Expanding the Capabilities of Genome Editing

Expanding the capabilities of genome editing

A dissertation presented by

Holly A. Rees

to

The Committee on Higher Degrees in Chemistry and Chemical Biology

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemistry and Chemical Biology

Harvard University

Cambridge, Massachusetts,

March 2019

Copyright Page

Abstract

The ability to precisely and efficiently edit DNA sequences within the genome of living cells has been a major goal of the life sciences since the first demonstration of restriction cloning. Recently, the use of RNA-programmable nucleases from CRISPR systems has been transformative for researchers seeking to knock out genes to interrogate their function. Nucleases stimulate gene knockout by generation of a double-stranded DNA break (DSB) in the target gene, a process that leads to incorporation of mutations – mostly in the form of insertions or deletions (indels) - around the cleavage site. The inherent stochasticity of this process means that DSB-based methods do not enable precise DNA changes with single nucleotide resolution. When such precise changes are required, researchers can supply a donor DNA that contains the desired mutation and homology arms surrounding the cleavage site. Through the cellular process of homology-directed repair (HDR), the desired mutation may be incorporated into the genome. However, because it requires a DSB, this process produces a great excess of undesired indels and is low in efficiency.

Previous work in the Liu lab established that a catalytically disabled Cas9 nucleases (a nickase, capable of cleaving only one of the two DNA strands) can be fused to a cytosine deaminase enzyme and a uracil DNA glycosylase inhibitor protein to generate a cytosine base editor (CBE), a construct capable of converting a cytosine to a uracil in genomic DNA. Upon DNA replication or repair, the uracil is repaired to a thymine, so CBEs convert a cytosine to a thymine (C-to-T). This tool was transformative thanks to its high efficiency in generation of C-to-T mutations and low prevalence of undesired additional mutations.

In this thesis, I begin by addressing two of the challenges associated with the original CBE - DNA target specificity and delivery. First, we engineer and characterize a high-fidelity version of the CBE that reduces undesired editing at off-target loci in the genome. Second, we establish a method to deliver the cytosine base editor as a ribonucleoprotein complex, and demonstrate its utility for editing DNA in the mouse inner ear.

Next, I present our characterization of a new class of base editor – the adenine base editor (ABE) – an editor that is capable of converting adenosine to inosine in the context of genomic DNA. Inosine is converted to guanine by DNA replication or repair. We characterize the DNA and RNA specificity of ABE, demonstrating that the ABE is surprisingly specific to its target site in DNA but can edit RNA in a non-targeted fashion at a low level. We directly compare the ABE to a recently developed method for HDR. The adenine base editor is capable of incorporating the tested point mutations with an >1000-fold improvement in the ratio of desired product: undesired insertion than HDR.

Finally, I developed and characterized a new method for performing HDR that does not rely on generation of DSBs. Instead, we develop a fusion construct between hRad51 and Cas9(D10A) nickase that can mediate efficient HDR at a nick site. This improves the HDR:indel ratio by up to 52-fold and maintains the efficiency of Cas9 DSB-based HDR. I discover that mutants of hRad51 unable to bind to its native binding partners (BRCA2 or hRad51) further improve the ratio of desired product: undesired insertion or overall efficiency of the tool.

Table of Contents

List of Figures

List of Tables

Abbreviations

AAV	Adeno-Associated Virus
ABE	Adenine Base Editor
ADAR	Adenosine Deaminase, Rna Specific
AID	Activation-Induced Cytidine Deaminase
BE	Base Editor
bp	Base Pair
BzG	Benzylguanine
CAMERA 2	CRISPR-mediated Analog Multi-Event Recording Apparatus" 2
Cas	CRISPR-associated
CBE	Cytosine Base Editor
CDA	Cytidine Deaminase
CFD	Cutting Frequency Determinant
DNA	Deoxyribonucleic Acid
DOMINO	DNA-Based Ordered Memory And Iteration Network Operator
DSB	Double-Stranded DNA Break
dsDNA	Double-Stranded DNA
FACS	Fluorescence-Activated Cell Sorting
GFP	Green Fluorescent Protein
GMO	Genetically Modified Organism
HDR	Homology-Directed Repair
hiPS	Human Induced Pluoripotent Stem
HSV	Herpes Simplex Virus
HTS	High-Throughput Sequencing
indels	Insertion And Deletions
MMEJ	Microhomology-Mediated End Joining
mRNA	Messenger RNA
NHEJ	Non-Homologous End Joining
NLS	Nuclear Localization Sequence
PACE	Phage-Assisted Continuous Evolution
PAM	Protospacer Adjacent Motif
PCR	Polymerase Chain Reaction
PFS	Protospacer Flanking Sequence
RDN	hRad51–Cas9(D10A) Nickase
REPAIR	Rna Editing for Programmable A-to-I Replacement
RNA	Ribonucleic Acid
RNP	Ribonucleoprotein
sgRNA	Single Guide Rna
SNP	Single Nucleotide Polymorphism
SSB	Single-Stranded DNA Break
ssDNA	Single-Stranded DNA
ssODN	Single Stranded Oligonucleotide
tsAAV	Trans-Splicing Aav
UGI	Uracil DNA Glycosylase
UNG	Uracil N-Glycosylase
WGS	Whole Genome Sequencing

Acknowledgements

The fact that PhDs are sequentially awarded to single individuals is something that is radically at odds with my experience as a graduate student. This dissertation is a compilation of work that could never have been conceived, let alone performed, without the help and advice of countless other people. The award of a degree to a single person seems unreflective of the collaborative nature of this effort.

Thank you to Professor David Liu, who has supported my scientific learning process from the day that I began working at Harvard. As well as providing us with wonderful lab space and financial support, Professor Liu has driven the completion of this thesis with his scientific insights and advice. I am immensely grateful for how he has created a supportive and collaborative lab environment, filled with wonderful colleagues.

Thank you to my family. To my Mum and Dad for their unwavering support, encouragement and enthusiasm throughout my life and, by extension, my education. From their examples I have learnt countless lessons. I think the most important lesson is that all people are equal, a value that has inspired me to strive for confidence and persistence and enjoy happiness in my daily endeavors. It is impossible to express to anyone except from my brother how grateful I am to my parents. To my brother, Peter, who has always challenged me to think differently about every aspect of our perception of culture and reality. It is easy to forget how incredibly fortunate I am to have such reliable and selfless people as my family.

Thank you to the Wayland E. Noland Fund in Chemistry and Chemical Biology and the Kilpatrick Educational fund for financial support.

Thank you to my GAC and DAC members – Professors. Keith Joung and Jack Szostak. I am extremely fortunate to have their scientific insight and support throughout my time in graduate school.

Thank you to my mentors, Professor Alexis Komor, Dr. Nicole Gaudelli and Dr. Ahmed Badran. These three mentors have shaped my scientific understanding and focus. Ahmed took

me on as a rotation student in 2014, and was a fantastic teacher who good-naturedly invited me to perform all kinds of molecular biology experiments, despite my lack of experience, speed and skill. Alexis was the driving force behind the conception of the high fidelity editor project. She taught me the critical skills involved with how to address an important but solvable scientific problem. Nicole helped me to expand on these skills, showing me by example how to address problems that many would believe to be impossible to solve. I have detailed more about the support of my mentors before each chapter.

Thank you to Ariel Yeh. I am immensely grateful to have been her rotation mentor four years ago. Although, on paper, I was supposed to be Ariel's mentor, our scientific relationship very quickly became a partnership. Ariel's scientific advice, technical brilliance and great help has been invaluable throughout the construction of the work detailed in this thesis. Furthermore, she is a wonderful friend.

Thank you to Dr. Jon Levy, Dr. Johnny Hu, Dr. Michael Packer, Shannon Miller and Dr. Michelle Richter. I have shared a bench or bay with all of these people throughout my time in the Liu research group. All of them have offered invaluable scientific advice and insights whenever I have been struggling, and have gamely suffered my running-related chit chat. John has offered so much scientific insight throughout my PhD, I can't thank him enough. He has inspired me to do the very best experiments that I can.

Thank you to Dr. Michael Packer, who trusted me with isolation of murine cells for a critical experiment. Separately, Michael taught me and helped me to troubleshoot and understand DNA analysis and alignments. with me, how to write custom alignment programs for my peculiar applications. Subsequently, thank you to Dr. Kendell Clement and Prof. Luca Pinello for listening to my concerns and requirements regarding complex analysis of high throughput sequencing reads.

Thank you to Ally Freedy. I was incredibly fortunate to be paired with Ally when she did a research rotation in the Liu lab. She has remained a great friend, and I am delighted that she

will be the next President of the Harvard Women in Chemistry Society (HWIC) after my graduation. I am grateful to the Chemistry and Chemical Biology department for their continued support of HWIC.

Thank you to every member of the Liu lab. I have been lucky to benefit from scientific insights and advice from extremely varied backgrounds, thanks to both our subgroup meetings and general conversations. In particular, I am grateful to Dr. Tina Wang, Dr. Bill Kim, Kevin Zhao, Beverly Mok, Dr. Zhen Chen, Dr. Weixin Tang, Dr. Julian Willis, Dr. Manda Arbab, Luke Koblan, Aditya Raguram and Dr. Ben Thuronyi for their suggestions and advice. I am excited to come to the lab every day because of the wonderful colleagues that are here.

Thank you to the running community of Cambridge and beyond. First, thank you to the Liu lab, in particular Alix Chan, Dr. Bill Kim, Dr. Philip Lichtor and Alex Peterson for sparking my enthusiasm for competition in the lab 5k, 2015. Thank you to Battle Road Track Club and New Balance for supporting my running and training ever since. In particular, thank you to my coach, Ryan Carrara, and my team mates for their constant and unwavering support, on and off the track.

Finally, thank you to Brian Crowley. I am incredibly grateful for the miles we have covered over the last year, and I am looking forward to the future.

Chapter One: Introduction to base editing

This chapter has been adapted from:

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

Contributions:

This chapter was written with Professor David R. Liu. Helpful advice and suggestions were received from J.K. Joung, F. Zhang, A. Raguram, W.-H. Yeh, T. Huang, K. Zhao and W. Tang.

I am extremely grateful to David Liu to have had the opportunity to write this chapter and the referenced article. Throughout the process of writing, rewriting and revising this article, I aspired to improve my writing skills for precision and clarity. David Liu both inspires and teaches excellence, and I am exceptionally grateful to have had such a patient mentor.

1.1 Overview

In 1973, the development of restriction cloning[1] highlighted the impressive power of targeted DNA manipulation techniques. Recently, RNA-programmable CRISPR-associated (Cas) nucleases have contributed to the pursuit of this goal[2-4] through their ability to generate a double-stranded DNA break (DSB) at a precise target location in the genome of a wide variety of cells and organisms[5-8] (reviewed extensively[9-12]). Catalytically inactivated Cas nucleases are also useful as programmable DNA-binding proteins that localize tethered proteins to target DNA loci[2,13-16].

Generation of a DSB does not directly lead to DNA editing; rather, editing following nuclease treatment occurs as a result of cellular responses to DSBs. Processes including non-homologous end-joining (NHEJ) and microhomology-mediated end-joining (MMEJ) can lead to gene disruption through the introduction of insertions, deletions, translocations, or other DNA rearrangements at the site of a DSB[17-19]. Alternatively, a precise DNA edit can be made by supplying a donor DNA template encoding the desired DNA change flanked by sequence homologous to the region upstream and downstream of the DSB. Cellular homology-directed repair (HDR) then results in the incorporation of sequence from the exogenous DNA template at the DSB site[20,21]. Although HDR is a flexible tool with the ability to make precise insertions, deletions, or any point mutation of interest, HDR is largely restricted to the G2 and S phases of the cell cycle, limiting efficient HDR to actively dividing cells, and even in cultured cell lines HDR efficiency can be modest[22-24]. Moreover, NHEJ and HDR are competing processes, and under most conditions NHEJ is more efficient than HDR. Thus, a majority of edited products will usually contain small insertions or deletions[24,25].

In mammalian cells, DSB-induced NHEJ is an effective way to disrupt a gene of interest. However, more reliable techniques that generate precise DNA or RNA modifications are necessary to make comparisons between alleles, study the effects of specific mutations within genes, or to treat genetic disease through gene correction. Although sampling bias due to the

extensive use of short-read sequencing to analyze genomic diversity is possible[26,27], the largest class of known human pathogenic mutations, by far, is the point mutation (also called single nucleotide polymorphism (SNP)) (Figure 1.1a)[28,29]. Installing or reversing pathogenic SNPs efficiently and cleanly is thus of great interest for the study and treatment of genetic disorders, and requires a method to specifically change the sequence of an individual base pair within a vast genome.

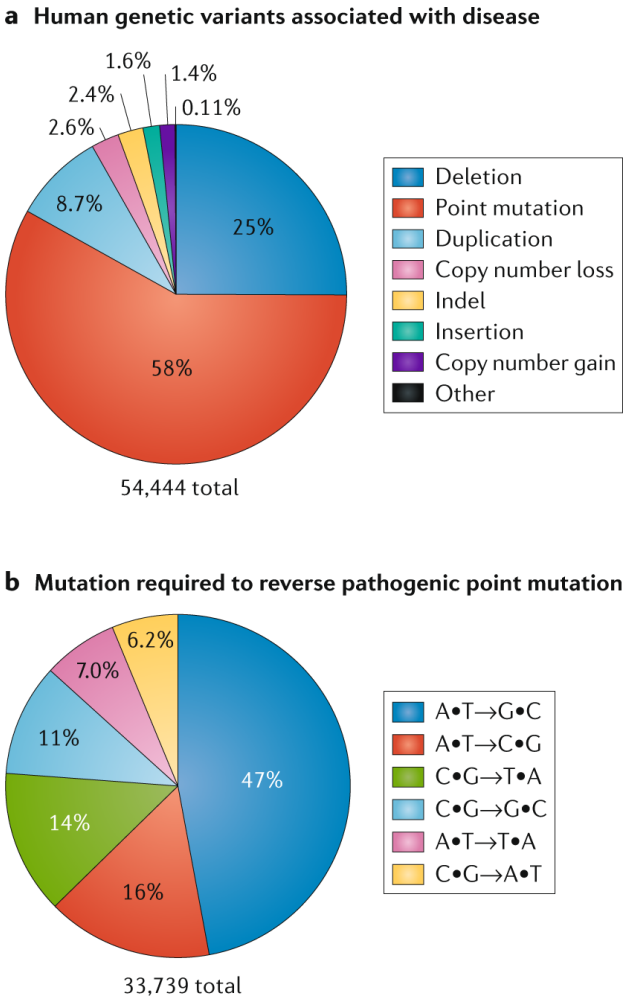


Figure 1.1: Distribution of human pathogenic genetic variants, including point mutations. **a)** Classification of human pathogenic genetic variants in the ClinVar database (accessed May 29, 2018)[28,29]. As noted in the text, sampling bias due to the extensive use of short-read sequencing to analyze genomic diversity is possible. **b)** Distribution of base pair changes needed to reverse the pathogenic point mutations represented in the red wedge in (a)[28,29]. The percentage represented by each base pair change is noted on the pie chart; transition mutations are labeled in white and transversion mutations are labeled in black.

DSBs created by nucleases such as Cas9 result in indels, translocations, and rearrangements[27,30-32] that are undesired byproducts when attempting to install a point mutation. Base editing is a genome editing method that directly generates precise point mutations in genomic DNA or in cellular RNA without directly generating DSBs, requiring a DNA donor template or relying on cellular HDR[33-35]. Since base editors do not normally create DSBs, they minimize the formation of DSB-associated byproducts[35,36]. Instead, DNA base editors (BEs) comprise fusions between a catalytically impaired Cas nuclease and a base-modification enzyme that operates on single-stranded DNA (ssDNA) but not double-stranded DNA (dsDNA). Upon binding to its target locus in DNA, base pairing between the guide RNA and target DNA strand leads to displacement of a small segment of single-stranded DNA in an "R-loop"[37]. DNA bases within this single-stranded DNA bubble are modified by the deaminase enzyme. To improve efficiency in eukaryotic cells, the catalytically disabled nuclease also generates a nick in the non-edited DNA strand, inducing cells to repair the non-edited strand using the edited strand as a template[33-35].

Two classes of DNA base editor have been described: cytosine base editors (CBEs) convert a C•G base pair into a T•A base pair[33,34,38], and adenine base editors (ABEs) convert an A•T base pair to a G•C base pair. Collectively, CBEs and ABEs can mediate all four possible transition mutations (C to T, A to G, T to C, and G to A) (Figure 1.1b)[35,39]. In RNA, targeted adenosine conversion to inosine has also been developed using both antisense[40-49] and Cas13-guided[39] RNA-targeting methods. In this Review, we describe the development of DNA and RNA base editors, their capabilities and limitations, and their current and future applications.

1.2 Development of cytosine base editors

The first DNA base editors convert a C•G base pair to a T•A base pair by deaminating the exocyclic amine of the target cytosine to generate uracil (Figure 1.2a). To localize

deamination activity to a small target window within the mammalian genome, Liu and coworkers used an APOBEC1 cytidine deaminase, that accepts ssDNA as a substrate but is incapable of acting on dsDNA[50]. Fusion of APOBEC1 to dead Cas9 from *Streptococcus pyrogenes* (dCas9, a mutant of Cas9 containing D10A and H840A) resulted in base editor 1 (BE1) (Table 1.1)[33]. When bound to its cognate DNA, dCas9 performs local denaturation of the DNA duplex to generate an R-loop in which the DNA strand not paired with the guide RNA exists as a disordered single-stranded bubble[2,37]. This feature enables BE1 to perform efficient and localized cytosine deamination in a test tube, with deamination activity restricted to a ~5-bp window of ssDNA (positions ~4-8, counting the protospacer adjacent motif (PAM) as positions 21-23) generated by dCas9. Fusion to dCas9 presents the target site to APOBEC1 in high effective molarity, enabling BE1 to deaminate cytosines located in a variety of different sequence motifs, albeit with differing efficacies[33] (Figure 1.2b).

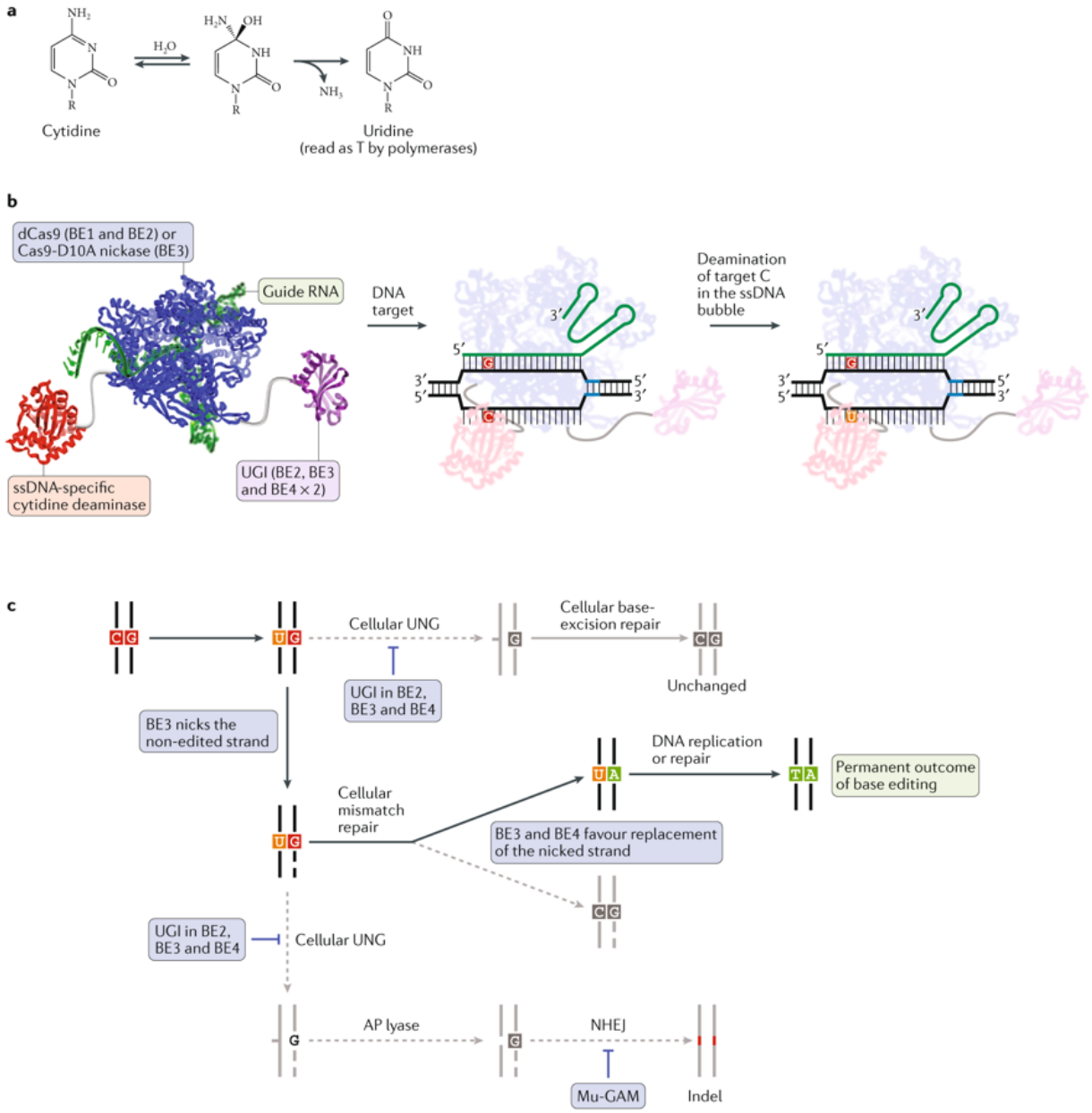


Figure 1.2: Cytosine base editing. (a) Cytosine deamination generates uracil, that base pairs as thymidine. R = 2′-deoxyribose in DNA, or ribose in RNA. **(b)** Cytosine base editing strategy by BE1, BE2, BE3, or BE4. R-loop formation exposes a region of single-stranded DNA to the cytidine deaminase domain. Target cytosines in this region are deaminated to uracil[33,36]. **(c)** Cellular response to cytosine base editing. Uracil DNA glycosylase-mediated excision of the uracil generated in genomic DNA is inhibited by BE2, BE3, and BE4. BE3 and BE4 are designed to nick the non-edited strand (containing the G of the original C•G target base pair), stimulating cellular DNA repair of that strand to replace the G with an A, completing the conversion of the original C•G base pair to a U•A or, following DNA replication or repair to a T•A base pair[33,36].

A major challenge for the use of base editors in mammalian cells is circumventing DNA repair processes that oppose target base pair conversion. Although BE1 mediates efficient, RNA-programmed deamination of target cytosines *in vitro*, it is not effective in human cells (deamination efficiency fell from 25-40% *in vitro* to 0.8-7.7% in cells)[33]. This decrease is largely due to effective cellular repair of the U•G intermediate in DNA[51]. Base excision repair (BER) of U•G in DNA is initiated by uracil N-glycosylase (UNG), that recognizes the U•G mismatch and cleaves the glyosidic bond between uracil and the deoxyribose backbone of DNA. BER will usually result in the reversion of the U•G intermediate created by BE1 back to a C•G base pair (Figure 1.2c)[51,52]. To inhibit UNG, Liu and co-workers fused uracil DNA glycosylase inhibitor (UGI), a small protein from bacteriophage PBS*,* to the C-terminus of BE1, generating BE2. UGI is a DNA mimic that potently inhibits both human and bacterial UNG[53]. BE2 mediates efficient base editing in bacterial cells[54] and moderately efficient editing in mammalian cells, enabling conversion of a C•G base pair to a T•A base pair through a U•G intermediate (Figure 1.2c)[33].

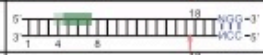
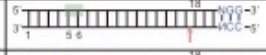
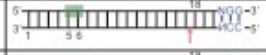
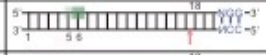
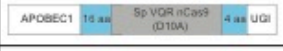
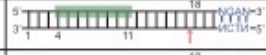
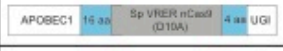
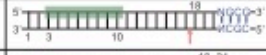
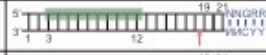
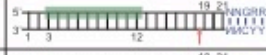
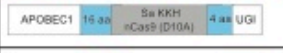
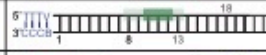
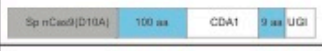
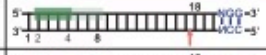
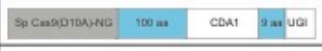
Base editor	Base editor architecture	Editing window and PAM
BE1		
BE2		
BE3		
HF–BE3		
BE4, BE4max		
BE4–GAM		
YE1–BE3		
EE–BE3		
YE2–BE3		
YEE–BE3		
VQR–BE3		
VRER–BE3		
Sa–BE3		
Sa–BE4		
SaBE4–Gam		
SaKKH–BE3		
Cas12a–BE		
Target–AID		
Target-AID-NG		
xBE3		
eA3A–BE3		
A3A-BE3		
BE–PLUS		
TAM		
CRISPR–X		
ABE7.9		
ABE7.10		
ABE7.10 *		
xABE		
ABESa		
VQR-ABE		
VRER-ABE		
SaKKH-ABE		

Table 1.1: DNA base editors and their approximate editing windows.
Cytosine base editor activity windows are shown in green (dark green indicates higher editing efficiency; light green denotes lower editing efficiency). The activity windows of editors that convert C to random mixtures of A, G or T are shown as a rainbow. Adenine base editors have windows shown in red. The location of an induced nick in the target DNA backbone is indicated by an arrow. Base positions are numbered relative to the PAM-distal end of the guide RNA; for example, the NGG PAM sequence of SpCas9 is numbered 21-23.

The base editing efficiency of BE2 is limited by its ability to edit only one strand of DNA. To direct cellular replacement of the G present in the non-targeted strand of DNA with A, Liu and co-workers designed third-generation base editors (BE3) that specifically nick the non-edited DNA strand (Figure 1.2c). Nicking the non-edited DNA strand biases cellular repair of the U•G mismatch to favor a U•A outcome, greatly elevating base editing efficiencies in mammalian cells. Restoration of His 840 in dCas9 generates a base editor that uses Cas9 "nickase" (D10A) instead of dCas9, resulting in nicking of the non-edited DNA strand (Figure 1.2c). The APOBEC1–Cas9 nickase–UGI fusion (BE3) yielded efficient editing in mammalian cells, averaging 37% across six loci in the initial report[33]. Notably, although indels are a detectable byproduct upon treatment with BE3, their frequency is typically small relative to the base edit (indel formation averaged 1.1% across the six reported loci), and much less frequent than indels induced by DSBs[33].

Nishida and co-workers described a similar system for cytosine base editing in yeast and mammalian cells, termed "Target-AID"[34]. In lieu of APOBEC1, they used the cytidine deaminase CDA1 in a Cas9 nickase–CDA1–UGI base editor construct named Target-AID. Target-AID displays a slightly shifted activity window relative to BE3 (Table 1.1). Nishida and co-workers also noted that base editing at certain bases is less precise than expected, demonstrating that C-to-G or C-to-A edits are, in some cases, significant byproducts of base editing[34], as was also observed with BE3[33]. Improvements to CBEs that minimize byproduct formation and increase editing efficiency are discussed below.

1.3 Development of adenine base editors

The distribution of pathogenic point mutations in living systems is not uniform across the six possible ways to exchange one base pair for another (Figure 1.1b). This uneven distribution is consistent with the relatively high rate of spontaneous cytosine deamination (estimated to be 100-500 deamination events per cell per day), that, if uncorrected, can mutate a G•C base pair to an A•T base pair[55,56]. A molecular machine capable of reversing such mutations by converting an A•T base pair into a G•C base pair is therefore of particular interest because it would enable correction of the most common type of pathogenic SNPs in the ClinVar database, representing ~47% of disease-associated point mutations (Figure 1.1b). Like cytosine, adenine contains an exocyclic amine that can be deaminated to alter its base pairing preferences. Deamination of adenosine yields inosine (Figure 1.3a). Although inosine in the third position of a tRNA anticodon is well-known to pair with A, U, or C in mRNA during translation, in the context of a polymerase active site, inosine exhibits the base-pairing preference of guanosine[57].

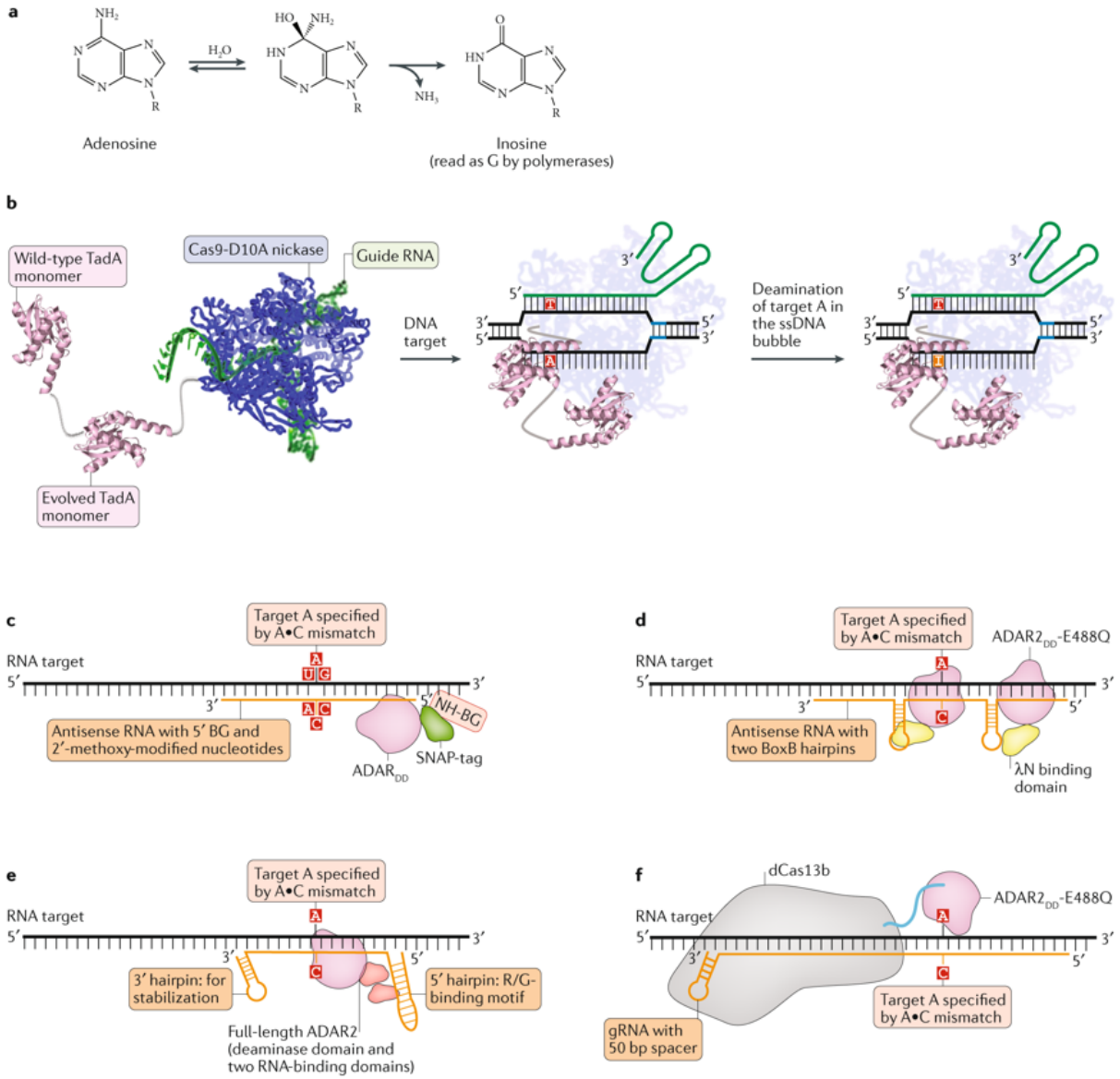


Figure 1.3: Adenine base editing in DNA and RNA. (a) Adenosine deamination generates inosine, which has the same base pairing preferences as a guanosine in the active site of a polymerase. R = 2′-deoxyribose in DNA, or ribose in RNA. **(b)** ABE-mediated DNA base editing to convert an A•T base pair to a G•C base pair. Current ABEs contain one wild-type TadA structural monomer and one evolved TadA catalytic monomer. R-loop formation exposes a small region of ssDNA, within which A is deaminated to I by the heterodimeric wild-type TadA–evolved TadA heterodimer.[35] **(c-e)** Antisense-RNA mediated RNA editing. ADAR variants are localized to the RNA transcript of interest through an antisense RNA with variable lengths of homology to the target transcript. The target A is specified with an A•C mismatch in the mRNA:antisense-RNA duplex. The antisense RNA is in orange, and the target mRNA is in black. **(c)** Antisense-directed RNA editing by covalent linkage of an ADAR deaminase domain ($ADAR_{DD}$)–SNAP tag fusion to a benzylguanine (BzG)-modified antisense RNA[47,49,58]. **(d)** ADAR-BoxB mediated RNA editing[40,41]. The $ADAR2_{DD}$ is fused to λN protein; λN binds to one of the two BoxB hairpins integrated into the antisense RNA, localizing ADAR-mediated deamination activity to the target adenine. **(e)** Full-length ADAR2-mediated RNA editing[43]. The

antisense RNA comprises a 5' R/G-binding motif hairpin, the native binding sequence for full length ADAR2, followed by a 19-nt antisense region complementary to the target RNA with the target adenine centrally located and specified by an A•C mismatch. Overexpression of full-length ADAR2 results in localization of the deaminase to the target base within the target transcript through the native ADAR2 double-stranded RNA binding domains. **(f)** RNA editing with REPAIR. A dCas13b:guide RNA complex is guided to a the target RNA by a 50-nt spacer. The target A is specified by an A•C mismatch centrally located within the 50-nt spacer.[39].

The major hurdle to the development of an ABE was the lack of any known adenosine deaminase enzymes capable of acting on ssDNA. Attempts to force RNA adenosine deaminases to act on DNA by installing them in place of APOBEC1 in BE3 resulted in no detectable adenine base editing[35]. To overcome this problem, Liu and co-workers evolved a deoxyadenosine deaminase enzyme that accepts ssDNA starting from an *Escherichia coli* tRNA adenosine deaminase enzyme, TadA[35]. *E. coli* cells were equipped with TadA mutants and defective antibiotic resistance genes. To grow in the presence of antibiotic, a mutant TadA–dCas9 fusion (TadA*–dCas9) must convert a deoxyadenosine to a deoxyinosine in the defective antibiotic resistance gene. Bacteria encoding TadA–dCas9 fusions capable of repairing the mutated resistance gene were isolated and then tested in a mammalian cell context.

Although TadA*–dCas9 fusions during this evolution and engineering process were capable of efficient A-to-I conversion in *E. coli*, simple TadA*–Cas9 nickase fusions resulted in only modest editing rates in mammalian cells. In its native (*E. coli*) context, TadA acts as a homodimer, with one monomer catalyzing deamination and the other monomer contributing to tRNA substrate binding[59]. In the *E. coli* selection, endogenous wild-type TadA could form dimers with the mutated TadA*–dCas9 construct *in trans*; however the absence of TadA in mammalian cells precludes TadA•TadA* heterodimerization. This challenge was addressed by engineering heterodimeric proteins that incorporate a wild-type non-catalytic TadA monomer, an evolved TadA* monomer and a Cas9 nickase (TadA–TadA*–Cas9 nickase) in a single polypeptide chain (Figure 1.3b). The single-chain heterodimeric construct greatly improved adenine base editing efficiency in mammalian cells when compared to the corresponding homodimeric TadA*–

TadA*–Cas9 nickase editor, suggesting that the mutations required to support deoxyadenosine deamination are incompatible with the structural role played by the N-terminal TadA monomer[35].

As with the CBEs, ABEs catalyze deamination within a small window of exposed ssDNA generated by Cas9:guide RNA binding to the target locus. ABE7.10, containing 14 amino acid substitutions in the catalytic TadA* domain, is the most efficient and sequence context-independent ABE reported to date, and performs A•T to G•C conversion within an editing window of protospacer positions ~4-7, counting the PAM as positions 21-23. Different ABE evolutionary relatives, such as ABE7.9 or ABE6.3, can offer higher editing efficiencies at positions closer to the PAM (such as positions 8 or 9; see Table 1.1). Together, ABEs represent powerful new tools that enable precise conversion of a target A•T base pair to G•C in the genomic DNA of living cells[35].

1.4 Base editing of RNA

Editing individual bases in RNA can also provide powerful capabilities for the life sciences and, potentially, for medicine. Due to its single-stranded nature, 12 possible base editors that operate on RNA, rather than six possible base editors that operate on dsDNA, are needed to cover all possible changes. To date, the only reported programmable oligonucleotide-directed transformation that changes Watson–Crick base pairing in RNA is deamination of A to I.

1.4.1 Antisense-oligonucleotide-directed A-to-I RNA editing

All RNA base editors characterized in mammalian cells thus far use adenosine deaminases from the ADAR family that natively catalyze hydrolytic adenosine deamination, converting an adenosine to an inosine[60,61]. Unlike most other RNA-editing enzymes, ADARs are not natively RNA-guided[62]. Instead, they contain a distinct RNA-binding domain that recognizes and localizes the enzyme to certain regions of double-stranded RNA[63,64].

Pioneering efforts by Stafforst, Rosenthal, Nakagawa, and their respective coworkers to generate a targetable adenine RNA editor tethered the catalytic domain of an ADAR enzyme to a guiding antisense RNA oligonucleotide[40-48]. These RNA editors rely on Watson–Crick base pairing between an antisense RNA and the target transcript to localize an ADAR deaminase domain ($ADAR_{DD}$) to the target RNA. At least three strategies have been developed to establish a physical linkage between the deaminase and the antisense RNA. First, fusing a SNAP tag to the ADAR and generating a benzylguanine-modified antisense RNA (BG-RNA)[65] enabled editing *in vitro*[48,58]. Delivery of the modified antisense RNA combined with overexpression of a SNAP–ADAR fusion in cells resulted in the covalent linkage between the SNAP-tagged ADAR and the antisense RNA[44-49] (Figure 1.3c). Second, appending the RNA-binding λ-phage N protein to the ADAR deaminase domain and fusing the antisense RNA with a 17-nt "Box B" hairpin that is bound by BoxB also enabled the association of the antisense RNA and the ADAR, enabling both the guiding RNA and deaminase construct to be genetically encoded[40,41] (Figure 1.3d). Third, Stafforst, Fukuda, and their respective coworkers showed that fusing the antisense RNA to the natural substrate for ADAR2 can localize ADAR2 to the antisense RNA for editing in cells[42,43] (Figure 1.3e).

Two key innovations improved the efficiency and specificity of these RNA-guided deamination systems. Stafforst and Schneider exploited the natural sequence preference of human ADAR1 and ADAR2, that preferentially deaminate an adenine that is mispaired with a cytosine in a double-stranded RNA substrate[66,67]. They designed an 17-nt antisense RNA sequence that placed a C opposite the target A to generate an A•C mismatch upon binding to the target RNA[48]. This use of the A•C mismatch to direct ADAR activity improved editing *in vitro* at the on-target adenine and in many of the motifs they tested, with no detectable editing at nearby adenines in the same RNA[48,58]. Rosenthal and coworkers combined the A•C mismatch strategy[40] with use of a hyperactive human ADAR2 mutant (E488Q) to further increase editing efficiency and demonstrated RNA editing in HEK293T cells[40], which was improved in efficiency

by using two BoxB recruitment domains[41] (Figure 1.3d). Despite these improvements, the use of antisense–deaminase conjugates remained challenging due to high rates of off-target deamination and strong context-dependent editing of adenine bases located in sequence motifs preferred by ADARs[40-42,44-48].

In the most recently reported antisense-guided RNA editing system, Stafforst and co-workers dramatically reduce off-target deamination that usually accompanies efficient RNA editing. They integrated an inducible SNAP–ADAR fusion construct into HEK293 cells and delivered chemically modified antisense 22-nt benzylguanine (BG)-linked RNAs by lipofection (Figure 1.3c)[49]. The SNAP tag spontaneously becomes covalently bound to the RNA. Editing efficiency was impressively high at six assayed endogenous target transcripts (15-90%) and could be multiplexed without efficiency loss. Significant improvements to the specificity of editing were also made through modifying all the nucleotides in the antisense RNA with a 2'-methoxy group other than the cytosine that specifies the target adenine through the previously described A•C mismatch and its two neighboring bases. This innovation minimized proximal off-target editing other than at adenine-rich triplet targets[49]. Distal, transcriptome-wide off-target editing was significant when hyperactive ADAR variants were used, but reduced to negligible levels with wild-type ADARs, although on-target editing rates were also lower with wild-type ADARs[49].

The most notable limitation of this method is its sequence context dependence; GAN (where N is any nucleotide) target sites are not efficiently edited with any assayed variant due to the native preference of ADAR1 and ADAR2. Future work may harness ADAR mutants, such as E488Q, that show a reduced sequence preference[68] into this system to overcome the targeting sequence limitation. For tolerated sequence motifs this approach represents a substantial improvement to efficiency and specificity of RNA editing when genomic integration of the RNA editor construct and delivery of a chemically modified antisense RNA can be performed[49].

1.4.2 Cas13-directed A-to-I RNA base editing

Zhang and co-workers developed a different approach to RNA-guided RNA base editing that uses a catalytically dead RNA-guided Cas13b enzyme (dPspCas13b) to localize an ADAR to the target RNA[39]. dPspCas13b is fused to the deamination domain (DD) of an ADAR ($ADAR_{DD}$) to generate an RNA-guided editor (Table 1.1; Figure 1.3f). This approach was termed RNA Editing for Programmable A-to-I Replacement (REPAIR)[39]. REPAIR incorporates two aspects of ADAR-mediated RNA editing described above: use of the hyperactive $ADAR2_{DD}$(E488Q) mutant, and specifying the target adenine with an A•C mismatch (Figure 1.3c-e)[39]. Notably, REPAIR may offer broad sequence context compatibility; when tested at all 16 possible NAN motifs in a luciferase reporter transcript, REPAIRv1 could edit all 16 codons, apparently overcoming the native ADAR preference through binding to the target site with high effective molarity[39].

Zhang and co-workers demonstrated that REPAIRv1 offers higher editing efficiency (89%) than two antisense-mediated strategies: BoxB-ADAR2 (50%)[40] and full-length ADAR2 (35%)[42] when targeted to a *Cluc* reporter transcript. However, in two endogenous transcripts tested with REPAIRv1, editing efficiency was reduced to 15-40%[39]. Transcriptome-wide RNA sequencing (RNA-seq) revealed that REPAIRv1 displays off-target editing that is comparable to that of the BoxB-ADAR strategy and significantly greater than that resulting from overexpression of full-length ADAR2[39]. Proximal off-target RNA base editing was also observed with REPAIRv1: adenine bases 50 bp up- or down-stream of the target adenine were edited at a frequency of approximately 10-20%[39]. Off-target RNA editing was attributed to overexpression of the hyperactive ADAR deaminase.

