme2-C.NR nuclease compared to SpRY and SpRY-HF1 adenine base editor and nuclease variants. The left-most column represents on-target activity. Values are listed for any sites at which ≥1% editing or indels was observed, and represent the average of *n* = 3 independent biological replicates.

For the Site 1 protospacer, five of the seven predicted sites sequenced well, and eNme2-C-ABE8e showed off-target base editing >1% at one of these five sequenced off-target sites, while eNme2-C.NR did not generate any off-target indels >1% (**Figure 3.23**). In contrast, SpRY-ABE8e and SpRY-HF1-ABE8e exhibited off-target base editing >1% at all five or four of five sites, respectively, despite having lower on-target efficiency than eNme2-C-ABE8e. As nucleases, SpRY and SpRY-HF1 showed higher fidelity, with only two of five or one of five off-target site(s) exhibiting indels >1%, respectively. Similar trends were observed for the Site 2 protospacer. No off-target base editing or indel formation >1% was observed at any of the twelve sequenced off-target sites for eNme2-C-ABE8e or eNme2-C.NR, whereas off-target base editing and indel formation >1% was observed at many sites for SpRY and SpRY-HF1. These data suggest that eNme2-C-ABE8e and eNme2-C.NR retain the high natural specificity of Nme2Cas9 and offer greater specificity than their SpRY and SpRY-HF1 counterparts, particularly for precision applications such as base editing.

To perform a more unbiased, genome-wide survey of potential off-targets, we used GUIDE-seq[95] to evaluate double-strand breaks generated by eNme2-C.NR compared to SpRY

variants at four protospacer-matched sites. Targeted sequencing of the on-target sites in treated

U2OS cells showed robust indel formation at all four sites for both SpRY nuclease and eNme2-

C.NR (30% and 40% indels for SpRY nuclease and eNme2-C.NR nuclease, respectively).

Surprisingly, despite 3 of the 4 sites containing NRN-PAMs, SpRY-HF1 nuclease only

generated >10% indels at the fourth site containing an NCN PAM. We also included the

nuclease-active version of eNme2-C, although as expected indel formation was inefficient

(<10%) at all but one site (**Figure 3.24**)**.** Across all four sites, eNme2-C.NR exhibited high

specificity, averaging 52-to-1 on-to-off-target reads, compared to SpRY which averaged a 1.2-

to-1 on-to-off-target ratio (**Figure 3.24**). These specificity values corresponded to a range of 7 to

22 putative off-target sites for eNme2-C.NR versus 14 to 591 putative off-target sites for SpRY

(**Figure 3.24, Figure 3.25, Figure 3.26, Figure 3.27, Figure 3.28**). At the site on which it was

active, eNme2-C similarly exhibited minimal off-target activity. In contrast, while SpRY-HF1

exhibited higher specificity than SpRY at the site on which it was active (Site 3), it still induced

substantial off-target editing compared to eNme2-C.NR (**Figure 3.24**). Together, these results

indicate that eNme2-C.NR and eNme2-C, afford improved genome-wide specificity relative to

SpRY variants
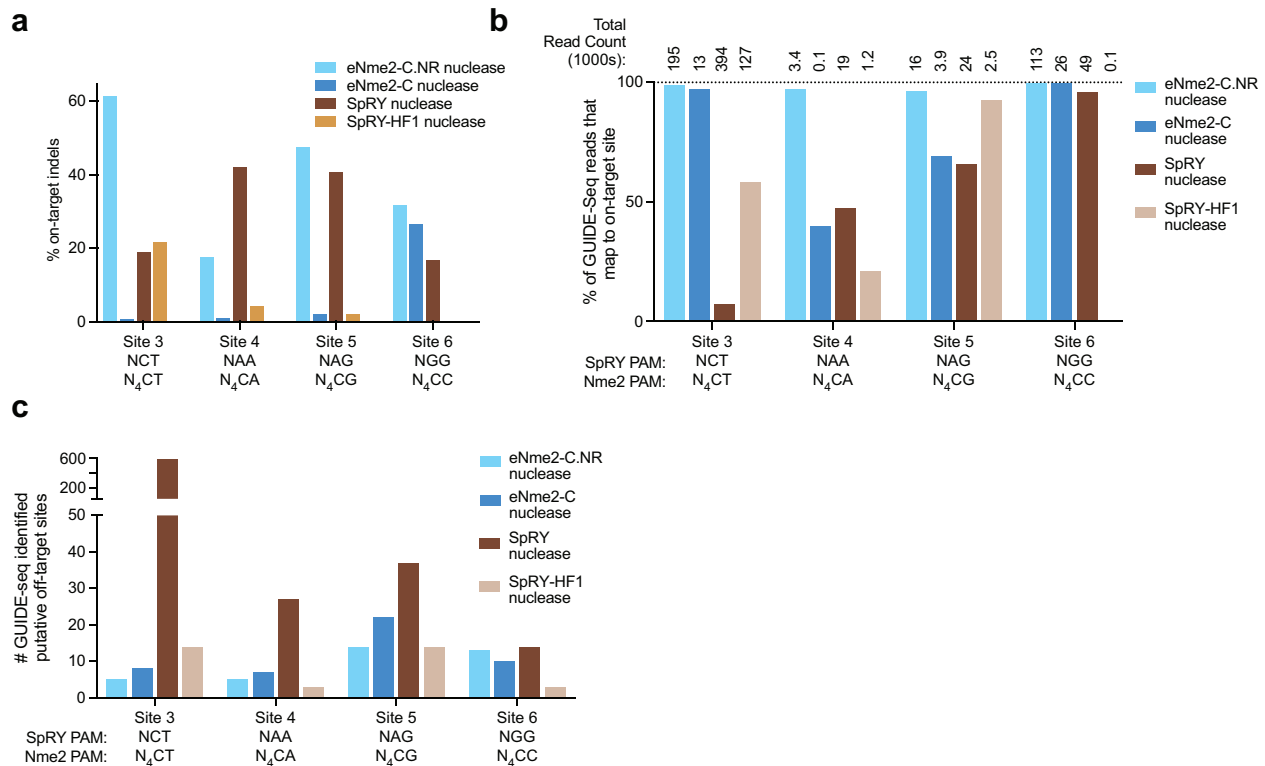
Figure 3.24. Unbiased genome-wide off-target comparison of eNme2-C, eNme2-C.NR, SpRY, and SpRY-HF1 nucleases using GUIDE-Seq.
(a) On-target indel formation of wild-type Nme2 nuclease, eNme2-C nuclease, and eNme2-C.NR nuclease compared to SpRY nuclease and SpRY-HF1 nuclease at each of the four protospacer-matched sites that were subsequently evaluated in GUIDE-Seq. Each bar represents the observed indel formation of one replicate in U2OS cells. **(b)** Percentage of on-target GUIDE-seq reads identified at four protospacer matched sites for eNme2-C nuclease, eNme2-C.NR nuclease, SpRY nuclease, and SpRY-HF1 nuclease. Total reads for the given nuclease are listed above each bar. **(c)** Total putative off-target sites identified by GUIDE-seq for eNme2-C nuclease, eNme2-C.NR nuclease, SpRY nuclease, and SpRY-HF1 nuclease at four protospacer-matched sites.

83
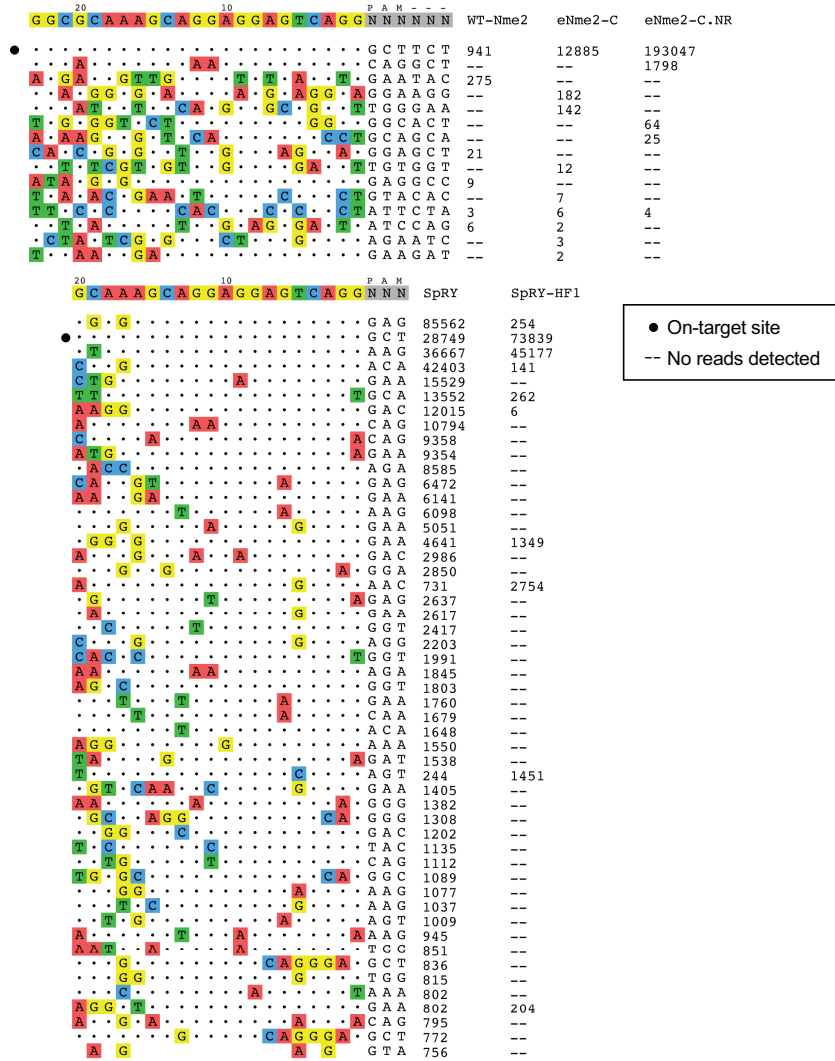
Figure 3.25. GUIDE-Seq identified off-targets and associated read counts for Nme2 variants or SpRY variants at Site 3.

84
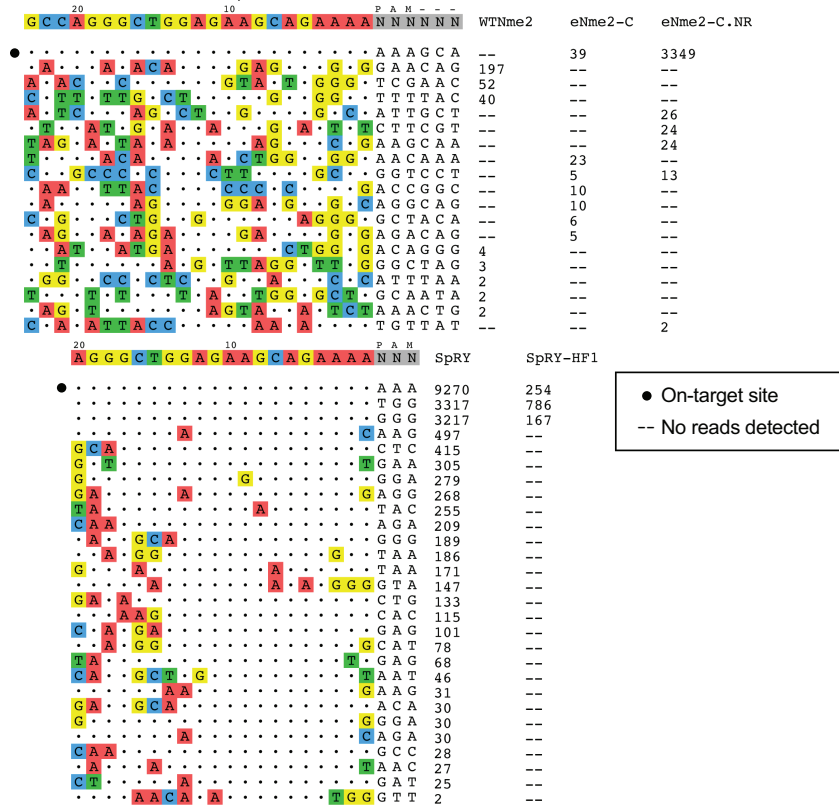
Figure 3.26. GUIDE-Seq identified off-targets and associated read counts for Nme2 variants or SpRY variants at Site 4.
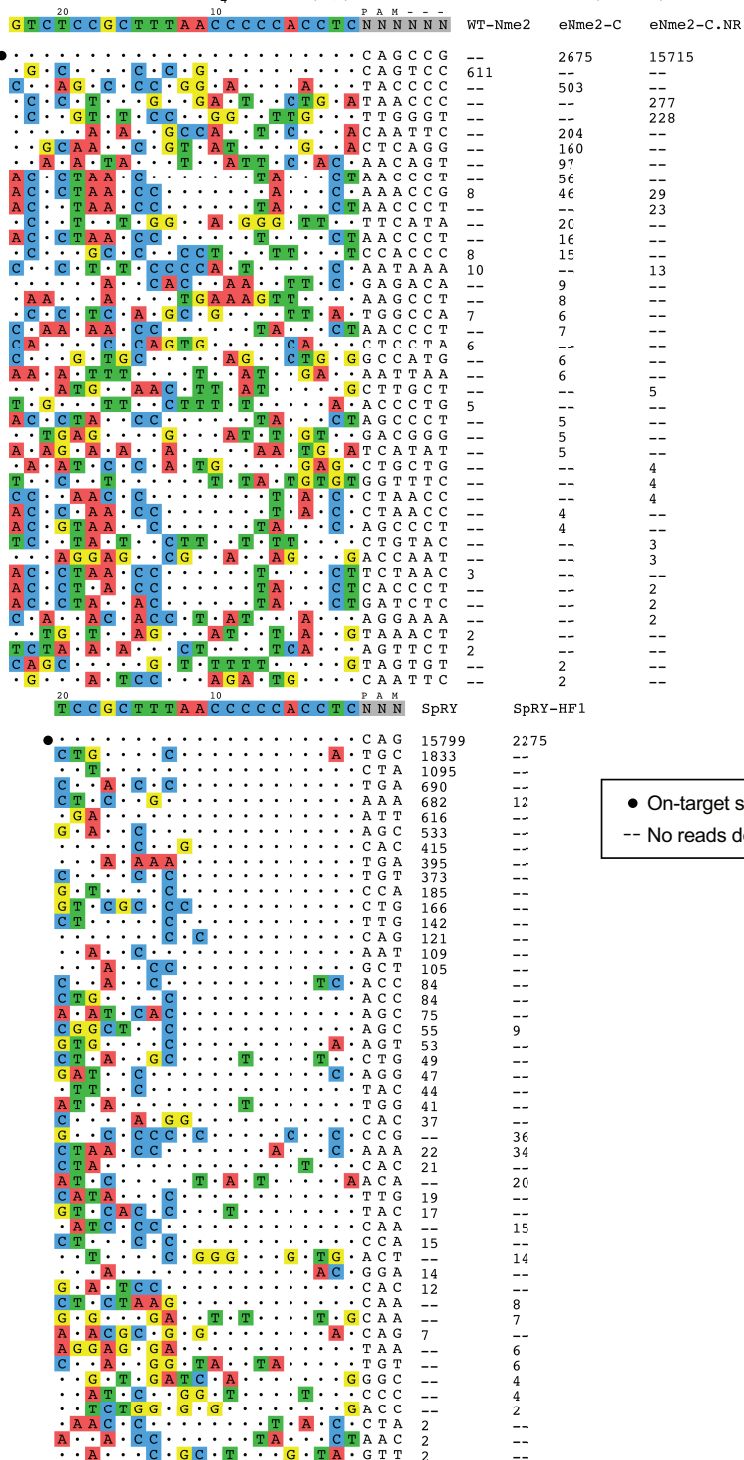
Figure 3.27. GUIDE-Seq identified off-targets and associated read counts for Nme2 variants or SpRY variants at Site 5.
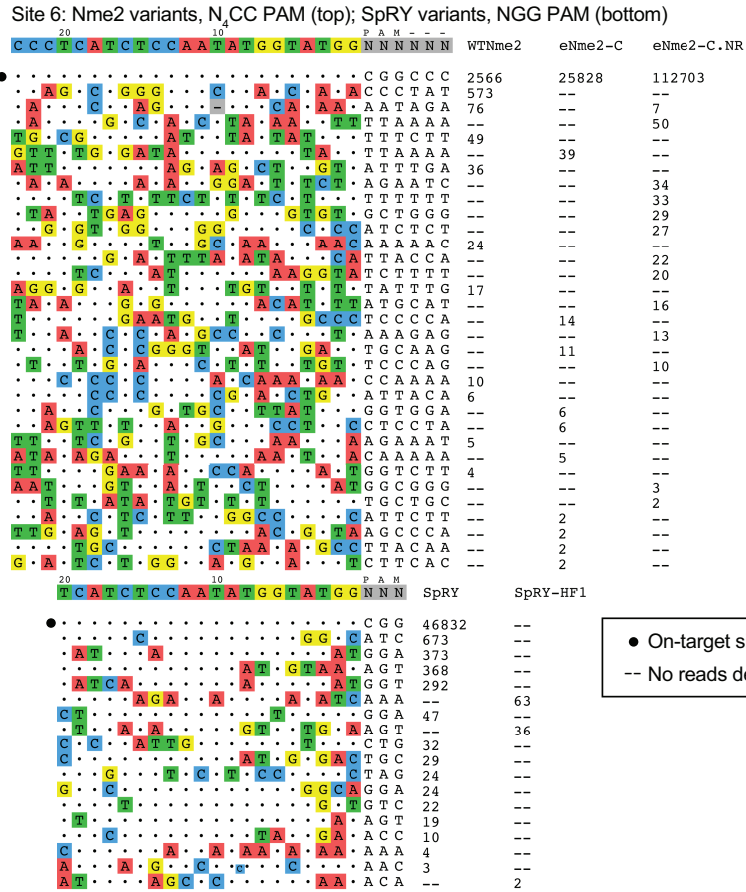
Figure 3.28. GUIDE-Seq identified off-targets and associated read counts for Nme2 variants or SpRY variants at Site 6.