To improve the specificity of REPAIRv1, Zhang and coworkers introduced mutations into $ADAR2_{DD}$(E488Q) designed to reduce the binding affinity between $ADAR_{DD}$ and non-target cellular RNA. Using $ADAR2_{DD}$(E488Q/T375G) in the REPAIRv1 architecture resulted in REPAIRv2. In transcriptome-wide sequencing assays using a guide programmed to edit *Cluc*,

REPAIRv2 yielded only 20 detectable off-target editing events, a 900-fold improvement relative to REPAIRv1. Although still detected, REPAIRv2 also dramatically reduced proximal off-target editing in the 100-nt region upstream or downstream of the target adenine. As expected due to its higher specificity, on-target editing efficiencies of REPAIRv2 were reduced relative to REPAIRv1 (from 89% to approximately 45% in the *Cluc* reporter), and it is possible that the sequence targeting scope of REPAIRv2 is reduced compared to REPAIRv1. Nevertheless, its high specificity makes REPAIRv2 a promising tool for A-to-I RNA base editing in the mammalian transcriptome[39].

1.4.3 Cellular decoding of inosine in mRNA

In DNA base editing of deoxyadenosine, the resulting deoxyinosine is decoded by a DNA or RNA polymerase either during DNA replication or during transcription. Inosine in RNA is functionally decoded by different machinery, such as the ribosome (when in protein-coding regions) or the spliceosome (when in splice sites). Whereas there is strong evidence that deoxyinosine in DNA is read as a G in the active site of a polymerase in human cells[57], an inosine in the wobble position of a tRNA pairs with A, C or U in mRNA, enabling a single tRNA to decode multiple cognate codons[69]. Indeed, in miRNAs the reduced binding strength between the I:C base pair compared to the G:C base pair is thought to be biologically significant for directing mRNA decay[70].

Inosine's ability to form base pairs with multiple bases raises concern that an inosine in an mRNA might be decoded as a mixture of bases in the context of a ribosome or spliceosome active site. Known examples of natural A-to-I editing in the coding regions of mRNA suggest that editing to an inosine at codon position 1 or 2 results predominantly in the inosine being read as a guanine, both in cells[71] and in vitro[72]. For applications involving RNA editing to modulate splicing, observations are also consistent with the spliceosome reading an inosine as a guanine, as A-to-I editing can directly generate or destroy splice sites as if the I were a G[73,74].

1.5 Base editor limitations and improvements

1.5.1 Base editing product purity

Initial reports of CBEs identified that at some genomic loci, unanticipated C-to-non-T edits are observed, reducing base editing product purity[33,34,75-77]. Liu and coworkers investigated the determinants of base editing product purity by performing cytosine base editing in cells lacking various genes including *UNG*, encoding uracil N-glycosylase. In $UNG^{-/-}$ cells, product purity improved from an average of 68% to >98% across 12 target cytosines, indicating that UNG is required for byproduct formation[36]. This insight was used to improve base editing outcomes. Fusing a second UGI domain onto the C-terminus of BE3 improved the editing purity in UNG-containing cell lines, likely due to increased inhibition of UNG. In addition, installation of a more flexible set of linkers improved efficiency of editing to generate a fourth-generation editor, BE4 (Figure 1.2c; Table 1.1)[36]. Overexpression of UGI *in trans* with a BE3 also improves product purity and reduces indel formation in mammalian cells[78], but this may be accompanied by a global increase in C to T mutation rates[79,80].

In some cases, the ability of a CBE with no fused UGI to mutate a target C to a mixture of T, A, and G provides a useful system for targeted random mutagenesis. Bassik, Chang, and their respective coworkers developed two such systems that exploit C-to-non-T editing abilities of base editors for targeted mutagenesis in mammalian cells. These approaches, targeted AID-mediated mutagenesis[77] and CRISPR-X[75], have been reviewed extensively here[81].

Adenine base editing by ABE typically exhibits very high product purity; indeed, there are no reports of significant A-to-non-G edits to date[35,82-85], perhaps because of the much weaker ability of cells to remove inosine from DNA than uracil. Consistent with this potential explanation, the use of ABE in cells deficient in alkyl adenine DNA glycosylase (AAG), an enzyme known to recognize and remove inosine in DNA[86], did not improve editing efficiency[35].

1.5.2 Generation of indels

DNA base editing can yield a low but detectable rate of indel formation. Liu and co-workers noted that as well as improved product purity profiles, UNG-knockout cells displayed reduced indel formation[36]. This observation is consistent with a model in which UNG-mediated creation of an abasic site following C-to-U deamination can lead to nicking of the targeted strand DNA-(apurinic or apyrimidinic site) lyase (AP lyase) (Figure 1.2c)[87]. If the opposite strand has been nicked by the Cas9 nickase component of the base editor, the resulting proximity of the two nicks results in a DSB, a species that is likely to be resolved by indel-prone end-joining processes (Figure 1.2c). Liu and co-workers showed that indel formation can be substantially reduced by fusing the bacteriophage Mu-derived Gam (Mu-GAM) protein to BE4 to generate BE4-Gam, which further reduces indels in treated HEK293T cells relative to BE4[36]. BE4-Gam treatment also resulted in higher product purity and reduced indel frequency compared to BE3 in rabbit embryos[88].

ABE typically leads to very low (in some cases undetectable) indel frequencies, typically well below 1%, for treated cells in culture[35,82,85,89,90], mice[82] and in plants[91]. The lower frequency of ABE-mediated indels is consistent with the requirement of a glycosylase or other enzyme involved in DNA repair to remove inosine and induce a nick in the edited strand to form an indel[35]. Since the removal of inosine is thought to be substantially less efficient than removal of uracil from DNA[86], fewer nicks in the target strand, fewer resulting DSBs, and fewer indels would be expected to follow adenine base editing compared to cytosine base editing.

1.5.3 Off-target editing with DNA base editors

As with all genome editing technologies, both cytosine and adenine DNA base editors have the potential to operate on DNA at off-target genomic loci[33-35,92]. Off-target base editing can be classified into “proximal off-target editing”, editing that takes place near (for example, within 200 bp of) the target locus but outside the activity window, and “distal off-target editing”, editing that takes place away from the target locus. While the off-target effects of DNA base editors

continue to be investigated, early evidence suggests that distal off-target base editing generally occurs only at a subset of loci that experience off-target editing from Cas9 nuclease[93]. In contrast to RNA editors (see above), current data[33 35] suggest that DNA base editors typically do not induce measureable proximal off-target edits, although an in-depth study of proximal off-target base editing has not yet been reported.

Since the Cas9 component mediates the DNA-targeting ability of base editors, off-target base edits have been interrogated through deep sequencing of genomic loci known to be edited by Cas9 nuclease[33-35,85,89,92]. As expected, off-target loci that contain a C positioned in the activity window of the editor are sometimes edited at a low but detectable frequency by CBEs. Since not all the Cas nuclease off-targets contain an editable cytosine, off-target profiles of CBEs are generally more favorable than that of the corresponding nucleases programmed with the same guide RNAs[33-35,85,89,92]. To improve the DNA specificity of cytosine base editing, high-fidelity versions of BE3 have been generated by incorporating mutations known to improve the editing fidelity of Cas9 nuclease into the Cas9 portion of BE3. Liu and co-workers used the mutations discovered by Joung and coworkers[94] to improve the DNA specificity of Cas9 nuclease, resulting in high-fidelity BE3 (HF-BE3)[94]. HF-BE3 shows a substantial reduction in off-target editing, even when paired with highly promiscuous guide RNAs[92] (Table 1.1). Soo-Kim and co-workers have generated an alternative high-fidelity base editor, called Sniper-BE3, using the same strategy with a different set of mutations[95].

Kim and co-workers developed an unbiased *in vitro* screen for identifying off-target edits by CBEs using purified genomic DNA and BE3ΔUGI (BE3b[36]) ribonucleoproteins (RNPs), finding that the off-target loci deaminated by rAPOBEC1–Cas9 nickase are indeed predominantly, but not entirely, a subset of the loci edited by Cas9 nuclease[93]. Although off-target adenine base editing has not been broadly interrogated, examination of off-target ABE activity at known off-targets of Cas9 nuclease when programmed with the same guide RNAs suggests that ABEs exhibit substantially lower off-target activity than Cas9 nucleases, and even

less than was observed from BE3[33,35]. Further studies to investigate off-target ABE activity in cells and *in vivo* are needed to fully characterize and explain the apparent higher DNA specificity of the ABEs compared to the CBEs.

In addition to the off-target editing that could be directed by the DNA-binding protein component of base editors, deamination of non-target ssDNA (such as within a transient bubble of ssDNA during transcription), or in RNA, may occur from DNA base editors. Misregulation or overexpression of endogenous deaminases has been linked to elevated mutation rates[96-98], and expression of the UGI component of CBEs could also lead to an elevated rate of C-to-T transitions in the genome through impeding repair of spontaneously generated uracils[79,80]. However, studies of CBE off-target editing to date do not report widespread C-to-T mutations upon CBE expression or treatment, and transient delivery methods such as RNP delivery are likely to further reduce the mutagenicity of UGI in the context of CBEs[33,34,76,85,89,92,93],

Whole-genome sequencing (WGS), when performed on the genomic DNA from sufficient numbers of independent cells, has the potential to detect all types of off-target base editing in cells or whole higher organisms. The WGS experiments reported to date on base-edited animals, however, have not been performed with sufficient power or controls to identify such events across an entire mammalian genome. Kim, Huang, and their respective co-workers performed WGS on mutant mice generated through treatment with ABE7.10 and a guide RNA targeted to the *Tyr* locus[82] or targeting the *Hoxd13* locus in a one-cell stage embryo[99]. Computational analysis indicated that none of the SNPs identified in the treated mice were likely to have arisen through off-target base editing. Together, these studies further suggest high DNA specificity of ABE7.10.

We should note that these studies do not exclude the possibility of deamination from base editors that is not directed by the DNA- or RNA-binding component of the editors, but instead by random encounters between the deaminase domain of base editors and transient single-stranded DNA. More data are required to characterize this possibility, including WGS of

treated and untreated littermate controls and of mice treated with base editor mutants with catalytically inactivated deaminases. The continued development of context-dependent base editors[100] or future base editor variants that lack the ability to bind ssDNA without assistance from the guide RNA represent potential solutions to further minimize the possibility of random non-directed off-target base editing.

1.5.4 Editing window and bystander edits

In the case of BE3, which incorporates *Streptococcus pyrogenes* (SpCas9) as the DNA-targeting moiety, the "activity window" in which efficient editing is observed, is approximately five nucleotides wide (positions 4-8, counting the PAM as positions 21-23)[33,36,92]. Bases located outside the activity window but within the ssDNA R-loop region may still be edited at a lower efficiency, particularly if they are located in a favorable editing motif (see below). For many genome-editing applications, only a single nucleotide is targeted for conversion, so an ideal base editor would have a narrow activity window that focuses activity only on the target base. However, such a narrow window necessitates that the base editor be targetable to a broad range of PAM sequences. As the repertoire of natural Cas nucleases with different PAM requirements and function in human cells (including SaCas9[101] LbCpf1 and AsCpf1[102], CjCas9[103], StCas9[104],and NmCas9[105], engineered CRISPR proteins[106,107], and laboratory-evolved CRISPR proteins[85]) continues to expand, the desirability of more precise base editor variants with narrower activity windows will increase.

For some target sites, multiple editable Cs or As exist within or nearby the activity window, a situation that can result in conversion of bases in addition to the target base. We use the term "bystander editing" to describe editing in the protospacer at a nucleotide other than the target nucleotide (Figure 1.4b). Bystander editing may be inconsequential, especially when base editing to disrupt promoters, splice sites, or other regulatory sequences, or when knocking out gene function by introducing premature stop codons. When editing protein-coding genes, within

a canonical 5-base editing window most, but not all, base editing cases will only result in the desired single amino acid change, in part because the genetic code dictates that almost all third-position transitions in a codon are silent.

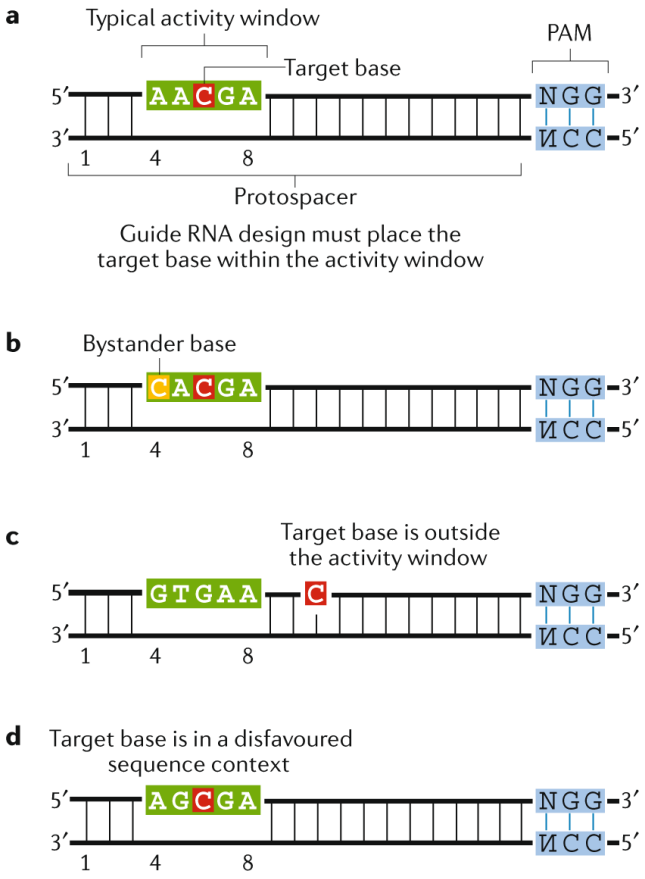


Figure 1.4: Overcoming targeting challenges associated with base editing. An example of a cytosine base editing site has been shown; these principles also apply to other classes of base editors **(a)** An ideal base editing target locus. The target base is located within the base editor activity window relative to the PAM site, there is only one target base in the activity window, and the target base is found in a motif (AC in this example) that is efficiently deaminated by most cytosine base editors[33,36]. **(b)** Example of a target site with a bystander base. If the bystander edit (deamination of the cytosine shown in yellow) is undesired, a narrowed-window[108] or context-specific base editor[100] may be used to preferentially edit the target base over the bystander base **(c)** The target base is located outside of the activity window. Base editing on this target may be possible with base editors that recognize different PAMs (see Table 1.1). **(d)** The target base is located within a sequence context that may not an efficient substrate for a particular deaminase[33,36]. Editing of the target may be improved by using an editor with a different deaminase[34,36], or an editor more tolerant of methylated DNA[109].

To minimize bystander editing, researchers have developed base editor variants with altered activity windows. Liu and coworkers engineered CBEs with mutations in the rAPOBEC1 domain that attenuate deamination activity, resulting in editors with reduced processivity and narrower activity windows (YE1-BE3, YE2-BE3, and YEE-BE3; see Table 1.1)[108]. These narrow-window CBEs enable selective editing of a target C over a neighboring C that is located within the standard editing window of BE3. For ABE7.10, which is generally the most efficient and widely used A-base editor, the activity window is approximately located from positions 4-7 in the protospacer (counting the PAM as positions 21-23). For certain targets, ABE7.9 or ABE6.3 may be more useful due to a slightly broader activity window enabling editing from positions 4-9[35]. Recent work by Kim and co-workers described how pairing a 5' extended guide RNA with ABE7.10 can increase editing to positions 2-3, although editing at these positions remains modest[82] (Table 1.1). The use of base editor variants that exhibit strong sequence context preference serves as a promising additional strategy to minimizing bystander base editing. These variants are discussed below (see "Base editing sequence context").

Conversely, Huang and coworkers expanded the width of the editing window by eingeering "BE-PLUS", a CBE variant in which a SunTag[110] was fused to the N-terminus of Cas9(D10A) nickase. Separately expressing a scFv–APOBEC–UGI fusion allows up to ~10 UGI domains to associate with each SunTag[111]. This construct enabled editing from protospacer positions 4-16, with reduced indel and C-to-non-T editing compared to BE3, likely due to the recruitment of many UGI domains[111] (Table 1.1). While base editors with enlarged editing windows are more prone to bystander editing, they also facilitate access of the target base pair, and may be especially useful when targeting non-protein-coding sites.

1.5.5 Targeting limitations

Successful DNA target binding by CRISPR-family nucleases requires a PAM, a conserved sequence up or downstream of the variable guide RNA protospacer sequence

(Figure 1.4a)[2,16]. For base editing, the PAM must be appropriately positioned relative to the target base to ensure efficient editing. Even though SpCas9 offers the least restrictive PAM among those CRISPR enzymes reported to function with high activity in mammalian cells, due to this requirement only ~26% of known pathogenic SNPs that are of the four types of base conversions (C to T, G to A, A to G, or T to C) that can be performed can be targeted by SpCas9-derived base editors (Figure 1.4c)[85]. This limitation creates the need to develop base editors with additional PAM compatibilities.

To increase the number of targetable bases, researchers have developed base editors incorporating different CRISPR-associated nuclease enzymes (Table 1.1). Liu and co-workers described a set of alternative CBEs with *Staphylococcus aureus* Cas9 (SaCas9) and engineered variants of SpCas9 and SaCas9 capable of efficient editing with non-NGG PAMs[108] including SaBE3, Sa(KKH)-BE3, VQR-BE3, VRER-BE3, and EQR-BE3. Chen and co-workers described a CBE derived from Cas12a (also known as Cpf1; PAM = TTTV, where V is A, C or G), which allows access to T-rich regions of genomic DNA[38]. Since there is no known mutation capable of transforming Cas12 into a nickase that cleaves only the target strand of DNA, Chen and co-workers characterized a dead LbCas12a-base editor, which nevertheless displays editing efficiencies averaging 22% across 10 target sites in HEK293T cells[38] (Table 1.1).

Recently, Liu and co-workers used phage-assisted continuous evolution (PACE) to evolve SpCas9 to recognize a broader range of PAMs. A resulting evolved variant, xCas9(3.7), harbors mutations allowing it to access some target sequences with some NG, GAA or GAT PAMs. Replacing Cas9 in the BE3 construct with xCas9(3.7) made xBE3, a CBE capable of editing some loci with NGN, GAA and GAT PAMs (Table 1.1)[85]. While xCas9 variants are capable of mediating DNA cleavage or base editing at several non-NGG PAMs, xCas9-mediated editing efficiency varies among different target sites, and like many engineered or evolved Cas9 variants, likely requires a high degree of perfection between the guide RNA and the target sequence, including a G at the 5' end of the guide RNA and at the corresponding first

position of the protospacer[100,112]. Surprisingly, in addition to its expanded PAM acceptance, xCas9 also displays higher editing fidelity than SpCas9[85].

Nureki and co-workers used a rational design approach to develop another SpCas9 variant with broadened PAM compatibility, termed "NG-Cas9"[113]. In mammalian cells, the relative activities of xCas9 and NG-Cas9 appears to be guide RNA-dependent; Nureki and co-workers reported that NG-Cas9 is more active than xCas9 at 15/15 NGC, 16/18 NGT and 15/19 NGA PAM sites, but NG-Cas9 exhibits a loss of efficiency at the canonical NGG PAM sites that is not observed with xCas9[113]. NG-Cas9 also does not exhibit the enhanced fidelity observed with xCas9, but tolerates inclusion of fidelity-enhancing mutations[113]. As a CBE, NG-Cas9 accepted a subset of NG PAM loci as substrates for efficient base editing (Table 1.1)[113].

Alternative-PAM ABEs have been developed by adapting SaCas9[83], Sa(KKH)Cas9[114,115], Sp(VQR)Cas9,[114,115] and Sp(VRER)Cas9[114] into the ABE7.10 architecture, resulting in efficient generation of mutant rice plants (Table 1.1). Additional ABE variants with altered PAM requirements would substantially augment the scope of targetable bases for adenine base editing.

1.5.6 Base editing sequence context

In addition to PAM- and activity window-imposed sequence restrictions, the particular deaminase enzyme variant used in a base editor may impose sequence context preferences that affect editing efficiency at a particular locus. For example, rAPOBEC1 exhibits poor processing of cytosines within some (but not all) GC motifs[33,36] (Figure 1.4d). By contrast, other cytidine deaminases such as activation induced deaminase (AID) or cytidine deaminase 1 (CDA1) do not display this particular sequence preference but exhibit lower editing efficiencies than rAPOBEC1 in most tested sequence contexts when tested in a BE3 architecture[36]. Yang and co-workers identified that rAPOBEC1-mediated base editing rates are reduced by DNA

methylation at CpG dinucleotides[109], and that human APOBEC 3A (hA3A) can edit cytosines found in CpG dinucleotides and in GC motifs more efficiently than rAPOBEC1[109].

Joung and co-workers harnessed the sequence preferences of different cytosine deaminase enzymes to engineer a mutant hA3A-based CBE that preferentially deaminates cytosines preceded by a T[100] (Table 1.1) as a strategy to reduce bystander editing. Structure-guided design and screening of hA3A deaminase mutants resulted in an enhanced variant (eA3A) with a single mutation (N57G) that deaminates the target motif (TC) but significantly reduces activity at Cs in other sequence contexts, resulting in a context-dependent base editor that maintains a 5-nucleotide activity window[100]. Importantly, Joung and co-workers performed a detailed analysis of the individual alleles that were generated upon successful base editing by eA3A-BE3, BE3, and other engineered variants (YEE-BE3, YE1-BE3 and YE2-BE3) to demonstrate that eA3A can make the desired allele at a high efficiency and purity[100]. A high-throughput sequencing data analysis package facilitated this detailed analysis of base editing outcomes[116].

Context-specific base editors such as those developed by Joung and coworkers represent an important advance that offers more precise base editing, with the trade-off of lower target site applicability since the target nucleotide must naturally exist in the preferred sequence context. Thus far, the data from mammalian-cell editing with ABE7.10 indicates that it is relatively free from motif-related sequence preferences in human cells[35], but Kim and co-workers have demonstrated that there is a preference for editing at TA motifs relative to GA, CA, or AA in *Arabidopsis thaliana*[91]. The development of additional context-specific ABE and CBE variants will be enabling for applications in which editing only a single base is paramount.

1.5.7 Improving intracellular expression and nuclear localization of base editors

For plasmid delivery of Cas9 nuclease, optimization of codon use for mammalian cell expression improves soluble protein levels and enhances editing efficiencies[112]. Optimization of

the nuclear localization sequence (NLS) also improves Cas9-medited editing *in vivo*[117]. Liu and co-workers identified that poor expression is also a bottleneck to the efficiency of base editors, and optimized codon usage and nuclear location sequences to generate improved cytosine and adenine base editors, resulting in BE4max and ABEmax from BE4 and ABE7.10 respectively[90]. The use of ancestral sequence reconstruction starting from the protein sequences of the hundreds of known APOBEC homologs, a process that has been demonstrated to improve protein expression[118], resulted in AncBE4max. All three optimized base editors offered substantially improved editing efficiency, especially under suboptimal conditions such as when delivery into cells is limiting[90].

In an elegant independent study, Dow and co-workers optimized CBE codon usage by removing premature poly(A) sites and rare mammalian codons, and improved CBE nuclear localization by adding a second NLS to the N-terminus of BE3, to generate an optimized FNLS-BE3 that results in much higher editing efficiencies than BE3. When packaged into lentivirus, FNLS-BE3 mediated efficient editing in murine intestinal organoids[119]. Hydrodynamic injection of the plasmid encoding FNLS, together with a guide RNA that programs the base editor to make a S45F mutation in *Ctnnb1* lead to significantly more efficient base editing and corresponding physiological changes (tumor nodule formation) in the livers of mice than BE3 treatment[119]. Dow and co-workers also generated lentiviral constructs with the corresponding optimized editor versions of BE4-Gam, that enable improved editing rates with reduced indel formation[119]. Ensuring optimal expression of the base editor construct in the target cell type is critical for applications that require high editing efficiency, and the above developments thus represent important advances.

1.6 Delivery of base editors

1.6.1 DNA delivery strategies: plasmid transfection and viral delivery

Since most proteins cannot spontaneously traverse cell membranes, a delivery method is required to facilitate cell entry. A common strategy is to deliver DNA encoding the target protein through chemical transfection[120], electroporation[121], or viral infection[122], and then rely on target cell transcription and translation to produce the desired protein.

For cell lines in culture (including HEK293T, HeLa, U2OS and murine NIH/3T3 cells), lipid-mediated transfection of plasmids encoding base editors has resulted in high editing efficiencies without selection for transfected cells[33-36,38,39,82,85,92]. For cell types resistant to plasmid lipofection, electroporation followed by fluorescence-activated cell sorting (FACS) to isolate transfected cells has yielded favorable editing efficiencies for lymphoblastoid cell lines (LCLs)[35] and mouse astrocytes[33]. Although plasmid-based delivery is a convenient delivery strategy, DNA delivery raises the risk of exogenous DNA recombination into the genome and protracted overexpression of genome-editing agents increases off-target editing rates[76,92,123-125].

The use of viruses to deliver DNA encoding base editors is a promising delivery modality for some *in vivo* research or therapeutic applications. Use of non-integrating vectors such as adeno-associated virus (AAV), herpes simplex virus (HSV), or adenoviral vectors reduces the potential for random integration of exogenous DNA into the host genome. Infection with adenovirus and HSV-1, however, may provoke inflammatory responses[122]. By contrast, AAV is thought to be both non-inflammatory and non-pathogenic[126]. When coupled with its broad tropism, well-studied serotypes, and ability to infect dividing cells, AAV is a particularly promising strategy for viral delivery of genome editing agents.

AAV-mediated delivery of many CRISPR genome-editing agents, including base editors, is challenging due to the 4.9 kbp packaging limit of AAV[127]. A CBE or ABE plus a guide RNA totals approximately 6 kbp. Kim and co-workers overcame this through use of two trans-RNA splicing AAVs (tsAAVs)[128] encoding each half of ABE7.10[82]. Dual tsAAV-mediated delivery of ABE7.10 into skeletal muscle in a mouse model of DMD corrected a premature stop codon[82].

After dual infection, homologous recombination between the identical inverted terminal repeat (ITR) sequences generates the full-length ABE7.10 transcript[82], enabling ABE7.10 protein production.

1.6.2 RNP delivery

DNA-free base editing enables precise and specific changes to genomic DNA without exposing a cell to exogenous DNA[89,92]. Sustained overexpression of genome-editing agents erodes DNA specificity. After successful editing, the target site is no longer a binding site for the editing agent, and residual editor can only act to mediate off-target editing. Thus, controlling the exposure to editing agents, including base editors, can greatly improve their DNA specificity[25,92,123,125].

Kim and co-workers established that purified Cas9 complexed with a guide RNA, forming an RNP complex, can be efficiently delivered into mammalian cells in culture by electroporation, and that RNP delivery of Cas9 leads to improved DNA specificity relative to plasmid-based delivery[125]. Liu and co-workers demonstrated that cationic lipid-mediated delivery of Cas9 RNP complexes can facilitate *in vivo* delivery of Cas9 near the site of administration, as well as efficient delivery into cells in culture, and resulted in greatly improved DNA specificity relative to plasmid-based lipofection[123,124,129].

BE3 protein has also been purified[76,92], and Liu and coworkers have packaged BE3:guide RNA RNP into cationic liposomes for lipid-mediated delivery to cultured cells, zebrafish embryos, and the inner ear of postnatal mice[89,92]. Analogous to the delivery of Cas9, cationic lipid-mediated delivery of BE3 dramatically improves DNA specificity in human cells compared to plasmid delivery[89,92]. BE3 RNPs have also been delivered through electroporation into mice[76], and through direct injection into *Xenopus laevis* embryos[130]. RNP delivery is also effective for alternative base editors; the engineered high-precision editor eA3A(N57Q) has been delivered as an RNP into human erythroid precursor cells via nucleofection of the RNP

complex to correct a mutant *HBB* allele, resulting in a 4-fold increase in *HBB* expression[100]. The advantages of RNP delivery include improving editing specificity and removing the reliance on intracellular transcription and translation to generate the editing agent.

1.6.3 mRNA delivery of base editors

Delivery of mRNA is a commonly used delivery strategy to deliver genome-editing agents into embryos. Kim, Huang, Lin, Liu, Zhang, Li and their respective coworkers have demonstrated that *in vitro* transcription followed by purification of an mRNA encoding BE3, when combined with a guide RNA, can be co-delivered into single cell mouse[76,131], human[132,133], rabbit[88], rat[134] or zebrafish zygote[135,136] by electroporation or direct injection to generate point mutations with high efficiency and DNA specificity. These studies establish mRNA delivery of base editors into embryos as a robust and efficient strategy for the generation of animals with tailor-made point mutations.

1.7 Applications of base editing

1.7.1 Base editing to install or correct pathogenic point mutations

Since point mutations are the largest class of known pathogenic genetic variants (Figure 1.1a) and CBEs and ABEs collectively have the potential to install or reverse up to ~60% of pathogenic point mutations (Figure 1.1b)[28,29], a major application of base editing is the study or treatment of disease-associated point mutations.

Examples of base-editor-induced gene correction in cultured cells are already numerous. Liu and coworkers showed that plasmid nucleofection of BE3 can convert the Alzheimer's disease associated allele *APOE4* to *APOE3r* in mouse astrocytes, and to correct the cancer-associated p53 mutation Tyr163Cys in breast cancer cells[33]. Subsequently, codon-optimized CBEs were delivered as plasmids in patient-derived fibroblasts to correct the Leu119Pro mutation in *MPDU1*[90] that causes the congenital disorder of glycosylation type 1f[137]. Liu and coworkers also showed that plasmid delivery of ABE7.10 can correct the hereditary-

haemochromatosis-causing mutation C282Y in an immortalized patient-derived LCLs, and to install a mutation known to increase fetal hemoglobin (*HBG*) expression in adults[35]. Joung, Huang, and their respective coworkers reported correction of a mutant *HBB* allele that causes beta-thalassemia in an engineered HEK293T cell line[100] and in patient-derived primary fibroblasts[133].

Direct injection of base editor-encoding mRNA along with a guide RNA has also proven effective for editing pathogenic alleles in human embryos. Direct injection of mRNA encoding BE3[132,133,138], YE1-BE3[138] or YEE-BE3[133] together with a guide RNA can generate homozygous mutants at a rate of up to 77% of embryos that survive to the blastomere stage[132,133].

Viral delivery of base editors is an effective method for correcting pathogenic mutations in mouse disease models *in vivo*. Kim and co-workers used AAV to deliver ABE7.10 with a guide RNA programmed to correct a premature stop codon in the *DMD* gene in a mouse model of muscular dystrophy. Although the correction rate was only 3.3% of sequenced cells, dystrophin expression was restored in 17% of muscle fibers[82], highlighting that low levels of editing can often lead to therapeutically-relevant phenotypic change. Separately, Musunuru and co-workers generated an adenoviral vector encoding BE3 and a guide RNA programmed to make the W159Stop mutation in murine *Pcsk9*. They measured a median rate of 25% editing in liver cells, and show a modest reduction in plasma PCSK9 protein levels and plasma cholesterol 4 weeks post-injection[139].

In vivo base editing has also been used to ascertain whether a genotype is causal for a particular phenotype. Dow and co-workers performed hydrodynamic transfection of an optimized BE3 plasmid construct, termed FNLS-BE3, with a guide RNA programmed to make the S45F cancer-associated mutation in *Ctnnb1*. They demonstrated efficient (nearly 100%) base editing in liver cells, and showed that treated mice treated with FNLS-BE3 plus the on-target guide RNA grew a significant number of visible tumor nodules compared to controls[119]. Lin and co-workers delivered of BE3 as an mRNA into one-cell-stage zebrafish embryos to

generate a P302S mutation in *tyr* that mimics a common mutation observed in human ocular albinism. This approach enabled investigation into the effects of such a mutation on ocular pigmentation[135]. These studies hint at the promise of BEs as potential therapeutics, and demonstrate their efficacy for researchers interested in ascertaining the phenotypic effects of precise genetic changes in cell culture and *in vivo*.

1.7.2 Base editing in post-mitotic cells

Liu and co-workers demonstrated that base editing can occur in the non-mitotic sensory supporting and hair cells[140] *in vivo* in the mouse inner ear[89]. BE3 combined with a guide RNA targeting β-catenin was used to control flux through the Wnt signaling pathway[89]. Blocking phosphorylation at S33 through an S33F mutation extends the cellular half-life of β-catenin, increasing Wnt signaling. For this target, maintaining low indel rates is critical as indels would likely disrupt the gene and reduce β-catenin levels, opposing the desired change[89]. Lipid-mediated delivery of BE3 complexed with the S33F guide RNA as an RNP into the inner ear of mice led to editing in post-mitotic somatic cells at efficiencies up to 8%. Dissection and staining of treated hair cells identified that BE3 treatment, unlike treatment with Cas9 nuclease and an HDR template, induced cellular reprogramming of other cells into cells resembling cochlear hair cells. These results establish the ability base editing to occur in post-mitotic cells that are resistant to DSB-stimulated HDR[23,141].

1.7.3 Cytosine base editing to introduce premature stop codons

CBEs (but not ABEs) can install premature stop codons to disrupt genes in a homogenous manner by precisely converting one of four codons (CAA, CAG, or CGA in the non-coding strand; or TGG in the coding strand) into stop codons. Kim and co-workers demonstrated this possibility by using BE3 to introducing a premature stop codon in *Dmd* in mouse embryos[76]. The CRISPR-Stop[142] and iSTOP[143] methods use this principle to enable high

throughput BE3-mediated gene inactivation without generation of DSBs and accompanying indels. Ciccia and co-workers generated a database describing a set of guide RNAs, that, when complexed with BE3, are capable of generating premature STOP codons in >98.6% of open reading frames (ORFs) in the human genome (reference genome assembly GRCh38). They published a freely accessible online database enabling researchers to find appropriate guide RNAs for iSTOP to use in eight species[143]. Adli and co-workers identified that this strategy results in a significant reduction in apoptosis when compared to Cas9 nuclease treatment[142], possibly due to lower DSB-induced toxicity[144-146]. While typically efficient and widely utilized, NHEJ-mediated knock out of genes following DSBs leads to a mixed population of cells, DNA translocations and rearrangements[27,147], and the induction of cell death[144-146], all of which in principle are avoided through the use of base editors to install precise stop codons. Flow cytometry of CRISPR-Stop treated cells indicated that stop-codon introduction is similar in efficiency to Cas9-mediated gene knockout[142].

Perez-Pinera and co-workers confirmed that base-editor induced C-to-T edits at the conserved splicing acceptor site can induce exon skipping[148]. Their method (termed CRISPR-SKIP) was similar in efficiency to Cas9 DSB-mediated exon skipping, but unlike nuclease treatment did not generate DSBs[148].

1.7.4 Base editing in embryos to generate animal models

A common goal of genome editing at the single-cell embryo stage is to generate model organisms. To minimize mosaicism and maximize the chance that editing occurs in the germ line, it is critical that editing occurs quickly and efficiently. Since nuclease-mediated editing strategies often fail to generate homozygous, non-mosaic progeny in the F0 generation[149,150] the high efficiency of base editing offers an attractive alternative. CBEs are particularly useful for generating loss-of-function animal models by inserting a premature stop codon into a gene of interest without generating DSBs or indels[76,135].

Kim and co-workers demonstrated that microinjection of mRNA encoding BE3 together with a guide RNA, or electroporation of the BE3:guide RNA RNP complex mediates efficient generation of premature stop codons in one-cell stage mouse embryos at two target sites: Q871Stop in *Dmd* or Q68Stop in *Tyr*[76]. Impressively, mRNA treatment yielded the target mutation in 11/15 or 10/10 blastocysts at the *Dmd* or *Tyr* loci, respectively. RNP delivery of BE3 was also effective; 2/7 of the embryos treated with a BE3 RNP pre-complexed with a guide RNA targeted to the *Tyr* locus were transplanted into surrogate mothers to yield homozygous, non-mosaic progeny with the expected albino phenotype[76]. Independently, Songyang and coworkers used either BE3 or a high-fidelity version of BE2 (BE2-HF2) mRNA to perform base editing in mouse zygotes, resulting in up to 50% of sequenced embryos harboring a C-to-T point mutation at the target locus[131]. Li and coworkers made rabbit models of human disease using mRNA injection of BE3, BE4-Gam or ABE7.10 into blastocysts[88]. They performed ABE-mediated editing generated the point mutation T297A in Exon9 of *Dmd* that is associated with cardio-specific XCLM in humans[88] and BE3 was used to make the mutation c.1821C>T in *Lmna*, generating a rabbit model of Hutchinson-Gilford progeria syndrome[88].

Huang and coworkers performed multiplexed base editing through co-injection of ABE7.10 and SaBE3 mRNA along with guide RNA sequences targeting *Tyr (*an *S. aureus* guide RNA was used to generate the Q58Stop mutation) and *Hoxd13* (a *Streptococcus pyogenes* guide generated the Q312R mutation) in one-cell mouse embryos. Impressively, A-to-G and C-to-T edits were simultaneously observed in blastocysts[99]. The same strategy has been used to deliver ABE mRNA and guide RNAs into rat embryos[115,134]. Zhang and coworkers showed that co-injection of two different guide RNAs efficiently generated two transmissible A to G point mutations in the F0 generation simultaneously[134] whilst Yin and co-workers used ABE to generate a rat model of Pompe disease[115]. These data demonstrate that base-editing is an enabling tool for generating mutant mice, rats, and rabbits for animal studies; previous

nuclease-based editing methods usually failed to generate non-mosaic mice with 100% mutation frequency in the F0 generation[151].

1.7.5 Base editors as cellular event recorders

In addition to its applications in biomedical research to install and correct point mutations, base editing has also been used as a synthetic biology tool to record cellular signaling and exposure to stimuli. Unlike the stochastic indels that result from Cas9 DNA cleavage, base editors generate predictable single point mutations. By coupling the stimulus of interest to the activity of the BE, the resulting stimulus-dependent single point mutations can be used to record exposure to signals into the genome. Liu and co-workers developed a ligand-responsive editing system by appending a blocking sequence to a guide RNA through a ligand-dependent hammerhead ribozyme[152]. This system facilitated ligand-dependent base editing in mammalian cells[152].

Subsequently, Liu and Tang demonstrated that controlling expression of a base editor or its accompanying guide RNA using stimulus-dependent promoters enables recording of a wide variety of stimuli—including exposure to light, nutrients, antibiotics, or virus—durably as point mutations into a cell's genome. This recording system was termed "CRISPR-mediated analog multi-event recording apparatus" 2 (CAMERA 2)[54]. Control of BE expression through small-molecule-responsive promoters enabled dose- and time-dependent base editing of four small molecules (aTc, IPTG, arabinose, and rhamnose) simultaneously in bacterial cells. Through careful design of two ratcheted protospacers, in which base editing from one guide RNA edits the binding site for the second guide RNA, the order of exposure could also be recorded[54]. The same principles were used in mammalian cells: signals including exposure to doxycycline, tetracycline, or IPTG were recorded as base edits in the *CCR5* safe-harbor locus. CAMERA 2 could also record changes in Wnt signaling in mammalian cells.

Independently, Lu and co-workers used cytosine base editors to develop a related platform for cellular reading and writing, named DOMINO (DNA-based Ordered Memory and Iteration Network Operator)[153]. As in CAMERA 2, expression of the base editor and guide RNAs are controlled with different small-molecule-responsive promoters in *E. coli.* DOMINO can directly couple stimulus-dependent base editing to a phenotypic readout. For example, successful DNA editing by two input guide RNAs could enable a third guide RNA to bind to a target DNA operator site upstream of a genomically-integrated *GFP* gene. Binding of the guide RNA:base editor complex to this operator resulted in GFP fluorescence reduction[153].

Lu and coworkers also used DOMINO as a self-reinforcing "molecular clock" in human HEK293 cells that records stimulus exposure time. They fused a CBE with the VP64 transcriptional activator to perform sequential editing of a repetitive operator region located just upstream of *GFP*. The circuit was designed such that over the course of 15 days the repetitive operator region was sequentially edited to generate more guide RNA binding sites, increasing localization of the editor to the operator region and thus increasing GFP expression. Both the number of GFP-positive cells and the C-to-T editing levels in the operator region reflected the number of days of exposure between the cell population and active editor construct[153]. Both DOMINO and CAMERA rely on the exquisite precision of base editing, as indel-generating methods would not be expected to predictably write new protospacer sequences. We anticipate that future cellular recording applications will use both CBEs and ABEs to develop more complex recording systems, since ABEs can erase signals written by CBEs, and vice-versa.

1.7.6 Base editing in plants

Base editing in plants could enable researchers and agriculturalists to rapidly generate novel plant mutants with an efficiency beyond that of conventional breeding[154]. Generation of precise, gain-of-function point mutations can improve many agronomic traits; for example, a point mutation in the plant *ALS* gene confers resistance to herbicides such as sulfonylureas and

imidazoliones[155]. Generating precise point mutations in plant cells remains challenging using DSB-induced HDR[156,157].

Multiple plant species of agronomic interest have been edited with CBEs and ABEs. Gao and co-workers demonstrated that BE3 generates efficient point mutations in maize, rice and wheat[158]. In a separate study, Kondo and co-workers showed that the Target-AID editor is capable of efficient editing in rice and tomato[159]. More recently, two independent reports from the Zhou and Zhu labs demonstrated that ABE editing is highly efficient in rice[83,84,114]. Gao and co-workers optimized the architecture of ABE7.10 for adenine base editing in rice and wheat, and used the resulting editor in protoplasts and in regenerated plants[160]. Kim and co-workers recently described two phenotypic changes generated through transient *Agrobacterium tumefaciens*-transfection of ABE7.10 into *A. thaliana* and *Brassica napus*[91]. Using a plant-optimized expression system, they performed editing in *A. thaliana* to generate a single codon change (Y85H) in the *FT* protein, generating a late-flowering phenotype, or to disrupt a splice acceptor site in the *PDS3* gene, generating a dwarf phenotype. After transformation, >85% of T1 plants showed >50% editing, and T2 seedlings isolated from T1 plants also displayed the same phenotypes, indicating that the editing was germline-transmissible[91].

These demonstrations establish that base editing is a promising approach for rapidly engineering of polyploid plant genomes. We anticipate that RNP delivery of BEs into crop species is particularly important from a regulatory and consumer perspective, because transgene integration from plasmid delivery results in plants with genetically modified organism (GMO) status. RNP-delivery of base editors would enable DNA-free precision editing that avoids the creation of GMO crops[155].

Chapter Two: Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery

This chapter has been adapted from:

Rees, H. A., Komor, A. C., Yeh, W.-H., Caetano-Lopes, J., Warman, M. Edge, A. S. B., Liu, D. R. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nature Communications* **8**, 15790 (2017)

Contributions:

This chapter of the thesis was written with the help of all of the co-authors. A.C.K., W.-H.Y and D.R.L. aided with development of the methodology and design of the experiments. W.-H.Y. performed all experimental procedures involving injections into the inner ear of mice and J.C.-L. performed injections into zebrafish.

I would like to thank Alexis Komor and David Liu in particular for their unwavering support, teaching and assistance. By carrying out this project under their guidance, I learnt a great deal about how to ask and answer interesting scientific questions. In addition, without the help, talent and enthusiasm of Wei-Hsi (Ariel) Yeh, we would never have been able to complete this project.

2.1 Introduction

Here, we describe two advances that greatly improve the DNA specificity of base editing and that allow base editing *in vitro* and *in vivo* without supplying exogenous DNA, which has been associated with a risk of recombination with the host genome and cytotoxicity[161,162]. First, we engineer a mutant form of BE3 incorporating mutations known to decrease the DNA affinity of Cas9[94] that reduces off-target editing events with only a modest decrease in on-target editing activity. Next, we demonstrate that lipid-mediated delivery of base editor proteins complexed with guide RNA results in even larger specificity enhancements with no apparent reduction in on-target base editing compared to plasmid DNA delivery. Delivery of base editors as RNPs typically reduces off-target editing to below measurable levels, even for a notoriously promiscuous guide RNA that targets a highly repetitive genomic DNA sequence, in cultured human and mouse cells. These advances enable us to demonstrate highly specific, DNA-free *in vivo* base editing in mice and zebrafish.

2.2 Results

2.2.1 Engineering a high-fidelity base editor

Cas9 nucleases and their associated fusion constructs have been shown to bind and cleave DNA at off-target genomic loci[31,163-165]. Joung and coworkers developed HF-Cas9, a high-fidelity SpCas9 variant containing four point mutations (N497A, R661A, Q695A, Q926A) that were designed to eliminate non-specific interactions between Cas9 and the phosphate backbone of the DNA target strand (Figure 2.1a)[94] consistent with the previous abrogation of non-specific DNA interactions in TALENs that greatly increased their DNA cleavage specificity[166]. Since base editors operate on the non-target strand within the single-stranded DNA bubble created by Cas9[33] we hypothesized that introducing these four point mutations

from HF-Cas9 into BE3 to generate “HF-BE3” might reduce off-target base editing without altering its base conversion capabilities (Figures. 2.1a, 2.1c).

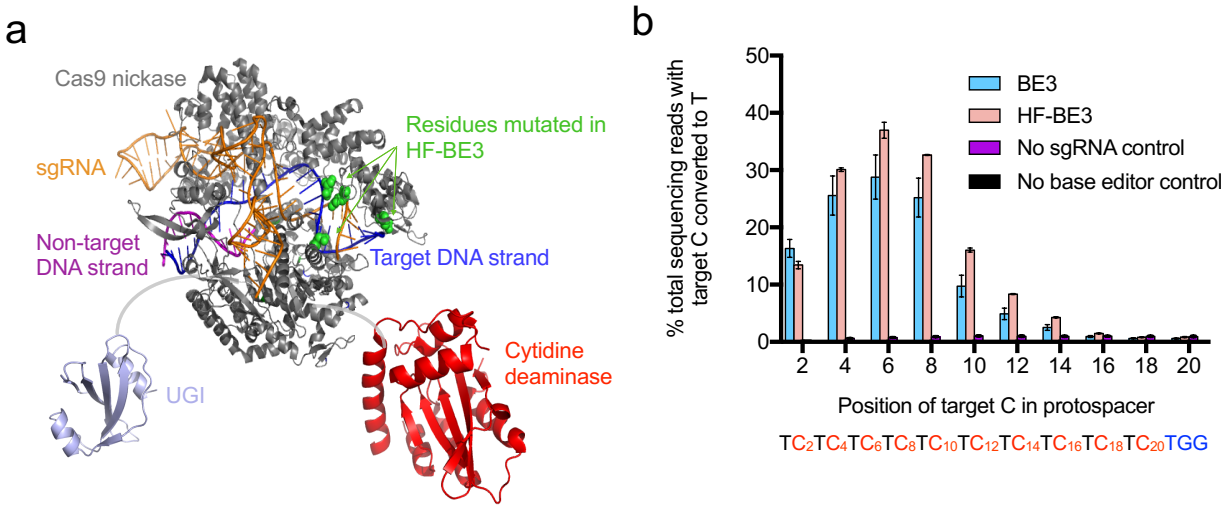


Figure 2.1: Engineering and *in vitro* characterization of a high fidelity base editor (HF-BE3). (a): Schematic representation of HF-BE3. Point mutations introduced into BE3 to generate HF-BE3 are shown in green. The representation used PDB structures 4UN3 (Cas9), 4ROV (cytidine deaminase) and 1UGI (uracil DNA glycosylase inhibitor). (b): *In vitro* deamination of synthetic substrates containing ‘TC’ repeat protospacers. Values and error bars reflect mean and range of two independent replicates performed on different days.

Plasmids encoding BE3 and HF-BE3 as His_6-tagged proteins were overexpressed in *E. coli* and purified first by nickel affinity chromatography and then by cation exchange chromatography (Figures. 2.2a, 2.2b). Following extensive optimization of expression and purification conditions, BE3 and HF-BE3 protein can be routinely produced at a yield of ~2 mg per liter of culture media (Figures. 2.2a, 2.2b, 2.2c).

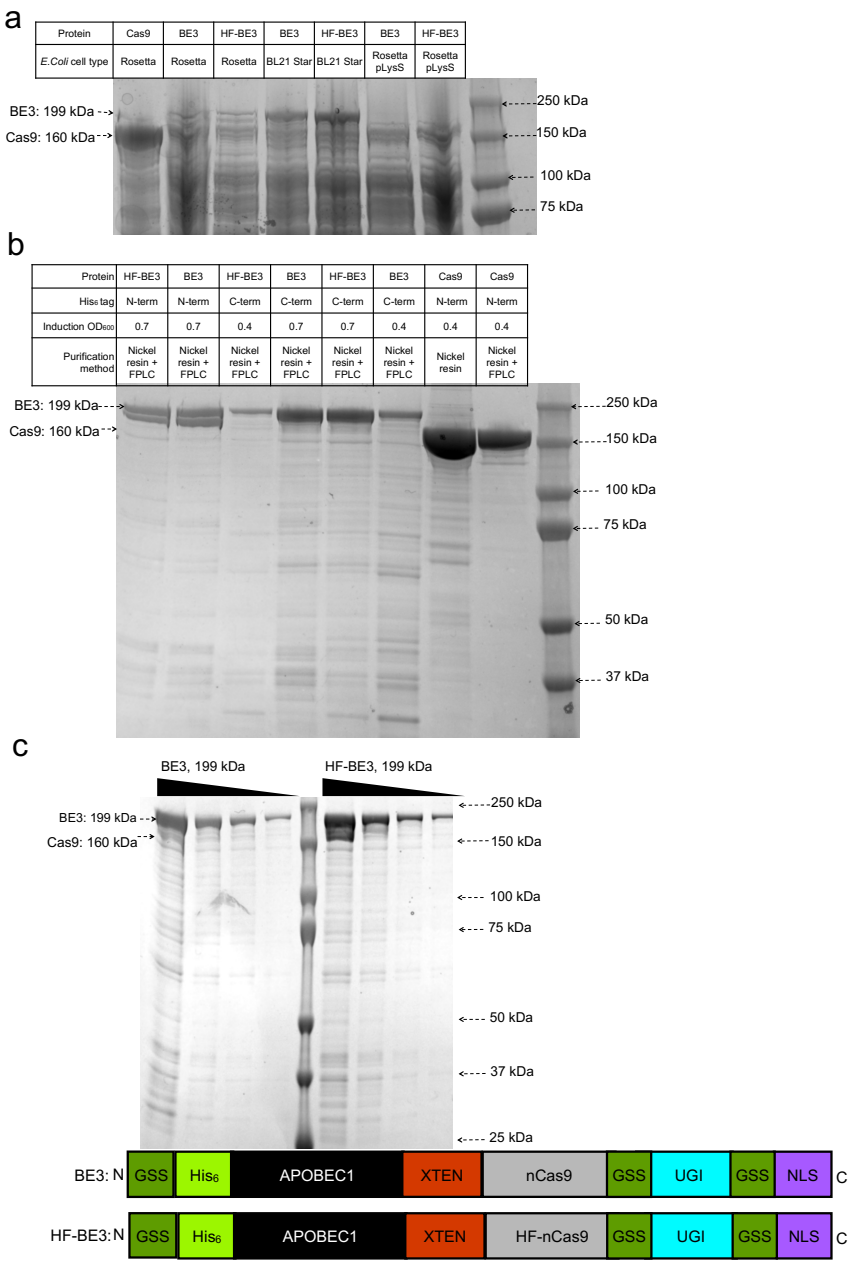


Figure 2.2: Purification of base editor proteins. (a): Selection of optimal *E. coli* strain for base editor expression. After IPTG-induced protein expression for 16 h at 18 °C, crude cell lysate was analyzed for protein content. BL21 Star (DE3) (Thermo Fisher) cells showed the most promising post-expression levels of both BE3 and HF-BE3 and were used for expression of base editors. (b): Purification of expressed base editor proteins. Placing the His_6 tag on the C-terminus of the base editors lead to production of a truncation product for both BE3 and HF-BE3 (lanes 1 and 2). Unexpectedly, this truncation product was removed by placing the His_6 tag on the N-terminus of the protein (lanes 3-6). Inducing expression of base editors at a cell density of OD_{600} = 0.7 (lanes 4-5), later than is optimal for Cas9 expression (OD_{600} = 0.4)[167], improves yield of base editor proteins. Purification was performed using a manual HisPur resin column followed by cation exchange FPLC (Akta). **(**c)**:** Purified BE3 and HF-BE3. Different concentrations of purified BE3 and HF-BE3 were denatured using heat and LDS and loaded

onto a polyacrylamide gel. Protein samples are representative of proteins used in this study. Gels in (a), (b), and (c) are BOLT Bis-Tris Plus 4-12% polyacrylamide (Thermo Fisher). Electrophoresis and staining were performed as described in Methods.

We used the purified base editor proteins to compare base editing efficiency and the width of the editing window of HF-BE3 and BE3 biochemically. We measured *in vitro* C to U conversion efficiencies in a synthetic dsDNA 79-mer with a protospacer comprised of TC repeats. The target dsDNA (250 nM) was incubated with BE3:sgRNA or HF-BE3:sgRNA (2 μM) for 30 min at 37 °C. After incubation, the edited DNA was amplified using a uracil-tolerant polymerase and sequenced by high-throughput DNA sequencing (HTS). We observed comparable editing efficiencies and activity window widths for HF-BE3 and BE3 *in vitro* (Figure 2.1b). These findings indicate that introduction of the high-fidelity mutations into BE3 does not compromise inherent on-target base editing efficiency or change the width of the editing window of the resulting HF-BE3 protein *in vitro*.

2.2.3 DNA transfection of HF-BE3 enhances editing specificity

Next, we compared base editing efficiencies, specificities, and editing window widths of BE3 and HF-BE3 in mammalian cells following plasmid DNA transfection. We chose four well-studied endogenous genomic loci (HEK293 site 3, FANCF, EMX1 and VEGFA site 2) to interrogate on- and off-target base editing in mammalian cells[31,33]. VEGFA site 2 is highly repetitive, containing 14 Cs out of 20 protospacer nucleotides, and is associated with notoriously high rate of known off-target genome editing[31,94,164,168]. We chose to include this site because it poses a formidable specificity challenge. In contrast with most nuclease-based genome editing applications, base editing relies on the precise location of the protospacer to place the target nucleotide within the editing window and usually little or no flexibility in the choice of guide RNA is available. Therefore, the development of base editors with enhanced specificities even for highly repetitive, promiscuous sgRNA targets is crucial[33,35].

We amplified by PCR and analyzed by HTS the on-target locus and known off-target loci following plasmid transfection[31] with each of the four base editor:sgRNA pairs. On-target editing in HEK293T cells for these four endogenous genomic loci was slightly reduced by introduction of the HF mutations; editing averaged 29±5% with BE3, and 21±3% (mean ± s.e.m. n=3 biological replicates) for HF-BE3 (Figures. 2.3a-d, 2.6a).

For each of the three standard, non-repetitive target sites (HEK293 site 3, FANCF, and EMX1), we examined the three most frequently modified off-target loci that contain a C within the editing window from the off-target loci previously reported to be modified from treatment with Cas9 and the same guide RNA[31]. When cells were transfected with BE3 plasmid, C→T conversion across the nine most frequently modified Cas9 off-target loci for HEK293 site 3, FANCF, and EMX1 averaged 1.1±0.3% (Figures 2.3a-c; mean ± s.d. n=3 biological replicates). Installation of the HF mutations reduced the absolute level of mean off-target editing by 37-fold to 0.03±0.005%, with only one instance of measureable off-target C→T conversion (Figure 2.3a; EMX1 C_5 at off-target 1).

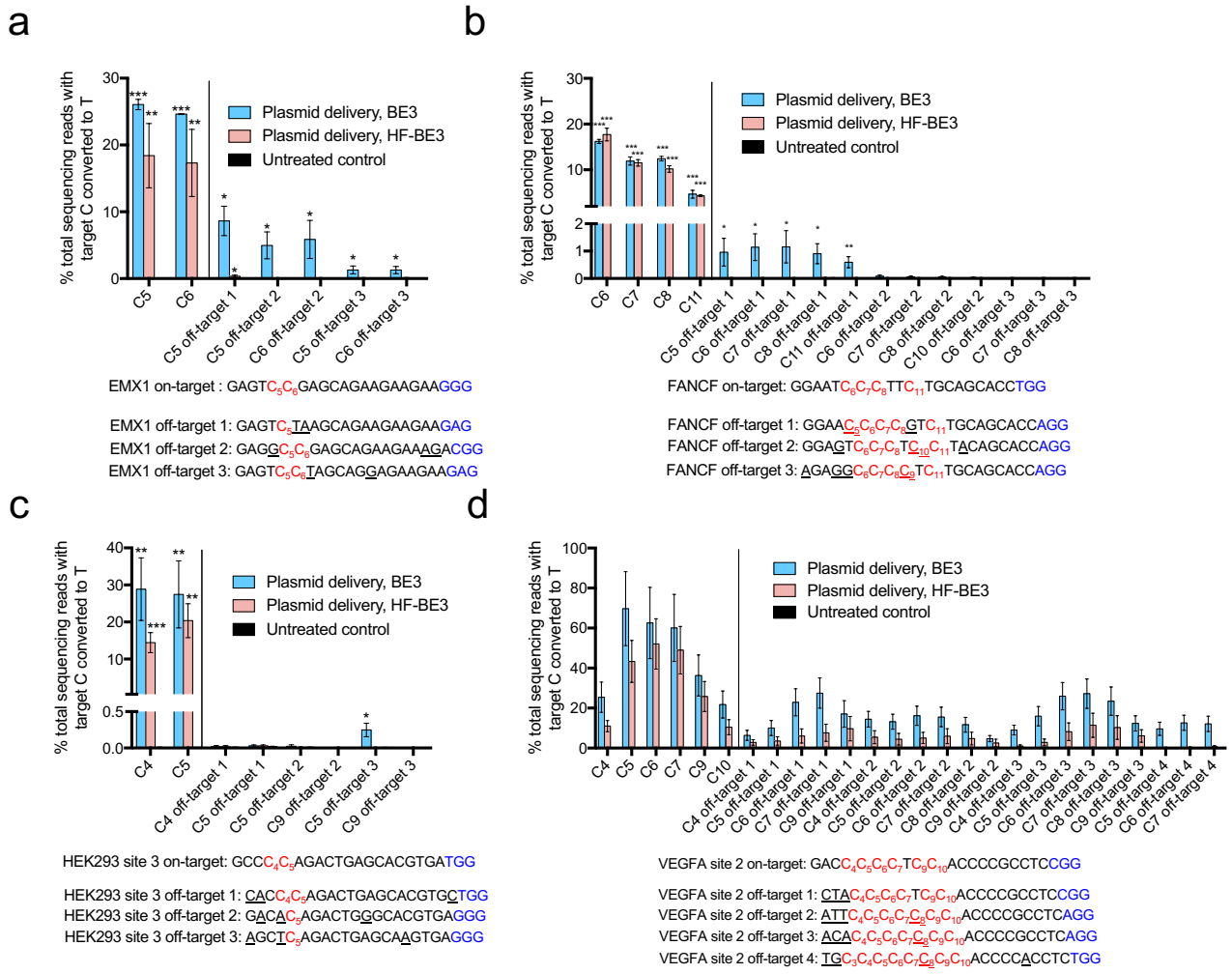


Figure 2.3: Activity of a high fidelity base editor (HF-BE3) in human cells. (a), (b) and (c): On- and off-target editing associated with plasmid transfection of BE3 and HF-BE3 was assayed using high-throughput sequencing of genomic DNA from HEK293T cells treated with sgRNAs targeting non-repetitive genomic loci EMX1 (a), FANCF (b), and HEK293 site 3 (c). On- and off-target loci associated with each sgRNA are separated by a vertical line. (d): On- and off-target editing associated with the highly repetitive sgRNA targeting VEGFA site 2. Values and error bars reflect mean ± S.D. of three independent biological replicates performed on different days. For (a), (b) and (c), stars indicate significant editing based on a comparison between the treated sample and an untreated control. * $p \leq 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$ (Student's two tailed t-test). For (d), asterisks are not shown since all treated samples displayed significant editing relative to the control. Individual *p*-values are listed in in Table 2.1.

To characterize HF-BE3 specificity on an extremely challenging site, we compared BE3 and HF-BE3 off-target activity when targeting the highly repetitive VEGFA site 2 locus. BE3 treatment lead to an average of 15±5% editing of cytosines located in the activity windows of

the four tested off-target sites associated with this sgRNA (all average values quoted in this paragraph represent mean ± s.d. n=3 biological replicates). In contrast, HF-BE3 lead to a 3-fold reduction in absolute off-target editing (5.0±2.3%) at the same off-target sites (Figure 2.3d). When compared to transfection of BE3, HF-BE3 significantly ($p < 0.05$, two-tailed Student's t test) reduced off-target editing at 27 of the 57 cytosines located at off-target loci (Table 2.1), while HF-BE3 treatment lead to a significant reduction ($p < 0.05$ two-tailed Student's t test) in on-target editing at only 3 of 16 the interrogated on-target cytosine residues.

Locus and cytosine position	*P* values (Student's two-tailed t-test) for comparisons between listed treatments									
	plasmid BE3 vs plasmid HF-BE3	plasmid BE3 vs protein BE3	plasmid BE3 vs protein HF-BE3	plasmid BE3 vs control	plasmid HF-BE3 vs protein BE3	plasmid HF-BE3 vs protein HF-BE3	plasmid HF-BE3 vs control	protein BE3 vs protein HF-BE3	protein BE3 vs control	protein HF-BE3 vs control
EMX1, C5	0.053	0.000	0.001	0.000	0.416	0.056	0.003	0.023	0.000	0.004
EMX1, C6	0.065	0.000	0.000	0.000	0.445	0.023	0.004	0.001	0.000	0.003
FANCF, C6	0.152	0.017	0.003	0.000	0.137	0.003	0.000	0.002	0.000	0.004
FANCF, C7	0.591	0.554	0.007	0.000	0.914	0.007	0.000	0.008	0.000	0.006
FANCF, C8	0.011	0.026	0.004	0.000	0.958	0.018	0.000	0.023	0.000	0.007
FANCF, C11	0.524	0.948	0.019	0.001	0.363	0.010	0.000	0.010	0.000	0.021
HEK site 3, C3	0.061	0.001	0.071	0.000	0.002	0.002	0.005	0.00199	0.001	0.003
HEK site 3, C4	0.048	0.924	0.010	0.004	0.001	0.001	0.001	0.00004	0.000	0.001
HEK site 3, C5	0.291	0.592	0.016	0.006	0.243	0.243	0.002	0.00022	0.000	0.001
VEGFA site 2, C3	0.060	0.416	0.239	0.010	0.042	0.280	0.002	0.475	0.002	0.018
VEGFA site 2, C4	0.036	0.191	0.047	0.004	0.032	0.803	0.002	0.066	0.000	0.005
VEGFA site 2, C5	0.098	0.650	0.028	0.003	0.044	0.169	0.002	0.004	0.000	0.001
VEGFA site 2, C6	0.452	0.781	0.118	0.004	0.165	0.239	0.002	0.013	0.000	0.001
VEGFA site 2, C7	0.401	0.683	0.172	0.003	0.120	0.454	0.002	0.026	0.000	0.001
VEGFA site 2, C9	0.225	0.308	0.254	0.004	0.504	0.828	0.004	0.624	0.000	0.002
VEGFA site 2, C10	0.064	0.061	0.023	0.005	0.732	0.257	0.009	0.057	0.000	0.003
EMX1, C5 off-target 1	0.003	0.003	0.002	0.002	0.119	0.036	0.035	0.269	0.255	0.643
EMX1, C5 off-target 2	0.013	0.013	0.013	0.013	0.158	0.294	0.521	0.390	0.058	0.054
EMX1, C6 off-target 2	0.024	0.024	0.024	0.024	0.285	0.560	0.954	0.420	0.103	0.306
EMX1, C5 off-target 3	0.022	0.019	0.019	0.019	0.297	0.297	0.300	>0.99999	0.882	0.815
EMX1, C6 off-target 3	0.017	0.015	0.015	0.015	0.296	0.296	0.328	>0.99999	0.051	0.025
FANCF, C5 off-target 1	0.031	0.031	0.031	0.031	0.314	0.530	0.333	0.337	0.349	0.618
FANCF, C6 off-target 1	0.016	0.016	0.016	0.016	0.347	0.786	0.930	0.338	0.344	0.678
FANCF, C7 off-target 1	0.028	0.028	0.028	0.027	0.374	0.039	0.106	0.353	0.346	0.639
FANCF, C8 off-target 1	0.014	0.014	0.014	0.014	0.341	0.932	0.685	0.343	0.318	0.605
FANCF, C11 off-target 1	0.007	0.007	0.007	0.007	0.374	0.001	0.000	>0.99999	0.475	0.016
FANCF, C6 off-target 2	0.099	0.099	0.032	0.036	0.599	0.475	0.912	0.914	0.393	0.060
FANCF, C7 off-target 2	0.080	0.080	0.027	0.030	0.898	0.638	0.819	0.539	0.859	0.530
FANCF, C8 off-target 2	0.123	0.123	0.045	0.050	0.789	0.538	0.960	0.047	0.539	0.252
FANCF, C10 off-target 2	0.093	0.093	0.029	0.033	0.630	0.509	0.847	0.768	0.670	0.482
FANCF, C11 off-target 2	0.264	0.264	0.127	0.107	0.599	0.658	0.326	>0.99999	0.047	0.345
FANCF, C6 off-target 3	0.872	0.872	0.492	0.108	0.239	0.493	0.129	0.469	0.584	0.252
FANCF, C7 off-target 3	>0.99999	>0.99999	0.859	0.016	0.537	0.866	0.116	0.572	0.272	0.595
FANCF, C8 off-target 3	0.886	0.886	0.246	0.757	>0.99999	0.001	0.648	0.495	0.780	0.650
FANCF, C10 off-target 3	0.566	0.566	0.284	0.202	0.053	0.387	0.260	0.453	0.913	0.541
FANCF, C11 off-target 3	0.422	0.422	0.145	0.145	0.495	0.230	0.230	0.658	0.658	>0.99999
HEK293 site 3, C3 off-target 1	>0.99999	0.910	0.412	0.326	0.910	0.412	0.326	0.293	0.223	0.480
HEK293 site 3, C4 off-target 1	>0.99999	0.994	0.437	0.391	0.994	0.437	0.391	0.495	0.451	0.614
HEK293 site 3, C5 off-target 1	>0.99999	0.616	0.814	0.337	0.616	0.814	0.337	0.459	0.116	0.481
HEK293 site 3, C3 off-target 2	0.285	0.473	0.100	0.473	0.141	0.735	0.141	0.038	>0.99999	0.038
HEK293 site 3, C5 off-target 2	0.375	0.294	0.177	0.294	0.687	0.428	0.687	0.064	>0.99999	0.064
HEK293 site 3, C9 off-target 2	0.053	0.624	0.374	0.624	0.554	0.154	0.554	0.872	>0.99999	0.872
HEK293 site 3, C3 off-target 3	0.067	0.116	0.768	0.435	0.519	0.230	0.561	0.349	0.768	0.643
HEK293 site 3, C5 off-target 3	0.011	0.011	0.011	0.011	0.016	0.643	0.435	0.184	0.025	0.346
HEK293 site 3, C9 off-target 3	>0.99999	0.374	0.652	0.811	0.132	0.539	0.776	0.609	0.643	0.893
VEGFA site 2, C4 off-target 1	0.101	0.015	0.014	0.012	0.041	0.032	0.025	0.117	0.012	0.001
VEGFA site 2, C5 off-target 1	0.060	0.013	0.012	0.010	0.078	0.062	0.044	0.201	0.009	0.012
VEGFA site 2, C6 off-target 1	0.019	0.005	0.005	0.004	0.080	0.062	0.045	0.087	0.002	0.012
VEGFA site 2, C7 off-target 1	0.017	0.004	0.004	0.003	0.080	0.060	0.037	0.076	0.001	0.002
VEGFA site 2, C9 off-target 1	0.230	0.088	0.037	0.011	0.667	0.256	0.051	0.134	0.004	0.007
VEGFA site 2, C10 off-target 1	0.535	0.136	0.106	0.035	0.283	0.211	0.050	0.717	0.028	0.010
VEGFA site 2, C4 off-target 2	0.038	0.004	0.003	0.003	0.087	0.051	0.048	0.063	0.048	0.134
VEGFA site 2, C5 off-target 2	0.033	0.004	0.004	0.004	0.078	0.061	0.059	0.028	0.020	0.248
VEGFA site 2, C6 off-target 2	0.026	0.005	0.005	0.004	0.051	0.038	0.038	0.043	0.038	0.783
VEGFA site 2, C7 off-target 2	0.053	0.006	0.005	0.005	0.072	0.056	0.055	0.078	0.064	0.704
VEGFA site 2, C8 off-target 2	0.071	0.006	0.006	0.006	0.079	0.065	0.065	0.118	0.107	0.703
VEGFA site 2, C9 off-target 2	0.193	0.008	0.007	0.006	0.103	0.090	0.084	0.068	0.007	0.217
VEGFA site 2, C10 off-target 2	0.063	0.003	0.003	0.002	0.116	0.107	0.090	0.545	0.016	0.346
VEGFA site 2, C4 off-target 3	0.005	0.003	0.003	0.003	0.091	0.031	0.030	0.116	0.107	0.158
VEGFA site 2, C5 off-target 3	0.011	0.007	0.005	0.005	0.211	0.048	0.042	0.220	0.177	0.001
VEGFA site 2, C6 off-target 3	0.020	0.005	0.003	0.003	0.142	0.038	0.033	0.193	0.149	0.015
VEGFA site 2, C7 off-target 3	0.045	0.006	0.003	0.003	0.101	0.035	0.030	0.093	0.060	0.083
VEGFA site 2, C8 off-target 3	0.069	0.007	0.005	0.005	0.087	0.045	0.039	0.120	0.067	0.041
VEGFA site 2, C9 off-target 3	0.093	0.006	0.005	0.005	0.041	0.032	0.028	0.396	0.195	0.005
VEGFA site 2, C10 off-target 3	0.342	0.011	0.008	0.007	0.109	0.081	0.069	0.273	0.098	0.036
VEGFA site 2, C3 off-target 4	0.001	0.001	0.001	0.001	0.374	0.374	0.230	0.271	0.358	0.633
VEGFA site 2, C4 off-target 4	0.007	0.006	0.006	0.006	0.137	0.137	0.137	0.592	0.862	0.690
VEGFA site 2, C5 off-target 4	0.007	0.007	0.007	0.007	0.026	0.017	0.018	0.461	0.655	0.279
VEGFA site 2, C6 off-target 4	0.005	0.004	0.004	0.004	0.021	0.018	0.018	0.398	0.546	0.149
VEGFA site 2, C7 off-target 4	0.007	0.006	0.006	0.006	0.051	0.048	0.050	0.373	0.720	0.029
VEGFA site 2, C8 off-target 4	0.007	0.006	0.006	0.006	0.092	0.092	0.092	0.325	0.014	0.275
VEGFA site 2, C9 off-target 4	0.016	0.007	0.007	0.007	0.150	0.150	0.150	0.502	1.000	0.615
VEGFA site 2, C10 off-target 4	0.213	0.009	0.009	0.009	0.261	0.261	0.261	0.653	0.575	0.660

Table 2.1: ***P*-values for differences in base editing under different treatment conditions at all loci evaluated in this study.** *p*-values were calculated using the Student's two tailed t-test

as described in the Materials and Methods. When the p-value indicated a significant difference ($p < 0.05$), the corresponding entry has been highlighted in red.

In addition to considering the differences between absolute editing at off-target loci, we also calculated the on-target:off-target editing specificity ratio by dividing the observed on-target efficiency by the off-target efficiency (Figures 2.4a and 2.4b). This metric takes into account any reduction in on-target editing associated with installation of the HF-mutations, and is useful for applications sensitive to both the efficiency and specificity of base editing. Off-target editing by HF-BE3 was below the detection limit of high-throughput sequencing for several off-target loci. For these cases, we assumed a conservative off-target editing efficiency equal to the upper limit of detection (0.025% C→T conversion; see Methods). Based on this analysis, the average improvement in specificity ratio upon installation of the HF mutations across all 34 target cytosines we examined was 19-fold, when plasmid delivery of the two constructs was performed. These results collectively establish that for non-repetitive sites (Figure 2.4a) as well as a highly repetitive site (Figure 2.4b), HF-BE3 results in substantially enhanced base editing specificity with only a modest reduction in on-target editing efficiency compared to BE3.

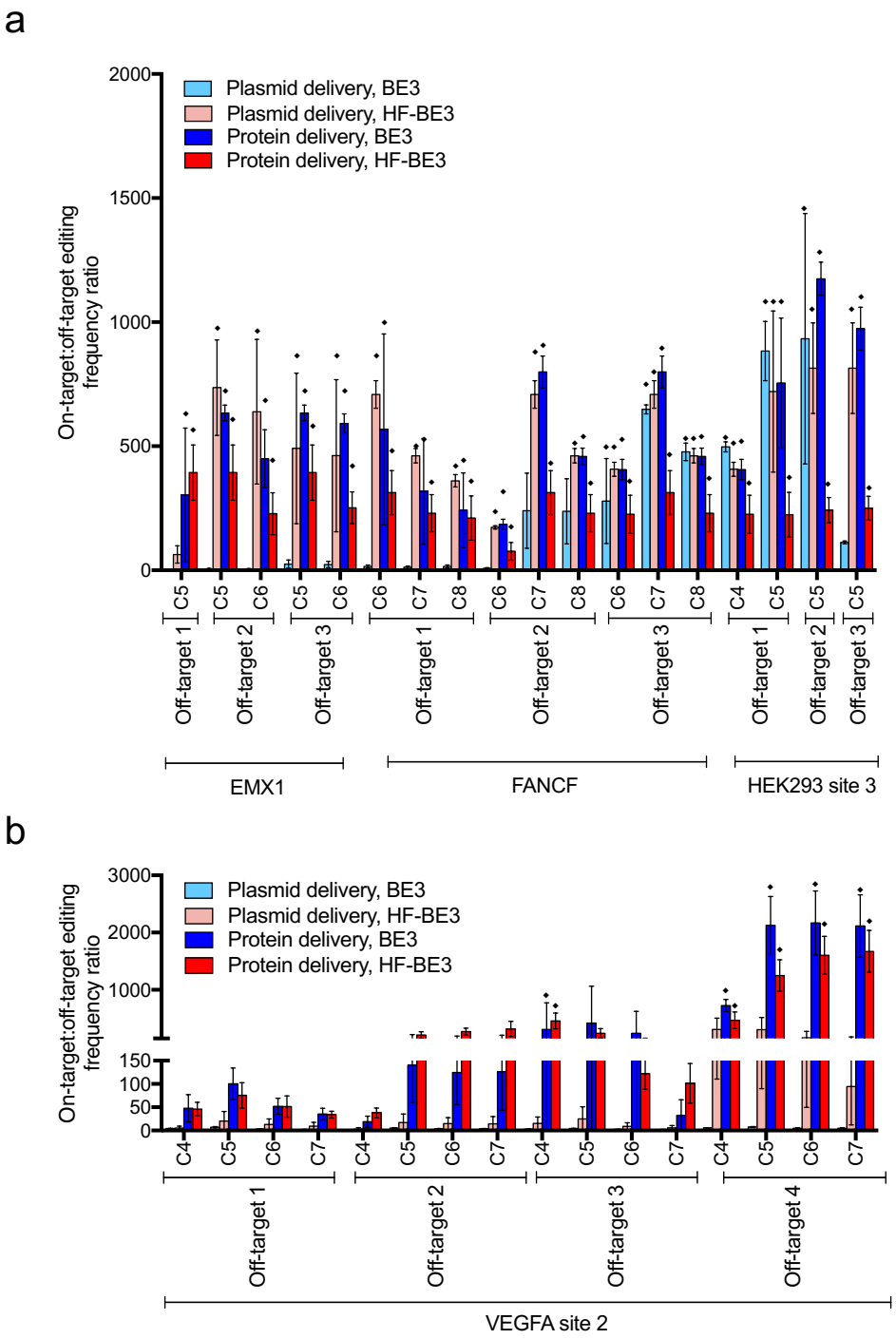


Figure 2.4: On-target:off-target base editing frequency ratios for plasmid and protein delivery of BE3 and HF-BE3. Base editing on-target:off-target specificity ratios were calculated by dividing the on-target editing percentage at a particular cytosine in the activity window by the off-target editing percentage at the corresponding cytosine for the indicated off-target locus – see Methods. When off-target editing was below the threshold of detection (0.025% of sequencing reads), we set the off-target editing to the limit of detection (0.025%) and divided the on-target editing percentage by this upper limit. In these cases, denoted by ♦, the specificity ratios shown represent lower limits. Specificity ratios are shown for non-repetitive sgRNAs

FANCF, HEK 293 site 3, and FANCF (a) and for the highly repetitive sgRNA VEGFA site 2 (b). Values and error bars reflect mean ± S.D. of three independent biological replicates performed on different days.

2.2.4 RNP delivery of BE3 enables DNA-free base editing

Next, we studied the ability of BE3 in DNA-free, RNP form to mediate base editing when directly delivered into cultured human cells. We and others recently established that cationic lipid reagents can potently deliver negatively charged proteins or protein:nucleic acid complexes into mammalian cells including ribonucleoprotein (RNP) complexes and that RNP delivery can substantially reduce off-target genome editing[124,167,169].

We combined the commercially available cationic lipid Lipofectamine 2000 with either purified BE3 protein or HF-BE3 protein after pre-complexation with a guide RNA targeting the EMX1, HEK293 site 3, FANCF, or VEGFA site 2 locus and incubated the resulting lipid:RNP complexes with HEK293T cells. After 72 h, we harvested genomic DNA and analyzed on-target and off-target base editing by high-throughput DNA sequencing. As with all Cas9-based technologies, we observed substantial variations in editing efficiency at different genomic loci (Figures. 2.3 and 2.5). To display trends associated with in on-target editing efficiency between different treatments, we calculated mean on-target base editing efficiencies at the four tested loci (Figure. 2.6a). Protein delivery of BE3 (200 nM) lead to on-target editing efficiencies comparable to those observed with plasmid transfection (26±4% vs. 29±5% respectively; mean ± s.e.m. n=3 biological replicates; Figure 2.6a).

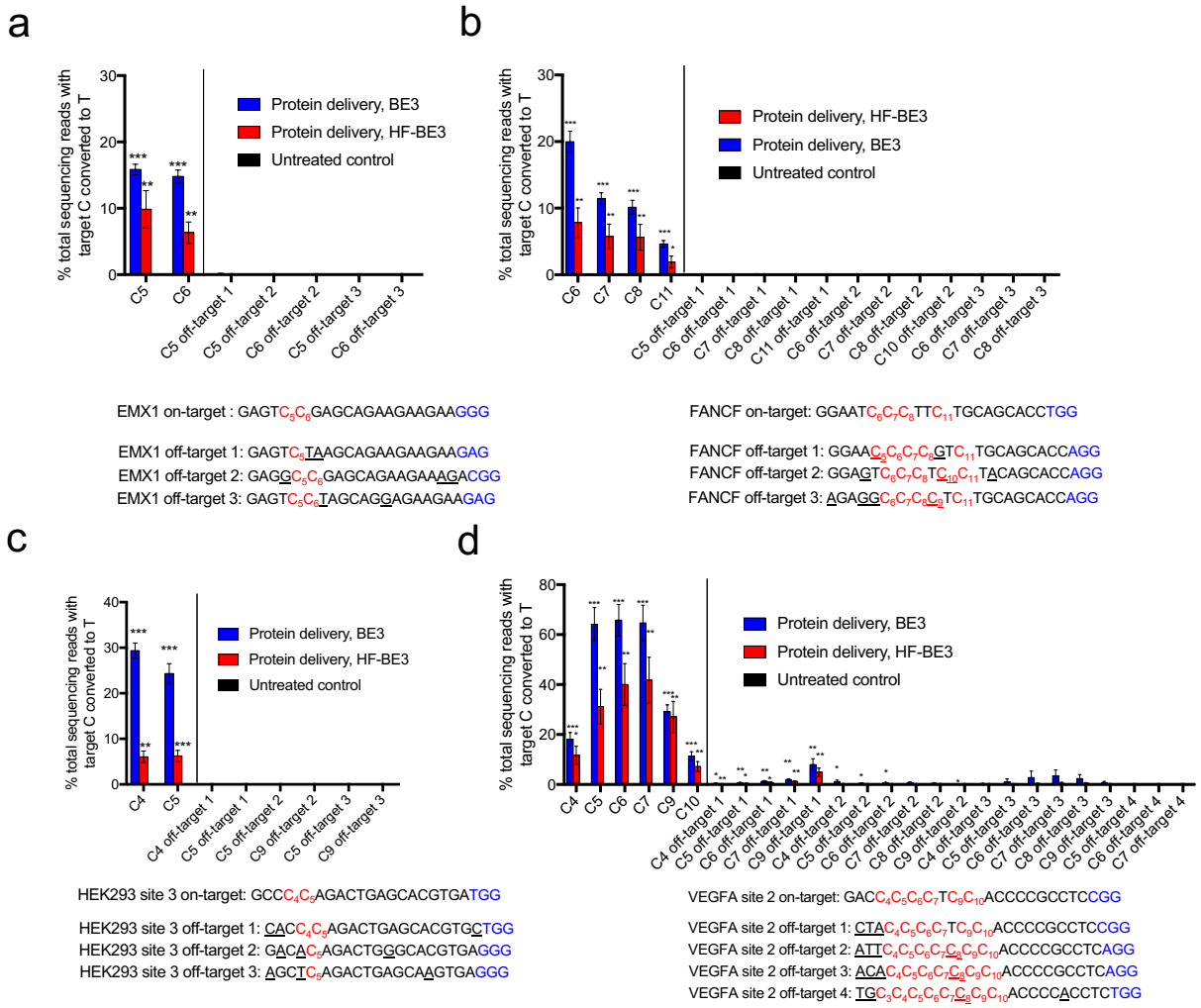


Figure 2.5: Protein delivery of base editors into human cells. (a-d**)**: On- and off-target editing associated with RNP delivery of base editors complexed with sgRNAs targeting EMX1 (a), FANCF (b), HEK293 site 3 (c) and VEGFA site 2 (d). Off-target base editing was undetectable at all of the sequenced loci for non-repetitive sgRNAs. Values and error bars reflect mean ± S.D. of three independent biological replicates performed on different days. Stars indicate significant editing based on a comparison between the treated sample and an untreated control. * $p \leq 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$ (Student's two tailed t-test).

In contrast, protein delivery of HF-BE3 reduced on-target editing compared to protein delivery of BE3 at the four genomic loci studied (average editing efficiency of 13±3% vs. 26±4%, respectively; mean ± s.e.m. n=3 biological replicates; Figure 2.6a). Since HF-BE3 and BE3 have comparable editing efficiencies in a test tube (Figure 2.1b) and editing is only slightly reduced

when HF-BE3 is expressed from plasmids in HEK293T cells (Figure 2.3a-d), it is tempting to speculate that the decreased efficiency of editing from HF-BE3 protein delivery may be a result of decreased HF-BE3 stability in mammalian cells. Lower stability could be offset by continual expression from a plasmid, but not following one-time protein delivery. This observation is consistent with a recent report of reduced on-target indel formation with purified HF-Cas9 compared to purified Cas9 when nucleofected into $CD34^+$ hematopoietic stem and progenitor cells[170]. While this work was in review, Kim et al demonstrated RNP delivery of BE3 into mouse embryos using electroporation[93]. To the best of our knowledge, our approach is the first DNA-free technique capable of generating precise changes to individual nucleotides in mammalian cells without electroporation, which has limited *in vivo* therapeutic relevance.

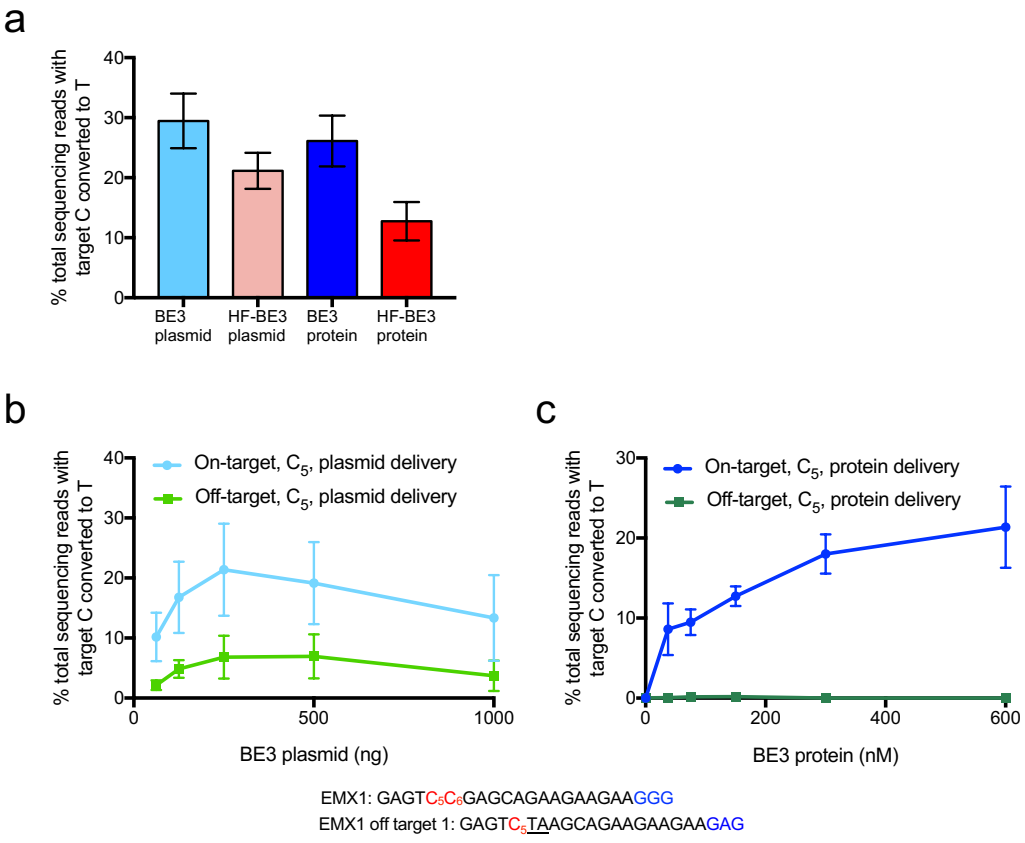


Figure 2.6: Effect of dosage of BE3 protein or plasmid on the efficiency of on-target and off-target base editing in human cells (a): On-target editing efficiency at each of the four genomic loci was averaged across all edited cytosines in the activity window for each sgRNA. Values and error bars reflect mean ± S.E.M of three independent biological replicates performed on different days. (b, c): On- and off-target editing at the EMX1 site arising from BE3 plasmid

titration (b) or BE3 protein titration (c) in HEK293T cells. Values and error bars reflect mean ± S.D. of three independent biological replicates performed on different days.