87

3.2.12 eNme2-C is active in multiple mammalian cell types and enables access to new targets

Having validated the high-efficacy and specificity of eNme2-C at target sites containing N$_4$CN PAMs, we next demonstrated its generalizability in multiple cell types. In an immortalized hepatocyte cell line, HUH7, eNme2-C-ABE8e retains its broad base editing activity across sites containing N$_4$CN PAMs, accessing all 15 sites tested with an average of 37% A•T-to-G•C base editing (**Figure 3.29**). Similarly, at 18 sites in U2OS cells, adenine base editing activity was seen at all sites, albeit at lower average efficiency (averaging 16% A•T-to-G•C editing) (**Figure 3.29**). In both cell types, eNme2-C-ABE8e outperforms SpRY-ABE8e and SpRY-HF1-ABE8e, although the extent varies. Finally, we nucleofected primary human dermal fibroblasts with eNme2-C-ABE8e mRNA, achieving 64% A•T-to-G•C base editing across seven endogenous sites (**Figure 3.29**). Notably, eNme2-C-ABE8e, SpRY-ABE8e, and SpRY-HF1-ABE8e appeared to perform equally well in this cell line with nucleofection, potentially due to the high efficacy of mRNA nucleofection[30,97]. Together, these data demonstrate that eNme2-C is a broadly applicable Cas protein enabling precision genome editing in multiple biologically relevant cell types.
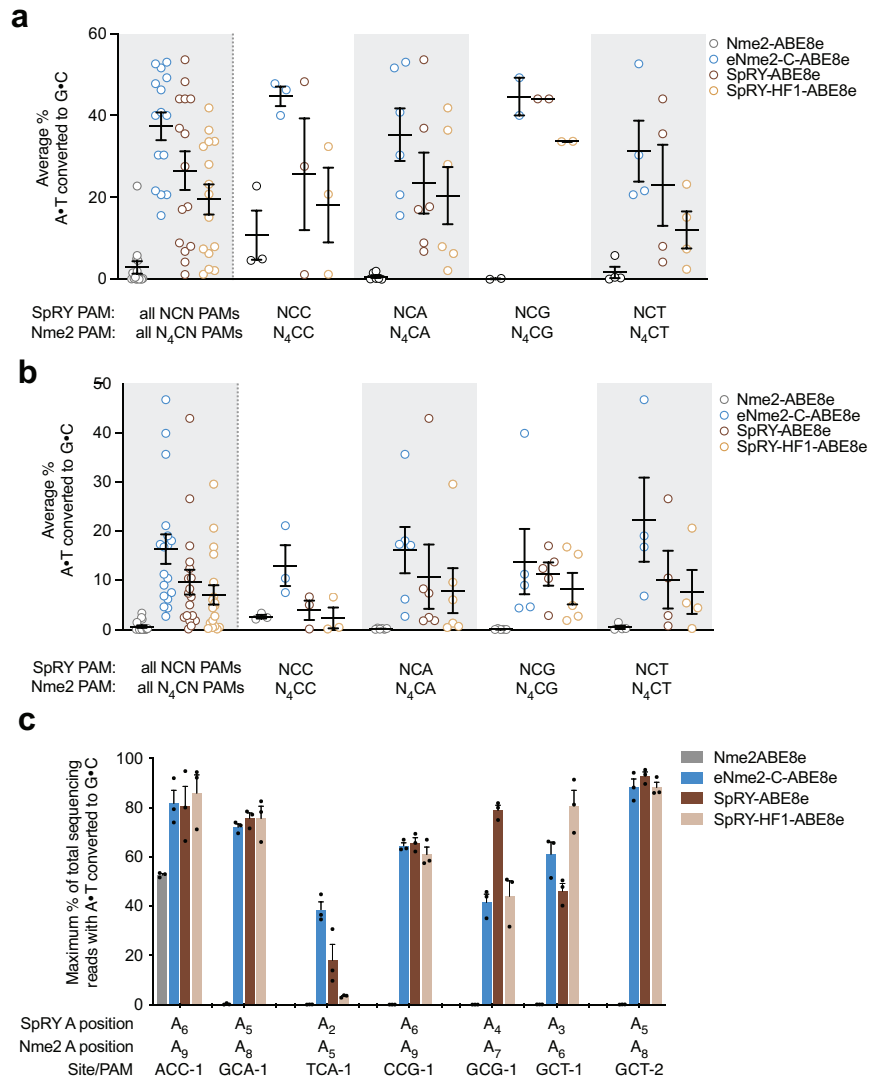
Figure 3.29. Comparison of eNme2-C-ABE8e to SpRY-ABE8e and SpRY-HF1-ABE8e in other mammalian cell types.
(a) Summary dot plots showing the activity of eNme2-C-ABE8e compared to SpRY-ABE8e and SpRY-HF1-ABE8e at 15 PAM-matched NCN/N₄CN sites in HUH7 cells. Left-most data represent a summary of all 15 sites, and subsequent columns represent a subdivision into specific PAMs. **(b)** Summary dot plots showing the activity of eNme2-C-ABE8e compared to SpRY-ABE8e and SpRY-HF1-ABE8e at 18 PAM-matched NCN/N₄CN sites in U2OS cells. Left-most data represent a summary of all 18 sites, and subsequent columns represent a subdivision into specific PAMs. For **(a,b)**, each point represents the average editing of *n* = 3 independent biological replicates measured at the maximally edited position within each given genomic site. Mean±SEM is shown and reflects the average activity and standard error of the pooled genomic site averages. **(c)** eNme2-C-ABE8e compared to SpRY-ABE8e and SpRY-HF1-ABE8e at eight PAM-matched NCN/N₄CN sites in HDFa cells. Bars represent mean±SEM of *n* = 3 independent biological replicates, with individual values shown as dots.

89

Because of its N_4CN PAM activity, eNme2-C is in theory perfectly complementary to single-G recognizing SpCas9 variants SpCas9-NG[84] and SpG[98], which are estimated to enable potential cleavage every ~2.2 bp in the human coding sequence[84]. As a cytosine or adenine base editor, eNme2-C enables access to 86% and 87% of pathogenic transition SNPs, respectively, recognized in the ClinVar database (**Figure 3.30**)[42,58]. Although SpRY base editors should access similar PAMs due to its near-PAMless nature, we hypothesized that differences in editing windows and specific PAM compatibilities would enable eNme2-C base editors to not only serve as higher-fidelity alternatives to SpRY base editors, but also facilitate access to new targets.
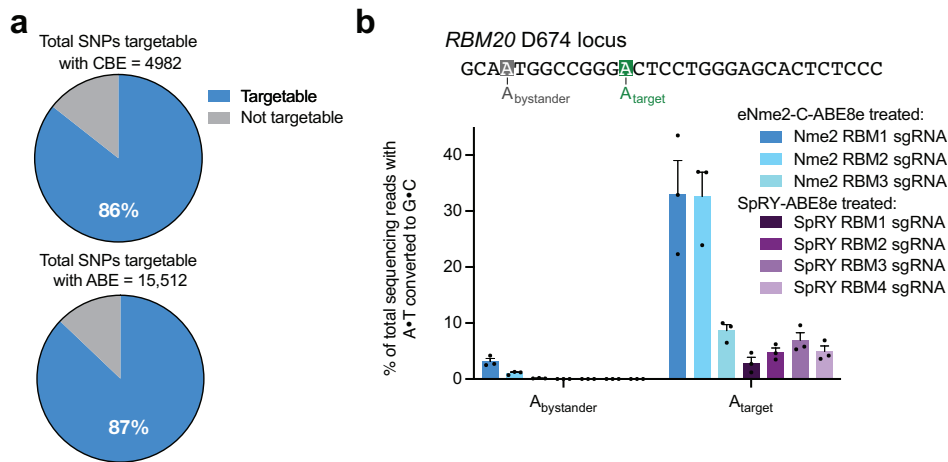
Figure 3.30. Disease-relevant SNPs targetable by eNme2-C base editors.
(a) ClinVar identified SNPs that can be targeted with an eNme2-C-ABE8e (top) or eNme2-C-BE4 (bottom). **(b)** Installation of a disease-relevant D674G mutation in the RBM20 gene. Tiled guides were used to install the mutation either with eNme2-C-ABE8e or SpRY-ABE8e. Bars represent mean±SEM *of n* = 3 independent biological replicates, with individual values shown as dots.

RBM20 is a gene encoding a *trans*-activating splicing factor, and mutations in the gene have been observed in 2-3% of familial dilated cardiomyopathy cases[131]. While many mutations have been identified in the coding sequence of *RBM20*, the individual effect of these mutations have not been well characterized, potentially due to the difficulty of installing some of these mutations in isolation. We used eNme2-C-ABE8e to install the D674G mutation, an A•T-to-G•C

transition in which the target base is upstream of a stretch of pyrimidine bases inaccessible to most characterized Cas variants. All three eNme2-C-ABE8e guides tested enabled editing of the target adenine, with the optimal guide reaching 33% A•T-to-G•C base editing. In contrast, none of the four SpRY guides placing the target adenine in the optimal editing window of SpRY (positions 4-7)[30] were able to achieve >10% A•T-to-G•C conversion (**Figure 3.30**). This data demonstrates that eNme2-C not only provides an efficient, high-fidelity alternative to SpRY in a variety of biological systems, but also enables the study and potential correction of previously inaccessible pathogenic SNPs.

3.3 Conclusions and outlook

By integrating a novel functional Cas enzyme selection (SAC-PACE) with high-throughput phage-assisted evolution platforms (PANCE & ePACE) and a high-throughput PAM profiling method (BE-PPA) to guide our evolutionary campaign, we demonstrated the first evolution of a non-*S. pyogenes* Cas protein to acquire single-nucleotide PAM recognition. We developed two highly efficient, highly specific Nme2Cas9 variants capable of targeting N_4CN PAM sequences across different gene editing modalities and two variants capable of adenine base editing at many N_4TN PAM sequences, affording unparalleled access to pyrimidine-PAM sequences. Together, these variants complement the suite of commonly used SpCas variants and will enable the study and potential correction of previously inaccessible or poorly accessible loci, while retaining the compact size and high genome-wide specificity of Nme2Cas9 that could be beneficial to downstream clinical applications.

In contrast to prior Cas9 evolutions which selected for novel PAM binding[77,97], SAC-PACE requires both novel PAM binding and subsequent activation steps necessary for base editing, increasing the likelihood of evolving desired editing properties. In addition to developing this new selection, we found that improvements analogous to those made to evolve high-activity SpCas9 variants could be easily incorporated into SAC-PACE, including limiting the concentration of active base editor through a split-intein system and requiring multiple editing events through the inclusion of additional base editing sites. Notably, the evolution campaign that resulted in eNme2-C generated substantially improved activity on N_4CC PAMs, the PAMs recognized by the wild-type protein, along with numerous mutations outside of the PID that appeared to contribute to this improved activity. This outcome supports the hypothesis that a functional selection enables improved evolution outcomes, in particular for Cas variants with lower starting activity[100]. Importantly, these selections should be broadly adaptable to the

evolution of any Cas ortholog towards novel PAMs, and the sequence-agnostic nature of the target site can be applied to evolving novel editing windows or disease-specific contexts.

While we provided ePACE lagoons with the opportunity to evolve activity on specific PAM variants (e.g. four separate lagoons for each N$_4$<u>CN</u> PAM), variants emerged that were broadly active on the PAM position 5 base that was targeted (C or T). This outcome is expected for selection schemes that select for novel activity but do not counter-select against undesired activities. Nevertheless, predicting which target PAM would yield eNme2-C, eNme2-T.1, or eNme2-T.2 *a priori* likely would have been difficult, as starting activity of wild-type Nme2Cas9 on any N$_4$<u>CN</u> or N$_4$<u>TN</u> is comparably low. This challenge highlights the strength of the ePACE platform, which enabled us to explore all trajectories in parallel, greatly enhancing the rate at which we were able to discover high activity variants (five ePACE versus 20 to 40 traditional PACE experiments). Subsequent incorporation of a counter-selection[132] against undesired PAMs in an ePACE-enabled parallel manner may result in highly PAM- or protospacer-specific Cas variants that further advance tailor-made genome modifying technologies.

3.4 Methods

3.4.1 General methods

Antibiotics (Gold Biotechnology) were used at the following working concentrations: carbenicillin - 50 μg/mL, chloramphenicol - 25 μg/mL, kanamycin - 50 μg/mL, tetracycline - 10 μg/mL, streptomycin - 50 μg/mL. Nuclease-free water (Qiagen) was used for PCR reactions and cloning. All PCR reactions were carried out using Phusion U Hot Start polymerase (Thermo Fisher Scientific) unless otherwise noted. All plasmids and SP described in this study were cloned by USER assembly unless otherwise noted. Primers and gene fragments used for cloning were ordered from Integrated DNA Technologies (IDT) or Eton Biosciences, as necessary. For cloning purposes, Mach1 (Thermo Fisher Scientific) cells were used, and subsequent plasmid purification was done with plasmid preparation kits (Qiagen or Promega). Illustra TempliPhi DNA Amplification Kits (Cytiva) were used to amplify cloned plasmids prior to Sanger sequencing. For all phage related experiments (phage cloning, phage propagation, PACE and PANCE experiments) were done in parent *E. coli* strain S2060.

3.4.2 Overnight phage propagation assay

Chemicompetent S2060 cells were transformed with the AP(s) and CP(s) of interest as previously described. Single colonies were subsequently picked and grown overnight in DRM media with maintenance antibiotics at 37°C with shaking, then back-diluted 200-1000 fold into fresh DRM media the next day and grown. Upon reaching OD_{600} 0.4-0.6, host cells are transferred into 500 μL aliquots and infected with 10 μL of desired SP (final titer 1×10^5 pfu/mL). Cells were then incubated for another 16-20 h at 37°C with shaking, then centrifuged at 3,600 g for 10 min. The supernatant containing phage is stored until use.