2.2.5 RNP delivery of base editors greatly enhances DNA specificity

Importantly, while RNP delivery of BE3 and HF-BE3 led to substantial on-target base editing, we observed no instances of measurable base editing (< 0.025%) at any of the nine tested off-target loci associated with EMX1, FANCF and HEK293 site 3, (Figures 2.5a-c). In contrast, plasmid delivery of BE3 lead to an average of 1.1±0.3% (mean ± s.d. n=3 biological replicates) off-editing across all sequenced cytosines within the base editing activity window, and detectable off-target editing at 11 of 16 off-target cytosines located at these nine off-target loci (Figures 2.3a-d). At off-target loci of the three non-repetitive loci tested, BE3 protein delivery lead to a 26-fold higher average specificity ratio than that of plasmid delivery (Figure 2.4a). These results reveal that RNP delivery of base editors dramatically increases the DNA specificity of base editing.

Protein delivery of either BE3 or HF-BE3 also resulted in greatly improved base editing specificity at the highly promiscuous VEGFA site 2 locus compared to plasmid delivery of either BE3 or HF-BE3 (compare Figures 2.3 and 2.5; see Table 2.1). Absolute frequencies of base editing at the off-target loci associated with this site were reduced upon protein delivery at least 10-fold for both BE3 (plasmid delivery: 15±4% off-target editing; protein delivery: 1.3±0.4% off-target editing; all averaged values in this paragraph represent mean ± s.d. n=3 biological replicates) and HF-BE3 (plasmid delivery: 5±2% off-target editing; protein delivery: 0.5±0.1% off-target editing). Across all four studied loci, base editing specificity ratios for on-target:off-target editing increased an average of 66-fold for protein delivery of BE3 compared with plasmid delivery of BE3 (Figure 2.4). Collectively, these results reveal that for both repetitive and non-repetitive target sites, RNP versus DNA delivery is a stronger determinant of base editing specificity than the presence or absence of the high-fidelity Cas9 mutations.

Neither introduction of the HF mutations nor delivery method substantially altered the low indel rates associated with base editing. Indel frequencies at all on-target loci across all treatment conditions in this study remained low (typically ≤ 5%; Figure 2.7a), and the editing:indel ratio remained higher in all cases tested (typically ≥ 10-fold; Figure 2.7b) than in previous studies using optimized HDR protocols[170-172]. For non-repetitive sgRNAs, very few indels were observed at off-target loci (Figure 2.7c), although we note that plasmid delivery of BE3 generated up to 5% indels for off-target loci associated with VEGFA site 2 (Figure 2.7c).

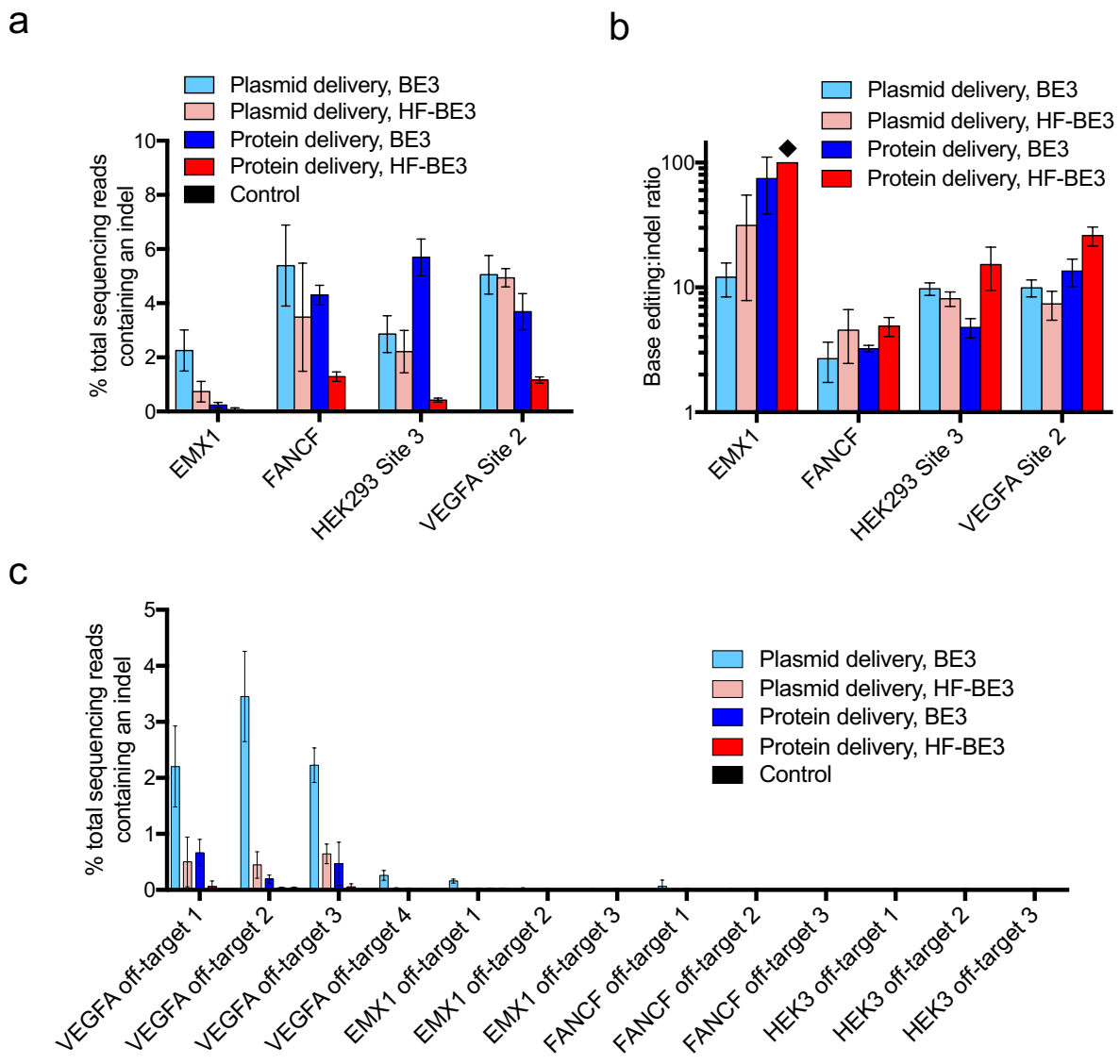


Figure 2.7: Indel formation associated with base editing at genomic loci. (a): Indel frequency at on-target loci for VEGFA site 2, EMX1, FANCF, and HEK293 site 3 sgRNAs. (b):

Ratio of base editing:indel formation. The diamond (♦) indicates no indels were detected (no significant difference in indel frequency in the treated sample and in the untreated control). (c): Indels observed at the off-target loci associated with the on-target sites interrogated in (a). Values and error bars reflect mean ± S.D. of three independent biological replicates performed on different days.

Taken together, these results establish that protein delivery of base editors maintains on-target base editing efficiency and greatly enhances editing specificity relative to delivery of plasmid DNA.

2.2.6 RNP delivery decouples on- and off-target editing

Given the striking enhancement of base editing specificity associated with protein delivery of BE3, we investigated if this improvement was a result of a reduction in the total quantity of active genome editing agent delivered into the cell. Using the sgRNA targeting EMX1, we performed a dose response study for plasmid (Figure 2.6b) and protein delivery (Figure 2.6c). To maximize transfection efficiency between treatment conditions, the volume of Lipofectamine 2000 was 1.5 µL for all tests, and the base editor protein:sgRNA molar ratio was maintained at 1:1.1 for protein delivery. For plasmid delivery, we used a mass ratio of sgRNA plasmid:BE3 plasmid of 1:3 (molar ratio ~1:1) and 1.5 µL of Lipofectamine 2000. We observed off-target base editing under all conditions tested for plasmid delivery (Figure 2.6b), but virtually no off-target editing under all protein delivery conditions tested (Figure 2.6c).

We performed linear regression analysis to assess the relationship between on- and off-target editing for plasmid and protein delivery. For plasmid delivery, off-target editing was closely associated with on-target editing rates (R^2 = 0.95, p = 0.0012 for non-zero slope, F-test), whereas there was no significant association between off-target and on-target editing using protein delivery (R^2 = 0.078, p = 0.59 for non-zero slope, F-test).

These data indicate that protein delivery of base editors offers an inherent specificity advantage that is independent of dosage. Together with our previous observations[167,173], these

findings support a model in which the higher DNA specificity of base editing from protein delivery compared to DNA delivery arises from the ability of protein delivery to avoid extended exposure of the genome to base editors, thereby minimizing the opportunity of base editors to process off-target loci after on-target loci have already been modified.

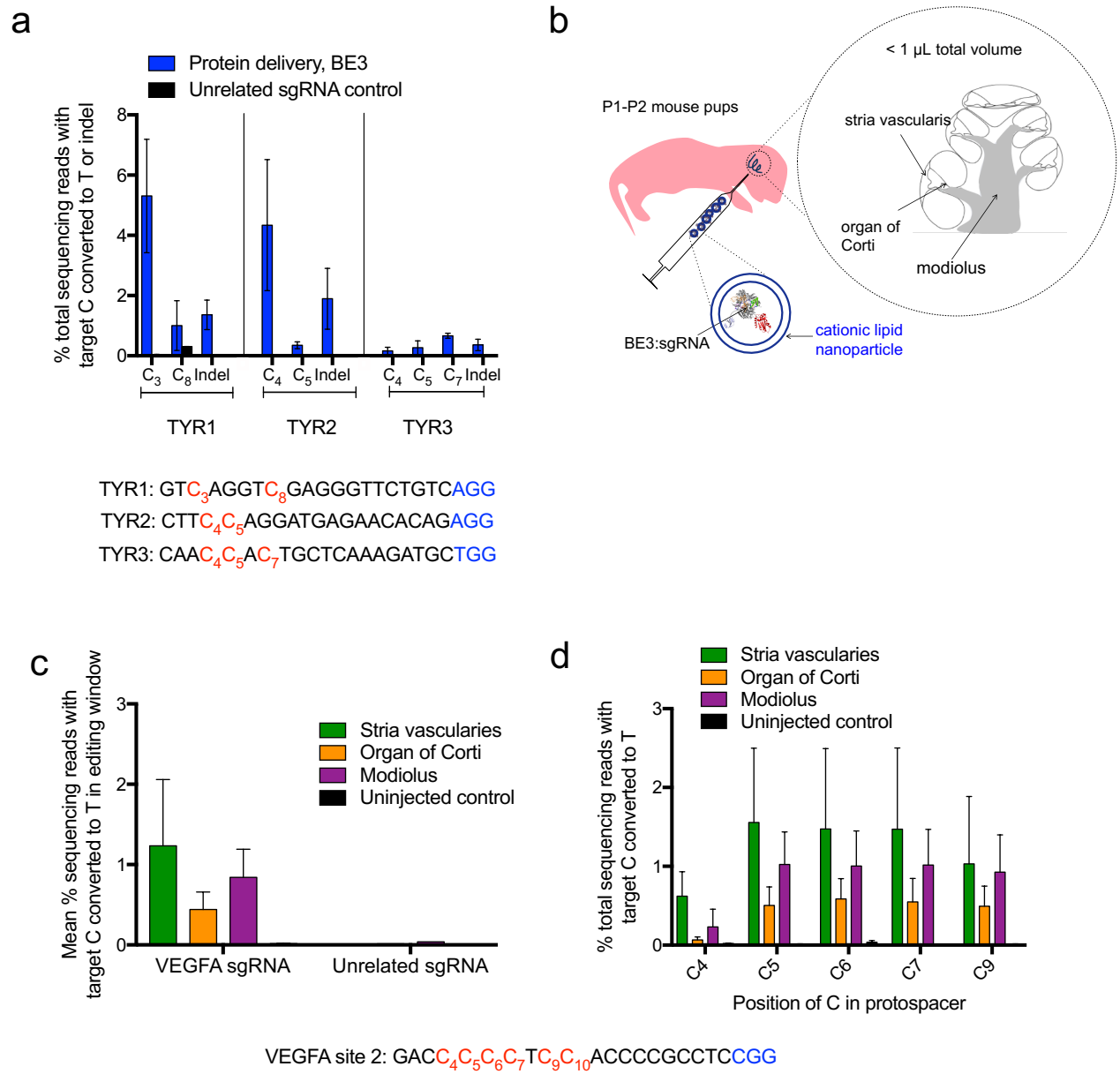


Figure 2.8: DNA-free *in vivo* base editing in zebrafish embryos and in the inner ear of live mice using RNP delivery of BE3. (a): On-target genome editing in zebrafish harvested 4 days after injection of BE3 complexed with indicated sgRNA. Values and error bars reflect mean ±

s.d. of three injected and three control zebrafish. Controls were injected with BE3 complexed with an unrelated sgRNA. (b): Schematic showing *in vivo* injection of BE3:sgRNA complexes encapsulated into cationic lipid nanoparticles (c): Base editing of cytosine residues in the base editor window at the VEGFA site 2 genomic locus (d): On-target editing at each cytosine in the base editing window of the VEGFA site 2 target locus. (d): (c and d): Values and error bars reflect mean ± S.E.M. of three mice injected with sgRNA targeting VEGFA Site 2, three uninjected mice and one mouse injected with unrelated sgRNA.

2.2.7 DNA-free base editing in zebrafish and mice

The above observations suggested the promise of protein delivery of BE3 to maintain on-target base editing while eliminating detectable off-target base editing. We therefore tested whether protein delivery of BE3 could be used to generate specific point mutations in zebrafish by injecting BE3:sgRNA complexes targeting the *tyrosinase* locus into fertilized zebrafish embryos. We harvested genomic DNA from the resultant zebrafish larvae 4 days post-injection and measured base editing and indel frequencies by high-throughput sequencing (Figure 2.8a). Two of the three BE3:sgRNA complexes tested induced substantial point mutations *in vivo* (TYR1: $C_3 \rightarrow T_3$ 5.3±1.8%, TYR2: $C_4 \rightarrow T_4$ 4.3±2.1%; mean ± s.d. of n=3 injected embryos; Figure 2.8a). Sequences of zebrafish loci are listed in Table 2.3.

Site	Sequence
TYR1	GTC_3AGGTC_8GAGGGTTCTGTCAGG
TYR2	CTTC_4C_5AGGATGAGAACACAGAGG
TYR3	CAAC_4C_5AC_7TGCTCAAAGATGCTGG

Table 2.3: Protospacer and PAM sequences for the zebrafish genomic loci studied in this work.

Finally, we applied these developments to achieve DNA-free, high-specificity base editing in mice. To maximize the likelihood of observing on- and off-target base editing *in vivo*,

we used the highly repetitive sgRNA targeting VEGFA site 2; conveniently, the murine and human genomes are identical at this target site.

Using cultured murine NIH/3T3 cells, we confirmed that BE3 protein delivery yielded efficient on-target base editing at this locus 34±11% (Figure 2.9a; all averaged editing percentages in this paragraph represent mean ± s.d. n=3 biological replicates). We used the Cutting Frequency Determinant (CFD) algorithm[167,173] to predict off-target loci in the mouse genome associated with the VEGFA site 2 sgRNA (Table 2.2). Using cultured NIH/3T3 cells, we confirmed that two of the top four predicted off-target loci are indeed modified by plasmid delivery of BE3 in cultured murine cells (CFD off-target locus 1, 9±5% editing; and CFD off-target locus 4, 3±2% editing). Consistent with our results from human cells, protein delivery of BE3 reduced off-target editing to levels similar to that of negative controls (Figure 2.9c, e). The mean base editing specificity ratio for CFD off-target loci 1 and 4 increased from 28±13 for plasmid delivery of BE3 to ≥ 780±300 for protein delivery of BE3 (values represent mean ± s.e.m.; n=3 biological replicates).

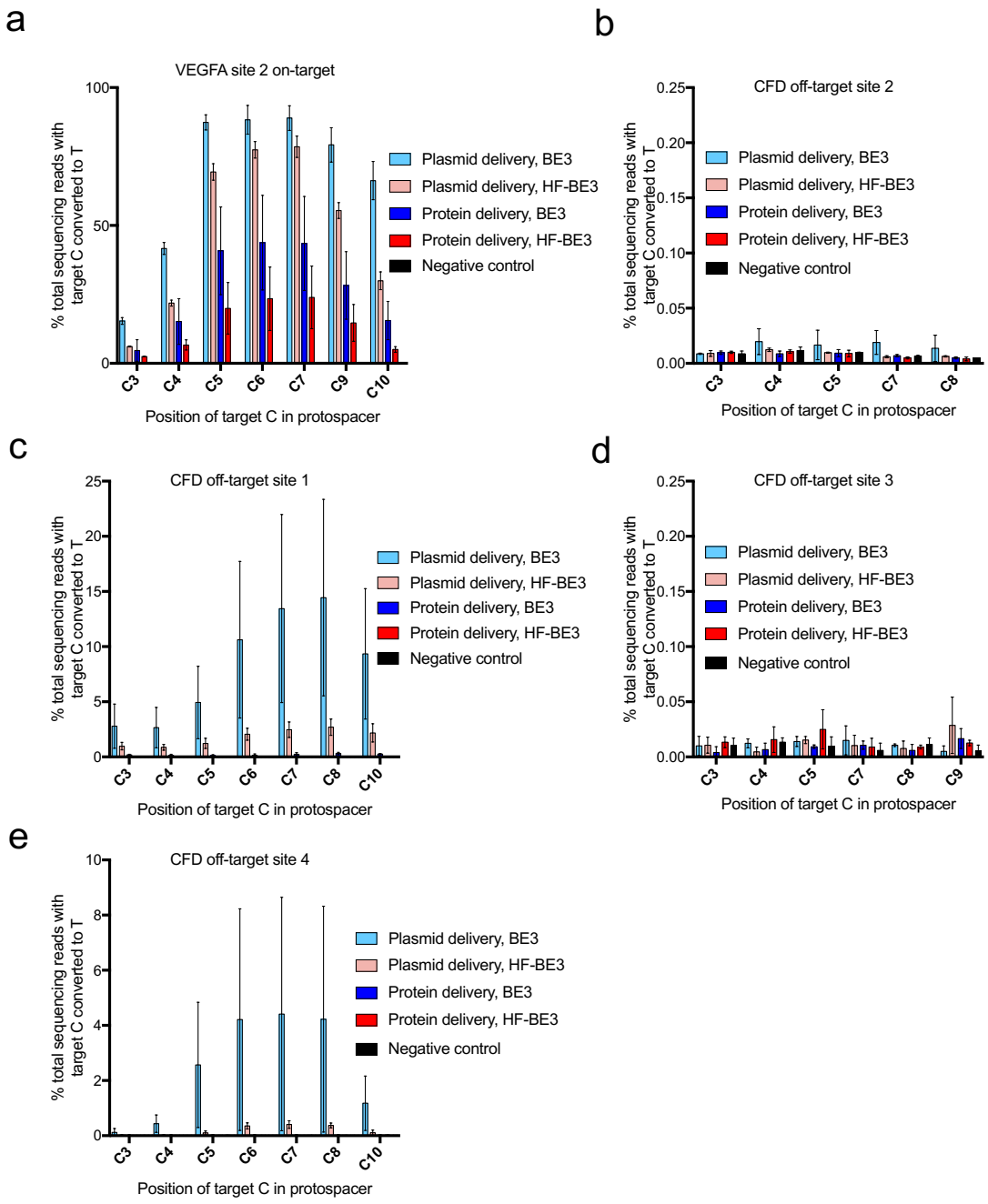


Figure 2.9: On- and off-target base editing in murine NIH/3T3 cells. (a): On-target base editing associated with the 'VEGFA site 2' sgRNA (See Figure 5E for sequences). The negative control corresponds to cells treated with plasmid encoding BE3 but no sgRNA. Values and error bars reflect mean ± S.D. of three independent biological replicates performed on different days. (B-E): Off-target editing associated with this site was measured using high-throughput DNA sequencing at the top four predicted off-target loci for this sgRNA (sequences shown in Figure 5E). (b): off-target 2, (c): off-target 1, (d): off-target 3, (e): off-target 4. Values and error bars reflect mean ± S.D. of three independent biological replicates performed on different days.

Site	Sequence	CFD score	Description of locus
On-target	GACCCCCTCCACCCCGCCTCCGG		VEGFA site 2
Off-target 1	TCCCCCCTCCACCCCACCTCCGG	0.7857	intergenic:mmu-mir-21c-Nrp1/Mir1903
Off-target 2	TGCCCACCTCACCCCGCCTCTGG	0.65	intron:Vipr1
Off-target 3	GCCCCTCCCAACCCCACCTCTGG	0.6323	intron:Nos1ap
Off-target 4	CACCCCCCTCACCCCGCCTCAGG	0.625	intergenic:Unc5b-mmu-mir-6408

Table 2.2: Protospacer and PAM sequences for the predicted off-target loci in the mouse genome associated with the VEGFA site 2 sgRNA. CFD scores [174] were calculated using CRISPOR [175]. Positions in the off-target protospacers that differ from the on-target sequence are underlined.

To establish DNA-free base editing in mice, we combined BE3:sgRNA complexes with Lipofectamine 2000 (Figure 2.8b) and performed intracochlear injections into mouse pups at P1-P2. Injected cochlear tissues were harvested 3-4 days post-injection and micro-dissected into 5-7 samples per cochlear region. Control cochlea from uninjected mice were harvested simultaneously. Genomic DNA was extracted from the harvested tissue, amplified by qPCR to late-exponential phase, and subjected to high-throughput DNA sequencing to measure C→T conversion. Although it is impossible to quantitate base editing efficiency among treated cells because it is not possible to retrieve DNA exclusively from cells exposed to base editor protein, we observed unambiguous base editing from tissue in three regions of the cochlea: the basal end of the organ of Corti, the stria vascularis and the modiolus (Figure 2.8c-d). We detected no significant indel formation in treated tissue samples (< 0.1% indels; Figure 2.10b).

The percentage of cochlear cells containing target C→T conversion (Figure 2.8c) was significantly lower than that observed in treated NIH/3T3 cells in culture (Figure 2.8a), consistent with the highly localized nature of lipid-based protein delivery and our inability to isolate DNA exclusively from cells exposed to base editor. Nonetheless, local delivery offers key advantages for accessible applications, including control over which cell types are edited, and ease of preparation and administration.

Finally, we analyzed off-target editing following intracochlear injection of BE3:sgRNA:lipid complexes. Analysis of all four predicted off-target loci, including the confirmed off-target sites CFD locus 1 and CFD locus 4, in genomic DNA from the cochlear tissue of mice injected with the BE3:VEGFA site 2 sgRNA:lipid complex revealed no detectable C→T conversion or indel formation above that observed in untreated controls samples for any of the off-target loci tested (Figure 2.10).

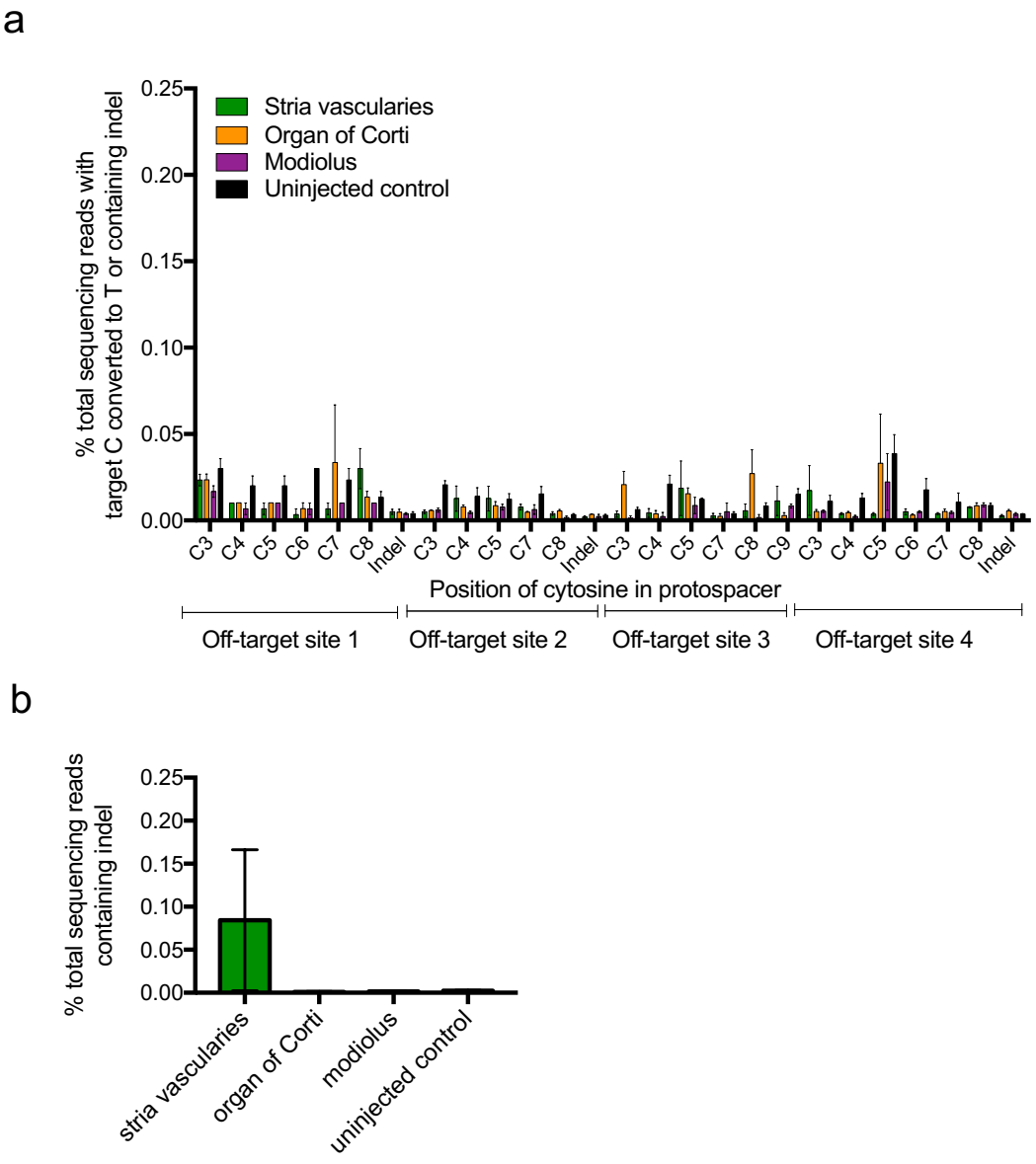


Figure 2.10: Off-target base editing and on-target indel analysis from *in vivo*-edited murine tissue. (a): Editing is plotted for each cytosine in the base editing window of off-target loci associated with VEGFA site 2. (b): Indel rates at the on-target base editor locus. Values and

error bars reflect mean ± S.E.M of three injected and three control mice.

Together, these *in vivo* base editing results establish a virus-free, DNA-free strategy for the precise conversion of individual nucleotides in the genomic DNA of animals with high DNA sequence specificity.

2.3 Conclusions and discussion

The strategies developed and implemented in this study expand the utility and applicability of base editing by removing or reducing off-target base editing and establishing a DNA-free delivery method that supports *in vivo* base editing. Protein delivery improves base editing specificity in human and murine cells compared with plasmid delivery of the same constructs (Figures 2.3,2.5 and 2.8), and enables specific base editing in zebrafish and in the mouse cochlea (Figure 2.8).

We generated a high-fidelity base editor by installing into BE3 mutations known to enhance the DNA specificity of Cas9[94]. The installation of these mutations into Cas9 was reported to result in undetectable indel formation at off target loci associated with non-repetitive sgRNAs, including the EMX1 locus interrogated here (Figure 2.3a)[94]. The specificity enhancements we observed in HF-BE3, while substantial, were more modest; HF-BE3 exhibited detectable off-target base editing at both repetitive and non-repetitive loci when delivered as plasmid DNA into mammalian cells (Figures 2.3a,d and Figure 2.8c,e). It is tempting to speculate that this specificity enhancement difference may arise from the fact that base editing, unlike Cas9-mediated indel formation, does not require DNA cleavage but only necessitates DNA-binding and R-loop formation[33], and some of the enhanced specificity of HF-Cas9 may arise from impaired DNA cleavage at already-bound off-target loci.

In a second attempt to reduce off-target base editing, we demonstrated that RNP delivery of base editors leads to decoupling of on- and off- target editing (Figure 2.6b,c). RNP

delivery ablated off-target editing at non-repetitive sites while maintaining on-target editing comparable to plasmid delivery (Figures 2.5a-c, 2.6a), and greatly reduced off-target editing even at the highly repetitive VEGFA site 2 (Figure 2.5d). RNP delivery of base editors may be especially useful for *in vivo* editing applications in which cellular dosage is typically difficult to control or characterize.

We and others previously used RNP delivery of Cas9 coupled with delivery of a donor DNA template to perform HDR-based genome editing in mammalian cells. These approaches, however, remain limited by low efficiency, cell-state dependence, and indel formation efficiencies typically exceeding those of desired HDR outcomes, especially for point mutation correction[167,170,171,176] DNA-free base editing, in contrast, generates a substantial excess of edited product relative to stochastic indels both *in vivo* and in cells (Figure 2.8a, Figures 2.10a and 2.10b). To the best of our knowledge, RNP delivery of base editors represents the first strategy for generating specific and precise modifications to genomic DNA without requiring exogenous DNA.

2.4 Methods

Cloning of plasmids

The plasmids in this study were generated by USER cloning. Phusion U Hot Start polymerase (Thermo Fisher) was used to install point mutations and construct protein expression plasmids from previously reported constructs[33]. Plasmids for expression of BE3 and HF-BE3 are available from Addgene.

Expression and purification of BE3 and HF-BE3

BL21 Star (DE3)-competent E. coli cells were transformed with plasmids encoding the bacterial codon optimized base editors with a His_6 N-terminal purification tag. A single colony was grown overnight in Luria-Bertani (LB) broth containing 50 μg mL-1 kanamycin at 37 °C. The

cells were diluted 1:200 into 2 L of the same media and grown at 37 °C until OD600 = 0.70-0.75. The cultures were incubated on ice for 60 min and protein expression was induced with 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG, GoldBio). Expression was sustained for 14-16 h with shaking at 18 °C. The subsequent purification steps were carried out at 4 °C. Cells were collected by centrifugation at 6,000 g for 20 min and resuspended in cell collection buffer (100 mM tris(hydroxymethyl)-aminomethane (Tris)-HCl, pH 8.0, 1 M NaCl, 20% glycerol, 5 mM tris(2-carboxyethyl)phosphine (TCEP; GoldBio), 0.4 mM phenylmethane sulfonyl fluoride (PMSF; Sigma Aldrich) and 1 cOmplete, EDTA-free protease inhibitor pellet (Roche) per 50 mL buffer used). Cells were lysed by sonication (6 min total, 3 s on, 3 s off) and the lysate cleared by centrifugation at 25,000 g (20 min).

The cleared lysate was incubated with His-Pur nickel nitriloacetic acid (nickel-NTA) resin (1 mL resin per litre of culture, Thermo Fisher) with rotation at 4 °C for 60-90 min. The resin was washed with 20 column volumes of cell collection buffer before bound protein was eluted with elution buffer ((100 mM tris(hydroxymethyl)-aminomethane (Tris)-HCl, pH 8.0, 0.5 M NaCl, 20% glycerol, 5 mM tris (2-carboxyethyl) phosphine (TCEP; GoldBio), 200 mM imidazole). The resulting protein fraction was further purified on a 5 mL Hi-Trap HP SP (GE Healthcare) cation exchange column using an Akta Pure FPLC. Protein-containing fractions were concentrated using a column with a 100,000 kDa cutoff (Millipore) centrifuged at 3,000 g and the concentrated solution was sterile filtered through an .22 μm PVDF membrane (Millipore). After sterile filtration, proteins were quantified with Reducing Agent Compatible Bicinchoninic acid (BCA) assay (Pierce Biotechnology), snap-frozen in liquid nitrogen and stored in aliquots at -80 °C.

In vitro transcription of sgRNA

Linear DNA fragments containing the T7 RNA polymerase promoter sequence upstream of the desired 20 bp sgRNA protospacer and the sgRNA backbone were generated by PCR (Q5

Hot Start MasterMix, New England Biolabs) and concentrated on minelute columns (Qiagen). sgRNA was transcribed with the HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs) at 16 °C for 14-16 h with 1 µg of linear template per 20 µL reaction. sgRNA was purified using the MEGAClear Transcription Clean Up Kit (Thermo Fisher), according to the manufacturer's instructions. Purified sgRNAs were stored in aliquots at -80 °C.

In vitro deamination assays

All oligonucleotides were purchased from IDT. Single-stranded oligonucleotides synthesized with complementary sequences were combined (5 µL of a 100 µM solution) in Tris buffer pH 8.0 and annealed by heating to 95 °C for 5 min, followed by a gradual cooling to 37 °C at a rate of 0.1 °C second-1 to generate 79 base pair (bp) dsDNA substrates. Freshly thawed base-editor proteins (2 µM final concentration in a 10 µL reaction volume) were complexed with the indicated sgRNA (2.2 µM final concentration) in Reaction Buffer (20 mM HEPES pH 7.5, 150 mM KCl, 0.5 mM DTT, 0.1 mM EDTA, 10 mM $MgCl2)^2$ for five minutes at room temperature. Annealed dsDNA substrates were then added to a final concentration of 250 nM. The reaction proceeded for 30 min at 37 °C before protein denaturation was performed by heating for 5 min at 99 °C. Addition of PB buffer (Qiagen, 100 µL) and isopropanol (25 µL) ensured protein was dissociated from the substrate DNA. DNA was purified with Minelute columns (Qiagen) and the resulting products amplified to the top of the linear range with 15 cycles of qPCR (12 ng input DNA, 50 µL reaction volume) using a U-tolerant polymerase (Phusion U Hot Start, ThermoFisher). Amplified DNA was purified using RapidTip2 (Diffinity Genomics) and barcoded with a second round of PCR (8 cycles, 5 ng input) before being prepared for sequencing on an Illumina MiSeq as described below.

Purification and sequencing of genomic DNA

Genomic DNA was isolated using Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter) according to the manufacturer's instructions. For the first PCR, DNA was amplified to the top of the linear range using Q5 Hot Start DNA Polymerase (NEB), according to the manufacturer's instructions but with the addition of 3% DMSO and SYBR Gold Nucleic Acid Stain (Thermo Fisher). For all amplicons, the PCR protocol used was an initial heating step of 2 min at 98 °C followed by an optimized number of amplification cycles (12 s at 98 °C, 25 s at 61 °C, 30 s at 72 °C). For zebrafish and for transfected cell samples, 30 ng of input DNA was used in a 50 µL reaction, for cochlear samples 20 ng was used in a 25 µL reaction. qPCR was performed to determine the optimal cycle number for each amplicon. Amplified DNA was purified using RapidTip2 (Diffinity Genomics) and barcoded with a further PCR (8 cycles, 5 ng input). The unique forward and reverse primers used in the first round PCR contained a constant region 5' to the annealing region, (forward: 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN-3', reverse: 5'-TGGAGTTCAGACGTGTGCTCTTCCGATCT-3') which facilitated binding of barcoding primers to amplified DNA for a second-round PCR. The second-round PCR used primers with three regions: a 5' constant region allowing the amplicon to bind to the Illumina flow cell (italicized), an 8-base barcoding region (X), and a 3' constant region allowing the barcoding primer to bind to the first-round PCR amplicon (in bold). Examples of primer sequences are:
forward: 5'-AATGATACGGCGACCACCGAGATCTACACXXXXXXXXACACTCTTTCCCTACACGAC-3'
reverse: 5'-CAAGCAGAAGACGGCATACGAGATXXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTC-3'

Sequencing adapters and dual-barcoding sequences are based on the TruSeq Indexing Adapters (Illumina). Barcoded samples were pooled and purified by gel extraction (Qiagen), and then purified using Ampure beads (Beckman Coulter) before quantification using the Qubit dsDNA HS Kit (Thermo Fisher) and qPCR (KAPA BioSystems) according to the manufacturer's

instructions. Sequencing of pooled samples was performed using a single-end read from 180-250 bases (depending on the amplicon size) on the MiSeq (Illumina) according to the manufacturer's instructions.

All oligonucleotides were obtained from IDT. The optimized number of PCR cycles for each amplicon in this study are as follows: VEGFA site 2 human genomic DNA (annealing temperature was 61 °C for 25 seconds for all extension steps): on-target: 29 cycles, off-target #1: 32 cycles, off-target #2: 28 cycles, off-target #3: 27 cycles, off-target #4: 27 cycles, VEGFA site 2 murine genomic DNA: on-target: 31 cycles, off-targets #1, #2, #3 and #4: 31 cycles. HEK293 site 3: off-targets #1: 29 cycles, off-target #2: 28 cycles, off-target #3: 28 cycles. FANCF off-target #1: 29 cycles, off-target #2: 28 cycles, off-target 3: 28 cycles. EMX1 off-targets #1, #2 and #3: 28 cycles. TYR1, TYR2 and TYR3 sgRNAs for amplification of zebrafish DNA: 32 cycles. Optimized protocols for the on-target amplification of the EMX1, FANCF, and HEK293 site 3 loci were followed as previously described[33].

Analysis and alignment of genomic DNA sequencing reads

Sequencing reads were analyzed as previously described[33]. In brief, sequencing reads were demultiplexed using MiSeq Reporter (Illumina), and individual FASTQ files were analyzed with a previously reported custom Matlab script[33]. Reads were aligned to the reference sequence using the Smith-Waterman algorithm. Base calls with Q-scores below 30 were replaced with a placeholder nucleotide (N). This quality threshold results in nucleotide frequencies with an expected error rate of 1 in 1,000. Indel frequencies were quantified with a previously published custom Matlab script which counts indels which occurring in a 30-base window around the nCas9 cleavage site and are a minimum of 2-base insertions or deletions[33]. Indels were defined as detectable if there was a significant difference (Student's two-tailed t-test, $p < 0.05$) between indel formation in the treated sample and untreated control.

For one of the sequenced amplicons, CFD off-target #3, associated with VEGFA site 2 sgRNA in the murine genome, it was not possible to accurately measure indel formation. The protospacer at this locus is directly preceded by 12 guanine bases, which makes PCR and high-throughput sequencing of this site prone to random insertion or deletions; deletion rates as high as 20% of sequencing reads were observed in multiple independent untreated control samples. Since no significant base editing was detected at this off-target locus under any treatment conditions (Figure 2.8 and Figure 2.9d), we suspect that indel formation is also negligible at this locus.

A phred.II Q30 score corresponds to an estimated 99.9% accuracy in basecalling[177]. A 0.1% probability of incorrect base calling at a given position corresponds to a lower limit for base calling of 0.1/4 = 0.025% if we assume base call errors are randomly distributed across the four bases. C→T editing percentages that fell beneath this threshold were classified as undetectable. Spontaneous deamination39 or polymerase error during PCR can also introduce artefactual C→T edits. In order to distinguish base editor-induced C→T editing from artefactual C→T editing rates, we sequenced untreated control cells for each amplicon and calculated whether the C→T editing under a particular condition was statistically significant using the Student's two-tailed t-test with $p < 0.05$ as the threshold. Off-target sites with statistically significant editing rates >0.025% were considered measurable.

Statistical analyses of genomic DNA sequence alignments

Unless otherwise noted, mean values cited throughout the main text are representative of n≥3 independent biological replicates and the mean ± standard deviation has been stated. The statistical analysis of the high-throughput sequencing data displayed in Figures 2.3 and 2.5 was performed by comparing on- and off-target editing percentages in treated samples to any editing measured in a negative control sample (untreated). The Student's two-tailed t test was used, and individual p-values can be found in Table 2.1. * $p \leq 0.05$, ** $p \leq 0.01$ and *** $p \leq$

0.001. When editing was below the detection limit (0.025%), significance was not calculated; all untreated control samples showed undetectable editing.

For Figure 2.6a, mean on-target base editing was calculated by averaging editing of cytosines in the base editing activity window (C4-C8 for HEK293 site 3 and EMX1, C4-C9 for FANCF and VEGFA site 2).

To account for sgRNA-dependent differences in base editing activity, the a base editing:indel ratio was calculated (Figure 2.7b). This ratio was generated by dividing the percentage of HTS reads with a C→T conversion (averaged across the base editing window for each site) by the percentage of HTS reads containing an indel. As described above, if the off-target editing for a particular locus was below the limit of detection we conservatively assumed the estimated upper bound of our detection method (0.025%) for the purpose of calculating specificity ratios.

Data analysis of in vitro edited DNA

Sequencing reads were automatically demultiplexed using MiSeq Reporter (Illumina.). Quality filtering was performed using the online package usegalaxy.org[178]. Individual bases with an Illumina quality score less than or equal to 30 were converted to the placeholder nucleotide 'N' using FASTQ Groomer followed by FASTA Masker[179]. The resulting quality-filtered FASTQ files were subsequently analyzed with a custom python script provided below. Sequencing reads were scanned for exact matches to two 14-base sequences that flank both sides of the target DNA sequence. If no exact matches were found, the read was excluded from analysis. If both 14-base sequences were located and the length of the sequence between them was equal to the expected protospacer length (20 bases), the protospacer sequence found between the flanking regions was saved and the bases called by high-throughput sequencing at each site within the protospacer were tallied.

Python script used to analyze quality-filtered *in vitro*-edited DNA.

```
from __future__ import print_function
from __future__ import division

import Bio #This will import the BioPython suite
from Bio import SeqIO #Necessary to read/write sequence handles
from Bio.Seq import Seq
import os
import collections
import csv

inputfile = "please_specify_your_input_file_here_containing_filtered_reads" #specify the filenames that contain sequences
filenames = []

for file in os.listdir(inputfile):
    if file.endswith(".fastqsanger"):
        filenames.append(file)

spacer = []
list_of_filenames = []

for file in filenames:
    site = #input site here
    output = open(file + ".txt", "w")
    list_of_filenames.append(file + ".txt") #allows calling of the txt files that come from fastq files later
    for rec in SeqIO.parse(file, "fastq"):
        split1=rec.seq.tostring().split("GTTCGCGGCGATCG") #14-base pair constant_region_before_protospacer
        if len(split1)>=2:
            split2=split1[1].split("TGGATCGCCTGGCA") #14-base pairc constant_region_after_protospacer
            site=split2[0]
            if len(site)==20:
                output.write(site + "\n")

BASES = 'ATGCN'
UNRECOGNIZED = 'X'
BASE_SEPERATOR = dict(zip(BASES, ',,,,\n'))
a_index = 0
t_index = 1
g_index = 2
c_index = 3
n_index = 4

def get_counts_by_column(base, count, library):
    current_count = library[count]
    if base == 'A':
        current_count[a_index] += 1
    elif base == 'T':
        current_count[t_index] += 1
```

```
48.     elif base == 'G':
49.         current_count[g_index] += 1
50.     elif base == 'C':
51.         current_count[c_index] += 1
52.     elif base == 'N':
53.         current_count[n_index] += 1
54.
55. def dna_counts(list_of_sequences, sample):
56.     first_oligo = list_of_sequences[0]
57.     for i in range (len(first_oligo)):
58.         sample.append([0,0,0,0,0])
59.     for j in range(len(first_oligo)):
60.         for i in range(len(list_of_sequences)):
61.             get_counts_by_column(list_of_sequences[i][j], j, libname)
62.
63.
64. for file in list_of_filenames:
65.     spacer_list = open(file).read().splitlines()
66.     output2=[]
67.     dna_counts(spacer_list, output2)
68.     with open(file  + ".csv", "wb") as f:
69.         writer = csv.writer(f)

        writer.writerows(output2)
```

Cell culture

Both HEK293T (ATCC CRL-3216) and NIH/3T3 (ATCC CRL-1658) were maintained in Dulbecco's Modified Eagle's Medium plus GlutaMax (ThermoFisher) supplemented with 10 % (v/v) fetal bovine serum (FBS), at 37 °C with 5 % CO2. Cells were obtained from ATCC and were authenticated and verified to be free of mycoplasma by ATCC upon purchase.

Plasmid transfection of base editors into HEK293T cells

HEK293T cells were seeded on 48-well collagen-coated BioCoat plates (Corning) in antibiotic free medium and transfected at approximately 70 % confluency. Unless otherwise noted, 750 ng of BE and 250 ng of sgRNA expression plasmids were transfected using 1.5 µl of Lipofectamine 2000 (Thermo Fisher) per well according to the manufacturer's protocol.

Protein transfection of base editors into HEK293T cells

HEK293T cells were seeded on 48-well collagen-coated BioCoat plates (Corning) in 250 µL antibiotic free medium and transfected at approximately 70 % confluency. Base editor protein and was incubated with 1.1X molar excess of the necessary sgRNA at room temperature for 5 min. The complex was then incubated with 1.5 µL Lipofectamine 2000 (Thermo Fisher) and transfected according to the manufacturer's protocol for plasmid delivery. Unless otherwise noted, BE protein was added to a final concentration of 200 nM (based on a total well volume of 275 µL).

Plasmid transfection of base editors into NIH/3T3 cells

NIH/3T3 cells were seeded on 48-well collagen-coated BioCoat plates (Corning) in antibiotic-free DMEM medium and transfected at approximately 75 % confluency. Unless otherwise noted, 600 ng of BE and 200 ng of sgRNA expression plasmids were transfected

using 1.4 μL of Lipofectamine 3000 with 1 μL of P3000 reagent (Thermo Fisher) per well according to the manufacturer's protocol.

Protein transfection of base editors into NIH/3T3 cells

NIH/3T3 cells were seeded on 48-well collagen-coated BioCoat plates (Corning) in antibiotic free DMEM medium and transfected at approximately 75 % confluency. Base editor proteins were incubated with 1.1-fold molar excess of the indicated sgRNA at 25 °C for 5 min. The complex was then incubated with 1.4 μL Lipofectamine 3000 (Thermo Fisher) and transfected according to the manufacturer's protocol for plasmid delivery. P3000 reagent was not used because its addition lead to protein precipitation and a reduction in base editing efficiency. Unless otherwise noted, BE protein was added to a final concentration of 400 nM (based on a total well volume of 275 μL).

Intracochlear delivery of BE3 protein:guide RNA encapsulated in cationic lipid

All animal experiments were approved by the Institutional Animal Care and the Use Committee of the Massachusetts Eye and Ear Infirmary. Intracochlear delivery was performed in P1-P2 mice of a mixed genetic background as described previously42. Mice were anesthetized by lowering body temperature before the surgical procedure. A postauricular incision was made near the right ear, and the bulla was lifted to expose the cochlea. BE3 protein (57.7 μM) was pre-complexed with the sgRNA (100 μM) in a 1:1.1 molar ratio and then mixed with Lipofectamine 2000 (Thermo Fisher) in a 1:1 volumetric ratio. The resulting solution (1.2-1.5 μL) was injected with a glass pipette (end diameter, 5 μm) through the cochlear capsule into scala media at the cochear basal turn that attached to a nanoliter micropump (WPI, UMP3 + Micro4 + NanoFil) at the rate of 250 nL min-1. After injection, the incision was closed and the mice were brought onto a heating pad to recover. After 3-4 days, the cochlea of mouse was dissected into the organ of Corti, stria vascularis, and modiolus. Each tissue was further micro-dissected into between 5 and 7 separate pieces and DNA extraction was performed separately for each

sample, followed by high-throughput sequencing as described above. The data presented in Figure 2.8 and Figure 2.10 show sequencing data resulting from extraction of one micro-dissected sample for each cochlear region.

Microinjection of BE3 protein:guide RNA into zebrafish embryo

Zebrafish (Tuebingen strain) were maintained under standard conditions in compliance with internal regulatory review at Boston Children's Hospital. One-cell stage zebrafish embryos were injected with approximately 2 nL of BE3 protein pre-complexed with the appropriate sgRNA or an unrelated sgRNA control in a 1:1 molar ratio (4.5 μM final concentration). Four days post-fertilization, DNA was extracted from larvae as previously described[180] in 50 mM NaOH for 30 minutes at 95 °C and the resulting solution was neutralized with Tris-HCl. Genomic DNA was quantified, amplified by PCR, and sequenced as described above.

Protein gel analyses

All protein gels shown were precast 4-12% polyacrylamide Bis-Tris Plus (Thermo Fisher). They were run in MOPS buffer (Thermo Fisher) at 180 V for 50 min. Samples were prepared for loading by heating to 99 °C in 100mM DTT and 1X lithium dodecyl sulfate (LDS) Sample Buffer for denaturation (Thermo Fisher) for 10 min. Gels were stained using Instant Blue Protein Stain (Expedion) according to manufacturer's instructions.
For cell lysate analysis, 2 mL of post-induction overnight culture was pelleted at 15,000 g before lysis in 100 μL B-PER (Thermo Fisher) according to the manufacturer's instructions.

Data Availability

High-throughput sequencing data that support the findings of this study have been deposited in the NCBI Sequence Read Archive database under Accession Number SRP097884. Plasmids encoding HF-BE3 and BE3 for protein expression, as well as HF-BE3 for

mammalian expression, are available from Addgene with Accession IDs 87439 (pCMV-HF-BE3), 87438 (pET42b-HF-BE3), 87437 (pET42b-BE3).

Chapter Three: Cellular Characterization of an DNA Adenine Base Editor (ABE)

This chapter has been adapted from:

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

Contributions:

Only a small portion of the data in the paper referenced above has been included in the thesis; the experiments and analyses which have been included here are those which H.A.R. was directly involved in generating. N.M.G., A.C.K., and D.R.L greatly aided with development of the methodology and design of the experiments. N.M.G. gave invaluable insight and guidance throughout the course of performing the experiments and the data analysis. N.M.G. lead the development of the A-base editor which enabled the experiments discussed here.

Having mentorship from Nicole Gaudelli whilst working on this project was a truly superb experience for me. Nicole taught me invaluable experimental skills, but, even more importantly, she inspired me to address difficult and seemingly intractable scientific problems. I am immensely grateful to her for her teaching and guidance.

3.1 Introduction

The development of an adenine base editor (ABE) has been greatly enabling for the field genome editing. The development of ABE, including the process by which an A-base editor was evolved for use on DNA, is discussed in Chapter 1 in the section entitled “Development of adenine base editors”.

Here, we have performed an analysis of the off-target cellular effects (in both DNA and RNA) of using a late-stage adenine base editor in cultured human cells. We also performed a comparison between editing using ABE 7.10 (the most efficient ABE construct for the majority of target loci) and homology-directed repair (HDR). Before the development of base editors, double stranded DNA break (DSB)-induced HDR was the most common method for introducing precise changes into the genome of eukaryotic cells[21]. Here, we directly compare the two methods for both their efficiency and product purity in HEK293T cells.

3.2 Results

3.2.1 Comparison between ABE and homology-directed repair (HDR)

We compared the efficiency and editing outcomes that arise from treatment with ABE 7.10 and a recently-developed method for HDR, “CORRECT”[24]. In the “CORRECT” method, the frequency of accurate HDR is improved through incorporation of mutation(s) to the PAM-binding region in addition to the desired point mutation, which is located in the sgRNA protospacer region[24,181]. HDR is stimulated by a Cas9-induced double stranded DNA break (DSB), which is repaired predominantly through processes which incorporate stochastic insertions and deletions (indels), such as non-homologous end joining (NHEJ) or micro-homology mediated end joining (MMEJ)[25].

We directly compared ABE-mediated A-to-G conversion with CORRECT HDR-mediated A-to-G conversion, using the same sgRNAs at five genomic loci in HEK293T cells. We observed average target point mutation frequencies ranging from 0.47% to 4.2% with 3.3% to 10.6% indels using the CORRECT HDR method under optimized 48-h conditions in HEK293T cells (Figure 3.1a). At the same five genomic loci, ABE7.10 resulted in average target mutation

frequencies of 10–35% after 48 h, and 55–68% after 120 h (Figure 3.1a), with fewer than 0.1% indels (Figure 3.1b). The target mutation/indel ratio averaged 0.43 for CORRECT HDR, and more than 500 for ABE7.10, representing an improvement of over 1,000-fold in product selectivity for ABE7.10. Although HDR is well-suited to introduce insertions and deletions into genomic DNA, these results demonstrate that ABE7.10 can introduce A•T to G•C point mutations with much higher efficiency and far fewer undesired products than a current Cas9 nuclease-mediated HDR method.

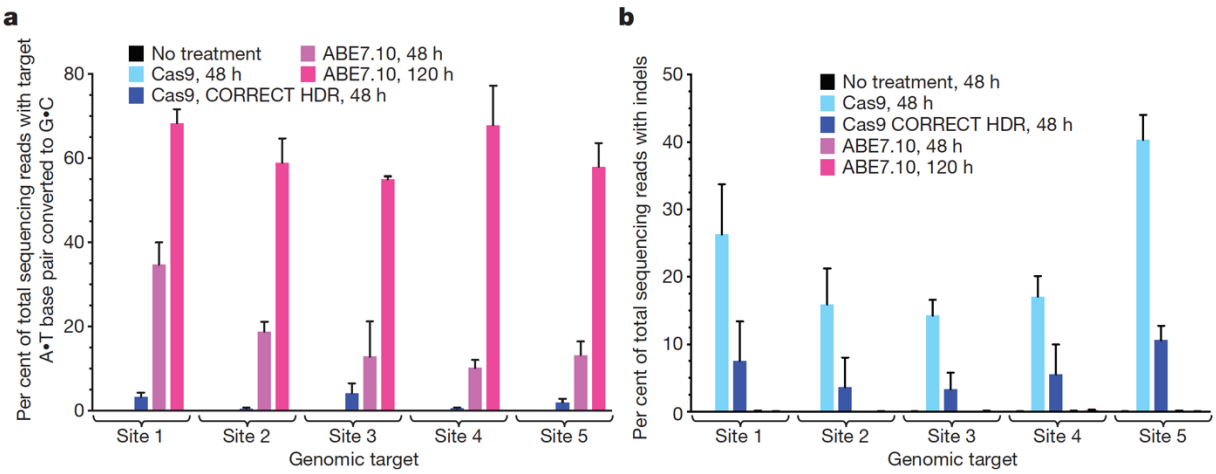


Figure 3.1 Comparison of ABE7.10-mediated base editing and Cas9-mediated HDR. a, A•T to G•C base editing efficiencies in HEK293T cells treated either with ABE7.10 or with Cas9 nuclease and an ssDNA donor template (following the CORRECT HDR method33) targeted to five human genomic DNA sites. **b**, Indel formation in HEK293T cells treated as described in **a**.

3.2.2 Analysis of off-target DNA editing induced by ABE

Next, we examined off-target editing by three ABE7 variants – all of which were obtained from the seventh round of TadA evolution, but which have slightly different editing windows (see Table 1.1). As no method yet exists to comprehensively profile the off-target activity of ABEs in cells, we assumed that off-target ABE editing occurred primarily at the off-target sites that are edited when Cas9 nuclease is complexed with the same guide RNA, as is the case with BE3[33,92,93]. We treated HEK293T cells with three well-characterized guide RNAs[31] and either Cas9 nuclease or ABE7 variants, and sequenced the on-target loci and the 12 most active off-

target human genomic loci associated with these guide RNAs as identified by the genome-wide GUIDE-Seq method[31]. The efficiency of on-target indels by Cas9 and the efficiency of on-target base editing by ABE7.10 both averaged 54% (Tables 3.1-3.3). We observed detectable modification (0.2% indels or more) by Cas9 nuclease at nine of the 12 (75%) known off-target loci (Tables 3.1-3.3). In contrast, when complexed with the same sgRNAs, ABE7.10, ABE7.9, or ABE 7.8 produced 0.2% or more off-target base editing at only four of the twelve (33%) known Cas9 off-target sites.

Moreover, the nine confirmed Cas9 off-target loci were modified with an average efficiency of 14% indels, whereas the four confirmed ABE off-target loci were modified with an average of only 1.3% A•T to G•C mutations (Tables 3.1-3.3). Although seven of the nine confirmed Cas9 off-target loci contained at least one adenine within the ABE activity window, three of these seven off-target loci were not detectably edited by ABE7.8, 7.9, or 7.10. Together, these data strongly suggest that ABE7 variants may be less prone to off-target genome modification than Cas9 nuclease, although a comprehensive, unbiased method of profiling the DNA specificity of ABEs is needed.

	% of total sequencing reads with target A•T base pair converted to G•C																				
HEK2 (on-target site)	G1	A2	A3	C4	A5	C6	A7	A8	A9	G10	C11	A12	T13	A14	G15	A16	C17	T18	G19	C20	Indel%
ABE 7.8		0.2	1.8		77.2		2.7	1.0	0.8			0.1		0.0		0.0					0.2
ABE 7.9		0.1	0.5		79.4		2.5	0.5	0.5			0.0		0.0		0.0					0.1
ABE 7.10		0.0	0.5		87.6		23.0	1.0	1.0			0.1		0.0		0.0					0.3
Cas9 nuclease		0.0	0.7		0.0		0.0	0.0	0.0			0.3		0.1		0.1					54.5
D10A Cas9 nickase		0.0	0.0		0.0		0.0	0.0	0.0			0.0		0.0		0.0					0.2
H840A Cas9 nickase		0.0	0.1		0.0		0.0	0.0	0.0			0.2		0.0		0.0					5.3
dCas9 (D10A + H840A)		0.0	0.0		0.0		0.0	0.0	0.0			0.0		0.0		0.0					0.0
no treatment		0.0	0.0		0.0		0.0	0.0	0.0			0.0		0.0		0.0					0.0

	% of total sequencing reads with target A•T base pair converted to G•C																				
HEK2 off-target site 1	G1	A2	A3	C4	A5	C6	A7	A8	T9	G10	C11	A12	T13	A14	G15	A16	T17	T18	G19	C20	Indel%
ABE 7.8		0.0	0.0		0.0		0.0	0.0				0.0		0.0		0.0					0.0
ABE 7.9		0.0	0.0		0.0		0.0	0.0				0.0		0.0		0.0					0.0
ABE 7.10		0.0	0.0		0.3		0.1	0.0				0.0		0.0		0.0					0.0
Cas9 nuclease		0.0	0.0		0.0		0.0	0.0				0.0		0.0		0.0					0.5
D10A Cas9 nickase		0.0	0.0		0.0		0.0	0.0				0.0		0.0		0.0					0.0
H840A Cas9 nickase		0.0	0.0		0.0		0.0	0.0				0.0		0.0		0.0					0.0
dCas9 (D10A + H840A)		0.0	0.0		0.0		0.0	0.0				0.0		0.0		0.0					0.0
no treatment		0.0	0.0		0.0		0.0	0.0				0.0		0.0		0.0					0.0

	% of total sequencing reads with target A•T base pair converted to G•C																				
HEK2 off-target site 2	A1	A2	A3	C4	A5	T6	A7	A8	A9	G10	C11	A12	T13	A14	G15	A16	C17	T18	G19	C20	Indel%
ABE 7.8	0.0	0.0	0.0		0.0		0.1	0.0	0.0			0.0		0.0		0.0					0.0
ABE 7.9	0.0	0.0	0.0		0.0		0.0	0.0	0.0			0.0		0.0		0.0					0.0
ABE 7.10	0.0	0.0	0.0		0.0		0.0	0.0	0.0			0.0		0.0		0.0					0.0
Cas9 nuclease	0.0	0.0	0.0		0.0		0.0	0.0	0.0			0.0		0.0		0.0					0.0
D10A Cas9 nickase	0.0	0.0	0.0		0.0		0.0	0.0	0.0			0.0		0.0		0.0					0.0
H840A Cas9 nickase	0.0	0.0	0.0		0.0		0.0	0.0	0.0			0.0		0.0		0.0					0.0
dCas9 (D10A + H840A)	0.0	0.0	0.0		0.0		0.0	0.0	0.0			0.0		0.0		0.0					0.0
no treatment	0.0	0.0	0.0		0.0		0.1	0.0	0.0			0.0		0.0		0.0					0.0

Table 3.1. Activities of ABE7.8, ABE7.9, and ABE7.10 at the HEK2 on-target and off-target sites previously characterized for S. pyogenes Cas9 nuclease.

	% of total sequencing reads with target A•T base pair converted to G•C																			
HEK3 (on-target site)	G1	G2	C3	C4	C5	A6	G7	A8	C9	T10	G11	A12	G13	C14	A15	C16	G17	T18	G19	A20	Indel%
ABE 7.8						8.4		14.7				0.1			0.0					0.0	0.1
ABE 7.9						12.2		9.4				0.2			0.0					0.0	0.2
ABE 7.10						62.7		18.6				0.2			0.0					0.0	0.2
Cas9 nuclease						0.0		0.0				0.0			0.2					0.2	64.8
D10A Cas9 nickase						0.0		0.0				0.0			0.0					0.0	2.0
H840A Cas9 nickase						0.0		0.0				0.0			0.0					0.0	0.4
dCas9 (D10A + H840A)						0.0		0.0				0.0			0.0					0.0	0.0
no treatment						0.0		0.0				0.0			0.0					0.0	0.0

	% of total sequencing reads with target A•T base pair converted to G•C																			
HEK3 off-target site 1	C1	A2	C3	C4	C5	A6	G7	A8	C9	T10	G11	A12	G13	C14	A15	C16	G17	T18	G19	C20	Indel%
ABE 7.8		0.0				0.0		0.0				0.0			0.0						0.0
ABE 7.9		0.0				0.0		0.0				0.0			0.0						0.0
ABE 7.10		0.0				0.0		0.0				0.0			0.0						0.0
Cas9 nuclease		0.0				0.0		0.0				0.0			0.0						1.0
D10A Cas9 nickase		0.0				0.0		0.0				0.0			0.0						0.0
H840A Cas9 nickase		0.0				0.0		0.0				0.0			0.0						0.0
dCas9 (D10A + H840A)		0.0				0.0		0.0				0.0			0.0						0.0
no treatment		0.0				0.0		0.0				0.0			0.0						0.0

	% of total sequencing reads with target A•T base pair converted to G•C																			
HEK3 off-target site 2	G1	A2	C3	A4	C5	A6	G7	A8	C9	C10	G11	G12	G13	C14	A15	C16	G17	T18	G19	A20	Indel%
ABE 7.8		0.0		0.0		0.0		0.1							0.0					0.0	0.0
ABE 7.9		0.0		0.0		0.0		0.0							0.0					0.0	0.0
ABE 7.10		0.0		0.0		0.0		0.0							0.0					0.0	0.0
Cas9 nuclease		0.0		0.0		0.0		0.0							0.0					0.0	1.6
D10A Cas9 nickase		0.0		0.0		0.0		0.0							0.0					0.0	0.0
H840A Cas9 nickase		0.0		0.0		0.0		0.0							0.0					0.0	0.0
dCas9 (D10A + H840A)		0.0		0.0		0.0		0.0							0.0					0.0	0.0
no treatment		0.0		0.0		0.0		0.0							0.0					0.0	0.0

	% of total sequencing reads with target A•T base pair converted to G•C																			
HEK3 off-target site 3	A1	G2	C3	T4	C5	A6	G7	A8	C9	T10	G11	A12	G13	C14	A15	A16	G17	T18	G19	A20	Indel%
ABE 7.8	0.0					0.0		0.0				0.0			0.0	0.0				0.0	0.0
ABE 7.9	0.0					0.0		0.0				0.0			0.0	0.0				0.0	0.0
ABE 7.10	0.0					0.0		0.0				0.0			0.0	0.0				0.0	0.0
Cas9 nuclease	0.0					0.0		0.0				0.0			0.0	0.0				0.0	0.2
D10A Cas9 nickase	0.0					0.0		0.0				0.0			0.0	0.0				0.0	0.0
H840A Cas9 nickase	0.0					0.0		0.0				0.0			0.0	0.0				0.0	0.0
dCas9 (D10A + H840A)	0.0					0.0		0.0				0.0			0.0	0.0				0.0	0.0
no treatment	0.0					0.0		0.0				0.0			0.0	0.0				0.0	0.0

	% of total sequencing reads with target A•T base pair converted to G•C																			
HEK3 off-target site 4	A1	G2	A3	C4	C5	A6	G7	A8	C9	T10	G11	A12	G13	C14	A15	A16	G17	A18	G19	A20	Indel%
ABE 7.8	0.0		0.0			0.0		0.0				0.0			0.0	0.0		0.0		0.0	0.0
ABE 7.9	0.0		0.0			0.0		0.0				0.0			0.0	0.0		0.0		0.0	0.0
ABE 7.10	0.0		0.0			0.0		0.0				0.0			0.0	0.0		0.0		0.0	0.0
Cas9 nuclease	0.0		0.0			0.0		0.0				0.0			0.0	0.0		0.0		0.0	0.0
D10A Cas9 nickase	0.0		0.0			0.0		0.0				0.0			0.0	0.0		0.0		0.0	0.0
H840A Cas9 nickase	0.0		0.0			0.0		0.0				0.0			0.0	0.0		0.0		0.0	0.0
dCas9 (D10A + H840A)	0.0		0.0			0.0		0.0				0.0			0.0	0.0		0.0		0.0	0.0
no treatment	0.0		0.0			0.0		0.0				0.0			0.0	0.0		0.0		0.0	0.0

	% of total sequencing reads with target A•T base pair converted to G•C																				
HEK3 off-target site 5	G1	A2	G3	C4	C5	A6	G7	A8	A9	T10	G11	A12	G13	C14	A15	C16	G17	T18	G19	A20	Indel%
ABE 7.8		0.0				0.0		0.0	0.0			0.0			0.0					0.0	0.1
ABE 7.9		0.0				0.0		0.0	0.0			0.0			0.0					0.0	0.1
ABE 7.10		0.0				0.0		0.0	0.0			0.0			0.0					0.0	0.0
Cas9 nuclease		0.0				0.0		0.0	0.0			0.0			0.0					0.0	0.1
D10A Cas9 nickase		0.0				0.0		0.0	0.0			0.0			0.0					0.0	0.0
H840A Cas9 nickase		0.0				0.0		0.0	0.0			0.0			0.0					0.0	0.1
dCas9 (D10A + H840A)		0.0				0.0		0.0	0.0			0.0			0.0					0.0	0.1
no treatment		0.0				0.0		0.0	0.0			0.0			0.0					0.0	0.0

Table 3.2. Activities of ABE7.8, ABE7.9, and ABE7.10 at the HEK3 site previously characterized for on-target and off-target modification by S. pyogenes Cas9 nuclease[31].

	% of total sequencing reads with target A•T base pair converted to G•C																				
HEK4 (on-target site)	G1	G2	C3	A4	C5	T6	G7	C8	G9	G10	C11	T12	G13	G14	A15	G16	G17	T18	G19	G20	Indel%
ABE 7.8				4.8											0.0						0.2
ABE 7.9				1.5											0.0						0.1
ABE 7.10				16.0											0.0						0.2
Cas9 nuclease				0.1											0.0						36.5
D10A Cas9 nickase				0.0											0.0						0.8
H840A Cas9 nickase				0.0											0.0						0.7
dCas9 (D10A + H840A)				0.0											0.0						0.0
no treatment				0.0											0.0						0.0

	% of total sequencing reads with target A•T base pair converted to G•C																				
HEK4 off-target site 1	T1	G2	C3	A4	C5	T6	G7	C8	G9	G10	C11	C12	G13	G14	A15	G16	G17	A18	G19	G20	Indel%
ABE 7.8				0.6											0.0			0.0			0.1
ABE 7.9				0.2											0.0			0.0			0.0
ABE 7.10				1.2											0.0			0.0			0.0
Cas9 nuclease				0.0											0.0			0.2			12.3
D10A Cas9 nickase				0.0											0.0			0.0			0.0
H840A Cas9 nickase				0.0											0.0			0.0			0.0
dCas9 (D10A + H840A)				0.0											0.0			0.0			0.0
no treatment				0.0											0.0			0.0			0.0

	% of total sequencing reads with target A•T base pair converted to G•C																				
HEK4 off-target site 2	G1	G2	C3	T4	C5	T6	G7	C8	G9	G10	C11	T12	G13	G14	A15	G16	G17	G18	G19	G20	Indel%
ABE 7.8															0.0						0.0
ABE 7.9															0.0						0.0
ABE 7.10															0.0						0.0
Cas9 nuclease															0.0						4.8
D10A Cas9 nickase															0.0						0.0
H840A Cas9 nickase															0.0						0.0
dCas9 (D10A + H840A)															0.0						0.0
no treatment															0.0						0.0

	% of total sequencing reads with target A•T base pair converted to G•C																				
HEK4 off-target site 3	G1	G2	C3	A4	C5	G6	A7	C8	G9	G10	C11	T12	G13	G14	A15	G16	G17	T18	G19	G20	Indel%
ABE 7.8				0.5			16.8								0.0						1.1
ABE 7.9				0.2			21.9								0.0						1.7
ABE 7.10				1.4			7.8								0.0						3.7
Cas9 nuclease				0.0			0.3								0.1						89.1
D10A Cas9 nickase				0.0			0.0								0.0						0.6
H840A Cas9 nickase				0.0			0.0								0.0						1.4
dCas9 (D10A + H840A)				0.0			0.0								0.0						0.0
no treatment				0.0			0.0								0.0						0.0

	% of total sequencing reads with target A•T base pair converted to G•C																				
HEK4 off-target site 4	G1	G2	C3	A4	T5	C6	A7	C8	G9	G10	C11	T12	G13	G14	A15	G16	G17	T18	G19	G20	Indel%
ABE 7.8				0.3			0.5								0.0						0.0
ABE 7.9				0.0			0.2								0.0						0.0
ABE 7.10				0.4			1.8								0.0						0.0
Cas9 nuclease				0.0			0.0								0.0						2.5
D10A Cas9 nickase				0.0			0.0								0.0						0.0
H840A Cas9 nickase				0.0			0.0								0.0						0.0
dCas9 (D10A + H840A)				0.0			0.0								0.0						0.0
no treatment				0.0			0.0								0.0						0.0

	% of total sequencing reads with target A•T base pair converted to G•C																				
HEK4 off-target site 5	G1	G2	C3	G4	C5	T6	G7	C8	G9	G10	C11	G12	G13	G14	A15	G16	G17	T18	G19	G20	Indel%
ABE 7.8															0.0						0.0
ABE 7.9															0.0						0.0
ABE 7.10															0.0						0.0
Cas9 nuclease															0.0						9.8
D10A Cas9 nickase															0.0						0.0
H840A Cas9 nickase															0.0						0.0
dCas9 (D10A + H840A)															0.0						0.0
no treatment															0.0						0.0

Table 3.3. Activities of ABE7.8, ABE7.9, and ABE7.10 at the HEK4 site previously characterized for on-target and off-target modification by S. pyogenes Cas9 nuclease. Although HEK4 off-target site 3 showed appreciable indel formation upon ABE treatment, this locus also showed unusually high (89%) indel formation by Cas9 nuclease and was the only tested off-target site exhibiting indel formation upon treatment with Cas9 nickases. We speculate that this locus is unusually fragile, and that indel formation here arises from simply nicking the site, rather than from ABE-mediated adenine deamination.

3.2.3 Analysis of direct RNA editing by ABE7.10

Wild type TadA (wtTadA) is a tRNA adenosine deaminase which acts on the single-stranded anticodon loop of $tRNA^{Arg}$ [182]. wtTadA acts as a homodimer in its native context in *E.Coli*[59], with one monomer performing catalysis (hydrolytic deamination of an adenine base located in the anticodon loop of $tRNA^{Arg}$), and one monomer performing a structural role (facilitating tRNA binding to TadA)[59]. ABE7.10 contains one wild type TadA monomer, which presumably plays a structural role in the deamination of ssDNA and one TadA* evolved monomer, which we anticipate is the monomer where ssDNA catalysis occurs.

Nevertheless, the inclusion of the wild type TadA monomer in ABE7.10 raises the liklihood that this monomer could act in a catalytic fashion on the RNA of the mammalian cell in which the ABE7.10 resides. In such an event, the evolved TadA* monomer may also remain capable of ssRNA deamination, or, if TadA* has lost this ability, TadA* may remain capable of assuming the structural role.

To assess the extent of ssRNA editing by ABE7.10, we performed plasmid expression of ABE7.10 in HEK293T cells for 5 days, isolated total mRNA and reverse transcribed this to cDNA. We performed deep sequencing of 217-265bp regions within selected amplicons using the Illumina MiSeq HTS sequencing system. We compared the sequencing reads from cells where ABE had been expressed to cells where there had been no treatment and cells where Cas9 had been expressed to enable a distinction between endogenous A-to-I editing and ABE7.10-induced A-to-I editing. Amplicons were chosen either because they are universally abundant human mRNAs (beta catenin, beta actin, GAPDH, RB1 and IP90) and one because it contained a region with homology to the native substrate of TadA (5'-GCUCGGCUACGGAACCG-3') (RSLD1).

Analysis of data from the cDNA sequencing was not trivial. We used our sequencing read depth (>10,000 reads per sample) and a per-base quality score filter of Q30 to set a lower limit of 0.2% A-to-G conversion as the requirement for a particular A in an amplicon to be

classified as “edited” (an A-to-I edit in mRNA would be called as a G after cDNA synthesis and PCR as inosine is recognized as guanine in the context of a polymerase active site[57]). Each condition (Cas9 treatment, ABE7.10 treatment and no treatment) was performed in duplicate, yielding 6 samples in total. We were interested in two measurements: first, whether the number of adenine residues which had detectable A-to-I editing would be increased upon ABE7.10 treatment, and second whether the extent of editing at adenines which are edited to inosines in human cells may be increased by ABE7.10 expression. To this end, for each adenine that occurs in the sequenced region of each amplicon, we tabulated whether there was any evidence of A-to-G conversion by HTS. If there was experimental condition in which A-to-G conversion exceeded the detection threshold of 0.2%, we tabulated editing at this residue under all experimental conditions (Figure 3.2) If there was no evidence of editing at a particular adenine residue, it does not appear in the analysis.

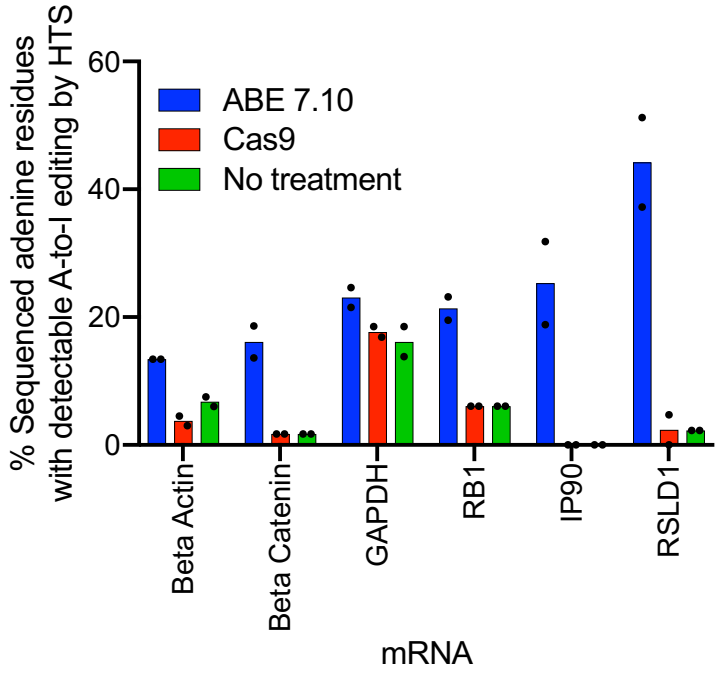


Figure 3.2. Analysis of A-to-I editing by deep sequencing of targeted regions of mRNA. After cDNA was generated from total cellular mRNA, regions from the six indicated mRNA sequences were amplified by qPCR and sequenced using HTS. The percentage of adenines with detectable (>0.2%) conversion to inosine (read as guanine after cDNA conversion and PCR) in each amplified region is shown. n=2 biological replicates for each condition.

Consistent with the hypothesis that the wild type TadA retains its capability to edit RNA when expressed as a component of ABE7.10 inside mammalian cells, we observed a greater number of edited adenines when ABE7.10 treatment was compared to either Cas9 or no treatment in all six of the mRNAs sequenced (Figure 3.2). However, the levels of ABE7.10-specific A-to-I editing were extremely low (generally below 2%) in all cases (Table 3.4), except from one of the adenines which resides in the RSLD1 mRNA region (position 152) which has high homology with the native TadA substrate, where the two ABE7.10 treatment replicates showed 3.1% and 6.4% editing and the Cas9 and no treatment controls showed a maxiumum of 0.01%.

Although we cannot rule out the possibility that this RNA editing arises from ABE7.10 editing the DNA which is being used for transcription of these RNAs, this seems highly unlikely given that no evidence of A-to-G editing outside the sgRNA binding site has been observed through DNA sequencing of targeted regions, including actively transcribed loci [35,183]. Further work must be done to assess the transcriptome-wide extent of ABE7.10-mediated mRNA editing, as well as the culprit; it is unclear whether the evolved TadA retains the capability of RNA editing or whether this activity is due to the wtTadA. The transient nature of RNA, coupled with the low editing levels observed in Table 3.4 indicate that RNA editing by ABE7.10 is unlikely to induce substantial biological off-target effects, but it should be fully characterized before a final conclusion as to its significance can be reached.

Beta-Actin

Amplicon sequence

GACAAAACCTAACTTGCGCAGAAAACAAGATGAGATTGGCATGGCTTTATTTGTTTTTTTTGTTTTGTTTTGG
TTTTTTTTTTTTTTTTGGCTTGACTCAGGATTTAAAAACTGGAACGGTGAAGGTGACAGCAGTCGGTTGGAG
CGAGCATCCCCCAAAGTTCACAATGTGGCCGAGGACTTTGATTGCACATTGTTGTTTTTTTAATAGTCATTCC
AAATATGAGATGCGTTGTTACAGGAAGTCCCTTGCCATCCTAAAAGC

Position of A in amplicon:	33	103	107	116	144	148	177	180	186	210	238
Treatment	% of A to G editing at indicated residue										
ABE-1	1.40	0.34	0.29	0.17	0.24	0.34	0.30	0.85	0.27	0.31	0.04
ABE-2	2.48	0.72	0.37	0.20	0.15	0.58	0.28	1.17	0.37	0.49	0.33
untreated-1	1.05	0.22	0.00	0.23	0.04	0.13	0.03	0.04	0.07	0.05	0.00
untreated-2	3.18	0.10	0.00	0.18	0.05	0.17	0.06	0.27	0.00	0.07	0.00
Cas9-1	1.83	0.27	0.00	0.23	0.05	0.31	0.11	0.32	0.08	0.04	0.01
Cas9-2	9.44	0.23	0.00	0.32	0.04	0.23	0.07	0.09	0.00	0.06	0.01

Beta-Catenin

Amplicon sequence

TTTGATGGAGTTGGACATGGCCATGGAACCAGACAGAAAAGCGGCTGTTAGTCACTGGCAGCAACAGTCT
TACCTGGACTCTGGAATCCATTCTGGTGCCACTACCACAGCTCCTTCTCTGAGTGGTAAAGGCAATCCTGAG
GAAGAGGATGTGGATACCTCCCAAGTCCTGTATGAGTGGGAACAGGGATTTTCTCAGTCCTTCACTCAAGA
ACAAGTAGCTGG

Position of A in amplicon:	28	50	54	72	128	130	144	156	158	174	177
Treatment	% of A to G editing at indicated residue										
ABE-1	0.74	0.16	0.32	0.31	0.49	0.18	0.31	0.14	0.46	0.84	0.21
ABE-2	0.85	0.26	0.50	0.77	0.68	0.23	0.37	0.30	0.74	1.50	0.31
untreated-1	1.14	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.01
untreated-2	2.43	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00
Cas9-1	2.43	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.02	0.01
Cas9-2	2.74	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

GAPDH

Amplicon sequence

GGCTACAGCAACAGGGTGGTGGACCTCATGGCCCACATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTG
GGGAGTCCCTGCCACACTCAGTCCCCCACCACACTGAATCTCCCCTCCTCACAGTTGCCATGTAGACCCCTTGAAGAGGGGAGGGGCCTAGGGAGCCGCACCTTGTCATGTACCAT
CAATAAAGTACCCTGTGCTCAACC

Position of A in amplicon:	37	53	67	79	85	87	88	93	98	102	128	174	178	180	204	208	222	226
Treatment	% of A to G editing at indicated residue																	
ABE-1	0.22	0.32	0.29	0.97	0.14	0.45	0.18	0.47	0.31	0.55	0.21	0.05	4.51	0.25	0.60	0.03	0.21	0.69
ABE-2	0.37	0.43	0.47	1.37	0.26	0.74	0.21	0.97	0.45	0.81	0.35	0.06	6.93	0.30	1.03	0.02	0.57	0.84
untreated-1	0.14	0.09	0.50	1.25	0.15	0.89	0.39	0.87	0.04	0.61	0.32	0.38	4.17	0.26	0.13	2.35	0.18	0.27
untreated-2	0.14	0.10	0.51	1.27	0.13	0.89	0.47	0.79	0.03	0.74	0.32	0.40	6.27	0.26	0.07	2.59	0.16	0.08
Cas9-1	0.13	0.06	0.41	1.15	0.03	0.65	0.12	0.84	0.01	0.55	0.23	0.00	4.77	0.21	0.05	0.08	0.15	0.28
Cas9-2	0.20	0.10	0.66	1.88	0.02	0.98	0.24	1.67	0.04	1.10	0.46	0.01	8.42	0.43	0.06	0.01	0.24	0.26

RB1

Amplicon sequence

GGAAGGATTATGATAGGGACAAGGATGATAAAATTTCCTGGGAAGAATACAAACAAGCCACCTATGGTTACTACCTAGGAAACCCCGCAGAGTTTCATGATTCTTCAGATCATCACACCTTTAA
AAAGATGCTGCCACGTGATGAGAGAAGATTCAAAGCTGCAGACCTCAATGGTGACCTGACAGCTACTCGGGAGGAGTTCACTGCCTTTCTGCATCCTGAAGAGTTTGAACATATGAAGGAAAT
TGTGG

Position of A in amplicon:	25	28	30	47	49	51	53	55	64	89	123	124	125	126	127	129	137	171	172
Treatment	% of A to G editing at indicated residue																		
ABE-1	0.48	0.14	0.52	0.25	0.72	0.33	0.19	0.12	0.70	0.33	1.16	0.24	0.43	0.22	0.36	0.24	0.71	0.92	0.23
ABE-2	0.73	0.28	0.84	0.54	1.32	0.81	0.33	0.39	1.42	0.32	2.04	0.36	0.61	0.42	0.53	0.40	0.76	0.89	0.30
untreated-1	0.01	0.01	0.02	1.04	0.01	1.02	0.01	0.02	0.01	0.29	0.04	0.03	0.02	0.01	0.02	0.01	0.85	0.98	0.03
untreated-2	0.01	0.01	0.02	1.95	0.02	1.91	0.01	0.01	0.01	0.29	0.01	0.02	0.01	0.01	0.03	0.00	0.75	0.90	0.00
Cas9-1	0.02	0.01	0.02	0.84	0.02	0.82	0.00	0.01	0.01	0.32	0.04	0.04	0.01	0.03	0.05	0.00	0.97	1.08	0.01
Cas9-2	0.01	0.02	0.02	5.07	0.01	5.00	0.01	0.00	0.01	0.32	0.01	0.01	0.01	0.02	0.02	0.01	0.89	1.24	0.01

IP90

Amplicon sequence

CTGGTTGACCAATCTGTGGTGAATAGTGGAAATCTGCTCAATGACATGACTCCTCCTGTAAATCCTTCACGTGAAATTGAGGACCCAGAAGACCGGAAGCCCGAGGATTGGGATGAAAGACCAAAAATCCCAGATCCAGAAGCTGTCAAGCCAGATGACTGGGATGAAGATGCCC
CTGCTAAGATTCCAGATGAAGAGGCCACAAAACCCGAAGGCTGGTTAGATGATGAGCCTGAGTACGTACCTGATCCAGACGCAG

Position of A in amplicon:	25	30	44	46	60	61	62	74	75	90	92	107	117	118	120	140	148	158	168	170	181	182	184	213	222	224	239
Treatment	% of A to G editing at indicated residue																										
ABE-1	0.58	0.20	0.21	0.10	1.44	0.13	0.38	0.20	0.16	0.18	0.18	0.15	0.20	0.42	0.60	0.19	0.22	0.29	0.55	0.14	0.28	0.16	0.21	1.13	1.08	0.34	0.50
ABE-2	1.01	0.27	0.36	0.25	2.53	0.25	0.69	0.33	0.26	0.25	0.33	0.21	0.27	0.70	1.02	0.20	0.36	0.50	0.90	0.28	0.43	0.26	0.37	2.05	1.84	0.48	0.84
untreated-1	0.01	0.02	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.02	0.01	0.01
untreated-2	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.03	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.06	0.04	0.01	0.01
Cas9-1	0.01	0.02	0.01	0.01	0.03	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.00	0.01	0.00	0.02	0.00	0.01	0.04	0.04	0.01	0.01
Cas9-2	0.01	0.02	0.00	0.01	0.02	0.01	0.01	0.02	0.01	0.01	0.01	0.00	0.01	0.02	0.01	0.00	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.10	0.07	0.01	0.01

RSLD1

Amplicon sequence

TTGGCTTTCCAAATCAGTGGGTCTGACTTGAGGTCTGTGATGTGACCCTTTTCCTCACCTGCTCAACCATTATTCACATGGACTCCATCATATTCATTTGTAGTCATTCCCAGAGTGGCCCAGTGAGGGTCTCGCTGTATGAGAGTCGGCTACGGAATT
TAGGAGAAACAGAAGTTTCTTGGCTTTCATGCTGAGCTTGTTGGTCTAAGCTTATGAG

Position of A in amplicon:	26	31	40	72	76	87	90	92	96	102	139	152	156	157	161	164	167	168	170	172	173	188	191	196
Treatment	% of A to G editing at indicated residue																							
ABE-1	0.32	0.60	0.27	0.27	0.15	0.19	0.42	1.98	0.41	0.32	0.21	3.11	0.76	0.46	0.64	0.08	0.09	0.01	0.01	0.29	0.25	0.22	0.00	0.00
ABE-2	0.35	0.88	1.11	0.70	0.23	0.76	1.00	3.81	0.83	0.50	0.72	6.42	1.84	1.38	1.17	0.49	0.29	0.25	0.32	0.52	0.64	0.34	0.00	0.00
untreated-1	0.40	0.03	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.00
untreated-2	0.28	0.02	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00
Cas9-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.15
Cas9-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.26	0.26

Table 3.4 A-to-I editing in six mRNA amplicons in HEK293T cells. As described in the Methods and the text, HEK293T cells treated with either ABE7.10, Cas9 or untreated cells from the same batch were subjected to RNA extraction, cDNA synthesis and targeted deep sequencing of 6 mRNA amplicon regions. When A-to-I (read as A-to-G after cDNA synthesis) editing was above the detection threshold of 0.2% this is highlighted in red in the table. Two replicates for each condition are shown, indicated as *treatment* -1 and -2 in the table. For the amplicon RSLD1, the adenine residues which fall within the region of homology to the TadA native site are highlighted in yellow.