3.4.3 Plaque assay

S2060 cells transformed with pJC175e (S2208[105]) were used for plaque assays unless otherwise stated. To prepare a cell stock, an overnight culture of S2208s was diluted 50-fold

into fresh 2xYT media with carbenicillin (50 ug/mL) and grown at 37°C to an OD_{600} ~0.6-0.8. SP were serially diluted (4 dilutions - 1:10 first dilution from concentrated phage stocks, then 1:100 remaining 3 dilutions) in DRM. 10 µL of each dilution is added to 150 µL of cells, followed by addition of 850 µL of liquid (55°C) top agar (2xYT media + 0.4% agar) supplemented with 2% Bluo-gal (1:50, final concentration 0.04%, Gold Biotechnology). These mixtures are then pipetted onto one quadrant of a quartered Petri dish containing 2 mL of solidified bottom agar (2xYT media + 1.5% agar, no antibiotics). Plates are allowed to briefly solidify before being incubated at 37°C overnight without inversion.

3.4.4   qPCR estimation of phage titer

When noted, phage titers were estimated by qPCR rather than plaque assay. SP pools (50 µL) were first heated at 80°C for 30 min to destroy polyphage. Polyphage genomes were then degraded by adding 5 µL of heated SP to 45 µL of 1x DNase I buffer containing 1 µL DNase I (New England Biolabs) and incubated at 37°C for 20 min followed by 95°C for 20 min. 1.5 µL of each prepared phage DNA stock is then added to a 25 µL qPCR reaction, prepared as follows: 10.5 µL H_2O, 12.5 µL 2x Q5 Mastermix (New England Biolabs), 0.25 µL Sybr Green (Thermo Fisher Scientific), 0.125 µL each primer (qPCR-Fw: 5'-CACCGTTCATCTGTCCTCTTT and qPCR-Rv: 5'-CGACCTGCTCCATGTTACTTAG). qPCR was then run with the following cycling conditions: 98°C for 2 min, 45 cycles of: [98°C for 10 s, 60°C for 20 s, and 72°C for 15 s]. Titers were calculated using a titration curve of an SP standard of known titer (by plaque assay). A limit of detection was set based on when primers amplified (without SP) or at the lowest titer prior to loss of linearity for the SP standard.

3.4.5   Phage-assisted noncontinuous evolution

Chemically competent S2060s were transformed with the AP(s) and CP(s) of interest along with a mutagenesis plasmid (MP6[124]), and plated on 2xYT agar containing maintenance antibiotics and 100 mM glucose. Three colonies are subsequently picked into DRM with

maintenance antibiotics and grown at 37°C with shaking to an OD_{600} ~0.4-0.6. Host cells are

then transferred into a 96-well plate in 500 μL aliquots, 10 mM arabinose is added to induce

mutagenesis, and SP dilutions from prior passages (or starting phage stocks) are added. Cells

are grown for 12-16 h at 37°C with shaking, and subsequent SP are isolated in the supernatant

following centrifugation at 3,600 g for 10 min. To increase and diversify phage titers when

necessary, SP were passaged in S2208s containing MP6; during such passages, cells were

only infected for 6-8 h. Starting phage stocks for PANCE1 (N1) and PANCE2 (N2) were all

diversified using this method prior to infection into the first PANCE passage. All SP titers were

estimated by qPCR as described above.

3.4.6   General ePACE methods

eVOLVER and PACE were run as previously described[105,106] with the following

modifications. Millifluidic devices controlling inducer flow into lagoons were sterilized before

connecting to the vials by filling lines and devices with 10% bleach letting sit for 30 minutes.

Bleach was subsequently flushed out with autoclaved di water, then lines purged with air and

connected to the vials and inducer bottles. Chemostats were inoculated to OD600 0.05 and run

at 30 ml total volume at 1 vol/hr. Cell OD was allowed to reach steady state before flow was

initiated into the lagoons. The volume of lagoons was set to 10 mL via continuous pumping of

waste with a high flow rate (45 ml/min) peristaltic pump (SQ2349291, FynchBio) from a 4''

hypodermic needle (Air-Tite N224) set in Port 2 of the custom ePACE vial cap. Cells were set to

pump in through Port 4 using a slow flow rate (1 ml/m) peristaltic pump (SQ2112453, FynchBio)

from a 3'' hypodermic needle (Air-Tite N163), and arabinose was pumped in through Port 1

using an IPP device. Before lagoon infection with phage, cells from the chemostats were flowed

through the vessel at 1 vol/hr with 250 mM arabinose flowing at 0.08 vol/hr for at least 1 hour.

Upon infection, cell flow rates were changed to the desired rate and arabinose flow rate set to

0.04 vol/h. Sampling and decisions on flow rate modifications were done as previously described[105]. Phage titer was quantified via qPCR method described above.

3.4.7   Millifluidic fabrication

All IPP and pressure regulator millifluidic devices were constructed as previously described[106]. Briefly, fluidic designs were drawn out in EAGLE (Autodesk) and patterned onto 1/4'' and 1/8'' acrylic using a 40W CO_2 laser cutter (Epilog Mini 24). The surface of the acrylic was then plasma treated for 1 minute with atmospheric gases at the maximum setting (Harrick Plasma, 30W Expanded Plasma Cleaner) to promote adhesion. These layers were then bonded together using an optically clear laminating adhesive sheet (3M, 8146-3) with a silicone membrane (0.01'', Rogers Corporation, BISCO HT-6240) between them that enables valve actuation.

3.4.8   IPP calibrations

To calibrate IPP devices, sealed bottles containing 1 L of water were attached to the input and pressurized to 1.5 psi. IPPs were controlled via 3-way solenoid valves (S10MM-31-12-3, Pneumadyne) connected to the custom eVOLVER pressure regulator system supplying 8 psi. Pumps were run at 4 different actuation frequencies long enough for at least 100 μl of water to flow, and then measured via pipette. A function of the form $y = kx^a$ is then fit to the resulting data and used to calculate the actuation frequency needed for a desired flow rate during experiments.

3.4.9   ePACE1

Host cells transformed with pTPH405 APs (each of the eight N3YTN PAMs) and MP6 were maintained in a chemostat as described above. Lagoons (8 total, 1 replicate of each PAM) were maintained as described above prior to infection with phage containing full-length wild-type Nme2-ABE8e in the SP391c architecture.

3.4.10 ePACE2

Host cells transformed with pTPH405 APs (each of the eight N_3YTN PAMs) and MP6 were maintained in a chemostat as described above. Lagoons (16 total, 2 replicates of each PAM) were maintained as described above prior to infection with pooled surviving phage from ePACE1 lagoons evolved on N_3CTC and N_3TTC PAMs.

3.4.11 ePACE3

Host cells transformed with pTPH405c (recoded gIII N-terminus) APs (each of the eight N_3YTN PAMs except N_3TTA PAM), pTPH412 TadA8e R26G-expressing CP, and MP6 were maintained in a chemostat as described above. Lagoons (14 total, 2 replicates of each PAM) were maintained as described above prior to infection with pooled surviving phage from ePACE1 and ePACE2 recoded into the split-phage SP404 architecture.

3.4.12 ePACE4

Host cells transformed with pTPH418b (recoded gIII N-terminus, dual PAM) APs (each of the six N_3WCD PAMs), pTPH412 TadA8e R26G-expressing CP, and MP6 were maintained in a chemostat as described above. Lagoons (16 total) were maintained as described above prior to infection with either pooled N1 replicate 1 & 2 passage 20 phage (6 lagoons), pooled N1 replicate 3 & 4 passage 20 phage (6 lagoons), or pooled N1 replicates 1-4 passage 20 phage (3 lagoons – N_3TCD PAMs). All lagoons were seeded with phage from corresponding N1 PAM lagoons.

3.4.13 ePACE5

Host cells transformed with pTPH418b (recoded gIII N-terminus, dual PAM) APs (each of the eight N_3YTN PAMs), pTPH412 TadA8e R26G-expressing CP, and MP6 were maintained in a chemostat as described above. Lagoons (16 total, 2 replicates of each PAM) were maintained as described above prior to infection with pooled N2 replicate 3 passage 7 phage from corresponding PAM lagoons.

3.4.14 Cloning of BE-PPA libraries

Cloning of the library plasmids (pTPH342 for CBE-PPA, pTPH424 for ABE-PPA) was done via one-piece USER assembly of purified PCR product amplified using a primer pool containing all desired PAM sequences (IDT). Purified PCR product was aliquoted into two 0.2 pmol USER reactions (~500 ng of a 4.2 kb fragment each), purified following USER digestion with PB buffer (Qiagen) and subsequent PE buffer washes (4x, Qiagen), and eluted into 15 µL H_2O. The entire amount was then transformed into electrocompetent 10B cells (New England Biolabs), enough to yield at minimum 14x coverage[133] of the expected library size. Electroporation was done in 25 µL aliquots using bacterial program X_13 in the 96-well Shuttle Device component of a 4D-Nucleofector system (Lonza). Transformed cells were immediately transferred to 1.5 mL (per 100 µL cells) of prewarmed SOC media. A serial dilution of the transformed cells (8 dilutions, 5-fold each, starting with undiluted cells) was immediately taken and plated on maintenance antibiotics, which was used to calculate effective library size. The remaining cells are allowed to recover at 37°C with shaking for 1 h prior to plating on 2xYT agar containing maintenance antibiotic. The following day, colonies were scraped and DNA was isolated using a Plasmid Plus Midi Kit (Qiagen).

3.4.15 Base editing-dependent PAM profiling assay

Chemicompetent 10B cells (New England Biolabs) were transformed with the base editor variants of interest. Three colonies of each base editor variant are seeded into 10 mL fresh DRM with maintenance antibiotic and grown at 37°C with shaking to an OD_{600} ~0.4-0.6. Upon reaching the desired cell density, cells were spun down at 5,000 xg for 10 minutes, washed 3x with ice-cold 10% (v/v) glycerol, then resuspended in a final volume of 100 µL 10% glycerol. 1 ug of library plasmid (pTPH342 or pTPH424) was added to these 100 µL aliquots, then transformed in 25 µL aliquots using bacterial program X_5 in the 96-well Shuttle Device component of a 4D-Nucleofector system. Transformed cells were immediately transferred to 1.5

99

mL (per 100 μL cells) of prewarmed SOC media. A serial dilution of the transformed cells (8 dilutions, 5-fold each, starting with undiluted cells) was immediately taken and plated on maintenance antibiotics, which was used to calculate effective library size. The remaining cells are allowed to recover at 37°C with shaking for 15 min, then diluted into 40 mL of prewarmed DRM containing maintenance antibiotics and 10 mM arabinose. Induced cells are then grown at 37°C with shaking for 22 h (ABE-PPA), or for 32 h with a 1:40 back-dilution at 16 h (CBE-PPA) before being harvested by centrifugation at 3,600 xg for 10 min. DNA is isolated from harvested cells using a Plasmid Plus Midi Kit (Qiagen).

3.4.16 High-throughput DNA sequencing of BE-PPA libraries

Library samples were prepared for high-throughput amplicon sequencing in two PCR steps. The first PCR (PCR1) was performed using forward primer BE-PPA-Fw (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAATACGCAACGCCTCTC-3') and reverse primer BE-PPA-Rv (5'-TGGAGTTCAGACGTGTGCTCTTCCGATCCTTGTCTGTAAGCGGATGC-3') at a 150 μL scale and 1 ug of template DNA. Cycling conditions were as follows: 98°C for 2 min, then 14 cycles of [98°C for 15 s, 60°C for 15s, 72°C for 20s], and a final extension at 72°C for 2 min. 14 cycles for PCR1 was observed to be within the linear amplification range for the libraries used in this study but may change for alternate library constructions. Following PCR1, PCR reactions were purified using the QIAquick PCR Purification Kit (Qiagen) and eluted in 16 μL nuclease-free H$_2$O. The second PCR (PCR2) was performed using forward and reverse Illumina barcoding primers at a 75 μL scale and half (8 μL) of the PCR1 purified product. Cycling conditions were as follows: 98°C for 2 min, then 8 cycles of [98°C for 15 s, 60°C for 15s, 72°C for 20s], and a final extension at 72°C for 2 min. 8 cycles for PCR2 was observed to be within the linear amplification range for the libraries used in this study but may change for alternate library constructions. PCR2 products were pooled, purified by electrophoresis with a 1% agarose gel

using a QIAquick Gel Extraction Kit (Qiagen), and eluted in nuclease-free H_2O. DNA

concentration was quantified with the KAPA Library Quantification Kit-Illumina (KAPA

Biosystems) and sequenced on an Illumina MiSeq instrument (paired-end read – R1: 210

cycles, R2: 0 cycles) according to the manufacturer's protocols.

3.4.17 Analysis of BE-PPA HTS data

Sequencing reads were demultiplexed using the Miseq Reporter (Illumina).

Demultiplexed files were subsequently analyzed for base editing activity using a custom

workflow combining the SeqKit[134] and CRISPResso2[96] packages. To analyze BE-PPA

sequenced files, the demultiplexed fastq files were filtered using the seqkit package/grep

function[134] to search for two flank sequences near either end of the amplicon. For ABE-PPA

profiled variants, groups of PAMs were UMI-tagged, and the specific UMI tag was used in place

of one of the flank sequences. Filtered files were then binned into individual fastq files per PAM

using the same function. The resulting PAM-specific fastq files were analyzed using standard

CRISPResso2[96] analysis.

3.4.18 Cell culture

HEK293T cells (ATCC CRL-3216) and HUH7 cells were cultured in Dulbecco's modified

Eagle's medium plus GlutaMax (DMEM, Thermo Fisher Scientific) supplemented with 10% (v/v)

fetal bovine serum (FBS, Thermo Fisher Scientific). U2OS cells were cultured in McCoy's 5A

Medium (Thermo Fisher Scientific) supplemented with 10% (v/v) FBS. Normal adult human

primary dermal fibroblasts (HDFa, ATCC PCS-201-012) were cultured in DMEM plus GlutaMax

supplemented with 20% (v/v) FBS. All cell types were cultured at 37°C with 5% CO_2. Cell lines

were authenticated by their suppliers and tested negative for mycoplasma.

3.4.19 HEK293T, HUH7, and U2OS cell line transfection protocols and genomic DNA isolation

HEK293T cells were seeded at a density of 2 x 10^4 cells per well on 96-well plates

(Corning) 16-20 h prior to transfection. Transfection conditions were as follows for HEK293T

cells: 0.5 µL Lipofectamine 2000 (Thermo Fisher Scientific), 250 ng of Cas effector plasmid (nuclease/base editor), and 83 ng of guide RNA plasmid were combined and diluted with Opti-MEM reduced serum media (Thermo Fisher Scientific) to a total volume of 10 µL and transfected according to the manufacturer's protocol. Cells were transfected at approximately 60-80% confluency. HUH7 cells and U2OS cells were seeded at a density of 2.5 x 10^4 cells per well on 96-well plates 16-20 h prior to transfection. Transfection conditions were as follows: 0.33 µL Lipofectamine 2000, 112.5 ng of Cas effector plasmid, and 37.5 ng of guide RNA plasmid were combined and diluted with Opti-MEM media to a total volume of 10 µL and transfected according to the manufacturer's protocol. Cells were transfected at approximately 80-100% confluency. Following transfection, all cell types were cultured for 3 days, after which the media was removed, the cells washed with 1x PBS solution, and genomic DNA harvested via cell lysis with 30 µL lysis buffer added per well (10 mM Tris-HCL, pH 8.0, 0.05% SDS, 20 ug/mL Proteinase K (New England Biolabs)). The cell lysis mixture was allowed to incubate for 1-2 h at 37°C before being transferred to 96-well PCR plates and enzyme inactivated for 30 min at 80°C. The resulting genomic DNA mixture was stored at -20°C until further use.

3.4.20 Base editor mRNA in vitro transcription

All base editor mRNA was generated from PCR product amplified from a template plasmid containing an expression vector for the base editor of interest cloned as described previously[135]. PCR product was amplified using forward primer IVT-F (5'-TCGAGCTCGGTACCTAATACGACTCACTATAAGGAAATAAGAGAGAAAAGAAG-3') and reverse primer IVT-R (5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTTCCTACTCAGGCTTTATTCAAAGACCA-3'), purified using the QIAquick PCR Purification Kit (Qiagen), and eluted in 15 µL nuclease-free H_2O. *In vitro* transcription was done using the HiScribe T7 High-Yield RNA

Synthesis Kit (New England Biolabs) according to the manufacturer's protocols but with full substitution of N1-methyl-pseudouridine (TriLink Biotechnologies) for uridine and cotranscriptional capping with CleanCap AG (TriLink Biotechnologies). mRNA isolation was performed using lithium chloride precipitation. Purified mRNA was stored at -20°C until further use.

3.4.21 *Human primary fibroblast nucleofection and genomic DNA extraction*

One day prior to nucleofection, 80-90% confluent HDFa cells were passaged at a 1:2 dilution ratio into fresh media. Nucleofection was performed by pooling 2.5×10^5 HDFa cells per condition and spun down at 300 xg for 10 minutes, washed with 1x PBS, spun again, then resuspended in P2 primary cell solution (10 µL per condition, Lonza). Concurrently, DNA mixtures were prepared by combining 50 pmol of chemically-synthesized guide RNA[30] (IDT or Synthego) with 1 ug of *in vitro* transcribed base editor mRNA and P2 primary cell solution into a total volume of 12 µL. Each 10 µL aliquot of HDFa cells is combined with DNA mixture to a total volume of 22 µL, and nucleofected with program DS-150 on 96-well Shuttle Device component of a 4D-Nucleofector system. Following nucleofection, cells were allowed to rest for 10 min before addition of 100 µL prewarmed media per well. 80 µL of each condition was subsequently taken and plated on a 48-well poly-D-lysine plate (Corning). Cells were cultured for 5 days post-nucleofection, with media replacement after the first day. Following removal of media and a wash with 1x PBS buffer, genomic DNA was isolated by addition of 100 µL lysis buffer following the same protocol as described for other cell lines. Genomic DNA was stored at -20°C until further use.

3.4.22 *High-throughput sequencing of genomic DNA*

High-throughput sequencing of genomic DNA from all cell lines was performed as previously described[30]. DNA concentrations were quantified with a Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific) or with a NanoDrop One Spectrophotometer (Thermo

Fisher Scientific) prior to sequencing on an Illumina MiSeq instrument (paired-end read – R1: 250-280 cycles, R2: 0 cycles) according to the manufacturer's protocols.

3.4.23 High-throughput sequencing data analysis

Individual sequencing runs were demultiplexed using the MiSeq Reporter (Illumina). Subsequent demultiplexed sequencing reads were analyzed using CRISPResso2[96] as described previously[30]. All editing values are representative of n = 3 independent biological replicates, with mean±SEM shown.

3.4.24 In silico prediction of off-target sites

Off-target site prediction was done using CHOPCHOPv3[130] and the "Paste Target" functionality with the following parameters: the Site 1 and Site 2 20 nt SpRY protospacers and corresponding 3 nt PAMs were used as search queries; under search options, the Cas9 PAM was set to custom "NNN", and mismatches within the protospacer was set to 2; self-complementarity parameters were removed; all other parameters were left as default. All resulting off-targets were then further screened manually, and sites with more than one mismatch within the PAM proximal region (≤10 bp from the PAM) were removed. Note that as the 23 nt Nme2Cas9 protospacer includes the 20 nt SpRY protospacer, any off-target for the Nme2Cas9 protospacer must also be an off-target for the SpRY protospacer.

3.4.25 U2OS nucleofection for GUIDE-Seq

One day prior to nucleofection, 80-90% confluent U2OS cells were passaged at a 1:2 dilution ratio into fresh media. Nucleofection was performed by pooling 3 x 10^5 U2OS cells per condition and spun down at 300 xg for 10 minutes, washed with 1x PBS, spun again, then resuspended in SE solution (10 µL per condition, Lonza). Concurrently, DNA mixtures were prepared by combining 750 ng of Cas9 plasmid, 250 ng of guide RNA plasmid, 5 pmol of the GUIDE-seq dsODN[95], and SE solution into a total volume of 12 µL. Each 10 µL aliquot of U2OS cells is combined with DNA mixture to a total volume of 22 µL, and nucleofected with program

DN-100 on the 96-well Shuttle Device component of a 4D-Nucleofector system. Following

nucleofection, cells were allowed to rest for 10 min before addition of 100 μL prewarmed media

per well. Each condition was then split into two 50 μL aliquots and plated on 24-well plates

(Corning). Cells were cultured for 5 days post-nucleofection, with media replacement after the

first day. Following removal of media and a wash with 1x PBS buffer, genomic DNA was

isolated using the DNAdvance Genomic DNA Isolation Kit (Agencourt), following the

manufacturer's protocols. Genomic DNA was stored at -20°C until further use.

3.4.26 Genomic DNA preparation and high-throughput sequencing for GUIDE-Seq

Genomic DNA was prepared for GUIDE-Seq as previously described[95], with the

following modifications. Genomic DNA shearing, end repair, dA-tailing, and adaptor ligation

were done in a one-pot mixture using the NEBNext Ultra II FS DNA Library Prep Kit for Illumina

(New England Biolabs), following the manufacturer's protocol for input DNA > 100 ng (without

size selection) and a desired fragment size distribution between 300 – 700 bp. During the

adaptor ligation step, the manufacturer-suggested NEBNext Adaptor for Illumina was replaced

with the custom GUIDE-Seq Y-adapter. DNA purification was done with AMPure XP beads

(Beckman Coulter). The subsequent PCR1, PCR2, library quantification, library normalization,

and high-throughput sequencing (paired-end Nextera sequencing – R1: 150, I1: 8, I2: 8, R2:

150) steps were done using the primers and protocols from the previously described protocol.

3.4.27 GUIDE-Seq analysis

Sequencing reads were demultiplexed using the MiSeq Reporter (Illumina), then

processed individually using the GUIDE-Seq analysis software, updated for Python 3 support

(https://github.com/tsailabSJ/guideseq). SpRY variants were analyzed using a mismatch

threshold of 8 and an NNN PAM. Nme2Cas9 variants were analyzed using a mismatch

threshold of 11 and an NNNNNN PAM. Visualization plots **in Figure 3.25, Figure 3.26, Figure**

3.27, Figure 3.28 were generated using a custom version of the original script, which has been

uploaded to the Khalil Lab GitHub repository (https://github.com/khalillab/guideseq).

Chapter 4. Improving Cas9 PAM specificity using negative selection

4.1 Introduction

Many natural, engineered, or evolved Cas orthologs have now been described, which together enable access to virtually all potential target genomic sites[30,66]. For some target sites, particularly those flanked by purine-containing PAMs, there are now multiple viable Cas variants. This increased targeting flexibility enables greater choice over the properties of an editing agent that factor into the correction of a specific genomic locus, such as the on-target efficiency, off-target and bystander editing characteristics, or deliverability of the selected editing agent. The expanded targetability of Cas9-based technologies has already enabled the correction or installation of SNPs responsible for reversing or curing sickle cell anemia (SCD)[43] and progeria[45] in disease-relevant models – targets that were previously poorly accessible by canonical SpCas9 and its limited NGG PAM accessibility (N = A, C, G, or T).

Efforts to expand the targeting capabilities of Cas proteins have so far primarily focused on improving efficiency, accessing novel sites, or both[82,84,87,98,101]. While these efforts have been largely successful, engineering or evolution of higher on-target activity or broadened PAM compatibility of Cas9 variant does not necessarily translate to other desirable properties, such as target specificity, deliverability, or immunogenicity, among others. In fact, unfocused activity enhancements via engineering or evolution have commonly been associated with promiscuity and increases in off-target activity, as previously observed in evolutions of proteases[136,137], deaminases[109,110], Cas proteins[97,98,138], and polymerases[132]. For Cas proteins specifically, PAM promiscuity generally leads to a drastic increase in off-target activity, as the PAM is the primary gatekeeper[139] for determining Cas-target interactions and subsequent specificity.

One common approach to addressing this specificity challenge is to introduce a counterselection campaign during engineering or evolution. A dual positive/negative selection ensures that improved activity is primarily restricted to the target(s) of interest by penalizing retained or improved activity on off-target substrates[132,140]. Counterselections have been applied

with great success to alter the specificity of proteases[136], polymerases[132], and early genome editing agents (zinc finger domains)[141], but have thus far been limited for Cas9, with only one example reported for altering the specificity of SpCas9[142].

In this chapter, we describe early efforts to develop a generalizable, sequence-agnostic dual positive/negative phage assisted continuous evolution (PACE) selection scheme for evolving genome editing agents with increased PAM or target specificity. We achieve this by introducing a negative selection component to the previously described SAC-PACE selection, such that genome editing at undesired PAMs incurs a survival penalty. In doing so, we can reduce the PAM promiscuity of previously evolved Nme2Cas9 variants while retaining efficiency at on-target, desired PAMs. Like the original SAC-PACE selection, this dual positive/negative selection is highly target/PAM generalizable, enabling the facile evolution of any desired Cas protein towards any specific protospacer or PAM.

4.2 Results

4.2.1 Development of a phage assisted continuous negative selection for Cas proteins

Phage assisted continuous evolution (PACE) is a valuable tool for tailoring the activity of desired proteins of interest (POIs), because of its rapid rate of diversification and selection relative to stepwise or library-based protein evolution methods[107]. In the context of large proteins with multiple complex activities like Cas enzymes, PACE is particularly powerful due to its ability to quickly and agnostically explore sequence space throughout the POI. This broad sequence exploration is in stark contrast to library-based rational engineering methods, which typically focus on small, promising regions of a given POI due to library-size and cost constraints[143]. Given the prior success of PACE in generating highly mutated variants of SpCas9[77,97] and Nme2Cas9 with increased PAM compatibility, we envisioned that a dual positive/negative PACE selection could be used to not broaden but shift PAM compatibility (**Figure 4.1**). By doing so, increased on-target activity at desired PAMs could be decoupled from the increased off-target activity that typically accompanies PAM-broadened variants (**Figure 4.2**).
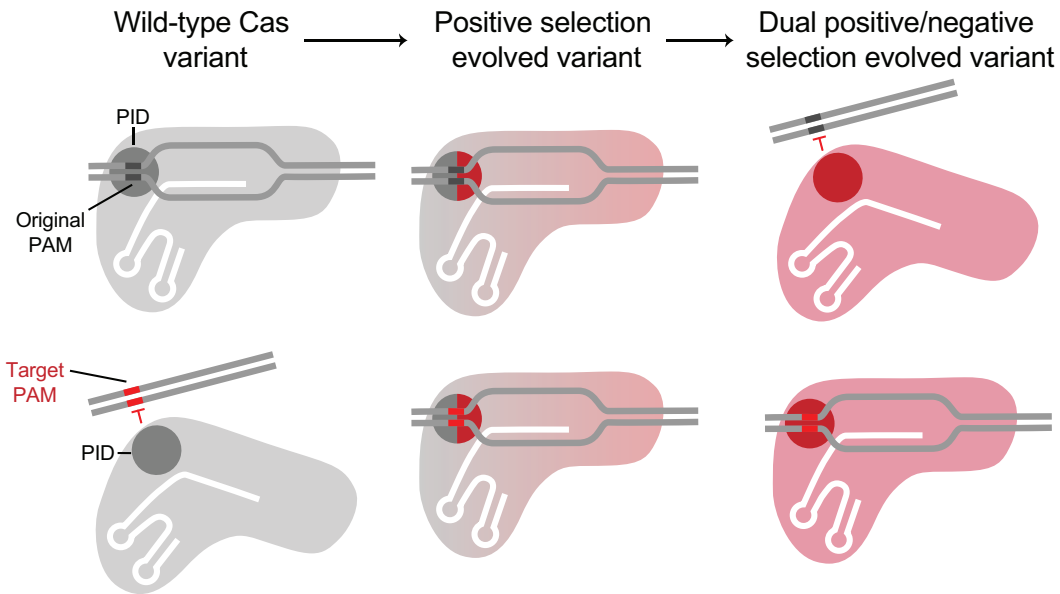
Figure 4.1. PAM activity of Cas variants across different selection schemes.
In selection schemes that only require evolving Cas variants to acquire novel PAM compatibility, a common outcome is PAM promiscuity, in which activity on new PAMs is observed while activity on wild-type or original PAMs is retained. This promiscuity can be reduced via introduction of a counterselection on the undesired, original PAM(s) while continuing to select for increased activity on novel, target PAMs. This type of dual positive/negative selection thereby enables the evolution of more PAM specific variants that may have more desirable on-to-off target activity profiles due to limited PAM promiscuity.

To develop a generalizable dual positive/negative selection for Cas protein PAM variants, we introduced a negative selection within the prior SAC-PACE framework. In SAC-PACE, the essential gene necessary for M13 bacteriophage propagation, gIII, is placed on an accessory plasmid (AP) and split by an *in cis* intein in which the N- and C-terminal intein halves are fused together with an arbitrary linker. This linker can be reprogrammed with any sequence context and contains one or more stop codons which prevent the expression of gIII. On the selection phage (SP), a Cas protein is expressed fused to either an adenosine deaminase or an intein that can undergo *in trans* splicing with an adenosine deaminase-intein fusion expressed from a complementary plasmid (CP) in host cells. Successful Cas engagement with a programmed protospacer/PAM containing the stop codons within the linker results in subsequent base editing that enables phage propagation. To enable negative selection, we

introduced a new negative accessory plasmid (AP_n), which has similar construction as the AP, except with gIII-neg, a dominant negative form of gIII that prevents phage propagation[132], instead of gIII (**Figure 4.2**). The construct expressing gIII-neg is split by a fused intein pair orthogonal to the one used in the AP[127], and the linker contains a stop codon flanked by an arbitrary, undesired PAM. Like the original SAC-PACE selection, the dual positive/negative SAC-PACE selection requires functional Cas activity, including R-loop formation and maintenance enabling subsequent base editing, while retaining high sequence generalizability such that any desired Cas protein can be evolved with this approach.
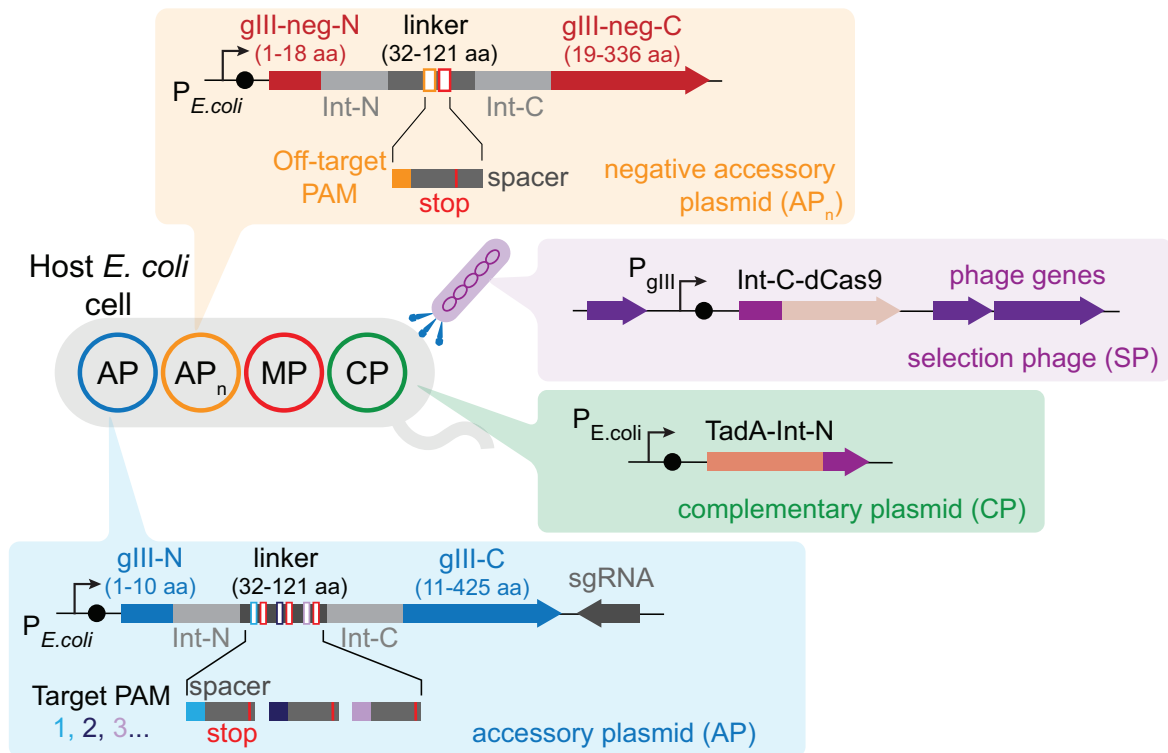
Figure 4.2. Schematic of the dual positive/negative sequence-agnostic Cas PACE (SAC-PACE) selection.
The selection circuit in the dual positive/negative SAC-PACE selection. The selection phage (SP) encodes an evolving Cas variant fused to the C-terminal intein fragment of a split-intein in place of gIII. In the host cells, a complementary plasmid (CP) expresses an adenosine deaminase fused to the N-terminal intein fragment of a split-intein, which splices together with the Cas variant from the SP to form competent adenine base editor. An accessory plasmid (AP) encodes a *cis* intein-split gIII, with a linker (31–121 aa) comprised of one or more target protospacer/PAM combinations that contain at least one stop codon each. Correction of the stop codons through recognition of a novel PAM and subsequent base editing results in excision of the *cis*-intein, production of functional gIII, and phage propagation. However, the host cells also contain a negative accessory plasmid (AP$_n$) which encodes an orthogonal *cis* intein-split gIII-neg, a dominant negative form of gIII that prevents phage propagation even in the presence of functional gIII. The intein-split gIII-neg fragments are fused *in cis* with a linker that contains a protospacer with a stop codon flanked by an undesired, off-target PAM. If evolving variants have activity on this undesired PAM, then phage propagation is limited even in the presence of strong activity on the target PAMs within the AP. Finally, the host cells also contain mutagenesis plasmid (MP), which diversifies the evolving phage.