3.3 Conclusions

This initial characterization of ABE7.10 DNA specificity has demonstrated that ABE7.10 is surprisingly specific in the context of Cas9-depenendet off target editing in DNA, particularly when compared to previously characterized CBEs (see Chapter 2). Subsequent to the completion of this work, other research groups have confirmed through both whole-genome sequencing of ABE7.10-treated mouse embryos[99,184] and through an in vitro assay using genomic DNA[183] that ABE7.10 is unusually specific for a Cas9-directed genome editing agent.

We have demonstrated in a preliminary set of experiments, that there is the possibility that ABE7.10 retains the ability to deaminate adenine bases located in RNA. The scope and biological implications of this activity are yet to be characterized. We anticipate that it will be possible, through either evolution or rational design, to generate ABE7.10 variants which are impaired in their RNA-editing capability. This may be particularly important for the development of therapeutic base editors, because endogenous A-to-I editing in RNA has been linked to carcinogenesis[185].

Finally, we have shown that ABE7.10 is dramatically more efficient and clean in its ability to install point mutations than the most commonly used alternative approach – CORRECT HDR[24,181]. The fact that CORRECT HDR necessitates generation of a DSB, while ABE7.10 can proceed with only a DNA nick means that the formation indels is negligible with ABE7.10-mediated base editing, but is the major product of CORRECT HDR[24,181]. Moreover, CORRECT HDR relies on endogenous cellular repair, limiting its utility to actively dividing cells and limiting the flux though the HDR pathway by the cellular abundance of HDR regulatory proteins.

In spite of the efficiency and product purity that can be achieved with both ABEs and CBEs, HDR remains the only method to generate transversion edits in genomic DNA. For this reason, the following chapter comprises our attempts to develop a Cas9-based construct capable of mediating HDR that is more efficient and offers improved product purity as compared with Cas9.

3.4 Methods

RNA isolation from HEK293T cells and analysis.

HEK293T cells were plated and a subset was transfected with ABE 7.10 as described above and incubated for five days before being removed from the plate using TrypLE Express (Thermo Fisher Scientific) and pelleted. RNA was extracted using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. cDNA was generated from the isolated RNA using the ProtoScript II First Strand cDNA Synthesis Kit (New England Biolabs) according to the manufacturer's instructions with a mixture of random primers and Oligo-dT primers. Amplification of the cDNA for high-throughput sequencing was performed to the top of the linear range (29 cycles for all amplicons) using qPCR as described above. High-throughput sequencing of the amplicons was performed as described above. Sequences were aligned to the reference sequence for each RNA, obtained from the NCBI.

Comparison between ABE 7.10 and HDR using the CORRECT method.

HEK293T cells grown in the absence of antibiotic were seeded on 48-well poly-D-lysine coated plates (Corning). After 12–14 h, cells were transfected at ~ 70% confluency with 750 ng Cas9 or base editor plasmid, 250 ng sgRNA expression plasmid, 1.5 μ l Lipofectamine 3000 (Thermo Fisher Scientific), and for HDR assays 0.7 μ g single-stranded donor DNA template (100 nt, PAGE-purified from IDT) according to the manufacturer's instructions.

Genomic DNA was harvested 48 h after transfection (as described[24]) using the Agencourt DNAdvance Genomic DNA isolation Kit (Beckman Coulter) according to the manufacturer's instructions. A size-selective DNA isolation step ensured that there was no risk of contamination by the single-stranded donor DNA template in subsequent PCR amplification and sequencing steps. We re-designed amplification primers to ensure that there was minimal risk of amplifying the donor oligo template. HTS of genomic DNA samples. Genomic sites of interest were amplified by PCR with primers containing homology to the region of interest and the appropriate Illumina forward and reverse adapters. Specifically, 25 µ l of a given PCR 1 reaction was assembled containing 0.5 µ M of each forward and reverse primer, 1 µ l genomic DNA extract and 12.5 µ l Phusion U Green Multiplex PCR Master Mix. PCR reactions were carried out as follows: 95 °C for 2 min, then 30 cycles of (95 °C for 15 s, 62 °C for 20 s, and 72 °C for 20 s), followed by a final 72 °C extension for 2 min. PCR products were verified by comparison with DNA standards (Quick-Load 100 bp DNA ladder) on a 2% agarose gel supplemented with ethidium bromide. Unique Illumina barcoding primer pairs were added to each sample in a secondary PCR reaction (PCR 2). Specifically, 25 µ l of a given PCR 2 reaction was assembled containing 0.5 µ M of each unique forward and reverse illumina barcoding primer pair, 2 µl unpurified PCR 1 reaction mixture, and 12.5 µ l Q5 Hot Start High-Fidelity 2× Master Mix. The barcoding PCR 2 reactions were carried out as follows: 95 °C for 2 min, then 15 cycles of (95 °C for 15 s, 61 °C for 20 s, and 72 °C for 20 s), followed by a final 72 °C extension for 2 min. PCR products were purified by electrophoresis with a 2% agarose gel using a QIAquick Gel Extraction Kit, eluting with 30 µ l H2O. DNA concentration was quantified with the KAPA Library Quantification Kit-Illumina (KAPA Biosystems) and sequenced on an Illumina MiSeq instrument according to the manufacturer's protocols.

General HTS data analysis.

Sequencing reads were demultiplexed in MiSeq Reporter (Illumina). Alignment of amplicon sequences to a reference sequence was performed as previously described using a Matlab script with improved output format[35]. In brief, the Smith–Waterman algorithm was used to align sequences without indels to a reference sequence; bases with a quality score of less than 30 were converted to 'N' to prevent base miscalling as a result of sequencing error. Indels were quantified separately using a modified version of a previously described Matlab script in which sequencing reads with more than half the base calls below a quality score of Q30 were filtered out. Indels were counted as reads which contained insertions or deletions of greater than or equal to 1 within a 30-bp window surrounding the predicted Cas9 cleavage site. To calculate the total number of edited reads as a proportion of the total number of successfully sequenced reads, the fraction of edited reads as measured by the alignment algorithm were multiplied by (1 – fraction of reads containing an indel).

Chapter Four: Development of a hRad51–Cas9 nickase fusion that mediates HDR without double-stranded breaks

This chapter has been adapted from:

Rees, H. A., Yeh, W-H., Liu, D. R. Development of a hRad51–Cas9 nickase fusion that mediates HDR without double-stranded breaks. *In preparation*

Contributions:

H.A.R. and D.R.L. designed the study; H.A.R. generated reagents; H.A.R and W-H. Y. performed experiments. D.R.L. supervised the research. All authors wrote the manuscript and assisted with proofing and revisions. We thank A. Hamidi, J. Levy, M. Arbab and L. Koblan for helpful discussions.

As with all the previous chapters, I am especially grateful to Professor David Liu for his advice, enthusiasm and mentorship throughout this project.

4.1 Introduction

Widely used genome editing strategies include gene disruption by generating insertions and deletions (indels) at a targeted locus following a double-stranded DNA break (DSB)[18], homology-directed repair (HDR) following a targeted DSB[21], and base editing, which enables the precise installation of transition point mutations (C to T, G to A, A to G, or T to C) without creating DSBs[10,33,35]. Among these three strategies, HDR offers access to the broadest possible range of desired changes to genomic DNA in mammalian cells (**Figure 4.1a**)[21]. The use of single-stranded DNA oligonucleotides containing PAM-blocking mutations as donor templates can improve HDR outcomes by preventing re-cutting of the target site after successful HDR (**Figure 4.1a**)[24]. Nevertheless, because HDR is usually initiated by a DSB, HDR is accompanied by undesired cellular side-effects including high levels of indel formation[22,24], DNA translocations[186], large deletions[187] and p53 activation[188,189].

We sought to improve ratios of desired:undesired HDR products by exploring the initiation of HDR from a DNA nick rather than a DSB. In contrast to DSBs, DNA nicks generally do not induce undesired genome modification,[56,190,191] a principle exploited by base editors to minimize editing byproducts[33,35,192]. Mutating catalytic residues in programmable nucleases can result in programmable nickases that cleave only one of the two strands of DNA at the target locus[193-196]. Although single nicks can lead to more favorable HDR:indel ratios than double-stranded DNA breaks,[8,195,197] nicks usually lead to much lower frequencies of genome editing when compared to DSBs (typically 5-20 fold)[6], making nickases substantially less useful than nucleases as genome editing tools[193,196-198].

In this study we achieved DSB-free HDR with minimal byproducts and reduced off-target editing by fusing a hRad51 variants to a programmable nickase to generate hRad51–Cas9(D10A) nickase fusions (RDN variants). RDN is capable of stimulating HDR at a DNA nick, resulting in a much higher ratio of HDR product:indel formation in human cells (up to 53-fold at the eight genomic loci tested here), substantially lower off-target editing, and HDR efficiency

that is generally similar to or better than that of DSB-mediated HDR. A known mutant of hRad51[199,200] that cannot bind BRCA2[199,200] can be used in RDN to further increase the HDR:indel ratio. A second known hRad51 mutant that cannot self-associate[199,200] increases overall HDR efficiency while slightly lowering HDR:indel ratios. RDN-mediated HDR is a one-step procedure that does not require inclusion of PAM-blocking mutations[24] and can use readily synthesized 100-mer single stranded DNA (ssDNA) oligonucleotides as donor templates. Although RDN remains limited by its dependence on cellular DNA repair processes underlying HDR, RDN may be useful for applications that require precise genome edits not accessible to base editing while minimizing undesired consequences of DSBs.

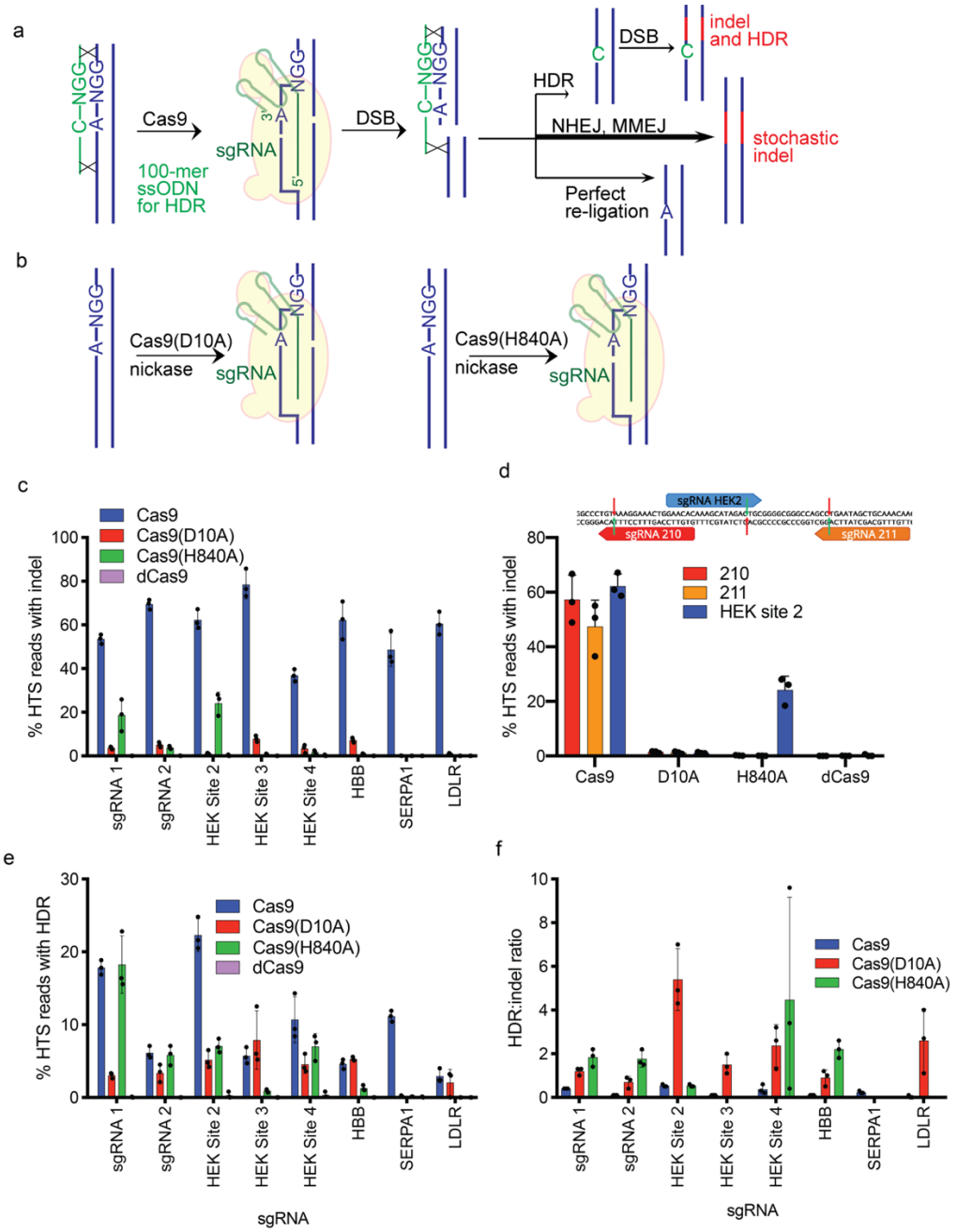


Figure 4.1. Indel formation and HDR in HEK293T cells mediated by Cas9 or Cas9 nickases. (a) DSB-mediated HDR using Cas9 and a 100-mer ssODN. **(b)** DNA nicks resulting from Cas9(D10A) or Cas9(H840A) nickase. **(c)** Indels resulting from Cas9 nuclease, Cas9 nickase, or dead Cas9 at eight loci in HEK293T cells. **(d)** Comparison of indel frequencies associated with three sgRNAs in close proximity. The sgRNA sequences used in are shown, with red arrows indicating nicks induced by Cas9(D10A) nickase, and green arrows showing nicks by Cas9(H840A) nickase. **(e)** Absolute frequencies of HDR, measured by high-throughput DNA sequencing in unsorted HEK293T cells at eight endogenous genomic loci. The position of the single point mutation that was incorporated by HDR is shown in the legend relative to the start of the sgRNA. **(f)** HDR:indel ratio associated with editing at eight loci. All data are shown as individual data points and mean ± s.d. for n=3 independent biological replicates, performed on different days.

4.2 Results

4.2.1 Indels caused by single Cas9 nickases

Cas9 contains two independent nuclease domains, either of which can be disabled to generate a nickase that selectively cleaves either the guide RNA-paired strand (Cas9(D10A) nickase) or the opposite strand (Cas9(H840A) nickase) (**Figure 4.1b**)[2]. We used high-throughput DNA sequencing (HTS) to systematically compare the editing outcomes of Cas9, Cas9(D10A), or Cas9(H840A) nickases at eight genomic loci in three human cell lines.

While both nickases resulted in substantially fewer indels than intact Cas9, nick-induced indel formation was highly strand- and locus-dependent (**Figure 4.1c**). The Cas9(D10A) and Cas9(H840A) nickases displayed different relative activities when paired with different sgRNAs; for example, at HEK site 2 the Cas9(H840A) nickase generated 24±5% indels and the Cas9(D10A) nickase generated only 1.1±0.2% indels, while at HEK site 3 Cas9(H840A) nickase resulted in only 0.73±0.38% indels but Cas9(D10A) nickase treatment generated 7.9±1.4% indels (**Figure 4.1c**). One of the eight sgRNAs we tested, sgRNA 171, did not lead to detectable indels when combined with either nickase despite robust indel formation when combined with Cas9 (**Figure 4.1c**). A similar pattern of indel formation at nicked sites was observed in HeLa and U2OS cells (**Figure 4.2a,b**), and with the ABEmax base editor, which contains a Cas9(D10A) nickase, although other base editors resulted in reduced indel frequencies compared to their component nickase domains alone (**Figure 4.3**).

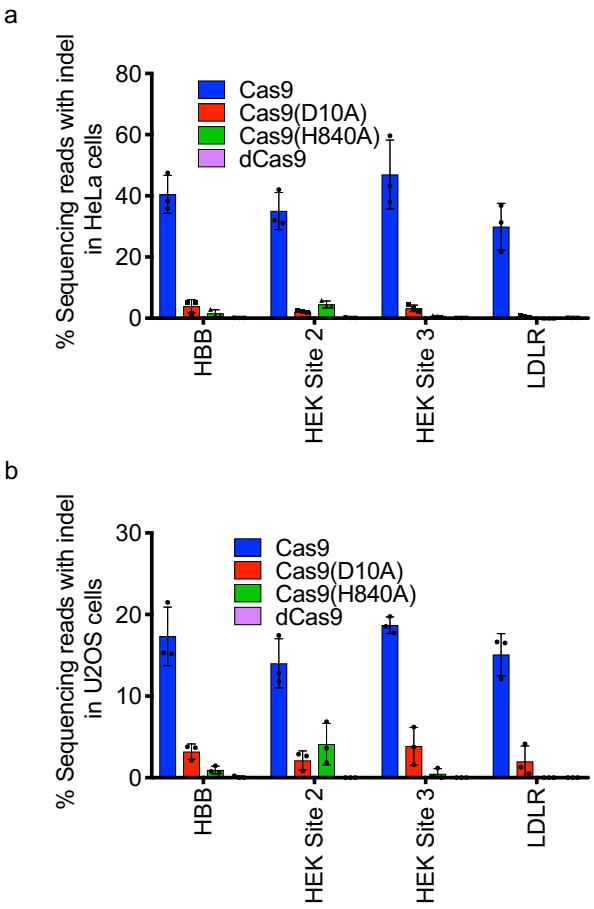


Figure 4.2. Frequency of nick-induced indels in HeLa and U2OS cells. Cells were lipofected with Cas9, D10A, H840A nickase or dCas9 plasmid and a plasmid expressing the indicated sgRNA. DNA was harvested and sequenced from unsorted cells and subjected to HTS. **(a)** Indel frequencies in HeLa cells. **(b)** Indel frequencies in U2OS cells. All data are shown as individual data points and mean ± s.d. for n=3 independent biological replicates, performed on different days.

Observed indel frequencies did not correlate with the presence of microhomology as predicted using inDelphi[201] (**Figure 4.4**). These results suggest that the cellular response to single nick generation is site-dependent and unpredictable by microhomology, though in general leads to substantially lower indel formation than the response to DSBs.

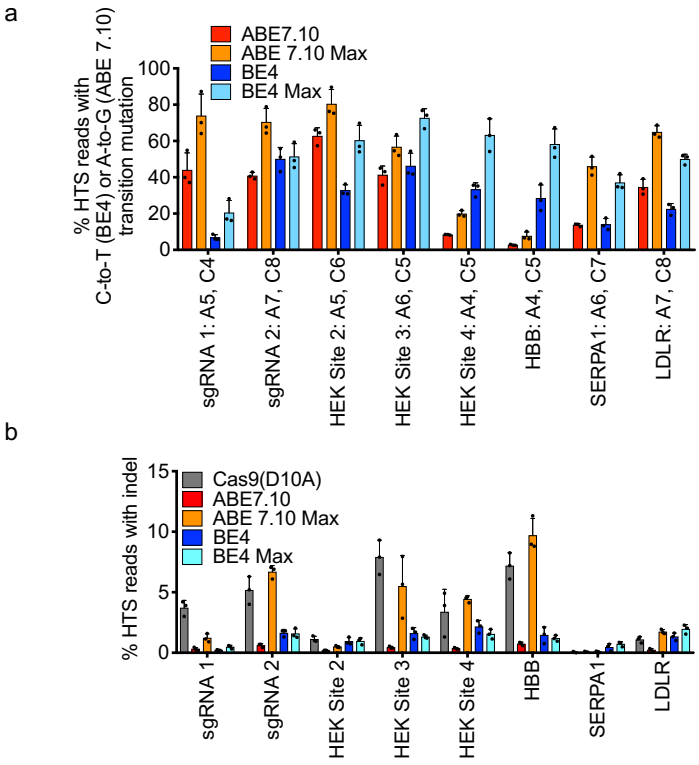


Figure 4.3. Indel formation and base editing in HEK293T cells at the same genomic loci as shown in Figure 4.1. (a) Indel frequencies associated with base editors and D10A nickase. **(b)** Base editing rates associated with base editors. All data are shown as individual data points and mean ± s.d. for n=3 independent biological replicates, performed on different days.

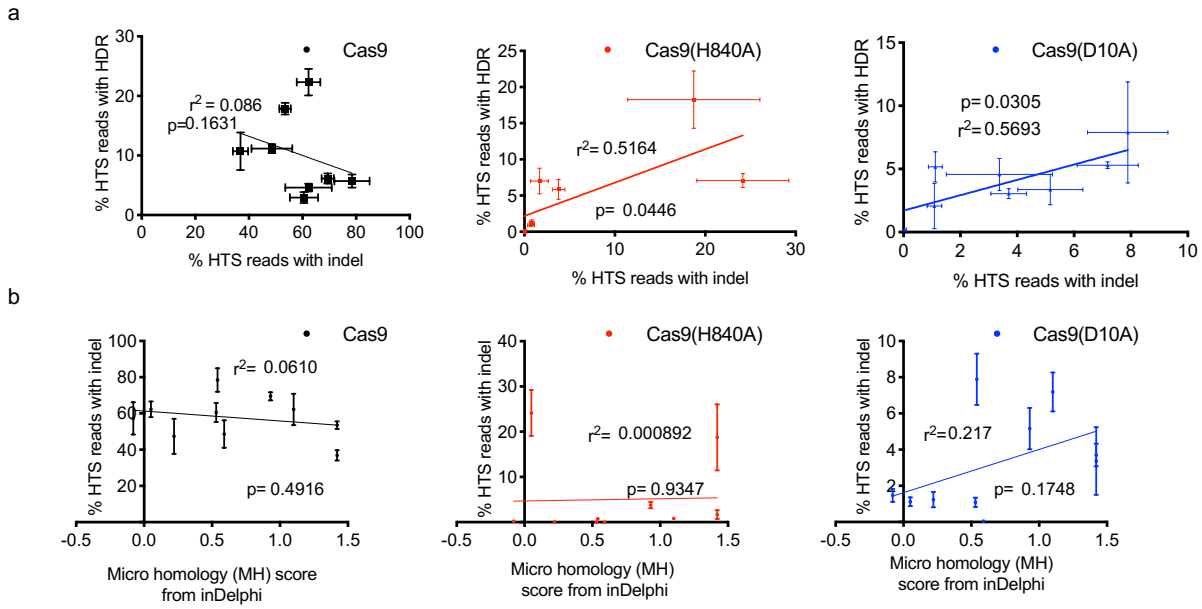


Figure 4.4. Correlation between HDR and indel frequencies and between indel frequencies and micro-homology with Cas9 nuclease and Cas9 nickases. (a) Indel

frequency in the absence of an ssODN plotted against HDR frequency. These data are also represented in Figure 1c and 1e. **(b)** Indel frequencies correlated to the micro homology score predicted by inDelphi[201] for each of the eight loci shown in Figure 1c and Figure 1d. For both **(a)** and **(b),** p-values were calculated in Prism. For **(a),** p-values represent a linear regression analysis to determine whether the slope is significantly non-zero. For **(b),** p-values represent a two-tailed test to determine whether the MH score is significantly correlated to the indel frequency. Data shows are the mean ± s.d. for n=3 independent biological replicates in HEK293T cells, performed on different days.

The Cas9 D10A nickase is a component of many DNA base editors, which are generally associated with low or undetectable indel rates[33,35]. We compared indel formation induced by D10A nickase to that associated with the recently reported expression-optimized base editors (ABEmax and BE4max)[202] and their predecessors (BE4[36] and ABE7.10[35]) (**Figure 4.3**). editors BE4-max, ABE7.10 and BE4 are associated with lower indel rates than the D10A-nickase alone (average indel generation across 8 loci was 3.7±2.8% for the D10A nickase, 1.2±0.5% for BE4-max, 1.2±0.7% for BE4 and 0.37±0.2% for ABE-7.10. To our surprise, ABE7.10-max generated very similar indel levels to the D10A-nickase alone – an average of 3.7±3.1%.

The basis for the elevated indel rates from optimizing ABE7.10 expression, which were not observed upon optimizing BE4 expression, is unclear, but may be attributed to increased levels of D10A nickase domain expression in ABE7.10max compared to ABE7.10, BE4max, or BE4[202]. These findings confirm that base editors generally induce lower indel rates than D10A nickase alone, and the elevated indel rates associated with ABEmax can be avoided by using ABE7.10[35].

We hypothesized that the site dependence of nickase-induced indels could be explained if the induced nicks were converted to DSBs by a separate, cellular process, such as DNA replication, that nicks the opposite DNA strand nearby. To test this possibility, we analyzed two sgRNAs (211 and 210) that target DNA either 28 bp upstream (sgRNA 210) or 18 bp downstream (sgRNA 211) of HEK site 2, a particularly asymmetric locus that results in high levels of Cas9(H840A) nickase-mediated indels but low levels of Cas9(D10A) nickase-induced indels. While Cas9(H840A) nickase and the HEK site 2 sgRNA resulted in high indel levels

(24±5%), nicking the same strand slightly upstream or downstream of HEK site 2 resulted in 17-fold lower indel formation (**Figure 4.1d**). These observations indicate that the high indel frequency generated by Cas9(H840A) nickase when paired with the HEK site 2 sgRNA is strongly dependent on the exact site being targeted. Together, these data suggest that the cellular response to nicks is distinct from the response to DSBs, and is highly sgRNA-dependent. The high degree of sgRNA-dependence associated with nick-induced indels may explain previously conflicting reports of the relative inactivity of the H840A nickase in human cells[203-205].

4.2.2 HDR stimulated by single Cas9 nickases

The use of HDR for precision genome editing in mammalian cells is limited by low efficiency in many cell types (T cells being a notable exception[206]), and the excess of indels and other undesired cellular outcomes that result from DSB formation. Previous work with Cas9 nickases[194,196-198,205], homing endonucleases converted to nickases[193], and zinc finger nickases[195] demonstrates that nicks can induce low levels of HDR when combined with a donor DNA template.

We wondered whether the observed variability among nick-induced indel formation also applies to nick-induced HDR. To assess this possibility, we designed 100-mer single-stranded DNA oligonucleotide (ssODN) templates for each of eight genomic loci and co-delivered them with Cas9 nuclease, Cas9 nickases, and catalytically dead Cas9 (dCas9). For three loci (*HBB*, *SERPA1* and *LDLR*), the ssODN encoded a single human pathogenic SNP located in the protospacer. For the remaining five loci, the donor templates were designed to incorporate an SNP within the protospacer as well as a PAM-altering SNP, as described in the CORRECT method for HDR donor template design[24]. We lipofected a plasmid encoding Cas9, Cas9 nickases, or dead Cas9, a plasmid expressing the indicated sgRNA, and the corresponding ssODN donor template into HEK293T cells. Four days post-lipofection, genomic DNA was

purified and analyzed by high throughput sequencing (HTS). We used Crispresso2[116] to filter out reads containing indels from our alignment prior to assessing HDR efficiency to ensure that reads containing both indels and HDR did not contribute to tabulated HDR efficiencies (See Methods).

At seven of eight sites, we detected HDR with one or both Cas9 nickases (**Figure 4.1e**). Regression analysis identified a weak positive correlation (R^2 = 0.57, p= 0.031 for the Cas9(D10A) nickase, R^2 = 0.51, p=0.045 for the H840A nickase) between indel formation and HDR frequencies with nickases, but no significant correlation with Cas9 nuclease (R^2 = 0.08 p=0.475) (**Figure 4.4a**). Although the absolute frequencies of HDR were 2.0- to 2.5-fold higher with Cas9 nuclease than with either Cas9 nickase (average across eight sites of 10% HDR product for Cas9, 5.0% for Cas9(H840A), and 4.0% for Cas9(D10A)), the HDR:indel ratio was 9.1- to 9.6-fold higher when using a nickase than Cas9 nuclease (the average HDR:indel ratio was 0.23 for Cas9, 2.1 for H840A, and 2.2 for Cas9(D10A)) (**Figure 4.1f**). Importantly, we did not detect HDR above a frequency of 0.2% when dCas9 was paired with the same sgRNAs and donor templates (**Figure 4.1e**), indicating that observed HDR frequencies are dependent on Cas9 nicking, and are not artefacts of the donor template acting as a primer during the PCR reaction prior to HTS, a source of artificially high apparent HDR frequencies (**Figure 4.5**). To ensure that the donor templates did not participate in the PCR reactions used a size-selective DNA purification step (**see Methods and Figure 4.5a and 4.5b**). These experiments establish that nick-induced HDR results in improved HDR:indel ratios compared to DSB-mediated HDR. However, the unpredictable nature of whether a nickase will be able to mediate HDR at a particular locus, as well as generally low efficiency, limits the utility of simple nickase-mediated HDR.

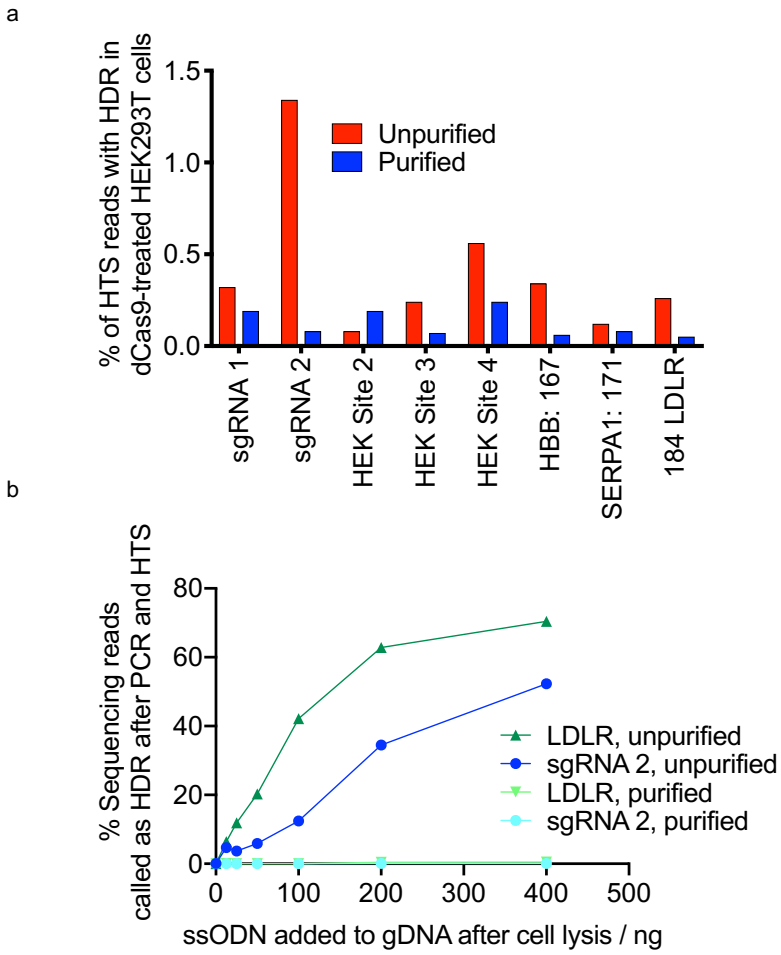


Figure 4.5. **Comparison of apparent HDR frequencies with and without magnetic bead based purification of genomic DNA. (a)** HEK293T cells were lipofected with a plasmid encoding dCas9, a plasmid encoding the indicated sgRNA, and a 50ng of a homologous 100-mer ssODN. Cells were lysed 4 days after treatment and crude cell lysate was saved before genomic DNA purification was performed with DNAdvance beads, as described in the Methods. The purified and unpurified genomic DNA samples were amplified by PCR and subjected to HTS, as described in the Methods. **(b)** Artefactual HDR frequencies recorded from addition of 100-mer ssODN to genomic DNA (gDNA) isolated from untreated HEK293T cells. The indicated ssODN was added to 600ng gDNA and the resulting mixture subjected to PCR and HTS as described in the methods ("unpurified samples"). A sample of each ssODN and gDNA mixture was purified using Agincourt DNAdvance magnetic beads as described in the Methods ("purified samples") to assess the extent to which bead-based purification can separate gDNA from ssODN donor.

4.2.3 Modulating HDR efficiencies by manipulating cellular repair proteins

To address these limitations, we sought to better understand the cellular proteins involved in catalyzing nick-induced HDR. To date, several studies have manipulated global DNA

repair processes to favor HDR over NHEJ[25,193,197,198,207,208]. Previous studies have identified key cellular DNA repair modulators that can be inhibited (such as p53 binding protein 1 (53BP1))[207,208] or overexpressed (such as Rad52[207]) to improve HDR:indel ratios in response to a targeted DSB. Knockdown of cellular hRad51, or inhibition of hRad51 by overexpression of the dominant negative mutant hRad51(K133R), increases both indel and HDR frequencies at targeted nicks[193,197,198]. Guided by these observations, we chose to manipulate these DNA repair modulators and study the resulting effects on DSB and nick-induced HDR.

We overexpressed either human hRad51 or hRad51(K133R) in conjunction with Cas9 or the Cas9(D10A) nickase (**Figure 4.6a**). Overexpression of hRad51 led to a significant ($p<0.05$; Student's two-tailed t test) decrease in HDR frequency at two of eight tested loci for Cas9(D10A) nick-mediated HDR (**Figure 4.6b**) and at five of eight loci for Cas9 DSB-mediated HDR (**Figure 4.6d**). Conversely, overexpression of hRad51(K133R), which inhibits cellular hRad51 activity, led to an increase in the efficiency of nick-induced HDR, but not DSB-induced HDR (**Figures 4.6b**, **d**). Finally, HDR:indel ratios remained largely unchanged by overexpression of hRad51 or hRad51(K133R). Together, these data demonstrate that hRad51 inhibition increased both HDR and indel frequencies at nick sites, but not at Cas9-induced DSBs (**Figures 4.6c, e**). Intriguingly, overexpression of hRad51(K133R) led to low but detectable levels of HDR at the previously refractory *SERPA1* site (sgRNA 171) (**Figure 4.6b**).

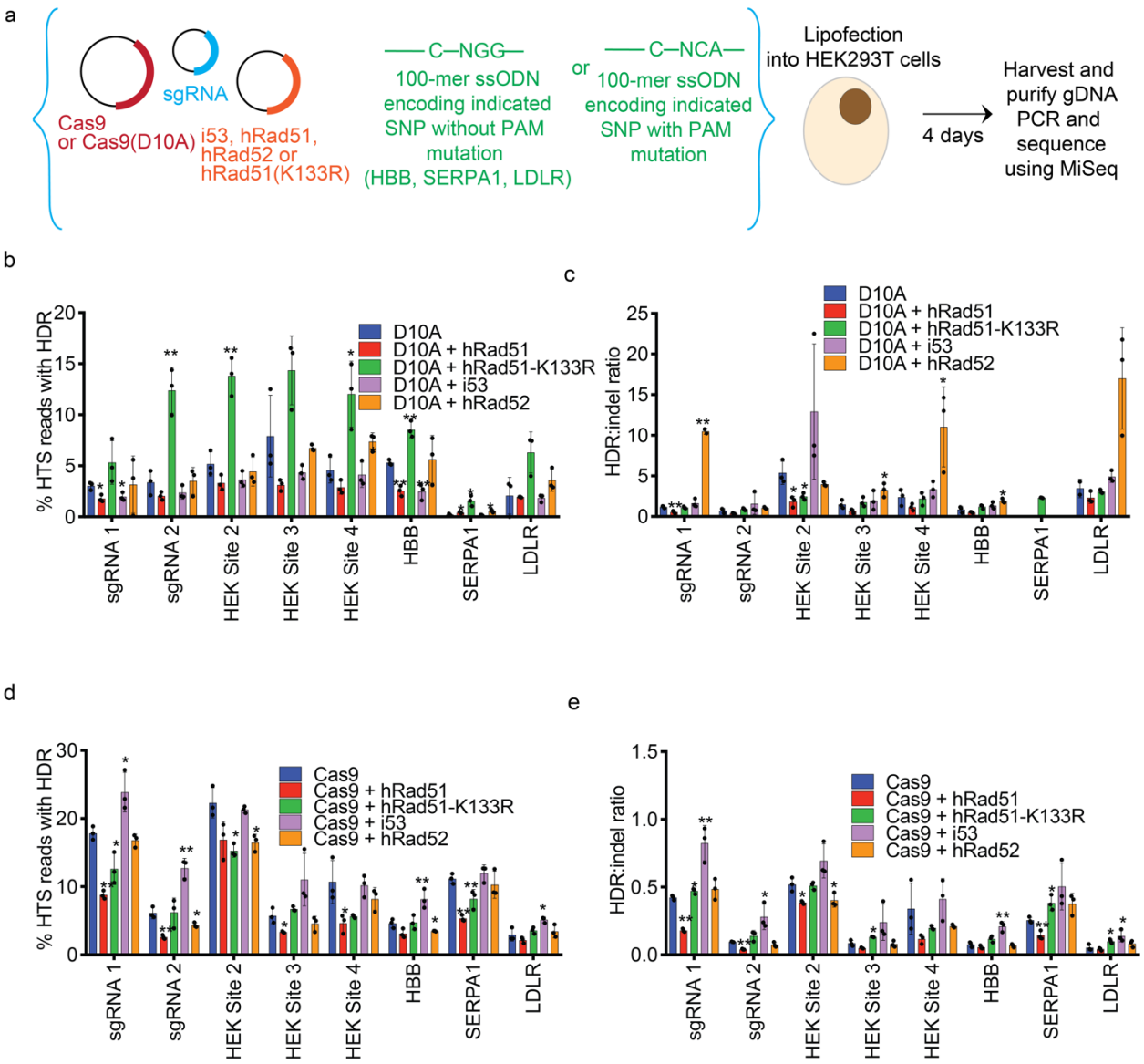


Figure 4.6. Manipulation of HDR frequency by global manipulation of cellular repair proteins. (a) Outline of experimental procedure. **(b)** and **(d)** Absolute frequencies of HDR, measured by high throughput sequencing in unsorted HEK293T cells at eight loci. **(c)** and **(d)** HDR:indel ratio at 8 loci. **(b)** and **(c)** show data associated with treatment of Cas9(D10A) nickase and **(d)** and **(e)** with Cas9 nuclease. All data are shown as individual data points and mean ± s.d. for n=3 independent biological replicates, performed on different days. Students two-tailed t-test was used to determine statistical significance between the indicated sample and in (**b**) Cas9(D10A) alone or in (**d**) Cas9 alone. (*): 0.01<p<0.05; (**): 0.001<p<0.01.

To test the potential effect of p53 binding protein 1 (53BP1) on nick-induced HDR, we overexpressed i53, a protein inhibitor of 53BP1[208]. 53BP1 directs DSBs towards NHEJ-mediated repair by preventing end resection, a key event on the HDR pathway[209].

Overexpression of i53 with Cas9 led to a significant (defined as $p<0.05$, Student's two-tailed t-test) increase in the absolute frequency of HDR at four of eight tested loci and an improvement in the HDR:indel ratio at six of eight loci compared to Cas9 alone (**Figures. 4.6d,e**). No such HDR improvements were observed when Cas9(D10A) nickase was used instead of Cas9 (**Figures 4.6b, c**), indicating that 53BP1 is unlikely to be a key modulator of nick-mediated HDR. Overexpression of Rad52, an interaction partner of hRad51, did not increase the efficiency of HDR arising from nicks or DSBs, but significantly improved the HDR:indel ratio at four of eight loci when HDR was stimulated by a nick (**Figures 4.6b-e**). Together, these findings suggest that global inhibition of cellular hRad51, but not inhibition of 53BP1 or elevating Rad52 levels, can increase the frequency of HDR in response to a DNA nick.

4.2.4 Development of Cas9(D10A)nickase fusions that promote HDR

Based on the above findings, we generated fusion constructs between the Cas9(D10A) nickase or the Cas9(H840A) nickase and hRad51(K133R). We anticipated that such fusions may be more effective and less perturbative than global inhibition of hRad51, which causes chromosomal instability[210]. We hypothesized that this fusion would lead to local inhibition of hRad51 at the target site. Under normal circumstances, cellular hRad51 would to bind to exposed genomic ssDNA after end-resection at the nick, leading to perfect, non-mutagenic repair of the nick[190,197]. This non-mutagenic repair process is inhibited by the dominant negative hRad51(K133R) mutant, which cannot hydrolyze ATP to initiate DNA strand invasion[211].

We began by optimizing the parameters for transfection by performing a titration of plasmid and donor template quantities and by measuring HDR and indel efficiencies at two loci with both the Cas9(D10A) nickase and the hRad51(K133R)–Cas9(D10A) fusion (**Figs. 4.7a-d**). To our surprise, a small quantity of ssODN (50 ng) was sufficient for efficient HDR, and increasing the ssODN amount to 400 ng reduced HDR efficiency. Fusion of hRad51(K133R) to the N-terminus of the Cas9(D10A) nickase increased HDR efficiency in HEK293T cells by an

average of 2.4-fold without altering the favorable HDR:indel ratio observed with the Cas9(D10A) nickase alone (**Figure 4.8b**). We refer to this fusion construct, hRad51(K133R)–Cas9(D10A) nickase, as RDN(K133R). Compared to RDN(K133R), moving the position of hRad51(K133R) to the C terminus of the Cas9(D10A) nickase did not significantly alter HDR frequencies (**Figure 4.8b**), nor did fusing an additional monomer of hRad51(K133R) to the N-terminus of Cas9(D10A) (**Figure 4.8b**). Fusion of one hRad51(K133R) monomer to the N-terminus and one to the C-terminus, however, reduced both HDR and indel formation, possibly due to the association of multiple fusion proteins into an extended multimer (**Figure 4.8b**). Consistent with the data showing that inhibition or overexpression of hRad51 does not have a substantial effect on DSB-mediated HDR, fusion between Cas9 and hRad51(K133R) led to a slight reduction to average HDR frequency at the loci tested (**Figure 4.8b**; **Figures. 4.11a,b**). Fusion between hRad51(K133R) and the Cas9(H840A) nickase also did not improve HDR frequency or HDR:indel ratios. The nickase preference of HDR enhancement upon hRad51(K133R) fusion may arise from the position of the nick introduced by Cas9(H840A) in the R-loop of displaced genomic DNA, compared with the position of the nick from Cas9(D10A) in the DNA:RNA duplex (**Figure 4.1b**).