To validate the AP$_n$, we first wanted to identify a new orthogonal intein pair not already in use in the complementary plasmid (CP, gp41-8 intein) used for expressing the adenosine deaminase TadA8e[109], or the AP (Npu intein). We selected three split intein pairs previously

113

shown to be orthogonal to both the Npu and gp41-8 intein pairs[127]. We constructed two test

AP_ns for each intein pair to test the maximum theoretical dynamic range of phage propagation

for each pair. One AP_n (positive control) contains gIII-neg with the intein pair fused *in cis* by an

arbitrary linker inserted after Ser18 and conditionally expressed by a phage shock promoter

(psp). The other AP_n (negative control) has the same construction except with one stop codon

inserted within the arbitrary linker. We tested these AP_ns in host cells containing split SAC-

PACE components: an AP containing one PAM/protospacer combination with two stop codons,

and a CP expressing a TadA8e R26G variant previously validated for split SAC-PACE. Notably,

all three intein pairs exhibited a robust on-to-off dynamic range ($>10^4$-fold) with a minimal

reduction in on-target propagation, suggesting that the AP_n architecture is compatible for

activating gIII-neg expression without drastically affecting the positive selection (**Figure 4.3**).
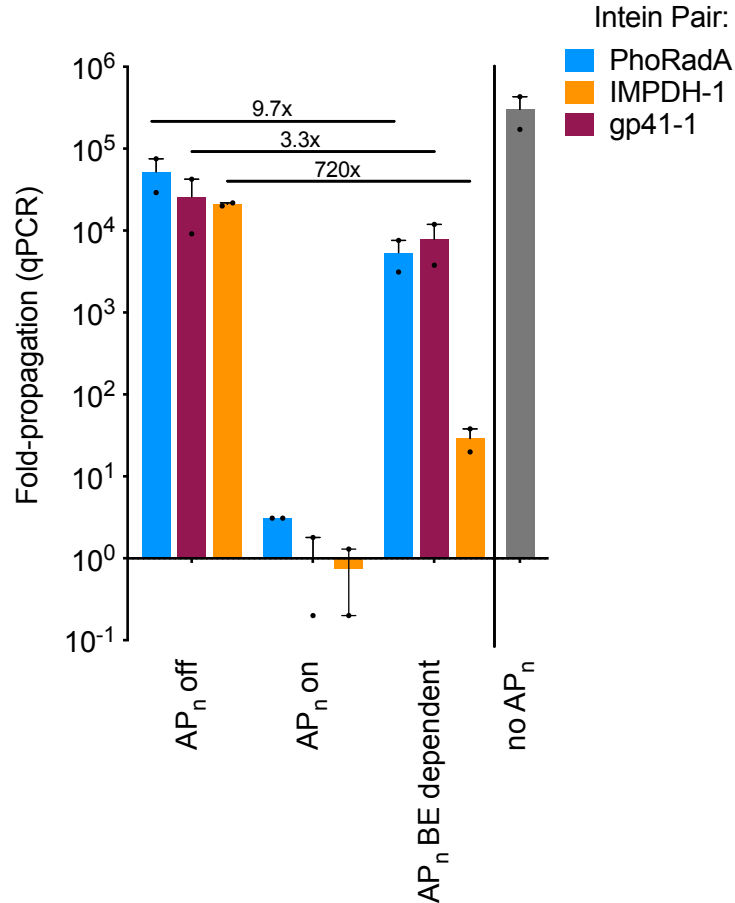
Figure 4.3. Validation of an orthogonal *in cis* intein pair for use in the AP_n.
Overnight propagation assay to test the ability of three different intein pairs to splice and enable functional gIII-neg expression from the AP_n in a base editing-dependent or independent manner. All host cells were infected with wild-type, Nme2Cas9 containing phage. The AP_n off condition represents an AP_n with a stop codon flanked by a PAM that cannot be targeted by wild-type Nme2Cas9. The AP_n on condition represents an AP_n without a stop codon in the linker. The AP_n BE dependent condition represents an AP_n with a stop codon flanked by a PAM that can be targeted by wild-type Nme2Cas9. The no AP_n condition does not contain an AP_n. Mean±SEM is shown and are representative of $n = 2$ independent biological replicates. Fold-propagation is calculated as the ratio of titer after overnight propagation over inoculating titer.

We next wanted to confirm that AP_n activation can be made dependent on undesired

PAM activity. We placed the same protospacer used in the AP within the AP_n linker, flanked by

a NNNNCC PAM that should be readily accessible to both wild-type Nme2Cas9 and evolved

eNme2-C. We then infected host cells containing split SAC-PACE components and this AP_n

with phage containing eNme2-C in an overnight propagation assay. Surprisingly, only the

IMPDH-1 intein pair retained a high base editing dependent dynamic range (722-fold difference between off-and-on states) (**Figure 4.3**). Both the PhoRadA and gp41-1 intein pairs had a <10-fold off-to-on ratio, perhaps suggesting that these inteins are less effective at generating competently spliced gIII-neg at lower expressed concentrations. Moving forward, we used the IMPDH-1 intein pair in our AP_n constructions.

4.2.2   *Preliminary evolution campaigns towards N₄TTN PAM-specific Nme2Cas9 variants*

Having validated a BE-dependent dual positive/negative SAC-PACE selection, we then wanted to select desirable PAMs to target during evolution. In our previous campaigns to evolve Nme2Cas9, one of the primary limitations of the positive-only SAC-PACE selection was the acquisition of broad, promiscuous PAM activity that translated to inconsistent activity in mammalian cells. This outcome was particularly evident in the N_4TN-PAM trajectory that required more drastic changes in PAM recognition for Nme2Cas9, which canonically interacts with N_4CC PAMs. We hypothesized that more PAM-specific variants would not only enable greater specificity, but also higher on-target activity as the variants are no longer capable of being sequestered at off-target PAM sites. We designed four evolution campaigns towards each of the four N_3TTN PAMs, with an initial AP_n penalizing activity on an N_3CCC PAM, a PAM accessible to wild-type Nme2Cas9 and still targeted in bacteria by variants previously evolved on N_4TN PAMs. For the positive selection, we further increased the selection stringency by including a third PAM (novel nucleobase identity at PAM positions 1-3 and 7) in addition to the two previously present in the dual PAM, split SAC-PACE strategy. Notably, Nme2Cas9 variants previously evolved on the dual PAM, split SAC-PACE selection (E5 phage) were still capable of propagating on these new triple PAM, split SAC-PACE APs at levels sufficient for PACE (**Figure 4.4**)[105].
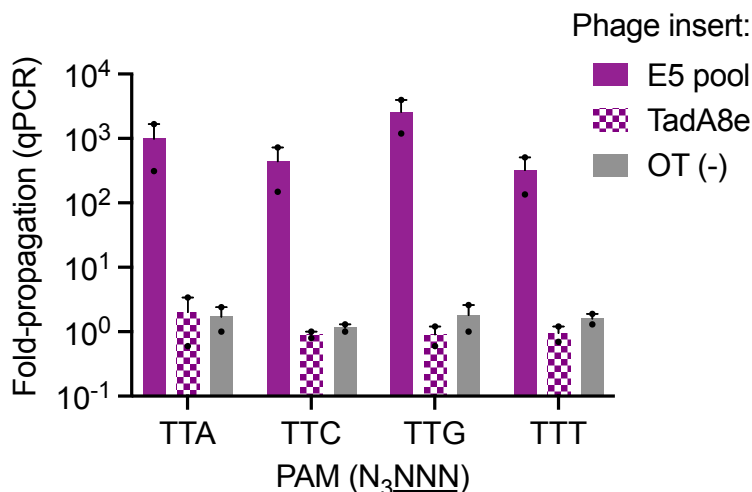
Figure 4.4. Validation of the triple PAM, split SAC-PACE APs.
Overnight propagation assay to test the ability of previously evolved Nme2Cas9 phage (E5 phage) to propagate on APs containing three copies of the target PAM requiring correction of three stop codons. E5 pooled phage were previously evolved on $N_3\underline{NTN}$ PAMs using the dual PAM, split SAC-PACE selection. The off-target phage (OT) contains neither a variant of Nme2Cas9 nor an adenosine deaminase. Mean±SEM is shown and are representative of $n = 2$ independent biological replicates. Fold-propagation is calculated as the ratio of titer after overnight propagation over inoculating titer.

As we wanted to explore multiple PAMs and negative selection stringencies in parallel, we began our evolution campaign towards PAM-specific Nme2Cas9 variants using phage-assisted non-continuous evolution (PANCE). Although PANCE is generally lower stringency than continuous evolution, the ability to multiplex in 96-well plates enables a high degree of control over evolution conditions[105]. This control is particularly powerful for negative selection, as the rate at which positive or negative stringency should be increased is difficult to predict in advance. We multiplexed 3 negative selection stringencies with the 4 target PAMs (positive selection: triple PAM, split SAC-PACE) for a total of 12 initial evolution conditions (N1, **Figure 4.5**). Promisingly, propagation of phage containing PAM-promiscuous variants of Nme2Cas9 (E5 phage) in overnight assays appeared to be directly correlated to the expression level of the AP_n, with higher promoter strengths on AP_n resulting in lower propagation (**Figure 4.6**).

Negative selection stringency
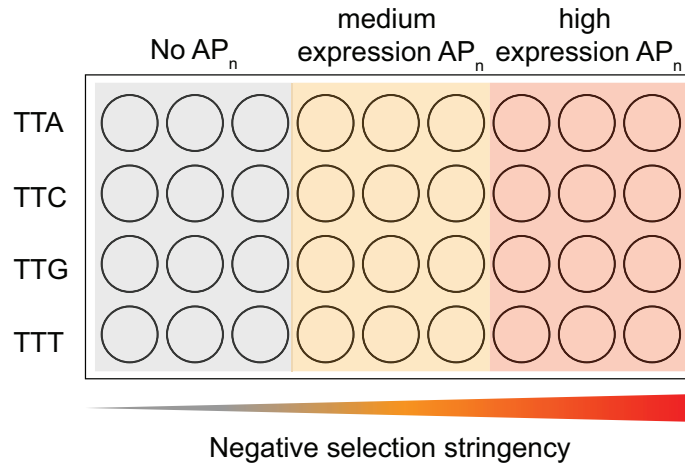
Figure 4.5. Layout of the N1 PANCE selection campaigns towards N_3TTN-specific variants of Nme2Cas9.
PANCE conditions used to evolve Nme2Cas9 towards N_3TTN-specific activity. Four PAMs were targeted in the positive selection (triple PAM, split base editor), multiplexed with three different negative selection stringencies (no AP_n, pro5 expression of gIII-neg, or proC expression of gIII-neg). All conditions were run in triplicate.
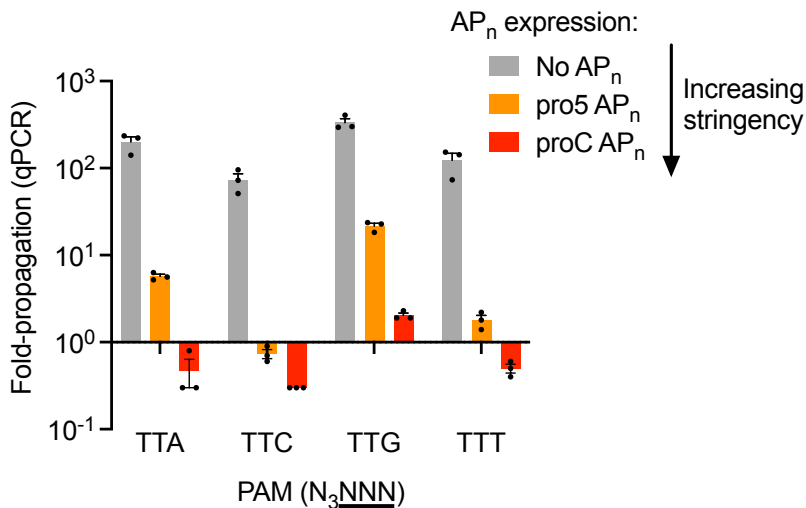
Figure 4.6. Overnight propagation of E5 phage on dual positive/negative SAC-PACE selections with varying negative selection stringencies.
Overnight propagation of previously evolved Nme2Cas9 (E5 phage) on the 12 different selection stringencies targeted in N1. Mean±SEM is shown and are representative of *n = 3* independent biological replicates. Fold-propagation is calculated as the ratio of titer after overnight propagation over inoculating titer.

After 13 initial rounds of PANCE, while all positive-only selections exhibited robust phage propagation, most of the high stringency negative selection trajectories were unable to sustain phage propagation except for conditions evolved on N_3TTG PAM-containing APs (**Figure 4.7**). This result would suggest that across the mutations present in the PAM-permissive variants of Nme2Cas9, few enable strong differentiation between N_3CCC and N_3TTN PAMs, unless if the PAM position 6 base is a G. We sequenced phage from each of the three negative selection stringencies targeting an N_3TTG PAM and found strong inter-lagoon convergence for phage evolved on the same negative selection stringency. Importantly, we observed a drastic shift in the identity of converged mutations within the PAM-interacting domain (PID) that corresponded to the presence of negative selection (**Table 4.1**).
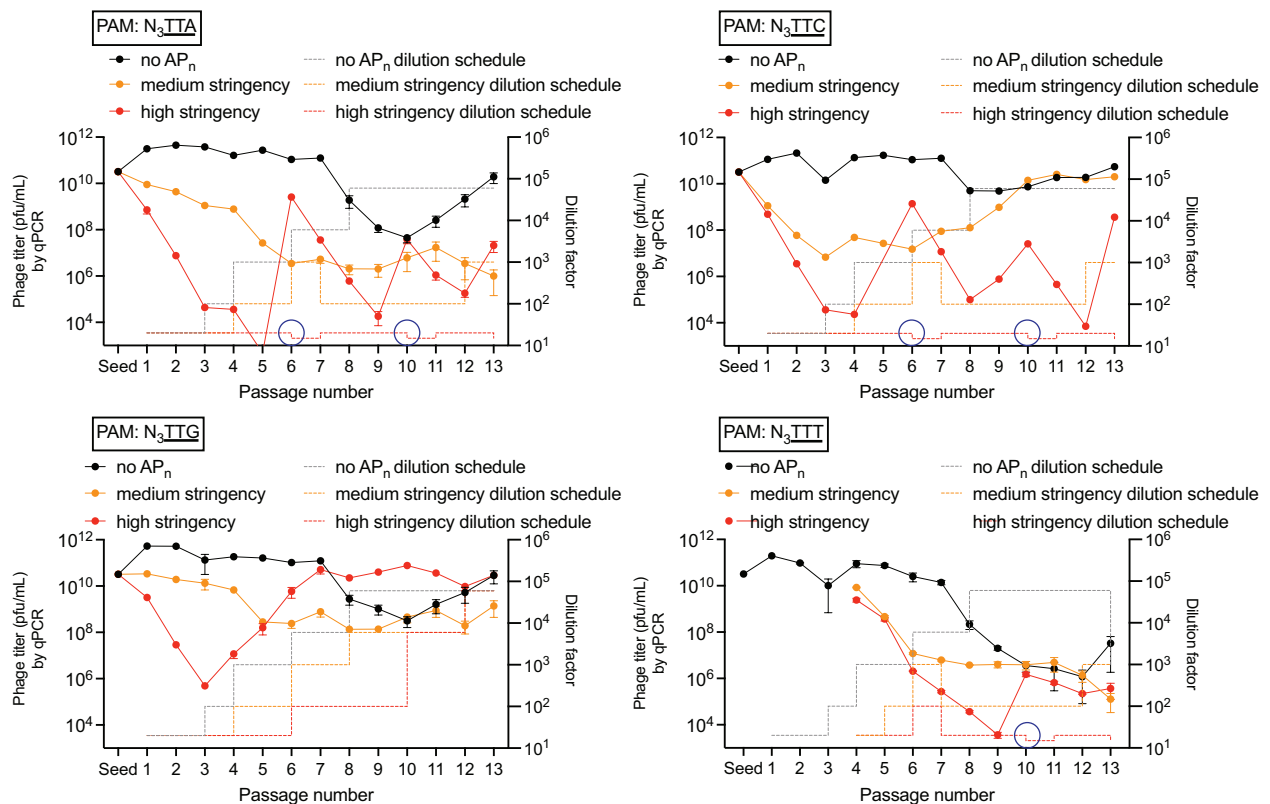
Figure 4.7. Dilution schedule and titers for N1.
SP containing pooled E5 phage were first diversified in *E.coli* host cells containing pJC175e[105] and MP6[105], isolated, then seeded into N1 (see conditions from **Figure 4.5**). Dilution schedules for each condition is shown in the plots, as are resulting titers (measured by qPCR). If lagoons needed to be reseeded, the passage is highlighted with a blue circle.

|	dNme2Cas9 (C-terminal: WED, PID)																						
Residue number	844	845	859	868	881	929	932	933	951	961	1005	1013	1025	1026	1028	1029	1031	1033	1043	1049	1051	1056	1080
wild-type	D	T	I	E	K	K	E	S	M	D	K	K	I	N	D	S	N	R	S	R	S	V	P
TTG-noAPn.N1-1	A						K		R	R					N	A		H		C		A	L
TTG-noAPn.N1-2	A						K		R	R					N	A		H		C		A	L
TTG-noAPn.N1-3	A		V	D					R	R					N	A		G		S			
TTG-noAPn.N1-4									R	R					N	A		H		C		A	L
TTG-noAPn.N1-5	A						K		R	R					N	A		H		C		A	L
TTG-mAPn.N1-6	V	I			R	R	R			G				S	Y	N	S	D	R				
TTG-mAPn.N1-7						R	R			G				S	Y	N	S	D	R				
TTG-mAPn.N1-8	V	I			R	R	R			G				S	Y	N	S	D	R				
TTG-mAPn.N1-9	V				R	R	R			G			R	S	Y	H	S	E	R		A		
TTG-mAPn.N1-10	V	I			R	R	R			G			R	S	Y	H	S	K			A		
TTG-mAPn.N1-11	V				R	R	R			G				S	Y	H	S	D	R				
TTG-mAPn.N1-12	V	I			R	R	R			G				S	Y	H	S	D	R				
TTG-mAPn.N1-13	V	I				R	R			G				S	Y	H	S	D	R				
TTG-mAPn.N1-14	V					R	R			G				S	Y	H	S	D	R				
TTG-hAPn.N1-15	V	I			R	R	R			G				S	Y	H	S	D					
TTG-hAPn.N1-16	V	I			R	R	R			G				S	Y	H	S	G	R				
TTG-hAPn.N1-17	V	I			R	R	R			G				S	Y	H	S	D	R				
TTG-hAPn.N1-18	V	I			R	R	R			G				S	Y	H	S	G	R				
TTG-hAPn.N1-19	V	I			R	R	R			G				S	Y	H	S	D	R				
TTG-hAPn.N1-20	V	I			R	R	R			G				S	Y	H	S	D	R				
TTG-hAPn.N1-21	V	I			R	R	R			G				S	Y	H	S	G	R				

Table 4.1. Mutation table of N1 variants evolved towards N₃TTG specificity.
Genotypes of individually sequenced plaques following 13 passages in N1 of the 3 conditions targeting N₃TTG PAMs in the positive selection. For clarity, only mutations in the PID are shown, with positions varying from wild-type displayed. Clones evolved without negative selection are highlighted in grey; clones evolved with medium stringency negative selection are highlighted in orange, and clones evolved with high stringency negative selection are highlighted in red.

To examine whether these Nme2Cas9 variants subjected to negative selection were indeed exhibiting altered PAM compatibilities, we used ABE-PPA to profile a representative variant evolved with high stringency negative selection (N1-21-ABE8e) or without negative selection (N1-5-ABE8e) (**Figure 4.8a**). At the targeted N₃TTG PAM, N1-5-ABE8e averaged 75% A•T-to-G•C conversion while N1-21-ABE8e average 67%, indicating comparable but slightly lower activity on the on-target PAM when negative selection is included. In contrast, however, at the negatively selected off-target N₃CCC PAM, the N1-5-ABE8e variant retains 57% A•T-to-G•C conversion, corresponding to a 1.3-to-1 on-to-off target activity ratio, while the N1-21-ABE8e variant virtually eliminates activity at this PAM, averaging 4.8% A•T-to-G•C conversion,

or a 14-to-1 on-to-off target activity ratio (**Figure 4.8b**). More broadly, the N1-21-ABE8e variant

appears to strongly disfavor sequences with a cytosine at PAM position 5 or 6, so long as the

position 6 base is not a guanine. In contrast, the N1-5-ABE8e variant retains the broad PAM

promiscuity observed for previous Nme2Cas9 variants evolved towards an N_4TN PAM. These

results would suggest that the dual positive/negative SAC-PACE selection can generate more

PAM-tailored variants of Nme2Cas9.
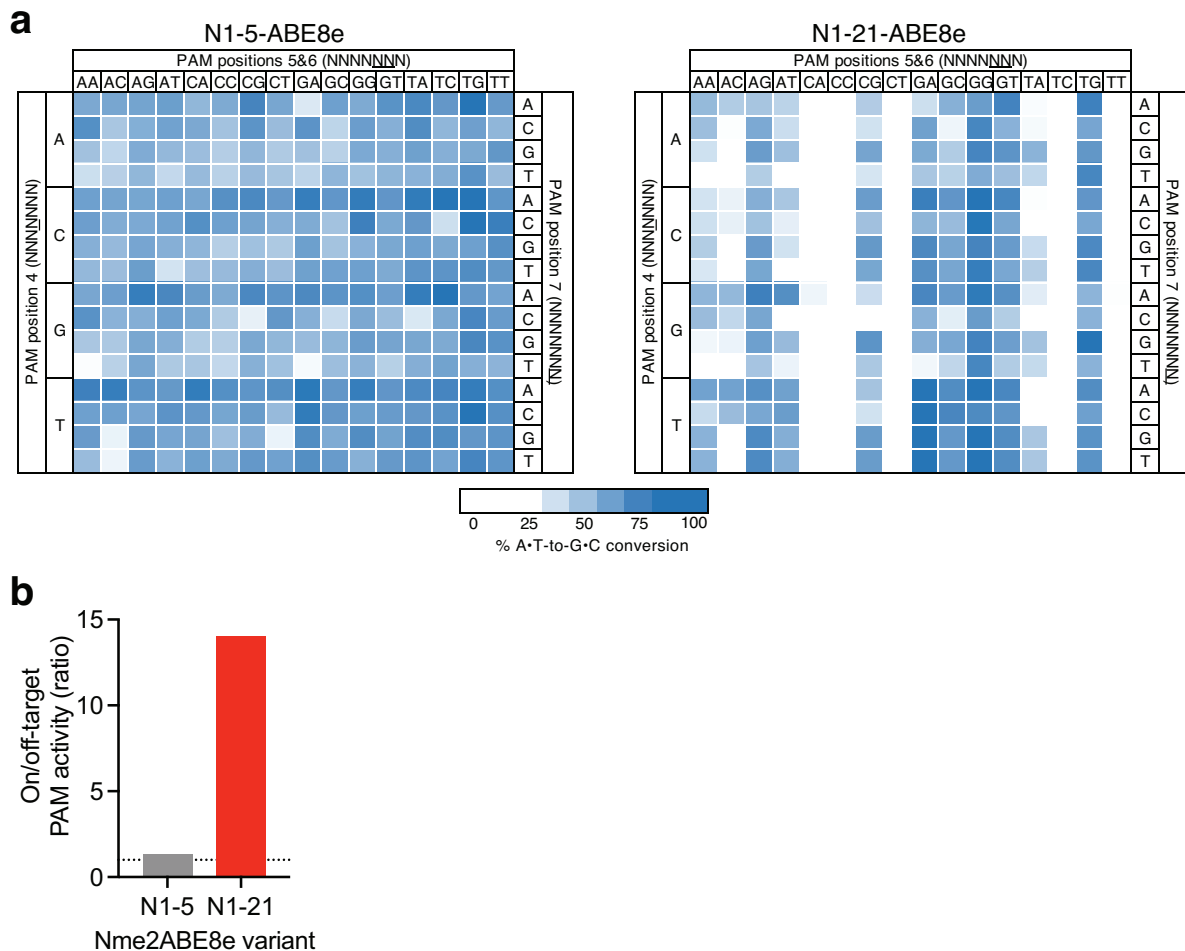
Figure 4.8. ABE-PPA activity of N1 evolved Nme2Cas9 variants.
(a) Heat maps showing ABE-PPA activity of N1-5-ABE8e (left), which was evolved without negative selection, and N1-21-ABE8e, which was evolved with high stringency negative selection, on the set of 256 N_3NNNN PAMs (PAM positions 1-3 fixed). Values are raw % A•T-to-G•C conversion observed for one replicate of each editor. **(b)** On/off-target PAM activity ratio of N1-5-ABE8e and N1-21-ABE8e. The ratio is calculated by taking the % A•T-to-G•C conversion observed for the positive selection PAM (N_3TTG) divided by the % A•T-to-G•C conversion observed for the negative selection PAM (N_3CCC).

Notably, an unintended outcome of this initial dual positive/negative selection was the emergence of an insufficiently limited PAM scope. While the N1-21-ABE8e variant did exhibit reduced PAM promiscuity relative to the N1-5-ABE8e variant, especially on N_4CN and N_4NC PAMs, the former variant exhibited improved activity at G-containing PAMs. At all N_4GN PAMs, N1-21-ABE8e averaged 64% A•T-to-G•C conversion compared to 60% for N1-5-ABE8e. This result would suggest that in the absence of explicit counterselection against all undesired PAM compatibilities, the dual positive/negative selection may not necessarily yield the desired PAM scope. In the case of this N_3TTG trajectory, counterselection against an N_3CCC PAM alone preferentially yielded variants that no longer had activity on N_4CN PAMs, but acquired or retained strong activity on N_4GN PAMs, a group of PAMs that may not be desired. Further optimization of the selection stringency via introduction of a multiple-PAM counterselection could be beneficial for evolving more specific variants.

We next compared the N1-21-ABE8e variant to the previously evolved, PAM-promiscuous eNme2-T.1-ABE8e variant in mammalian cells at six genomic sites containing N_3NTN PAMs (**Figure 4.9**). Indeed, the improved specificity of the N1-21-ABE8e variant for N_3NTG PAMs observed in ABE-PPA appears to translate to the initial panel of endogenous genomic sites tested. At the four tested sites containing N_3NTG PAMs, N1-21-ABE8e exhibited comparable or slightly improved adenine base editing activity to eNme2-T.1-ABE8e (16.0% vs. 14.2% A•T-to-G•C conversion). In contrast, at the two tested sites containing N_3NTW PAMs (where W = A or T), N1-21-ABE8e is not active, exhibiting 0.1% A•T-to-G•C conversion compared to the robust activity of eNme2-T.1-ABE8e at these same sites (36.0% A•T-to-G•C conversion). Together, this initial mammalian cell data coupled with the observed PAM profile from ABE-PPA indicates that the dual positive/negative SAC-PACE selection is capable of generating more PAM-specific variants of Cas proteins.
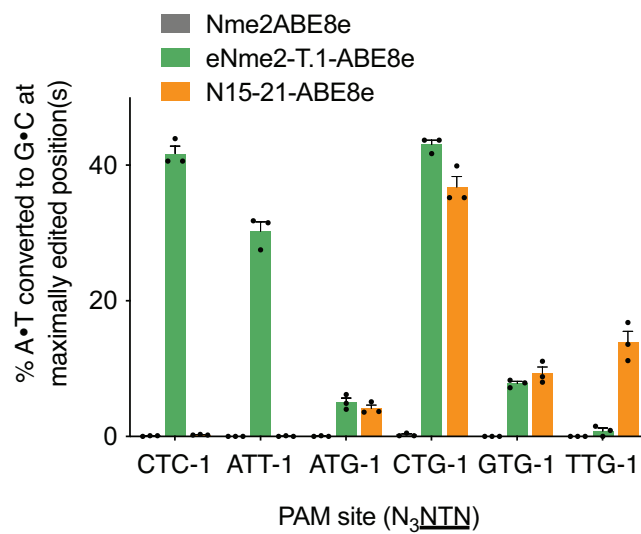
Figure 4.9. Comparison of N15-21-ABE8e to eNme2-T.1-ABE8e at N₃NTN PAM sites in HEK293T cells.

Adenine base editing activity of the dual positive/negative SAC-PACE evolved N15-21-ABE8e variant compared to a prior, PAM-promiscuous eNme2-T.1-ABE8e variant at six N_3NTN sites in HEK293T cells. Mean±SEM is shown and reflects the average activity and standard error of n = 3 independent biological replicates measured at the maximally edited position within each given genomic site.

4.3 Conclusions and outlook

In this chapter, we have demonstrated preliminary results that validate the potential to couple negative selection with the previously developed SAC-PACE selection to evolve bespoke, PAM-specific variants of Cas proteins. To achieve a robust counterselection, we introduced a negative accessory plasmid encoding gIII-neg. The expression of gIII-neg is activated by off-target PAM activity, which enables base editing of a stop codon within the coding sequence of gIII-neg that then results in full-length expression. Like the original SAC-PACE accessory plasmid, the AP_n is sequence agnostic and requires functional Cas protein activity outside of simple PAM binding, enabling generalizable, robust evolution of high efficiency, high specificity PAM activity for any desired Cas protein of interest.

However, the initial evolution campaigns to evolve PAM-specific variants of Nme2Cas9 did reveal some aspects of the selection that may require additional developments. The N1-21-ABE8e variant evolved to more specifically target N_3TTG PAMs did indeed retain activity on N_3TTG PAMs while ablating activity on the counterselected N_3CCC PAMs. However, this selectively appeared to be due more so to an increased preference for a G at PAM positions 5 or 6 ($N_3N\underline{NN}$), rather than full specificity on a TTG motif. This outcome is perhaps because acquiring a strong G preference is simpler than acquiring TTG specificity, and this is possible as there was not an explicit negative selection against binding G-containing PAMs. One potential solution to this undesired outcome is to include a multiplexed negative selection, much like how multiplexed positive selection was used to better evolve robust, PAM-broadened variants of Nme2Cas9. To develop a multiplexed negative selection, we cannot simply introduce additional protospacer/PAM combinations within the *cis* intein split gIII-neg construct, as this would decrease negative selection stringency since now an evolving variant must have off-target activity on both PAMs to yield gIII-neg. Instead, it may be worthwhile to explore introducing two copies of the gIII-neg construct, each with a unique off-target PAM. Nevertheless, the

124

developments described in this chapter provide the basis for a robust, generalizable platform for

evolving PAM-specific variants of any Cas protein, a capability that will be invaluable to the

continued improvements of precision genome editing technologies towards therapeutics.