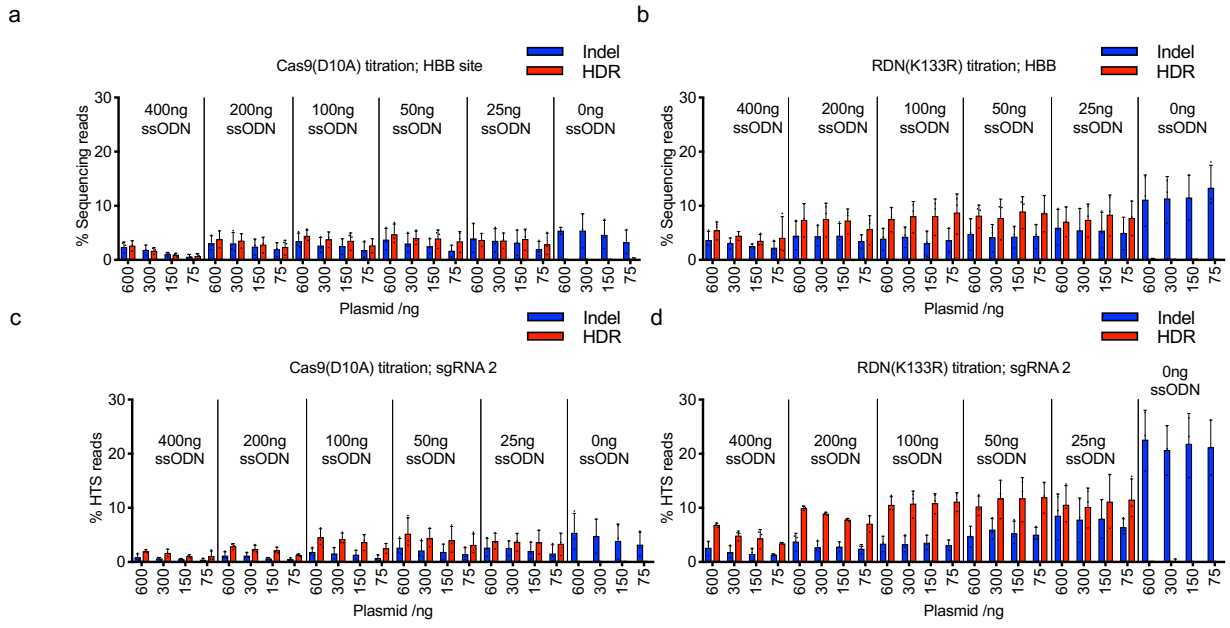
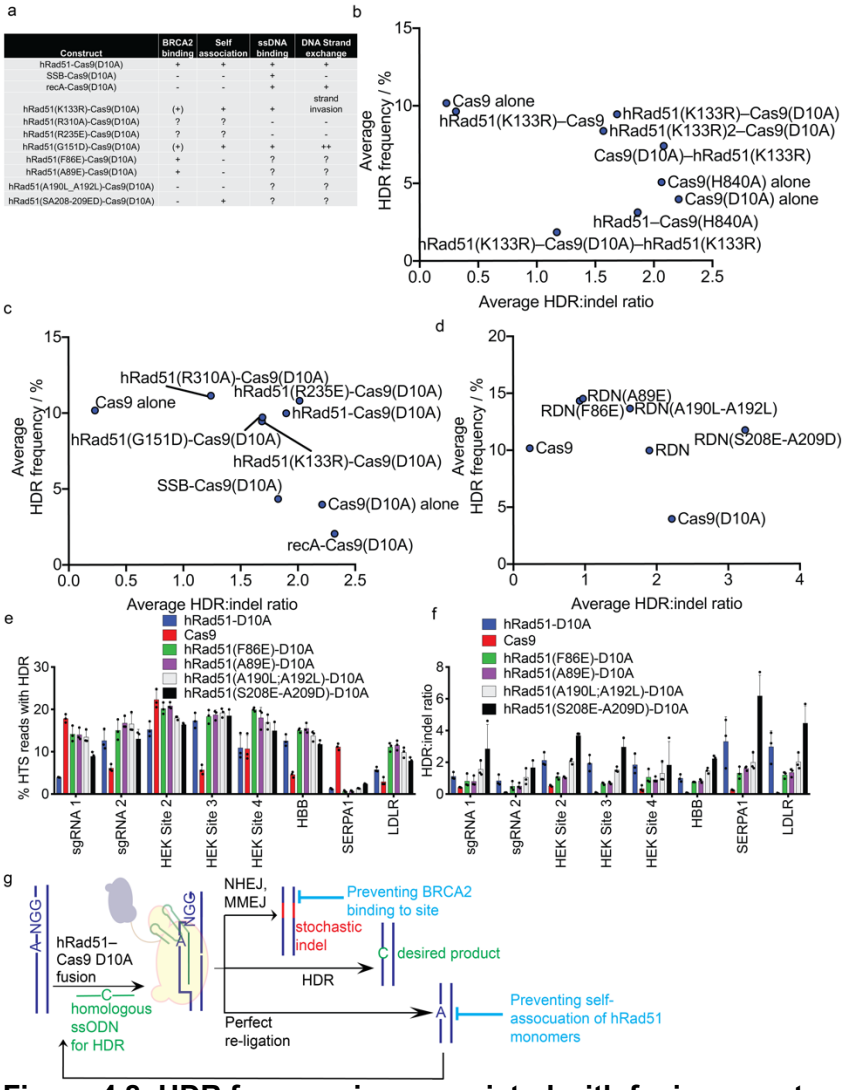


Figure 4.7. Titration of plasmid and ssODN quantities for lipofection-mediated transfection. HDR and indel frequencies associated with the indicated quantities of plasmid or ssODN, targeted to the indicated genomic locus. **(a)** and **(c)** show HDR and indel rates associated with D10A nickase; **(b)** and **(d)** show HDR and indel frequencies associated with the hRad51(K133R)-D10A fusion. 1.4 µL Lipofectamine 2000 was used for all conditions. All data are shown as individual data points and mean ± s.d. for n=3 independent biological replicates, performed on different days.

Surprisingly, fusion of wild-type hRad51 to Cas9(D10A), hereafter referred to as RDN, also resulted in increased HDR efficiency (**Figure 4.8c**), even though overexpression of hRad51 *in trans* with the Cas9(D10A) nickase lead to slightly decreased HDR efficiency (**Figure 4.6b**). These results indicate that increased HDR frequency mediated by RDN results from a mechanism distinct from global inhibition of hRad51. Together, these data demonstrate that localizing hRad51 to a targeted DNA nick through the RDN fusion increases nick-mediated HDR efficiency without inhibition of strand invasion mediated by cellular hRad51.



Construct	BRCA2 binding	Self association	ssDNA binding	DNA Strand exchange
hRad51-Cas9(D10A)	+	+	+	+
SSB-Cas9(D10A)	-	-	+	-
recA-Cas9(D10A)	-	-	+	+
hRad51(K133R)-Cas9(D10A)	(+)	+	+	strand invasion
hRad51(R310A)-Cas9(D10A)	?	?	-	-
hRad51(R235E)-Cas9(D10A)	?	?	-	-
hRad51(G151D)-Cas9(D10A)	(+)	+	+	++
hRad51(F86E)-Cas9(D10A)	+	-	?	?
hRad51(A89E)-Cas9(D10A)	+	-	?	?
hRad51(A190L_A192L)-Cas9(D10A)	-	-	?	?
hRad51(SA208-209ED)-Cas9(D10A)	-	+	?	?

Figure 4.8. HDR frequencies associated with fusion constructs between hRad51 and its mutants and Cas9 or Cas9 nickases. **(a)** Catalytic activity and protein-protein binding interactions associated with hRad51, mutants of hRad51 and the homologous protein recA. **+** indicates activity has been validated, – indicates absence of activity has been validated, ? indicates activity is unknown, (+) indicates activity has not been explicitly validated but is expected from structural considerations, ++ indicates improved activity relative to wild type **(b),** **(c)** and **(d)** Dot plots depicting the average frequencies of HDR and the average HDR:indel ratio associated with the indicated construct measured by high throughput sequencing in unsorted HEK293T cells at eight loci. **(b)** Comparison between fusion constructs between Cas9(D10A) and hRad51(K133R) with different fusion architectures. **(c)** Comparison between catalytic mutants of hRad51 bound to the N-terminus of Cas9(D10A). **(d)** Comparison between binding mutants of hRad51 bound to the N-terminus of Cas9(D10A). **(e)** HDR frequencies associated with hRad51 and the mutants depicted in **(d),** plotted by genomic locus. **(f)** HDR:indel ratio associated with editing at 8 loci. For **(e)** and **(f)**, data are shown as individual data points and mean ± s.d. for n=3 independent biological replicates, performed on different days. **(g)** Diagrammatic representation of our understanding of HDR induced by hRad51–Cas9(D10A) nickase fusions.

Next, we sought to understand if the HDR frequency enhancement associated with RDN and RDN(K133R) arises from steric occlusion of DNA repair proteins from accessing the nick or whether the affinity of hRad51 for single-stranded DNA leads to localization of the single-stranded DNA donor to the nick. To illuminate these possibilities, we designed two fusions between the Cas9(D10A) nickase and RecA or bacteriophage T4-derived single-stranded binding protein (SSB). RecA is a bacterial homolog of hRad51 that catalyzes strand invasion between homologous strands of DNA. Neither RecA–Cas9(D10A) nor SSB–Cas9(D10A) resulted in HDR enhancement (**Figure 4.8c**). Furthermore, incorporation of three additional hRad51 mutants (R310A, R235E and G151D) into RDN to generate RDN(R310A), RDN(R235E) and RDN(G151D) all displayed HDR enhancement frequencies indistinguishable from that of RDN and RDN(K133R) (**Figure 4.8c, and Figures 4.8i and 4.8j**), in spite of their differing catalytic and DNA-binding characteristics (**Figure 4.8a**)[212-214]. Taken together, these observations reveal that neither the fusion orientation of hRad51 relative to Cas9(D10A) nor the strand invasion and strand exchange activities of hRad51 are critical for the ability of RDN to mediate HDR.

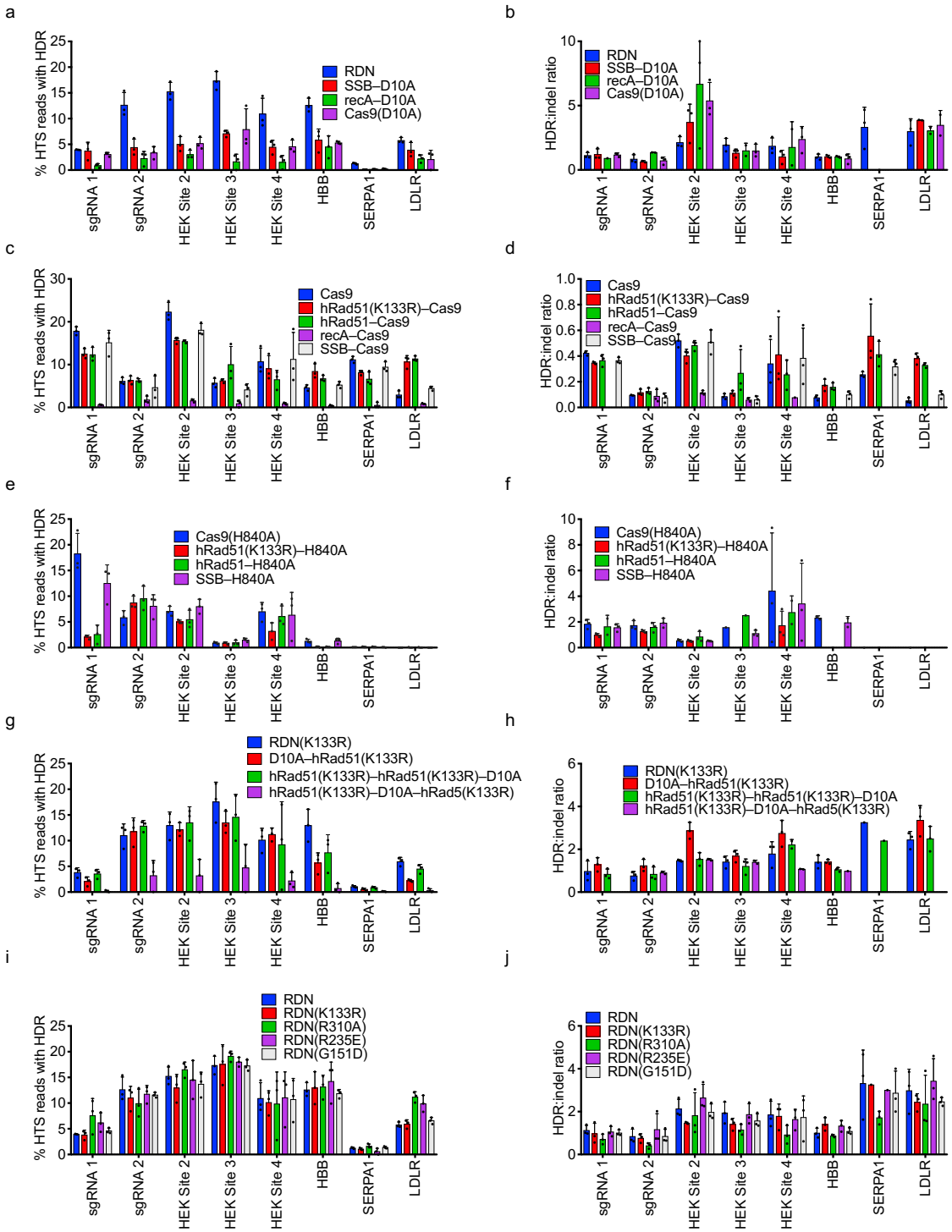


Figure 4.9. Site-by-site plots of HDR frequency and HDR:indel ratios in HEK293T cells, as described in Figure 4.8. (a), (c), (e), (g), (i) Site-by-site plots of HDR frequency. **(b), (d), (f), (h)** and **(j)** Site-by-site plots of HDR:indel ratio. These data have been processed and previously plotted in Figure 3b and 3c. Data are shown as individual data points and mean ± s.d. for n=3 independent biological replicates, performed on different days.

4.2.5 Donor template optimization

Including a PAM-altering mutation together with the target mutation in a donor template is an effective approach to improve HDR efficiency[24,181]. Mutating the PAM can prevent re-cutting and subsequent modification of the desired HDR product. HDR efficiencies are highly dependent on the distance between the DNA cleavage site and the mutation that is being incorporated[24,181]. Since indels are generated much less efficiently with nick-induced HDR compared to DSB-induced HDR (**Figure 4.1e**), we sought to test whether PAM-blocking mutations are necessary for nick-induced HDR and to define the region between the PAM and target mutation that can support efficient HDR.

We designed a series of eight ssODN templates targeting the HEK site 3 locus, each containing an SNP located in a different position within the protospacer from position 7 to 25, counting the PAM as positions 21-23. Two sets of donor templates were used. The first set of ssODNs incorporated a PAM mutation (replacing the TGG PAM with TTT) alongside the target mutation, while the second set only encoded each target mutation. Indeed, we observed an increase in the frequency of HDR mediated by Cas9 was when the PAM-blocking template was used compared to the non-PAM-blocking template (**Figure 4.10a**). By contrast, incorporating a PAM mutation into the donor ssODN did not lead to increased HDR frequency for nick-induced HDR, mediated either by Cas9(D10A) or RDN(K133R), as long as the target mutation is located within the sgRNA protospacer sequence (**Figure 4.10a**).

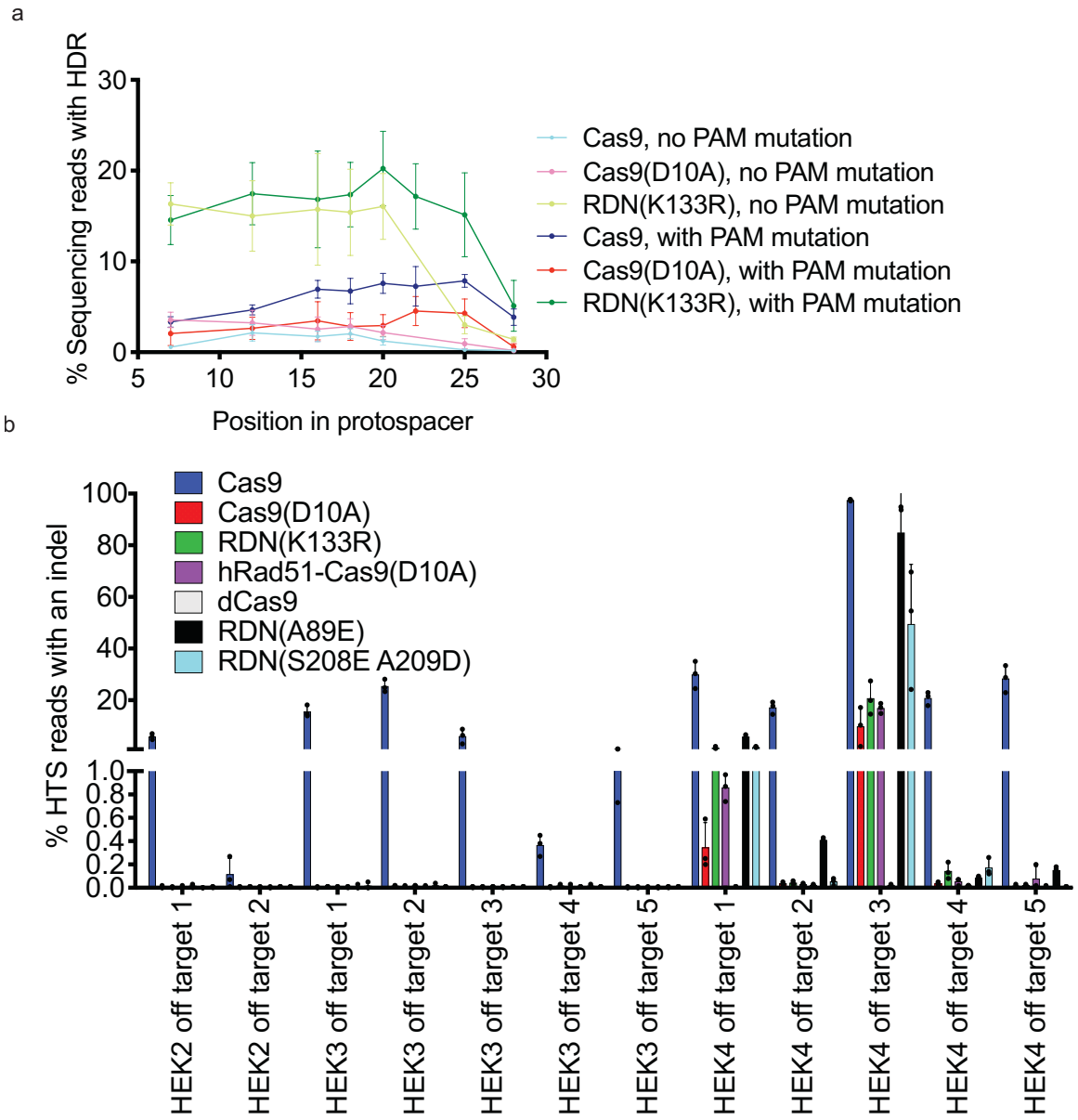


Figure 4.10. Characterization of the activity window and off-target profile of nick-mediated HDR. **(a)** Absolute frequencies of HDR, measured by high throughput sequencing in unsorted HEK293T cells using ssODNs with point mutations distributed along the sgRNA protospacer sequence of the HEK 3 sgRNA site. **(b)** Indel frequencies at off-target genomic loci in cells treated with Cas9 nuclease, Cas9(D10A) nickase, or Cas9(D10A) fusions with hRad51 or the indicated mutants thereof. Dead Cas9 (dCas9) treated cells were included as a negative control. All data are shown as individual data points and mean ± s.d. for n=3 independent biological replicates, performed on different days.

Unlike DSB-induced HDR, in which HDR efficiency steeply declines as the distance between the DSB and the incorporated mutation increases[24,181] (**Figure 4.10a**), we observed

comparable HDR efficiencies when RDN(K133R) was paired with different donor templates that introduced mutations from position 7 to 18 in the protospacer (**Figure 4.10a**). This greater apparent independence of HDR efficiency from the relationship between the location of the mutation to be installed and the location of the protospacer suggests that RDN may offer more flexibility with regards to guide RNA choice than Cas9 nuclease-mediated HDR.

We also tested donor template oligonucleotides that were oriented in the same sense as the sgRNA (forward template, which was used for all other experiments in this study) and in the opposite sense (reverse template). We did not observe any difference in the resulting HDR efficiencies mediated by Cas9(D10A), Cas9, Cas9(H840A), or RDN(K133R) (**Figure 4.11**), indicating that ssODN orientation is not a substantial determinant of HDR efficiencies under the conditions tested.

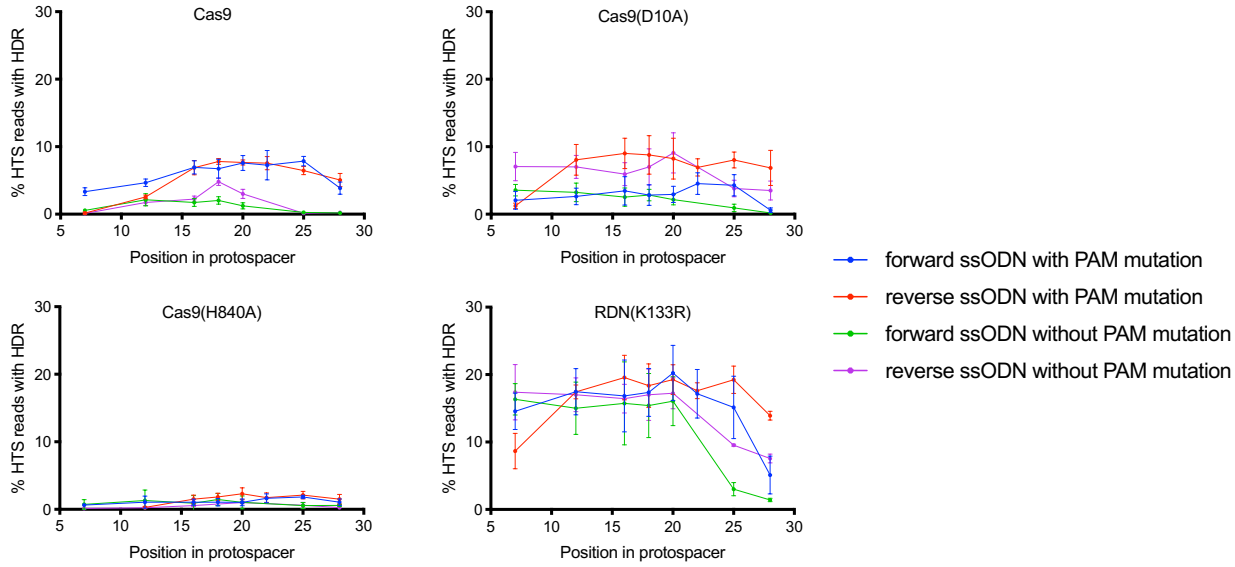


Figure 4.11 Assessment of the effect of ssODN sense on HDR editing rates, in HEK293T cells. Related to Figure 4a: ssODN sense (forward or reverse) was varied in the context of introducing single point mutations at different locations at the HEK 3 locus. Forward ssODN indicates that the ssODN-donor is in the same sense as the sgRNA; Reverse ssODN indicates that the ssODN-donor is in the reverse sense relative to the sgRNA (See Table 4.1). Data are shown as mean ± s.d. for n=3 independent biological replicates, performed on different days.

4.2.6 RDN with additional hRad51 mutants

Although the development of RDN as a tool to mediate HDR led to consistently improved HDR:indel ratios, when compared to Cas9 nuclease-mediated HDR, the overall HDR frequency associated with the two constructs is very similar (**Figure 4.8c**). In an attempt to improve overall HDR efficiency further while maintaining favorable HDR:indel ratios, we assessed four additional mutants of hRad51 in RDN constructs.

In addition to their role in catalyzing DNA strand invasion, hRad51 monomers directly bind to BRCA2[215-217], or to other hRad51 monomers[200,218]. Mutants of hRad51 that have lost either or both of these capabilities have been engineered[199,200] (**Figure 4.8a**). We installed these mutations into the RDN context and assayed HDR and indel outcomes of the resulting constructs to assess whether these binding interactions influence editing outcomes (**Figures. 4.8d-f**). The results revealed that using hRad51 mutants incapable of self-association, but which maintain BRCA2 binding, increased HDR efficiency in HEK293T cells at the 8 tested sites to an average of 14% (F86E mutant, RDN(F86E)) or 15% (A89E mutant, RDN(A89E)), compared to 10% for RDN. However, both of these mutants were associated with a modest reduction in HDR:indel ratio, from an average of 1.9 for RDN to 0.93 for RDN(F86E) or 0.98 for RDN(A89E).

In contrast, removing the BRCA2-binding ability of hRad51 using the double mutant (RDN(S208E, A209D)) only slightly improved HDR efficiency relative to RDN (to an average of 12%), but substantially improved the HDR:indel ratio (to 3.3), suggesting that abolishing recruitment of BRCA2 to the nick promotes more favorable HDR:indel partitioning. We should note that even with these improvements, the efficiency of nick-induced HDR remains more sgRNA-dependent than from DSB-induced HDR. For example, pairing original or mutant RDN constructs with sgRNA 171 leads to modest (<3%) HDR frequencies compared with Cas9 (11.1±0.6%).

We tested a final hRad51 mutant that lacks both BRCA2-binding and hRad51 self-association ability. The A190L A192L mutant was used instead of hRad51 to make RDN(A190L

A192L). This construct mediated HDR with an average efficiency of 14% and an HDR:indel ratio of 1.6, offering intermediate levels of HDR efficiency and HDR:indel ratio compared to the above RDN variants.

These analyses inform potential mechanisms by which RDN can mediate efficient HDR with favorable HDR:indel ratios. The data are consistent with a model in which self-association of hRad51 is important to maintain a high HDR:indel ratio but also limits HDR efficiency by promoting perfect repair of the DNA nick. In contrast, recruitment of BRCA2 to the nick site reduces the rate of perfect repair of the nick (**Figure 4.8g**). For applications that benefit most from maintaining the highest possible HDR efficiency, RDN(A89E) is the most useful, whereas applications that require maximizing the HDR:indel ratio will benefit from use of the RDN (S208E, A209D) variant.

4.2.7 Off-target modification induced by RDN variants

Cas9 nuclease[31] and Cas9-derived proteins such as base editors[33,35,92] can induce off-target editing in an sgRNA-dependent fashion. We characterized off-target editing at known off-target sites associated with three well-studied sgRNAs[31]: HEK site 2, HEK site 3, and HEK site 4, which is a notoriously promiscuous sgRNA[85,219]. Unlike with base editors, which can induce C-to-T or A-to-G point mutations as well as indels at off-target loci, the homology required between the target genomic locus and the ssODN prevents significant off-target HDR products from being generated by Cas9 combined with a ssODN. However, indel formation from Cas9 nuclease activity at off-target sites under these conditions is common; compared to a dCas9 negative control we observed statistically significant ($p<0.05$, Student's two tailed t-test, compared to dCas9 treatment) off-target indel formation with Cas9 treatment, at 11 of 12 tested known Cas9 off-target sites averaging 21% indels across the 12 sites (**Figure 4.10b**). In contrast, Cas9(D10A) nickase and RDN significantly edited 0/12 and 8/12 off-target loci, respectively (**Figure 4.10b**). Average off-target editing frequency induced by Cas9(D10A)

nickase at the 12 known Cas9 off-target sites was 1.8%, while that of the RDN and RDN(A89E) were 1.5% and 7.6%, respectively. These results confirm that RDN-mediated HDR offers substantially lower off-target DNA modification than nuclease-based HDR, even with RDN(A89E), which typically results in higher on-target HDR frequencies than Cas9.

4.2.8 HDR in other human cell types

HEK293 cells are particularly amenable to ssODN-mediated HDR[220]. Indeed, some other commonly used immortalized cell lines including HeLa and U2OS are thought to be completely refractory to ssODN-mediated HDR[220]. We compared RDN- and Cas9-mediated HDR outcomes in other immortalized cell lines and in primary human cells, including HeLa cells, U2OS cells, human induced pluoripotent stem (hiPS) cells and K562 cells.

In HEK293T cells, we observed that RDN(A89E) offers the highest HDR frequency (**Figure 4.8e**) and RDN(S208E, A209D) offered the highest HDR:indel ratio (**Figure 4.8f**) of all the constructs tested, so we tested these two constructs in the wider range of cell types. For this comparison, we used oligonucleotides designed without PAM mutations to maximize the generality of the results and due to our conclusions that nick-mediated HDR does not benefit from PAM blocking mutations (**Figure 4.10a**). Unless otherwise specified, we report results from unsorted cells and percentages are noted as a percentage of the entire cell population, not as percentages of “edited” or “modified” cells, that would greatly increase apparent editing efficiencies.

RDN (containing wild-type hRad51) led to substantially reduced HDR frequencies when compared to Cas9 in all non-HEK293T cell types tested. For example, in K562 cells the average reduction in efficiency was from a mean of 16% with Cas9 to 3.8% with RDN (**Figure 4.12a**). The mean HDR:indel ratio, however, was improved 87-fold in K562 cells and 3-fold in HeLa cells (**Figure 4.12b, d**). RDN(S208E, A209D) demonstrated slightly improved HDR:indel ratios

when compared to RDN, but the overall efficiency of HDR remained low compared to that achieved by Cas9 (**Figure 4.12**).

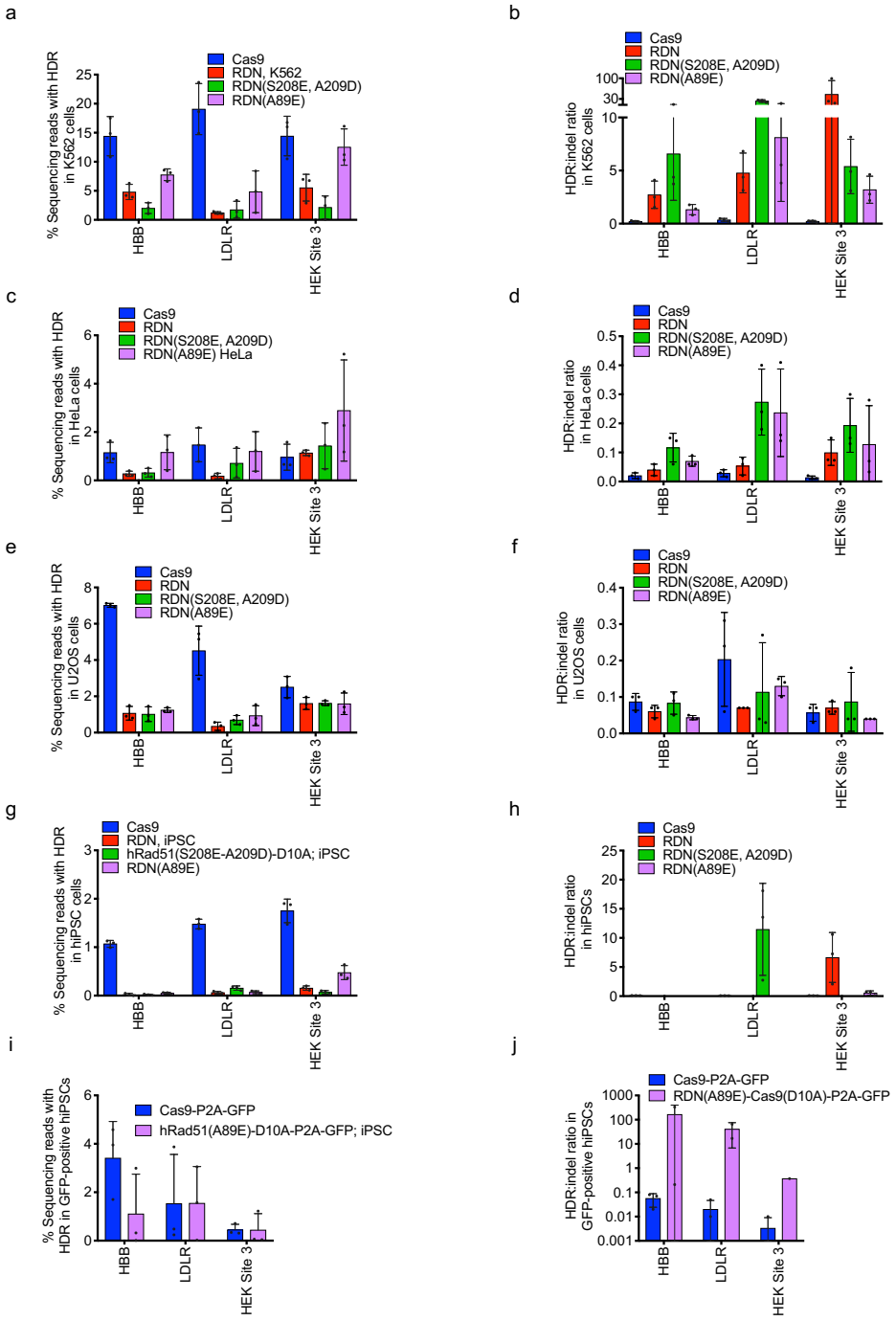


Figure 4.12. hRad51–Cas9(D10A) nickase activity in K562, U2OS, HeLa and hiPS cells. **(a), (c), (e), (g)** Absolute frequencies of HDR, measured by high throughput sequencing in unsorted, nucleofected cells at three loci; (**b**), **(d), (f), (h)**: HDR:indel ratio associated with editing at the same 3 loci. **(i)** and **(j)** HDR frequency and HDR:indel ratios in iPS cells nucleofected with P2A-GFP tagged constructs and sorted for GFP-positive cells. All data are shown as individual data

points and mean ± s.d. for n=3 biological replicates, performed independently.An HDR:indel ratio was not reported if the HDR frequency was <0.1% (see Methods).

When RDN(A89E) was used, however, the average HDR efficiency was substantially improved, with mean HDR frequencies of 8.3% in K562, 1.3% in U2OS, and 1.8% in HeLa cells (**Figure 4.12a, c, e**). HDR efficiencies in these three non-HEK293T cell types were on average 2.1-fold lower than those following Cas9 treatment. RDN(A89E) was associated with a 15-fold improvement in HDR:indel ratio in K562 cells and 7-fold in HeLa cells compared to Cas9 treatment. This improvement was not observed in U2OS cells, where a slight reduction in HDR:indel ratio was observed when RDN(A89E) was used (**Figure 4.12f**). In hiPS cells, only one of the three tested loci was amenable to RDN(A89E)-mediated HDR, demonstrating that this method may be more site-dependent in hiPS cells than in immortalized cell lines. To test if this limitation was due to poor expression of RDN(A89E) in hiPS cells, we generated Cas9 and RDN(A89E) constructs tagged with P2A GFP to enable isolation of Cas9- or RDN(A89E)-expressing cells. With Cas9–P2A–GFP, isolating GFP-positive cells resulted in 1.8% average HDR efficiencies in hiPS cells with an average HDR:indel ratio of 0.03 (**Figure 4.12i** and **j**). Among GFP-positive cells expressing RDN(A89E)–P2A–GFP, average HDR efficiencies were 1.0%, with an average HDR:indel ratio of 46 (**Figure 4.12i and 5j**), reflecting a modest decrease in HDR efficiency but a >1,000-fold improvement in HDR:indel ratio. (**Figure 4.12i, j**) (See **Figure 4.13** for FACS plots). Among GFP-positive cells isolated with the RDN(A89E)-P2A-GFP construct, average indel frequency was 1.6% and the vast majority showed no target site modification. This observation suggests that the majority of nicks induced by RDN(A89E) construct are perfectly repaired in hiPS cells; in contrast, GFP-positive cells containing Cas9–P2A–GFP contained an average of 77% indels.

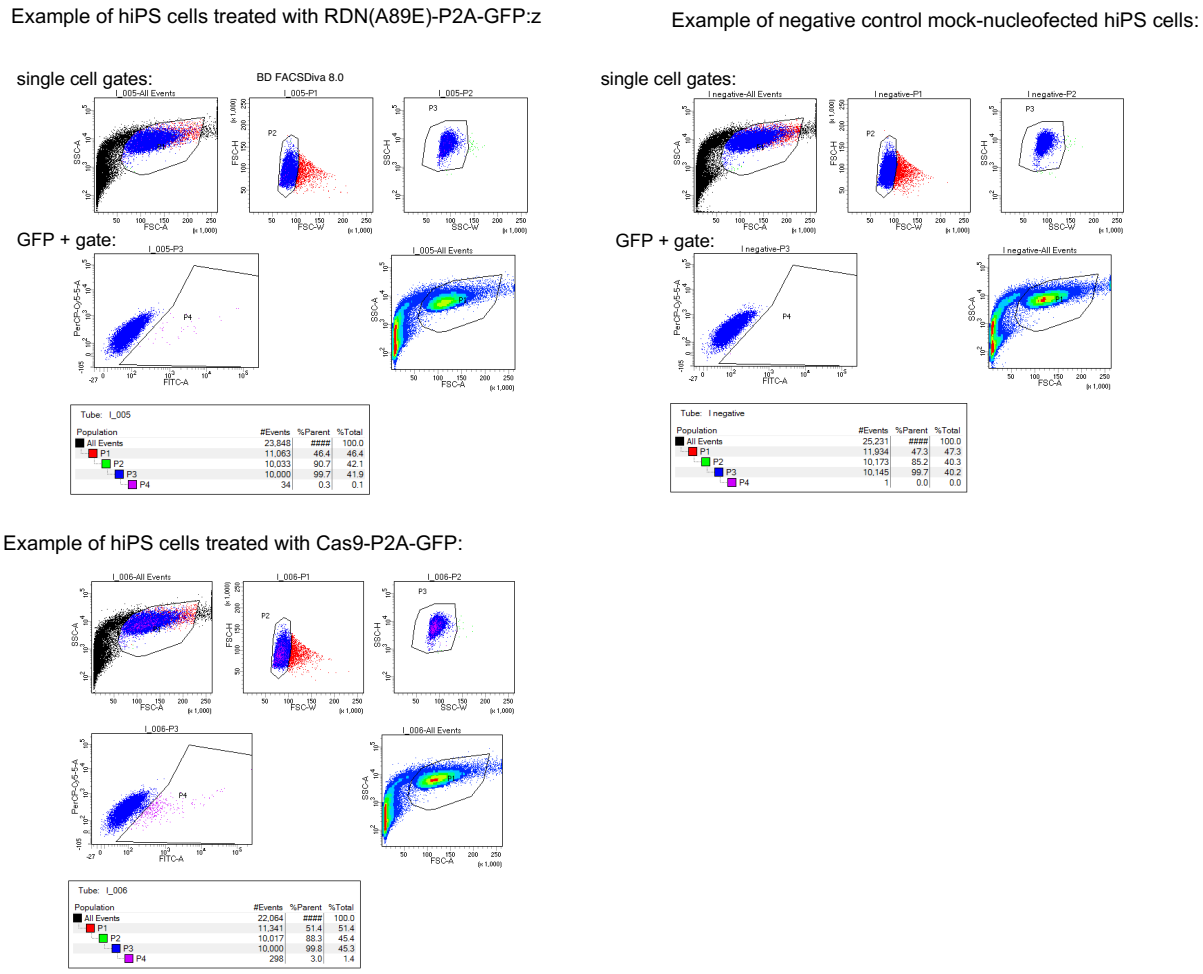


Figure 4.13. Gating examples for flow sorting human iPSC cells (hiPSC). Related to Figure 4.12i and 4.12j. Examples of flow sorting gates for single cells and for GFP+ cells are shown.

These data together reveal that RDN(A89E) mediates more efficient HDR than Cas9 nuclease in HeLa and HEK293T cells, maintains similar levels of HDR efficiency in K562 cells, and offers improved HDR:indel ratios in HeLa, HEK293T, K562, and hiPS cells. Neither an efficiency nor a product purity advantage from any tested RDN variant was observed in U2OS cells, possibly as a result of unusual regulation of DNA repair in U2OS cells[220,221]. This variability is likely due to the reliance of RDN on cellular repair processes that are highly cell type-dependent.

4.3 Discussion

The method developed in this study enables efficient, precise and specific changes to be made to genomic DNA through the cellular process of homology-directed repair, without generating a double stranded DNA break. Use of the fusion construct hRad51–Cas9(D10A) (RDN) or variants of this construct in which hRad51 has been replaced by hRad51 mutants, can address some of the challenges associated with using HDR to make precise changes to genomic DNA in some human cell types.

The HDR:indel ratio generated by RDN is generally improved compared to that which can be achieved using a DSB. This improvement in the purity of editing outcomes is particularly important for genome editing applications in which gene knockout resulting from indel formation opposes desired biological outcomes, or in which mixtures of many different edited genotypes—the typical cellular response to DSBs—is undesired. The RDN(S208E, A209D) construct is particularly useful under such circumstances since it offers ~3.2-fold greater HDR:indel ratios (**Figure 4.8d**). In addition, the efficiency of HDR mediated by RDN and RDN(A89E) is superior to that of Cas9 in some (but not all) cell types (**Figure 4.8e, Figure 4.12a**), although HDR efficiency remains modest, likely limited by dependence on cellular repair processes. RDN and its variants also offer substantially higher DNA specificity (lower off-target indel formation) compared to Cas9 nucleases combined with the same sgRNAs, even when applied to a notoriously promiscuous guide RNA with many known off-target loci (**Figure 4.10b**). RDN with wild-type hRad51 offers the greatest degree of DNA specificity among the mutants tested, but this difference was only notable at the promiscuous HEK Site 4, as were not able to detect off-target editing at frequencies above 0.2% at any other tested loci following use of RDN, RDN(A89E) or RDN(S208E, A209D) (**Figure 4.10b**). Finally, since RDN variants cannot generate DSBs, we anticipate that the likelihood of inducing translocations, large deletions, or p53 activation will be greatly reduced compared to nuclease-based genome editing methods.

Additional studies using are needed to fully characterize the scope of cellular responses to targeted nicks compared to targeted DSBs.

We anticipate that RDN(A89E) or RDN(S208E, A209D) will be most useful for applications in which efficiency and cleanliness of genome editing are critical. Recent work whereby saturation genome editing was performed to investigate variants of unknown significance in *BRCA1*[222] highlight the utility of a tool with the ability to generate mutations with single nucleotide resolution. Nuclease-mediated approaches to saturation editing can only be performed on essential genes because of the requirement that cells in which indels are induced must be excluded from the analysis. The favorable HDR:indel ratio and HDR efficiency offered by RDN may permit mutagenesis with nucleotide-level resolution on non-essential genes. Finally, we hope that the principles illuminated in this work will be enabling for researchers seeking to develop new methods for studying and manipulating cellular DNA damage and repair.

4.4 Methods

Plasmid cloning

All mammalian cell expression plasmids were constructed by USER cloning from gBlock gene fragments (Integrated DNA Technologies), as previously described[223]. Phusion U Green Multiplex PCR Master Mix (ThermoFisher) was used for amplification of DNA. sgRNA plasmids were constructed by blunt end ligation of a linear PCR product generated by encoding the 20-nt variable protospacer sequence onto the 5' end of an amplification primer and treating the resulting piece to KLD Enzyme Mix (New England Biolabs) according to the manufacturers' instruction. Mach1 chemically competent *E. coli* (ThermoFisher) cells were used.

Preparation of plasmids for mammalian cell transfection

To obtain endotoxin-free plasmids for transfection, 45 mL of Mach1 cells expressing freshly-transformed plasmid were pelleted by centrifugation (6000 g, 10 mins, 4 °C) and purified using ZymoPURE II Plasmid Midi Prep Kits (Zymo Research), according to the manufacturer's instructions with the inclusion of the optional step of passing the plasmid across the EndoZero Spin column (Zymo Research). Plasmid yield was quantified using a Nanodrop and by electrophoresis on a 1% agarose Tris/Borate/EDTA gel supplemented with ethidium bromide.

Mammalian cell culture

All cells were cultured and maintained at 37 °C with 5 % CO_2. Antibiotics were not used for cell culture. HEK293T cells (ATCC CRL-3216) and HeLa cells (ATCC CCL-2) were cultured in Dulbecco's modified Eagle's medium (DMEM) plus GlutaMax (ThermoFisher) supplemented with 10 % (v/v) fetal bovine serum (FBS). K562 cells (ATCC CCL-243) were cultured in Roswell Park Memorial Institute (RPMI) 1640 Medium plus GlutaMax (ThermoFisher) supplemented with 10 % (v/v) fetal bovine serum (FBS). U2OS cells were cultured in MyCoy's 5A Medium plus GlutaMax (ThermoFisher) supplemented with 10 % (v/v) fetal bovine serum (FBS).

hiPS cells (human episomal iPS cell line; A18945; ThermoFisher) were cultured in Essential 8 Flex Medium (ThermoFisher) supplemented with RevitaCell after passaging (ThermoFisher) according to the manufacturer's directions. Versene (Thermo Fisher) was used for cell passaging and dissociation. Prior to nucleofection, cells were harvested with Accutase (ThermoFisher).

For data shown in Figure 5i and 5j, nuclease expression plasmids were constructed whereby the Cas-enzyme construct (Cas9 or RDN(A89E)) was proceeded by P2A-GFP to enable isolation of transfected cells. iPS cells were flow sorted at the MIT FACS core 3-5 days after nucleofection and genomic DNA was isolated directly after sorting.

Mammalian cell lipofection and genomic DNA isolation

HEK293T cells were seeded on 48-well poly-D-lysine coated plates (Corning) 16-20 hours before lipofection. Lipofection was performed at a cell density of 65%. Unless otherwise stated, cells were transfected with 231 ng of nuclease- or base-editor expression plasmid DNA, 69 ng of sgRNA expression plasmid DNA, 50 ng (1.51pmol) 100-nt ssODN (PAGE-purified; Integrated DNA Technologies) and 1.4 µL Lipofectamine 2000 (ThermoFisher) per well. For experiments where global inhibition or overexpression of a cellular HDR-component was performed 100 ng of the appropriate plasmid was included. Cells were harvested 4 days post-transfection and genomic DNA isolation and purification was performed with Agincort DNAdvance Kit (Beckman Coulter), according to the manufacturer's protocol. Size-selective DNA purification was necessary to prevent contamination of gDNA with donor ssODN HDR templates. For analysis of indel formation in Figure 4.2, HeLa and U2OS cells were transfected according to the above protocol except they were transfected at a density of 80% with 1.4 µL Lipofectamie 3000 and 1 µL of P3000 (ThermoFisher) per well.

Nucleofection of U2OS, HeLa, hiPS cell and K562 cells

For data generated in Figure 5, nucleofection of K562, HeLa and U2OS cells was performed. For these three cell types, 350 ng nuclease-expression plasmid, 150 ng sgRNA-expression plasmid and 200 pmol (6.6 µg) 100-nt ssODN (PAGE-purified; Integrated DNA Technologies) was nucleofected in a final volume of 20 µL per sample in a 16-well Nucleocuvette strip (Lonza). K562 cells were nucleofected using the SF Cell Line 4D-Nucleofector X Kit (Lonza) with 5 x 10^5 cells per sample (program FF-120), according to the manufacturer's protocol. U2OS cells were nucleofected using the SE Cell Line 4D-Nucleofector X Kit (Lonza) with 3-4 x 10^5 cells per sample (program DN-100), according to the manufacturer's protocol. HeLa cells were nucleofected using the SE Cell Line 4D-Nucleofector X Kit (Lonza) with 2 x 10^5 cells per sample (program CN-114), according to the manufacturer's protocol. Cells were harvested 48 hours after nucleofection; genomic DNA was purified using the Agincort DNAdvance Kit (Beckman Coulter), according to the manufacturer's protocol.

hiPS cells were nucleofected with 400 ng nuclease-expression plasmid, 400 ng sgRNA-expression plasmid and 200 pmol (6.6 µg) 100-nt ssODN (PAGE-purified; Integrated DNA Technologies) in a final volume of 20 µL per sample in a 16-well Nucleocuvette strip (Lonza) using the CB-150 program in the P3 Primary Cell 4D-Nucleofector X Kit (Lonza) with 0.75-1.5x10^6 cells per sample.

Preparation of genomic DNA for High Throughput Sequencing (HTS)

Sites of interest were amplified using the primers listed (Tab;e 4/2). Amplification primers for the first PCR reaction (PCR1) were designed with primer2 and had 5' extensions to enable amplification with an Illumina barcoding primer in a second PCR reaction (PCR2). Phusion U Green Multiplex PCR Master Mix (ThermoFisher) was used for both PCR1 and PCR2. For PCR1, each reaction contained 0.5 µM of the appropriate forward and reverse primer (Table 4.2) and 30-100 ng of genomic DNA was as a template. Cycling conditions were 98 °C for 1 min 30 s, then 30 cycles of (98 °C for 10 s, 61 °C for 15 s, and 72 °C for 15 s) followed by a final

extension of 1 min at 72 °C per 30 µL reaction. PCR1 products were verified on a 2% agarose gel Tris/Borate/EDTA gel supplemented with ethidium bromide. For PCR2, 1 µL of unpurified PCR1 plus 0.5 µM of each of a unique forward and reverse barcoding primer pair were added to each sample for a final volume of 30 µL. Cycling conditions were 98 °C for 1 min 30 s, then 7 cycles of (98 °C for 10 s, 61 °C for 15 s, and 72 °C for 15 s) followed by a final extension of 1 min at 72 °C. PCR2 products were purified by gel electrophoresis on a 2% agarose gel using the QIAquick Gel Extraction Kit (Qiagen). Purified product was passed over a second Minelute column (Qiagen) for a further round of purification before quantification with QBit ssDNA HS Assay Kit (ThermoFisher) and sequenced using an Illumina MiSeq with 230-270-bp single end reads according to the manufacturer's instructions.

Primers for amplification of genomic DNA	
LDLR forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCCCTGCTTCTTTTTCTCTGGT
LDLR reverse	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCATTAACGCAGCCAACTTCA
HBB forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCTTCTCTGTCTCCACATGCC
HBB reverse	TGGAGTTCAGACGTGTGCTCTTCCGATCTTAGGGTTGGCCAATCTACTCCC
HEK site 3 and sgRNA 2 forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGAAACGCCCATGCAATTAGTC
HEK site 3 and sgRNA 2 reverse	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTTGTCAACCAGTATCCCGGTG
HEK site 2 forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGAATGGATTCCTTGGAAACAATG
HEK site 2 reverse	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGCCCCATCTGTCAAACT
HEK site 4 forward	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTTCAACCCGAACGGAG
HEK site 4 reverse	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCTGGTCTTCTTTCCCCTCC
sgRNA 1 forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGTTACTGCTCAGACATGTAA
sgRNA 1 reverse	TGGAGTTCAGACGTGTGCTCTTCCGATCTGACCTCGTGATCCACCTGCC
SERPA1 forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTGTTGAACTTGACCTCGGGG
SERPA1 reverse	TGGAGTTCAGACGTGTGCTCTTCCGATCTCATCAGCCAAAGCCTTGAGGAG

Table 4.2. List of primers used for amplification of genomic DNA prior to HTS

Analysis of HTS data

Demultiplexing of pooled sequencing reads was performed using the MiSeq Reporter software (Illumina). Crispresso-v2[224] was used to perform alignments between sequenced amplicons and reference amplicons. Indels were quantified in a 10-bp window surrounding the expected cut site for each sgRNA. For quantification of HDR, we discarded reads that contained indels from the alignment to the reference sequence using "discard-indel-reads" filter. This

approach ensured that we did not erroneously count reads that contained both an SNP incorporated through HDR and an indel as an HDR event, as has been previously described[24]. The resulting alignment contained only reads that do not contain an indel within the 10-bp window around the sgRNA cleavage site. Separately from the alignment matrix, the output of Crispresso-v2 reported the percentage of reads that had been excluded from the alignment because they contained an indel (%cells with indel). For each target point mutation that was incorporated via HDR, the alignment alone could be used to determine the % of non-indel containing cells (% indel-free cells with target mutation) that had successfully incorporated the target mutation. In order to assess the % of all cells that had the target mutation, the following correction was performed:

$$\%\ cells\ with\ target\ mutation = \%\ indelfree\ cells\ with\ target\ mutation * \frac{100\% - \%cells\ with\ indel}{100}$$

For calculation of HDR:indel ratio, the % cells with indel-free HDR at the indicated sequence was divided by the % cells with an indel in the 10-bp window surrounding the cleavage site. For experiments with HEK293T cells, where robust (>1%) HDR and indel percentages were detectable for many conditions, HDR:indel ratios were not calculated if HDR frequency was less than 1% for a particular sample, to avoid reporting artificially high HDR:indel ratios that could accompany very low frequency events. For the data shown in Figure 5, HDR and indel frequencies were measured in cell types less able than HEK293T cells to support HDR. For these instances, an HDR:indel ratio was not reported if the HDR frequency was <0.1% for the same reason. For calculations in the text in which averages across sites were made, if an HDR:indel ratio was not calculated due to a low HDR rate, then the HDR:indel ratio was set to zero and included in the calculation of the mean to avoid artificially inflating HDR:indel ratios.

Data availability

Plasmids encoding the constructs used in this study will be available on Addgene after publication. High-throughput DNA sequencing data has been deposited in the NCBI Sequence Read Archive with BioProject accession number PRJNA515942.

Chapter Five: Conclusions and future perspectives

The first part of this chapter has been adapted from:

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

The ability to efficiently and cleanly install changes to genetic information in living systems at the highest-resolution level—that of the individual base pair—resembled science fiction even only recently. The major developments summarized in this review have rapidly established base editing of individual nucleotides as a robust technology with the potential to broadly impact the life sciences and medicine.

5.1 Base editing

The two classes of DNA base editors described thus far have repeatedly proven effective for making precise point mutations in the genome of a wide variety of living cells and organisms. That said, cytosine and adenine base editors make only two of the six possible changes of one base pair to another. Much additional work is needed to develop base editors that can install transversion mutations, and possibly other DNA or RNA changes, at programmable target loci. Success will likely benefit a deep understanding and creative manipulation of cellular mechanisms controlling base modification and DNA repair in mammalian cells.

Although early examples of *in vivo* base editing are highly encouraging, challenges associated with delivery of large proteins into specific tissues remain an important focus of ongoing efforts, including the use of base editing to treat human genetic diseases. Thus the development of novel base editor delivery systems, including those that target specific tissues, is likely to be another major focus in the coming years. Detailed analyses of the off-target editing activities of base editors *in vivo* under a variety of conditions relevant to ongoing research and therapeutic applications are also needed, together with assessments of the potential biological consequences of making off-target point mutations *in vivo*. For example, since base editors in general do not create DSBs that can lead to indels, translocations, or large DNA rearrangements, can the clinically relevant consequences of off-target base editing be adequately assessed by monitoring the DNA sequences of a defined set of oncogenesis-

associated genes and their regulatory regions? Experimentally testing such possibilities in animals would represent important steps towards advancing base editing into the clinic.

The continued development of additional editing technologies that maximize base editing efficiency and targeting scope, while minimizing off-target base editing, will continue to propel the field towards increasingly ambitious and sophisticated applications. For the vast majority of base editing applications described here, the target sequence is known in advance. Thus, the development of many distinct classes of future base editors that each convert a target DNA base pair or RNA base exclusively in a particular sequence context, or in a protospacer containing a particular PAM, is likely to play an important role in maximizing the precision and specificity of base editing.

5.2 Nick-induced HDR

The data presented in Chapter Four indicate that some cell types, but not all, are amenable to the approach of nick-induced HDR. In cell types where this approach is effective, the constructs we have developed can enable efficient and clean installation of a single or a library of point mutations in a particular gene. This technology may prove useful for developing selections in mammalian cells, using the tool for mutagenesis in genomic DNA. Alternatively, it may enable allele-specific phenotypes to be interrogated in screening assays, as is often performed with siRNA or nuclease-mediated gene knockout.

I hope that more work may illuminate the genetic programs limiting the efficacy of RDN in some human cell types. Simple pull-down and Western Blotting experiments, or more involved CRISPR screens or transcriptomics analyses may tease apart the reasons behind these differences. Alone, this would be interesting to discern, but may also enable more effective tools to be developed for HDR-based precision editing.

The future for precision genome editing, through both HDR and base editing, will be an exciting one, both to watch and remain involved with.

Bibliography

1 Cohen, S. N. C., A.C.Y; Boyer, H.W; Helling, R.B. Construction of Biologically Functional Bacterial Plasmids In Vitro. *Proc Natl Acad Sci U S A* **70**, 3240-3244 (1973).

2 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).

3 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 (2002).

4 Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71, doi:10.1038/nature09523 (2010).

5 Cho, S. W., Kim, S., Kim, J. M. & Kim, J. S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**, 230-232, doi:10.1038/nbt.2507 (2013).

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

7 Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471, doi:10.7554/eLife.00471 (2013).

8 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).

9 Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278, doi:10.1016/j.cell.2014.05.010 (2014).

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

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

12 Sternberg, S. H. & Doudna, J. A. Expanding the Biologist's Toolkit with CRISPR-Cas9. *Mol Cell* **58**, 568-574, doi:10.1016/j.molcel.2015.02.032 (2015).

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

14 Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* **41**, 7429-7437, doi:10.1093/nar/gkt520 (2013).

15 Cheng, A. W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res* **23**, 1163-1171, doi:10.1038/cr.2013.122 (2013).

16 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).

17 Jeggo, P. A. DNA breakage and repair. *Adv Genet* **38**, 185-218 (1998).
18 Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol* **14**, 8096-8106 (1994).
19 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 (1994).
20 Rudin, N., Sugarman, E. & Haber, J. E. Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. *Genetics* **122**, 519-534 (1989).
21 Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc Natl Acad Sci U S A* **91**, 6064-6068 (1994).
22 Chapman, J. R., Taylor, M. R. & Boulton, S. J. Playing the end game: DNA double-strand break repair pathway choice. *Mol Cell* **47**, 497-510, doi:10.1016/j.molcel.2012.07.029 (2012).
23 Cox, D. B., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat Med* **21**, 121-131, doi:10.1038/nm.3793 (2015).
24 Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125-129, doi:10.1038/nature17664 (2016).
25 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).
26 Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**, 333-351, doi:10.1038/nrg.2016.49 (2016).
27 Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol*, doi:10.1038/nbt.4192 (2018).
28 Landrum, M. J. *et al.* spec *Nucleic Acids Res* **44**, D862-868, doi:10.1093/nar/gkv1222 (2016).
29 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).
30 Shin, H. Y. *et al.* CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat Commun* **8**, 15464, doi:10.1038/ncomms15464 (2017).
31 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).
32 Zhang, L. *et al.* Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. *PLoS One* **10**, e0120396, doi:10.1371/journal.pone.0120396 (2015).

33 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).
34 Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, doi:10.1126/science.aaf8729 (2016).
35 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).
36 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).
37 Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949, doi:10.1016/j.cell.2014.02.001 (2014).
38 Li, X. *et al.* Base editing with a Cpf1-cytidine deaminase fusion. *Nat Biotechnol* **36**, 324-327, doi:10.1038/nbt.4102 (2018).
39 Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science* **358**, 1019-1027, doi:10.1126/science.aaq0180 (2017).
40 Montiel-Gonzalez, M. F., Vallecillo-Viejo, I., Yudowski, G. A. & Rosenthal, J. J. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc Natl Acad Sci U S A* **110**, 18285-18290, doi:10.1073/pnas.1306243110 (2013).
41 Montiel-Gonzalez, M. F., Vallecillo-Viejo, I. C. & Rosenthal, J. J. An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Res* **44**, e157, doi:10.1093/nar/gkw738 (2016).
42 Fukuda, M. *et al.* Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-to-I RNA editing. *Sci Rep* **7**, 41478, doi:10.1038/srep41478 (2017).
43 Wettengel, J., Reautschnig, P., Geisler, S., Kahle, P. J. & Stafforst, T. Harnessing human ADAR2 for RNA repair - Recoding a PINK1 mutation rescues mitophagy. *Nucleic Acids Res* **45**, 2797-2808, doi:10.1093/nar/gkw911 (2017).
44 Vogel, P., Hanswillemenke, A. & Stafforst, T. Switching Protein Localization by Site-Directed RNA Editing under Control of Light. *ACS Synth Biol* **6**, 1642-1649, doi:10.1021/acssynbio.7b00113 (2017).
45 Hanswillemenke, A., Kuzdere, T., Vogel, P., Jekely, G. & Stafforst, T. Site-Directed RNA Editing in Vivo Can Be Triggered by the Light-Driven Assembly of an Artificial Riboprotein. *J Am Chem Soc* **137**, 15875-15881, doi:10.1021/jacs.5b10216 (2015).
46 Vogel, P. & Stafforst, T. Site-directed RNA editing with antagomir deaminases--a tool to study protein and RNA function. *ChemMedChem* **9**, 2021-2025, doi:10.1002/cmdc.201402139 (2014).
47 Vogel, P., Schneider, M. F., Wettengel, J. & Stafforst, T. Improving site-directed RNA editing in vitro and in cell culture by chemical modification of the guideRNA. *Angew Chem Int Ed Engl* **53**, 6267-6271, doi:10.1002/anie.201402634 (2014).
48 Stafforst, T. & Schneider, M. F. An RNA-deaminase conjugate selectively repairs point mutations. *Angew Chem Int Ed Engl* **51**, 11166-11169, doi:10.1002/anie.201206489 (2012).

49 Vogel, P. *et al.* Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. *Nat Methods* **15**, 535-538, doi:10.1038/s41592-018-0017-z (2018).
50 Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell* **10**, 1247-1253 (2002).
51 Kunz, C., Saito, Y. & Schar, P. DNA Repair in mammalian cells: Mismatched repair: variations on a theme. *Cell Mol Life Sci* **66**, 1021-1038, doi:10.1007/s00018-009-8739-9 (2009).
52 Pearl, L. H. Structure and function in the uracil-DNA glycosylase superfamily. *Mutat Res* **460**, 165-181 (2000).
53 Mol, C. D. *et al.* Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* **82**, 701-708 (1995).
54 Tang, W. & Liu, D. R. Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* **360**, doi:10.1126/science.aap8992 (2018).
55 Krokan, H. E., Drablos, F. & Slupphaug, G. Uracil in DNA--occurrence, consequences and repair. *Oncogene* **21**, 8935-8948, doi:10.1038/sj.onc.1205996 (2002).
56 Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715, doi:10.1038/362709a0 (1993).
57 Yasui, M. *et al.* Miscoding properties of 2'-deoxyinosine, a nitric oxide-derived DNA Adduct, during translesion synthesis catalyzed by human DNA polymerases. *J Mol Biol* **377**, 1015-1023, doi:10.1016/j.jmb.2008.01.033 (2008).
58 Schneider, M. F., Wettengel, J., Hoffmann, P. C. & Stafforst, T. Optimal guideRNAs for re-directing deaminase activity of hADAR1 and hADAR2 in trans. *Nucleic Acids Res* **42**, e87, doi:10.1093/nar/gku272 (2014).
59 Losey, H. C., Ruthenburg, A. J. & Verdine, G. L. Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. *Nat Struct Mol Biol* **13**, 153-159, doi:10.1038/nsmb1047 (2006).
60 Bass, B. L. & Weintraub, H. An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* **55**, 1089-1098 (1988).
61 Matthews, M. M. *et al.* Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat Struct Mol Biol* **23**, 426-433, doi:10.1038/nsmb.3203 (2016).
62 Yu, Y. T. *et al.* Internal modification of U2 small nuclear (sn)RNA occurs in nucleoli of Xenopus oocytes. *J Cell Biol* **152**, 1279-1288 (2001).
63 Bass, B. L. RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* **71**, 817-846, doi:10.1146/annurev.biochem.71.110601.135501 (2002).
64 Bazak, L. *et al.* A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res* **24**, 365-376, doi:10.1101/gr.164749.113 (2014).
65 Keppler, A. *et al.* A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* **21**, 86-89, doi:10.1038/nbt765 (2003).

66 Herbert, A. & Rich, A. The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1. *Proc Natl Acad Sci U S A* **98**, 12132-12137, doi:10.1073/pnas.211419898 (2001).
67 Lehmann, K. A. & Bass, B. L. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* **39**, 12875-12884 (2000).
68 Kuttan, A. & Bass, B. L. Mechanistic insights into editing-site specificity of ADARs. *Proc Natl Acad Sci U S A* **109**, E3295-3304, doi:10.1073/pnas.1212548109 (2012).
69 Crick, F. H. Codon--anticodon pairing: the wobble hypothesis. *J Mol Biol* **19**, 548-555 (1966).
70 Kume, H., Hino, K., Galipon, J. & Ui-Tei, K. A-to-I editing in the miRNA seed region regulates target mRNA selection and silencing efficiency. *Nucleic Acids Res* **42**, 10050-10060, doi:10.1093/nar/gku662 (2014).
71 Seeburg, P. H. & Hartner, J. Regulation of ion channel/neurotransmitter receptor function by RNA editing. *Curr Opin Neurobiol* **13**, 279-283 (2003).
72 Yang, J. H., Sklar, P., Axel, R. & Maniatis, T. Editing of glutamate receptor subunit B pre-mRNA in vitro by site-specific deamination of adenosine. *Nature* **374**, 77-81, doi:10.1038/374077a0 (1995).
73 Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem* **79**, 321-349, doi:10.1146/annurev-biochem-060208-105251 (2010).
74 Athanasiadis, A., Rich, A. & Maas, S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol* **2**, e391, doi:10.1371/journal.pbio.0020391 (2004).
75 Hess, G. T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat Methods* **13**, 1036-1042, doi:10.1038/nmeth.4038 (2016).
76 Kim, K. *et al.* Highly efficient RNA-guided base editing in mouse embryos. *Nat Biotechnol* **35**, 435-437, doi:10.1038/nbt.3816 (2017).
77 Ma, Y. *et al.* Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat Methods* **13**, 1029-1035, doi:10.1038/nmeth.4027 (2016).
78 Wang, L. *et al.* Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. *Cell Res* **27**, 1289-1292, doi:10.1038/cr.2017.111 (2017).
79 Di Noia, J. & Neuberger, M. S. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**, 43-48, doi:10.1038/nature00981 (2002).
80 Radany, E. H. *et al.* Increased spontaneous mutation frequency in human cells expressing the phage PBS2-encoded inhibitor of uracil-DNA glycosylase. *Mutat Res* **461**, 41-58 (2000).
81 Hess, G. T., Tycko, J., Yao, D. & Bassik, M. C. Methods and Applications of CRISPR-Mediated Base Editing in Eukaryotic Genomes. *Mol Cell* **68**, 26-43, doi:10.1016/j.molcel.2017.09.029 (2017).

82 Ryu, S. M. *et al.* Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat Biotechnol*, doi:10.1038/nbt.4148 (2018).
83 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).
84 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).
85 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).
86 Lau, A. Y., Wyatt, M. D., Glassner, B. J., Samson, L. D. & Ellenberger, T. Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc Natl Acad Sci U S A* **97**, 13573-13578, doi:10.1073/pnas.97.25.13573 (2000).
87 Kouzminova, E. A. & Kuzminov, A. Patterns of chromosomal fragmentation due to uracil-DNA incorporation reveal a novel mechanism of replication-dependent double-stranded breaks. *Mol Microbiol* **68**, 202-215, doi:10.1111/j.1365-2958.2008.06149.x (2008).
88 Liu, Z. *et al.* Highly efficient RNA-guided base editing in rabbit. *Nat Commun* **9**, 2717, doi:10.1038/s41467-018-05232-2 (2018).
89 Yeh, W. H., Chiang, H., Rees, H. A., Edge, A. S. B. & Liu, D. R. In vivo base editing of post-mitotic sensory cells. *Nat Commun* **9**, 2184, doi:10.1038/s41467-018-04580-3 (2018).
90 Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol*, doi:10.1038/nbt.4172 (2018).
91 Kang, B. C. *et al.* Precision genome engineering through adenine base editing in plants. *Nat Plants*, doi:10.1038/s41477-018-0178-x (2018).
92 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).
93 Kim, D. *et al.* Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat Biotechnol* **35**, 475-480, doi:10.1038/nbt.3852 (2017).
94 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).
95 Lee, J. K. *et al.* Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat Commun* **9**, 3048, doi:10.1038/s41467-018-05477-x (2018).
96 Yamanaka, S. *et al.* Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proc Natl Acad Sci U S A* **92**, 8483-8487 (1995).
97 Okazaki, I. M. *et al.* Constitutive expression of AID leads to tumorigenesis. *J Exp Med* **197**, 1173-1181, doi:10.1084/jem.20030275 (2003).

98 Burns, M. B. *et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**, 366-370, doi:10.1038/nature11881 (2013).
99 Liu, Z. *et al.* Efficient generation of mouse models of human diseases via ABE- and BE-mediated base editing. *Nat Commun* **9**, 2338, doi:10.1038/s41467-018-04768-7 (2018).
100 Gehrke, J. M. *et al.* An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol*, doi:10.1038/nbt.4199 (2018).
101 Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186-191, doi:10.1038/nature14299 (2015).
102 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).
103 Kim, E. *et al.* In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat Commun* **8**, 14500, doi:10.1038/ncomms14500 (2017).
104 Muller, M. *et al.* Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. *Mol Ther* **24**, 636-644, doi:10.1038/mt.2015.218 (2016).
105 Lee, C. M., Cradick, T. J. & Bao, G. The Neisseria meningitidis CRISPR-Cas9 System Enables Specific Genome Editing in Mammalian Cells. *Mol Ther* **24**, 645-654, doi:10.1038/mt.2016.8 (2016).
106 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).
107 Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-485, doi:10.1038/nature14592 (2015).
108 Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol* **35**, 371-376, doi:10.1038/nbt.3803 (2017).
109 Wang, X. *et al.* Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nat Biotechnol*, doi:10.1038/nbt.4198 (2018).
110 Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635-646, doi:10.1016/j.cell.2014.09.039 (2014).
111 Jiang, W. *et al.* BE-PLUS: a new base editing tool with broadened editing window and enhanced fidelity. *Cell Res*, doi:10.1038/s41422-018-0052-4 (2018).
112 Kim, S., Bae, T., Hwang, J. & Kim, J. S. Rescue of high-specificity Cas9 variants using sgRNAs with matched 5' nucleotides. *Genome Biol* **18**, 218, doi:10.1186/s13059-017-1355-3 (2017).
113 Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* (2018).
114 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).
115 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).

116 Pinello, L. *et al.* Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol* **34**, 695-697, doi:10.1038/nbt.3583 (2016).
117 Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144-149, doi:10.1038/nature20565 (2016).
118 Wheeler, L. C., Lim, S. A., Marqusee, S. & Harms, M. J. The thermostability and specificity of ancient proteins. *Curr Opin Struct Biol* **38**, 37-43, doi:10.1016/j.sbi.2016.05.015 (2016).
119 Zafra, M. P. *et al.* Optimized base editors enable efficient editing in cells, organoids and mice. *Nat Biotechnol*, doi:10.1038/nbt.4194 (2018).
120 Midoux, P., Pichon, C., Yaouanc, J. J. & Jaffres, P. A. Chemical vectors for gene delivery: a current review on polymers, peptides and lipids containing histidine or imidazole as nucleic acids carriers. *Br J Pharmacol* **157**, 166-178, doi:10.1111/j.1476-5381.2009.00288.x (2009).
121 Bodles-Brakhop, A. M., Heller, R. & Draghia-Akli, R. Electroporation for the delivery of DNA-based vaccines and immunotherapeutics: current clinical developments. *Mol Ther* **17**, 585-592, doi:10.1038/mt.2009.5 (2009).
122 Thomas, C. E., Ehrhardt, A. & Kay, M. A. Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet* **4**, 346-358, doi:10.1038/nrg1066 (2003).
123 Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat Biotechnol* **33**, 73-80, doi:10.1038/nbt.3081 (2015).
124 Wang, M. *et al.* Efficient delivery of genome-editing proteins using bioreducible lipid nanoparticles. *Proc Natl Acad Sci U S A* **113**, 2868-2873, doi:10.1073/pnas.1520244113 (2016).
125 Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* **24**, 1012-1019, doi:10.1101/gr.171322.113 (2014).
126 Miao, C. H. *et al.* Nonrandom transduction of recombinant adeno-associated virus vectors in mouse hepatocytes in vivo: cell cycling does not influence hepatocyte transduction. *J Virol* **74**, 3793-3803 (2000).
127 Dong, J. Y., Fan, P. D. & Frizzell, R. A. Quantitative analysis of the packaging capacity of recombinant adeno-associated virus. *Hum Gene Ther* **7**, 2101-2112, doi:10.1089/hum.1996.7.17-2101 (1996).
128 Lai, Y. *et al.* Efficient in vivo gene expression by trans-splicing adeno-associated viral vectors. *Nat Biotechnol* **23**, 1435-1439, doi:10.1038/nbt1153 (2005).
129 Gao, X. *et al.* Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. *Nature* **553**, 217-221, doi:10.1038/nature25164 (2018).
130 Park, D. S. *et al.* Targeted Base Editing via RNA-Guided Cytidine Deaminases in Xenopus laevis Embryos. *Mol Cells* **40**, 823-827, doi:10.14348/molcells.2017.0262 (2017).
131 Liang, P. *et al.* Effective gene editing by high-fidelity base editor 2 in mouse zygotes. *Protein Cell* **8**, 601-611, doi:10.1007/s13238-017-0418-2 (2017).

132 Li, G. *et al.* Highly efficient and precise base editing in discarded human tripronuclear embryos. *Protein Cell* **8**, 776-779, doi:10.1007/s13238-017-0458-7 (2017).
133 Liang, P. *et al.* Correction of beta-thalassemia mutant by base editor in human embryos. *Protein Cell* **8**, 811-822, doi:10.1007/s13238-017-0475-6 (2017).
134 Ma, Y. *et al.* Highly efficient and precise base editing by engineered dCas9-guide tRNA adenosine deaminase in rats. *Cell Discov* **4**, 39, doi:10.1038/s41421-018-0047-9 (2018).
135 Zhang, Y. *et al.* Programmable base editing of zebrafish genome using a modified CRISPR-Cas9 system. *Nat Commun* **8**, 118, doi:10.1038/s41467-017-00175-6 (2017).
136 Tanaka, S. *et al.* In vivo targeted single-nucleotide editing in zebrafish. *Sci Rep* **8**, 11423, doi:10.1038/s41598-018-29794-9 (2018).
137 Schenk, B. *et al.* MPDU1 mutations underlie a novel human congenital disorder of glycosylation, designated type If. *J Clin Invest* **108**, 1687-1695, doi:10.1172/JCI13419 (2001).
138 Zeng, Y. *et al.* Correction of the Marfan Syndrome Pathogenic FBN1 Mutation by Base Editing in Human Cells and Heterozygous Embryos. *Molecular Therapy*, doi:10.1016/j.ymthe.2018.08.007 (2018).
139 Chadwick, A. C., Wang, X. & Musunuru, K. In Vivo Base Editing of PCSK9 (Proprotein Convertase Subtilisin/Kexin Type 9) as a Therapeutic Alternative to Genome Editing. *Arterioscler Thromb Vasc Biol* **37**, 1741-1747, doi:10.1161/ATVBAHA.117.309881 (2017).
140 Roccio, M., Hahnewald, S., Perny, M. & Senn, P. Cell cycle reactivation of cochlear progenitor cells in neonatal FUCCI mice by a GSK3 small molecule inhibitor. *Sci Rep* **5**, 17886, doi:10.1038/srep17886 (2015).
141 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).
142 Kuscu, C. *et al.* CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat Methods* **14**, 710-712, doi:10.1038/nmeth.4327 (2017).
143 Billon, P. *et al.* CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons. *Mol Cell* **67**, 1068-1079 e1064, doi:10.1016/j.molcel.2017.08.008 (2017).
144 Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med*, doi:10.1038/s41591-018-0049-z (2018).
145 Ihry, R. J. *et al.* p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat Med*, doi:10.1038/s41591-018-0050-6 (2018).
146 Aguirre, A. J. *et al.* Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov* **6**, 914-929, doi:10.1158/2159-8290.CD-16-0154 (2016).
147 Roukos, V. & Misteli, T. The biogenesis of chromosome translocations. *Nat Cell Biol* **16**, 293-300, doi:10.1038/ncb2941 (2014).
148 Gapinske, M. *et al.* CRISPR-SKIP: programmable gene splicing with single base editors. *Genome Biol* **19**, 107, doi:10.1186/s13059-018-1482-5 (2018).

149 Hur, J. K. *et al.* Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. *Nat Biotechnol* **34**, 807-808, doi:10.1038/nbt.3596 (2016).
150 Sung, Y. H. *et al.* Knockout mice created by TALEN-mediated gene targeting. *Nat Biotechnol* **31**, 23-24, doi:10.1038/nbt.2477 (2013).
151 Sung, Y. H. *et al.* Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. *Genome Res* **24**, 125-131, doi:10.1101/gr.163394.113 (2014).
152 Tang, W., Hu, J. H. & Liu, D. R. Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nat Commun* **8**, 15939, doi:10.1038/ncomms15939 (2017).
153 Farzadfard, F. *et al.* Single-Nucleotide-Resolution Computing and Memory in Living Cells. *bioRxiv*, doi:10.1101/263657 (2018).
154 Yin, K., Gao, C. & Qiu, J. L. Progress and prospects in plant genome editing. *Nat Plants* **3**, 17107, doi:10.1038/nplants.2017.107 (2017).
155 Voytas, D. F. & Gao, C. Precision genome engineering and agriculture: opportunities and regulatory challenges. *PLoS Biol* **12**, e1001877, doi:10.1371/journal.pbio.1001877 (2014).
156 Svitashev, S. *et al.* Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA. *Plant Physiol* **169**, 931-945, doi:10.1104/pp.15.00793 (2015).
157 Endo, M., Mikami, M. & Toki, S. Biallelic Gene Targeting in Rice. *Plant Physiol* **170**, 667-677, doi:10.1104/pp.15.01663 (2016).
158 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).
159 Shimatani, Z. *et al.* Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat Biotechnol* **35**, 441-443, doi:10.1038/nbt.3833 (2017).
160 Li, C. *et al.* Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol* **19**, 59, doi:10.1186/s13059-018-1443-z (2018).
161 Li, J., Sun, Y., Du, J., Zhao, Y. & Xia, L. Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system. *Mol Plant* **In Press**, doi:10.1016/j.molp.2016.12.001 (2016).
162 Koyama, S., Ishii, K. J., Coban, C. & Akira, S. Innate immune response to viral infection. *Cytokine* **43**, 336-341, doi:10.1016/j.cyto.2008.07.009 (2008).
163 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31**, 827-832, doi:10.1038/nbt.2647 (2013).
164 Fu, Y. F. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* **31**, 822-826, doi:10.1038/nbt.2623 (2013).
165 Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology* **31**, 839-843, doi:10.1038/nbt.2673 (2013).
166 Guilinger, J. P. *et al.* Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nature methods* **11**, 429-435, doi:10.1038/nmeth.2845 (2014).

167 Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature biotechnology* **33**, 73-80, doi:10.1038/nbt.3081 (2015).
168 Singh, R., Kuscu, C., Quinlan, A., Qi, Y. J. & Adli, M. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic acids research* **43**, 677-683, doi:ARTN e118
10.1093/nar/gkv575 (2015).
169 Liang, X. Q. *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol* **208**, 44-53, doi:10.1016/j.jbiotec.2015.04.024 (2015).
170 DeWitt, M. A. *et al.* Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci Transl Med* **8**, 360ra134, doi:10.1126/scitranslmed.aaf9336 (2016).
171 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).
172 Lin, Y. C. *et al.* Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature communications* **5**, 4767, doi:10.1038/ncomms5767 (2014).
173 Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat Chem Biol* **11**, 316-318, doi:10.1038/Nchembio.1793 (2015).
174 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology* **34**, 184-191, doi:10.1038/nbt.3437 (2016).
175 Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17**, 148, doi:10.1186/s13059-016-1012-2 (2016).
176 Lin, S., Staahl, B., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife*, e04766, doi:10.7554/eLife.04766 (2014).
177 Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**, 186-194 (1998).
178 Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**, W3-W10, doi:10.1093/nar/gkw343 (2016).
179 Blankenberg, D. *et al.* Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**, 1783-1785, doi:10.1093/bioinformatics/btq281 (2010).
180 Meeker, N. D., Hutchinson, S. A., Ho, L. & Trede, N. S. Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *Biotechniques* **43**, 610, 612, 614, doi:10.2144/000112619 (2007).
181 Kwart, D., Paquet, D., Teo, S. & Tessier-Lavigne, M. Precise and efficient scarless genome editing in stem cells using CORRECT. *Nat Protoc* **12**, 329-354, doi:10.1038/nprot.2016.171 (2017).

182 Kim, J. *et al.* Structural and kinetic characterization of Escherichia coli TadA, the wobble-specific tRNA deaminase. *Biochemistry* **45**, 6407-6416, doi:10.1021/bi0522394 (2006).
183 Liang, P. *et al.* Genome-wide profiling of adenine base editor specificity by EndoV-seq. *Nat Commun* **10**, 67, doi:10.1038/s41467-018-07988-z (2019).
184 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).
185 Fritzell, K., Xu, L. D., Lagergren, J. & Ohman, M. ADARs and editing: The role of A-to-I RNA modification in cancer progression. *Semin Cell Dev Biol* **79**, 123-130, doi:10.1016/j.semcdb.2017.11.018 (2018).
186 Richardson, C. & Jasin, M. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* **405**, 697-700, doi:10.1038/35015097 (2000).
187 Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* **36**, 765-771, doi:10.1038/nbt.4192 (2018).
188 Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med* **24**, 927-930, doi:10.1038/s41591-018-0049-z (2018).
189 Ihry, R. J. *et al.* p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat Med* **24**, 939-946, doi:10.1038/s41591-018-0050-6 (2018).
190 Maizels, N. & Davis, L. Initiation of homologous recombination at DNA nicks. *Nucleic Acids Res* **46**, 6962-6973, doi:10.1093/nar/gky588 (2018).
191 Caldecott, K. W. Single-strand break repair and genetic disease. *Nat Rev Genet* **9**, 619-631, doi:10.1038/nrg2380 (2008).
192 Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet*, doi:10.1038/s41576-018-0059-1 (2018).
193 Davis, L. & Maizels, N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proc Natl Acad Sci U S A* **111**, E924-932, doi:10.1073/pnas.1400236111 (2014).
194 Davis, L., Zhang, Y. & Maizels, N. Assaying Repair at DNA Nicks. *Methods Enzymol* **601**, 71-89, doi:10.1016/bs.mie.2017.12.001 (2018).
195 Ramirez, C. L. *et al.* Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res* **40**, 5560-5568, doi:10.1093/nar/gks179 (2012).
196 Kan, Y., Ruis, B., Takasugi, T. & Hendrickson, E. A. Mechanisms of precise genome editing using oligonucleotide donors. *Genome Res* **27**, 1099-1111, doi:10.1101/gr.214775.116 (2017).
197 Davis, L. & Maizels, N. Two Distinct Pathways Support Gene Correction by Single-Stranded Donors at DNA Nicks. *Cell Rep* **17**, 1872-1881, doi:10.1016/j.celrep.2016.10.049 (2016).
198 Bothmer, A. *et al.* Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat Commun* **8**, 13905, doi:10.1038/ncomms13905 (2017).

199 Yu, D. S. *et al.* Dynamic control of Rad51 recombinase by self-association and interaction with BRCA2. *Mol Cell* **12**, 1029-1041 (2003).
200 Pellegrini, L. *et al.* Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. *Nature* **420**, 287-293, doi:10.1038/nature01230 (2002).
201 Shen, M. W. *et al.* Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature*, doi:10.1038/s41586-018-0686-x (2018).
202 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).
203 Shen, B. *et al.* Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods* **11**, 399-402, doi:10.1038/nmeth.2857 (2014).
204 Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380-1389, doi:10.1016/j.cell.2013.08.021 (2013).
205 Miyaoka, Y. *et al.* Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci Rep* **6**, 23549, doi:10.1038/srep23549 (2016).
206 Roth, T. L. *et al.* Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature* **559**, 405-409, doi:10.1038/s41586-018-0326-5 (2018).
207 Paulsen, B. S. *et al.* Ectopic expression of RAD52 and dn53BP1 improves homology-directed repair during CRISPR–Cas9 genome editing. *Nature Biomedical Engineering* **1**, 878-888, doi:10.1038/s41551-017-0145-2 (2017).
208 Canny, M. D. *et al.* Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *Nat Biotechnol* **36**, 95-102, doi:10.1038/nbt.4021 (2018).
209 San Filippo, J., Sung, P. & Klein, H. Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem* **77**, 229-257, doi:10.1146/annurev.biochem.77.061306.125255 (2008).
210 Schlacher, K. *et al.* Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* **145**, 529-542, doi:10.1016/j.cell.2011.03.041 (2011).
211 Stark, J. M. *et al.* ATP hydrolysis by mammalian RAD51 has a key role during homology-directed DNA repair. *J Biol Chem* **277**, 20185-20194, doi:10.1074/jbc.M112132200 (2002).
212 Prasad, T. K., Yeykal, C. C. & Greene, E. C. Visualizing the assembly of human Rad51 filaments on double-stranded DNA. *J Mol Biol* **363**, 713-728, doi:10.1016/j.jmb.2006.08.046 (2006).
213 Jennifer Mason, Yuen-Ling Chan, Weichselbaum, R. W. & Bishop, D. K. Non-enzymatic roles of human RAD51 at stalled replication forks. *BioRxiv* (2018).
214 Marsden, C. G. *et al.* The Tumor-Associated Variant RAD51 G151D Induces a Hyper-Recombination Phenotype. *PLoS Genet* **12**, e1006208, doi:10.1371/journal.pgen.1006208 (2016).

215 Yang, H. *et al.* BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science* **297**, 1837-1848, doi:10.1126/science.297.5588.1837 (2002).
216 Yang, H., Li, Q., Fan, J., Holloman, W. K. & Pavletich, N. P. The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA-ssDNA junction. *Nature* **433**, 653-657, doi:10.1038/nature03234 (2005).
217 Liu, J., Doty, T., Gibson, B. & Heyer, W. D. Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA. *Nat Struct Mol Biol* **17**, 1260-1262, doi:10.1038/nsmb.1904 (2010).
218 Ma, C. J., Kwon, Y., Sung, P. & Greene, E. C. Human RAD52 interactions with replication protein A and the RAD51 presynaptic complex. *J Biol Chem* **292**, 11702-11713, doi:10.1074/jbc.M117.794545 (2017).
219 Vakulskas, C. A. *et al.* A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat Med* **24**, 1216-1224, doi:10.1038/s41591-018-0137-0 (2018).
220 Richardson, C. D. *et al.* CRISPR-Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat Genet* **50**, 1132-1139, doi:10.1038/s41588-018-0174-0 (2018).
221 Tay, Y., Tan, S. M., Karreth, F. A., Lieberman, J. & Pandolfi, P. P. Characterization of dual PTEN and p53-targeting microRNAs identifies microRNA-638/Dnm2 as a two-hit oncogenic locus. *Cell Rep* **8**, 714-722, doi:10.1016/j.celrep.2014.06.064 (2014).
222 Findlay, G. M. *et al.* Accurate classification of BRCA1 variants with saturation genome editing. *Nature* **562**, 217-222, doi:10.1038/s41586-018-0461-z (2018).
223 Badran, A. H. *et al.* Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. *Nature* **533**, 58-63, doi:10.1038/nature17938 (2016).
224 Clement, K. *et al.* Analysis and comparison of genome editing using CRISPResso2. *bioRxiv*, 392217, doi:10.1101/392217 (2018).