4.4 Methods

4.4.1   General methods

Antibiotics (Gold Biotechnology) were used at the following working concentrations: carbenicillin - 50 µg/mL, chloramphenicol - 25 µg/mL, kanamycin - 50 µg/mL, tetracycline - 10 µg/mL, streptomycin - 50 µg/mL. Nuclease-free water (Qiagen) was used for PCR reactions and cloning. All PCR reactions were carried out using Phusion U Hot Start polymerase (Thermo Fisher Scientific) unless otherwise noted. All plasmids and SP described in this study were cloned by USER assembly unless otherwise noted. Primers and gene fragments used for cloning were ordered from Integrated DNA Technologies (IDT) or Eton Biosciences, as necessary. For cloning purposes, Mach1 (Thermo Fisher Scientific) cells were used, and subsequent plasmid purification was done with plasmid preparation kits (Qiagen or Promega). Illustra TempliPhi DNA Amplification Kits (Cytiva) were used to amplify cloned plasmids prior to Sanger sequencing. For all phage related experiments (phage cloning, phage propagation, PANCE experiments) were done in parent *E. coli* strain S2060.

4.4.2   Overnight phage propagation assay

Chemicompetent S2060 cells were transformed with the AP(s), AP_n(s), and CP(s) of interest as previously described. Single colonies were subsequently picked and grown overnight in DRM media with maintenance antibiotics at 37°C with shaking, then back-diluted 200-1000 fold into fresh DRM media the next day and grown. Upon reaching OD_{600} 0.4-0.6, host cells are transferred into 500 µL aliquots and infected with 10 µL of desired SP (final titer 1×10^5 pfu/mL). Cells were then incubated for another 16-20 h at 37°C with shaking, then centrifuged at 3,600 g for 10 min. The supernatant containing phage is stored until use.

4.4.3   Plaque assay

S2060 cells transformed with pJC175e (S2208[105]) were used for plaque assays unless otherwise stated. To prepare a cell stock, an overnight culture of S2208s was diluted 50-fold

126

into fresh 2xYT media with carbenicillin (50 ug/mL) and grown at 37°C to an OD_{600} ~0.6-0.8. SP were serially diluted (4 dilutions - 1:10 first dilution from concentrated phage stocks, then 1:100 remaining 3 dilutions) in DRM. 10 µL of each dilution is added to 150 µL of cells, followed by addition of 850 µL of liquid (55°C) top agar (2xYT media + 0.4% agar) supplemented with 2% Bluo-gal (1:50, final concentration 0.04%, Gold Biotechnology). These mixtures are then pipetted onto one quadrant of a quartered Petri dish containing 2 mL of solidified bottom agar (2xYT media + 1.5% agar, no antibiotics). Plates are allowed to briefly solidify before being incubated at 37°C overnight without inversion.

4.4.4   qPCR estimation of phage titer

When noted, phage titers were estimated by qPCR rather than plaque assay. SP pools (50 µL) were first heated at 80°C for 30 min to destroy polyphage. Polyphage genomes were then degraded by adding 5 µL of heated SP to 45 µL of 1x DNase I buffer containing 1 µL DNase I (New England Biolabs) and incubated at 37°C for 20 min followed by 95°C for 20 min. 1.5 µL of each prepared phage DNA stock is then added to a 25 µL qPCR reaction, prepared as follows: 10.5 µL H_2O, 12.5 µL 2x Q5 Mastermix (New England Biolabs), 0.25 µL Sybr Green (Thermo Fisher Scientific), 0.125 µL each primer (qPCR-Fw: 5'-CACCGTTCATCTGTCCTCTTT and qPCR-Rv: 5'-CGACCTGCTCCATGTTACTTAG). qPCR was then run with the following cycling conditions: 98°C for 2 min, 45 cycles of: [98°C for 10 s, 60°C for 20 s, and 72°C for 15 s]. Titers were calculated using a titration curve of an SP standard of known titer (by plaque assay).

4.4.5   Phage-assisted noncontinuous evolution

Chemically competent S2060s were transformed with the AP(s), AP_n(s) and CP(s) of interest along with a mutagenesis plasmid (MP6[124]), and plated on 2xYT agar containing maintenance antibiotics and 100 mM glucose. Three colonies are subsequently picked into DRM with maintenance antibiotics and grown at 37°C with shaking to an OD_{600} ~0.4-0.6. Host

cells are then transferred into a 96-well plate in 500 µL aliquots, 10 mM arabinose is added to induce mutagenesis, and SP dilutions from prior passages (or starting phage stocks) are added. Cells are grown for 12-16 h at 37°C with shaking, and subsequent SP are isolated in the supernatant following centrifugation at 3,600 g for 10 min. To increase and diversify phage titers when necessary, SP were passaged in S2208s containing MP6; during such passages, cells were only infected for 6-8 h. Starting phage stocks for PANCE (N1) were diversified using this method prior to infection into the first PANCE passage. All SP titers were estimated by qPCR as described above.

4.4.6 Base editing-dependent PAM profiling assay

Chemicompetent 10B cells (New England Biolabs) were transformed with the base editor variants of interest. Three colonies of each base editor variant are seeded into 10 mL fresh DRM with maintenance antibiotic and grown at 37°C with shaking to an OD_{600} ~0.4-0.6. Upon reaching the desired cell density, cells were spun down at 5,000 xg for 10 minutes, washed 3x with ice-cold 10% (v/v) glycerol, then resuspended in a final volume of 100 µL 10% glycerol. 1 ug of library plasmid (pTPH424) was added to these 100 µL aliquots, then transformed in 25 µL aliquots using bacterial program X_5 in the 96-well Shuttle Device component of a 4D-Nucleofector system. Transformed cells were immediately transferred to 1.5 mL (per 100 µL cells) of prewarmed SOC media. A serial dilution of the transformed cells (8 dilutions, 5-fold each, starting with undiluted cells) was immediately taken and plated on maintenance antibiotics, which was used to calculate effective library size. The remaining cells are allowed to recover at 37°C with shaking for 15 min, then diluted into 40 mL of prewarmed DRM containing maintenance antibiotics and 10 mM arabinose. Induced cells are then grown at 37°C with shaking for 22 h (ABE-PPA) before being harvested by centrifugation at 3,600 xg for 10 min. DNA is isolated from harvested cells using a Plasmid Plus Midi Kit (Qiagen).

4.4.7   High-throughput DNA sequencing of BE-PPA libraries

Library samples were prepared for high-throughput amplicon sequencing in two PCR

steps. The first PCR (PCR1) was performed using forward primer BE-PPA-Fw (5'-

ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAATACGCAACGCCTCTC-3') and

reverse primer BE-PPA-Rv (5'-

TGGAGTTCAGACGTGTGCTCTTCCGATCCTTGTCTGTAAGCGGATGC-3') at a 150 μL scale

and 1 ug of template DNA. Cycling conditions were as follows: 98°C for 2 min, then 14 cycles of

[98°C for 15 s, 60°C for 15s, 72°C for 20s], and a final extension at 72°C for 2 min. 14 cycles for

PCR1 was observed to be within the linear amplification range for the libraries used in this study

but may change for alternate library constructions. Following PCR1, PCR reactions were

purified using the QIAquick PCR Purification Kit (Qiagen) and eluted in 16 μL nuclease-free

H_2O. The second PCR (PCR2) was performed using forward and reverse Illumina barcoding

primers at a 75 μL scale and half (8 μL) of the PCR1 purified product. Cycling conditions were

as follows: 98°C for 2 min, then 8 cycles of [98°C for 15 s, 60°C for 15s, 72°C for 20s], and a

final extension at 72°C for 2 min. 8 cycles for PCR2 was observed to be within the linear

amplification range for the libraries used in this study but may change for alternate library

constructions. PCR2 products were pooled, purified by electrophoresis with a 1% agarose gel

using a QIAquick Gel Extraction Kit (Qiagen), and eluted in nuclease-free H_2O. DNA

concentration was quantified with the KAPA Library Quantification Kit-Illumina (KAPA

Biosystems) and sequenced on an Illumina MiSeq instrument (paired-end read – R1: 210

cycles, R2: 0 cycles) according to the manufacturer's protocols.

4.4.8   Analysis of BE-PPA HTS data

Sequencing reads were demultiplexed using the Miseq Reporter (Illumina).

Demultiplexed files were subsequently analyzed for base editing activity using a custom

workflow combining the SeqKit[134] and CRISPResso2[96] packages. To analyze BE-PPA

sequenced files, the demultiplexed fastq files were filtered using the seqkit package/grep

function[134] to search for two flank sequences near either end of the amplicon. For ABE-PPA

profiled variants, groups of PAMs were UMI-tagged, and the specific UMI tag was used in place

of one of the flank sequences. Filtered files were then binned into individual fastq files per PAM

using the same function. The resulting PAM-specific fastq files were analyzed using standard

CRISPResso2[96] analysis.

4.4.9   Cell culture

HEK293T cells (ATCC CRL-3216) and HUH7 cells were cultured in Dulbecco's modified

Eagle's medium plus GlutaMax (DMEM, Thermo Fisher Scientific) supplemented with 10% (v/v)

fetal bovine serum (FBS, Thermo Fisher Scientific). Cells were cultured at 37°C with 5% CO_2.

Cell lines were authenticated by their suppliers and tested negative for mycoplasma.

4.4.10 Transfection protocols and genomic DNA isolation

HEK293T cells were seeded at a density of 2 x 10^4 cells per well on 96-well plates

(Corning) 16-20 h prior to transfection. Transfection conditions were as follows for HEK293T

cells: 0.5 µL Lipofectamine 2000 (Thermo Fisher Scientific), 250 ng of Cas effector plasmid

(nuclease/base editor), and 83 ng of guide RNA plasmid were combined and diluted with Opti-

MEM reduced serum media (Thermo Fisher Scientific) to a total volume of 10 µL and

transfected according to the manufacturer's protocol. Cells were transfected at approximately

60-80% confluency. Following transfection, cells were cultured for 3 days, after which the media

was removed, the cells washed with 1x PBS solution, and genomic DNA harvested via cell lysis

with 30 µL lysis buffer added per well (10 mM Tris-HCL, pH 8.0, 0.05% SDS, 20 ug/mL

Proteinase K (New England Biolabs)). The cell lysis mixture was allowed to incubate for 1-2 h at

37°C before being transferred to 96-well PCR plates and enzyme inactivated for 30 min at 80°C.

The resulting genomic DNA mixture was stored at -20°C until further use.

4.4.11 High-throughput sequencing of genomic DNA

High-throughput sequencing of genomic DNA from all cell lines was performed as previously described[30]. DNA concentrations were quantified with a Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific) or with a NanoDrop One Spectrophotometer (Thermo Fisher Scientific) prior to sequencing on an Illumina MiSeq instrument (paired-end read – R1: 250-280 cycles, R2: 0 cycles) according to the manufacturer's protocols.

4.4.12 High-throughput sequencing data analysis

Individual sequencing runs were demultiplexed using the MiSeq Reporter (Illumina). Subsequent demultiplexed sequencing reads were analyzed using CRISPResso2[96] as described previously[30]. All editing values are representative of n = 3 independent biological replicates, with mean±SEM shown.

References

1. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* **93**, 1156-1160, doi:10.1073/pnas.93.3.1156 (1996).
2. Wolfe, S. A., Nekludova, L. & Pabo, C. O. DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct* **29**, 183-212, doi:10.1146/annurev.biophys.29.1.183 (2000).
3. Bibikova, M., Beumer, K., Trautman, J. K. & Carroll, D. Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**, 764, doi:10.1126/science.1079512 (2003).
4. Miller, J. C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* **25**, 778-785, doi:10.1038/nbt1319 (2007).
5. Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763, doi:10.1126/science.1078395 (2003).
6. Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326**, 1509-1512, doi:10.1126/science.1178811 (2009).
7. Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501, doi:10.1126/science.1178817 (2009).
8. Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757-761, doi:10.1534/genetics.110.120717 (2010).
9. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology* **29**, 143-148, doi:10.1038/nbt.1755 (2011).
10. Mahfouz, M. M. *et al.* De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc Natl Acad Sci U S A* **108**, 2623-2628, doi:10.1073/pnas.1019533108 (2011).
11. Li, T. *et al.* TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res* **39**, 359-372, doi:10.1093/nar/gkq704 (2011).
12. 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).
13. Jansen, R., Embden, J. D., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**, 1565-1575, doi:10.1046/j.1365-2958.2002.02839.x (2002).
14. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**, E2579-2586, doi:10.1073/pnas.1208507109 (2012).
15. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).
16. Jeggo, P. A. in *Advances in Genetics* Vol. 38 (eds Jeffery C. Hall, Jay C. Dunlap, Theodore Friedmann, & Francesco Giannelli) 185-218 (Academic Press, 1998).
17. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and Cellular Biology* **14**, 8096-8106, doi:10.1128/mcb.14.12.8096 (1994).
18. Lukacsovich, T., Yang, D. & Waldman, A. S. Repair of a specific double-strand break generated within a mammalian chromosome by yeast endonuclease I-SceI. *Nucleic Acids Res* **22**, 5649-5657, doi:10.1093/nar/22.25.5649 (1994).
19. Yeh, C. D., Richardson, C. D. & Corn, J. E. Advances in genome editing through control of DNA repair pathways. *Nature Cell Biology* **21**, 1468-1478, doi:10.1038/s41556-019-0425-z (2019).

20. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine* **24**, 927-930, doi:10.1038/s41591-018-0049-z (2018).
21. Ihry, R. J. *et al.* p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nature Medicine* **24**, 939-946, doi:10.1038/s41591-018-0050-6 (2018).
22. Shen, M. W. *et al.* Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646-651, doi:10.1038/s41586-018-0686-x (2018).
23. van Overbeek, M. *et al.* DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Molecular Cell* **63**, 633-646, doi:https://doi.org/10.1016/j.molcel.2016.06.037 (2016).
24. Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotechnology* **32**, 267-273, doi:10.1038/nbt.2800 (2014).
25. Jasin, M. & Rothstein, R. Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol* **5**, a012740, doi:10.1101/cshperspect.a012740 (2013).
26. Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125-129, doi:10.1038/nature17664 (2016).
27. Rees, H. A., Yeh, W.-H. & Liu, D. R. Development of hRad51–Cas9 nickase fusions that mediate HDR without double-stranded breaks. *Nature Communications* **10**, doi:10.1038/s41467-019-09983-4 (2019).
28. Richardson, C. D., Ray, G. J., Dewitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature Biotechnology* **34**, 339-344, doi:10.1038/nbt.3481 (2016).
29. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766, doi:10.7554/eLife.04766 (2014).
30. Huang, T. P., Newby, G. A. & Liu, D. R. Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nature Protocols* **16**, 1089-1128, doi:10.1038/s41596-020-00450-9 (2021).
31. 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).
32. 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).
33. Kurt, I. C. *et al.* CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nature Biotechnology* **39**, 41-46, doi:10.1038/s41587-020-0609-x (2021).
34. Zhao, D. *et al.* Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat Biotechnol* **39**, 35-40, doi:10.1038/s41587-020-0592-2 (2021).
35. Koblan, L. W. *et al.* Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nature Biotechnology* **39**, 1414-1425, doi:10.1038/s41587-021-00938-z (2021).
36. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149-157, doi:10.1038/s41586-019-1711-4 (2019).
37. Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S. & Sternberg, S. H. Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature* **571**, 219-225, doi:10.1038/s41586-019-1323-z (2019).
38. Strecker, J. *et al.* RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48-53, doi:10.1126/science.aax9181 (2019).

39.   Vo, P. L. H. *et al.* CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nature Biotechnology* **39**, 480-489, doi:10.1038/s41587-020-00745-y (2021).

40.   Anzalone, A. V. *et al.* Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nature Biotechnology*, doi:10.1038/s41587-021-01133-w (2021).

41.   Ioannidi, E. I. *et al.* Drag-and-drop genome insertion without DNA cleavage with CRISPR-directed integrases. *bioRxiv*, 2021.2011.2001.466786, doi:10.1101/2021.11.01.466786 (2021).

42.   Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980-D985, doi:10.1093/nar/gkt1113 (2013).

43.   Newby, G. A. *et al.* Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature* **595**, 295-302, doi:10.1038/s41586-021-03609-w (2021).

44.   Geurts, M. H. *et al.* CRISPR-Based Adenine Editors Correct Nonsense Mutations in a Cystic Fibrosis Organoid Biobank. *Cell Stem Cell* **26**, 503-510.e507, doi:10.1016/j.stem.2020.01.019 (2020).

45.   Koblan, L. W. *et al.* In vivo base editing rescues Hutchinson–Gilford progeria syndrome in mice. *Nature* **589**, 608-614, doi:10.1038/s41586-020-03086-7 (2021).

46.   Nishimasu, H. *et al.* Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156**, 935-949, doi:https://doi.org/10.1016/j.cell.2014.02.001 (2014).

47.   Huang, T. P. *et al.* Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nature Biotechnology* **37**, 626-631, doi:10.1038/s41587-019-0134-y (2019).

48.   Jiang, W. *et al.* BE-PLUS: a new base editing tool with broadened editing window and enhanced fidelity. *Cell Research* **28**, 855-861, doi:10.1038/s41422-018-0052-4 (2018).

49.   Ma, Y. *et al.* Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nature Methods* **13**, 1029-1035, doi:10.1038/nmeth.4027 (2016).

50.   Hess, G. T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nature Methods* **13**, 1036-1042, doi:10.1038/nmeth.4038 (2016).

51.   Liu, L. D. *et al.* Intrinsic Nucleotide Preference of Diversifying Base Editors Guides Antibody *Ex Vivo* Affinity Maturation. *Cell Reports* **25**, 884-892.e883, doi:10.1016/j.celrep.2018.09.090 (2018).

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

53.   Grünewald, J. *et al.* Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433-437, doi:10.1038/s41586-019-1161-z (2019).

54.   Tan, J., Zhang, F., Karcher, D. & Bock, R. Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nature Communications* **10**, 439, doi:10.1038/s41467-018-08034-8 (2019).

55.   Pluciennik, A. *et al.* PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. *Proceedings of the National Academy of Sciences* **107**, 16066-16071, doi:10.1073/pnas.1010662107 (2010).

56.   Heller, R. C. & Marians, K. J. Replisome assembly and the direct restart of stalled replication forks. *Nature Reviews Molecular Cell Biology* **7**, 932-943, doi:10.1038/nrm2058 (2006).

57.   Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* **44**, D862-D868, doi:10.1093/nar/gkv1222 (2015).

58. Landrum, M. J. *et al.* ClinVar: improvements to accessing data. *Nucleic Acids Res*, doi:10.1093/nar/gkz972 (2019).
59. Lin, Q. *et al.* High-efficiency prime editing with optimized, paired pegRNAs in plants. *Nat Biotechnol* **39**, 923-927, doi:10.1038/s41587-021-00868-w (2021).
60. Choi, J. *et al.* Precise genomic deletions using paired prime editing. *Nature Biotechnology* **40**, 218-226, doi:10.1038/s41587-021-01025-z (2022).
61. Zhao, Z., Anselmo, A. C. & Mitragotri, S. Viral vector-based gene therapies in the clinic. *Bioeng Transl Med* **7**, e10258-e10258, doi:10.1002/btm2.10258 (2021).
62. Colella, P., Ronzitti, G. & Mingozzi, F. Emerging Issues in AAV-Mediated In Vivo Gene Therapy. *Mol Ther Methods Clin Dev* **8**, 87-104, doi:10.1016/j.omtm.2017.11.007 (2017).
63. Xu, Z. *et al.* Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnology* **13**, 87, doi:10.1186/1472-6750-13-87 (2013).
64. Kolb, A. F. Genome engineering using site-specific recombinases. *Cloning Stem Cells* **4**, 65-80, doi:10.1089/153623002753632066 (2002).
65. Peters Joseph, E., Makarova Kira, S., Shmakov, S. & Koonin Eugene, V. Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proceedings of the National Academy of Sciences* **114**, E7358-E7366, doi:10.1073/pnas.1709035114 (2017).
66. Anzalone, A. V., Koblan, L.W., Liu, D.R. Genome Editing with CRISPR-Cas Nucleases, Base Editors, Transposases, and Prime Editors. *Nature Biotechnology*, submitted (2020).
67. Zong, Y. *et al.* Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat Biotechnol* **35**, 438-440, doi:10.1038/nbt.3811 (2017).
68. Zeng, Y. *et al.* Correction of the Marfan Syndrome Pathogenic FBN1 Mutation by Base Editing in Human Cells and Heterozygous Embryos. *Mol Ther*, doi:10.1016/j.ymthe.2018.08.007 (2018).
69. Rees, H. A. & Liu, D. R. Base Editing: Precision Chemistry on the Genome and Transcriptome of Living Cells. *Nature Reviews Genetics* (2018).
70. 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).
71. Rees, H. A. *et al.* Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat Commun* **8**, 15790, doi:10.1038/ncomms15790 (2017).
72. Yan, F. *et al.* Highly Efficient A.T to G.C Base Editing by Cas9n-Guided tRNA Adenosine Deaminase in Rice. *Mol Plant* **11**, 631-634, doi:10.1016/j.molp.2018.02.008 (2018).
73. Kang, B. C. *et al.* Precision genome engineering through adenine base editing in plants. *Nat Plants* **4**, 427-431, doi:10.1038/s41477-018-0178-x (2018).
74. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* **44**, D862-868, doi:10.1093/nar/gkv1222 (2016).
75. 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).
76. Li, X. *et al.* Base editing with a Cpf1-cytidine deaminase fusion. *Nat Biotechnol* **36**, 324-327, doi:10.1038/nbt.4102 (2018).
77. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63, doi:10.1038/nature26155 (2018).
78. Hua, K., Tao, X., Yuan, F., Wang, D. & Zhu, J. K. Precise A.T to G.C Base Editing in the Rice Genome. *Mol Plant* **11**, 627-630, doi:10.1016/j.molp.2018.02.007 (2018).

79. Hua, K., Tao, X. & Zhu, J. K. Expanding the base editing scope in rice by using Cas9 variants. *Plant Biotechnol J*, doi:10.1111/pbi.12993 (2018).

80. Yang, L. *et al.* Increasing targeting scope of adenosine base editors in mouse and rat embryos through fusion of TadA deaminase with Cas9 variants. *Protein Cell* **9**, 814-819, doi:10.1007/s13238-018-0568-x (2018).

81. Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol* **36**, 843-846, doi:10.1038/nbt.4172 (2018).

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

83. 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).

84. Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259-1262, doi:10.1126/science.aas9129 (2018).

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

86. Gu, T. *et al.* Highly efficient base editing in Staphylococcus aureus using an engineered CRISPR RNA-guided cytidine deaminase. *Chem Sci* **9**, 3248-3253, doi:10.1039/c8sc00637g (2018).

87. 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).

88. 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).

89. Zafra, M. P. *et al.* Optimized base editors enable efficient editing in cells, organoids and mice. *Nat Biotechnol* **36**, 888-893, doi:10.1038/nbt.4194 (2018).

90. Zong, Y. *et al.* Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat Biotechnol*, doi:10.1038/nbt.4261 (2018).

91. Ryu, S. M. *et al.* Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat Biotechnol* **36**, 536-539, doi:10.1038/nbt.4148 (2018).

92. Oakes BL, F. C., Rishi H, Taylor KL, Ren SM, Nadler D, Yokoo R, Arkin A, Doudna JA, Savage DF. CRISPR-Cas9 circular permutants as programmable scaffolds for genome modification. *Cell* (2018).

93. Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867-871, doi:10.1126/science.aad8282 (2016).

94. Huai, C. *et al.* Structural insights into DNA cleavage activation of CRISPR-Cas9 system. *Nat Commun* **8**, 1375, doi:10.1038/s41467-017-01496-2 (2017).

95. 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).

96. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* **37**, 224-226, doi:10.1038/s41587-019-0032-3 (2019).

97. Miller, S. M. *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nature Biotechnology*, doi:10.1038/s41587-020-0412-8 (2020).

98. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290-296, doi:10.1126/science.aba8853 (2020).

99. Fedorova, I. *et al.* PpCas9 from Pasteurella pneumotropica — a compact Type II-C Cas9 ortholog active in human cells. *Nucleic Acids Res* **48**, 12297-12309, doi:10.1093/nar/gkaa998 (2020).

100. Mir, A., Edraki, A., Lee, J. & Sontheimer, E. J. Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. *ACS Chem Biol* **13**, 357-365, doi:10.1021/acschembio.7b00855 (2018).
101. Kleinstiver, B. P. *et al.* Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nature Biotechnology* **37**, 276-282, doi:10.1038/s41587-018-0011-0 (2019).
102. Xu, X. *et al.* Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Mol Cell* **81**, 4333-4345.e4334, doi:10.1016/j.molcel.2021.08.008 (2021).
103. Edraki, A. *et al.* A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for In Vivo Genome Editing. *Mol Cell* **73**, 714-726.e714, doi:10.1016/j.molcel.2018.12.003 (2019).
104. Liu, Z. *et al.* Efficient and high-fidelity base editor with expanded PAM compatibility for cytidine dinucleotide. *Science China Life Sciences* **64**, 1355-1367, doi:10.1007/s11427-020-1775-2 (2021).
105. Miller, S. M., Wang, T. & Liu, D. R. Phage-assisted continuous and non-continuous evolution. *Nature protocols* **15**, 4101-4127, doi:10.1038/s41596-020-00410-3 (2020).
106. Wong, B. G., Mancuso, C. P., Kiriakov, S., Bashor, C. J. & Khalil, A. S. Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. *Nat Biotechnol* **36**, 614-623, doi:10.1038/nbt.4151 (2018).
107. 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).
108. Shams, A. *et al.* Comprehensive deletion landscape of CRISPR-Cas9 identifies minimal RNA-guided DNA-binding modules. *Nature Communications* **12**, 5664, doi:10.1038/s41467-021-25992-8 (2021).
109. Richter, M. F., Zhao, K.T., Eton, E., Lapinaite, A., Newby, G.A., Thuronyi, B.W., Wilson, C., Zeng, J., Bauer, D.E., Doudna, J.A, Liu, D.R. Continuous evolution of an adenine base editor with enhanced Cas domain compatibility and activity. *Nature Biotechnology*, in press (2020).
110. Thuronyi, B. W. *et al.* Continuous evolution of base editors with expanded target compatibility and improved activity. *Nature Biotechnology* **37**, 1070-1079, doi:10.1038/s41587-019-0193-0 (2019).
111. Shah, N. H. & Muir, T. W. Inteins: Nature's Gift to Protein Chemists. *Chemical science* **5**, 446-461, doi:10.1039/C3SC52951G (2014).
112. Gogarten, J. P., Senejani, A. G., Zhaxybayeva, O., Olendzenski, L. & Hilario, E. Inteins: structure, function, and evolution. *Annu Rev Microbiol* **56**, 263-287, doi:10.1146/annurev.micro.56.012302.160741 (2002).
113. Zettler, J., Schütz, V. & Mootz, H. D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett* **583**, 909-914, doi:10.1016/j.febslet.2009.02.003 (2009).
114. Wang, T., Badran, A. H., Huang, T. P. & Liu, D. R. Continuous directed evolution of proteins with improved soluble expression. *Nat Chem Biol* **14**, 972-980, doi:10.1038/s41589-018-0121-5 (2018).
115. Brissette, J. L., Weiner, L., Ripmaster, T. L. & Model, P. Characterization and sequence of the Escherichia coli stress-induced psp operon. *J Mol Biol* **220**, 35-48, doi:10.1016/0022-2836(91)90379-k (1991).
116. Chen, F. *et al.* Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat Commun* **8**, 14958, doi:10.1038/ncomms14958 (2017).
117. DeBenedictis, E. A. *et al.* Systematic molecular evolution enables robust biomolecule discovery. *Nature Methods* **19**, 55-64, doi:10.1038/s41592-021-01348-4 (2022).

118. Zhong, Z. *et al.* Automated Continuous Evolution of Proteins in Vivo. *ACS Synthetic Biology* **9**, 1270-1276, doi:10.1021/acssynbio.0c00135 (2020).

119. Marshall, R. *et al.* Rapid and Scalable Characterization of CRISPR Technologies Using an E. coli Cell-Free Transcription-Translation System. *Mol Cell* **69**, 146-157.e143, doi:10.1016/j.molcel.2017.12.007 (2018).

120. Jung, C. *et al.* Massively Parallel Biophysical Analysis of CRISPR-Cas Complexes on Next Generation Sequencing Chips. *Cell* **170**, 35-47.e13, doi:10.1016/j.cell.2017.05.044 (2017).

121. Leenay, R. T. *et al.* Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Mol Cell* **62**, 137-147, doi:10.1016/j.molcel.2016.02.031 (2016).

122. Arbab, M. *et al.* Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. *Cell* **182**, 463-480.e430, doi:10.1016/j.cell.2020.05.037 (2020).

123. Zhang, Y., Rajan, R., Seifert, H. S., Mondragón, A. & Sontheimer, E. J. DNase H Activity of Neisseria meningitidis Cas9. *Mol Cell* **60**, 242-255, doi:10.1016/j.molcel.2015.09.020 (2015).

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

125. Sun, W. *et al.* Structures of Neisseria meningitidis Cas9 Complexes in Catalytically Poised and Anti-CRISPR-Inhibited States. *Mol Cell* **76**, 938-952.e935, doi:10.1016/j.molcel.2019.09.025 (2019).

126. Carvajal-Vallejos, P., Pallissé, R., Mootz, H. D. & Schmidt, S. R. Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. *J Biol Chem* **287**, 28686-28696, doi:10.1074/jbc.M112.372680 (2012).

127. Pinto, F., Thornton, E. L. & Wang, B. An expanded library of orthogonal split inteins enables modular multi-peptide assemblies. *Nature Communications* **11**, 1529, doi:10.1038/s41467-020-15272-2 (2020).

128. Gong, S., Yu, H. H., Johnson, K. A. & Taylor, D. W. DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Reports* **22**, 359-371, doi:https://doi.org/10.1016/j.celrep.2017.12.041 (2018).

129. Ivanov, I. E. *et al.* Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. *Proceedings of the National Academy of Sciences* **117**, 5853-5860, doi:10.1073/pnas.1913445117 (2020).

130. Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* **47**, W171-W174, doi:10.1093/nar/gkz365 (2019).

131. Lennermann, D., Backs, J. & van den Hoogenhof, M. M. G. New Insights in RBM20 Cardiomyopathy. *Curr Heart Fail Rep* **17**, 234-246, doi:10.1007/s11897-020-00475-x (2020).

132. 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).

133. Bosley, A. D. & Ostermeier, M. Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomol Eng* **22**, 57-61, doi:10.1016/j.bioeng.2004.11.002 (2005).

134. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE* **11**, e0163962, doi:10.1371/journal.pone.0163962 (2016).

135. Gaudelli, N. M. *et al.* Directed evolution of adenine base editors with increased activity and therapeutic application. *Nature Biotechnology* **38**, 892-900, doi:10.1038/s41587-020-0491-6 (2020).

136. Blum, T. R. *et al.* Phage-assisted evolution of botulinum neurotoxin proteases with reprogrammed specificity. *Science* **371**, 803-810, doi:10.1126/science.abf5972 (2021).
137. Packer, M. S., Rees, H. A. & Liu, D. R. Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nature Communications* **8**, 956, doi:10.1038/s41467-017-01055-9 (2017).
138. Zhang, W. *et al.* In-depth assessment of the PAM compatibility and editing activities of Cas9 variants. *Nucleic Acids Res* **49**, 8785-8795, doi:10.1093/nar/gkab507 (2021).
139. Collias, D. & Beisel, C. L. CRISPR technologies and the search for the PAM-free nuclease. *Nature Communications* **12**, 555, doi:10.1038/s41467-020-20633-y (2021).
140. Tracewell, C. A. & Arnold, F. H. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr Opin Chem Biol* **13**, 3-9, doi:10.1016/j.cbpa.2009.01.017 (2009).
141. Oakes, B. L. *et al.* Multi-reporter selection for the design of active and more specific zinc-finger nucleases for genome editing. *Nature Communications* **7**, 10194, doi:10.1038/ncomms10194 (2016).
142. Goldberg, G. W. *et al.* Engineered dual selection for directed evolution of SpCas9 PAM specificity. *Nature Communications* **12**, 349, doi:10.1038/s41467-020-20650-x (2021).
143. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nature Reviews Genetics* **16**, 379-394, doi:10.1038/nrg3927 (2015).