

# UC San Diego

## UC San Diego Electronic Theses and Dissertations

### Title

Development and Characterization of Multiplexed Orthogonal Base Editing Systems: Tools to Install Distinct Point Mutations in DNA

### Permalink

<https://escholarship.org/uc/item/8zg4h5kc>

### Author

Cowan, Quinn Tadao

### Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Development and Characterization of Multiplexed Orthogonal Base Editing Systems:  
Tools to Install Distinct Point Mutations in DNA

A Dissertation submitted in partial satisfaction of the requirements  
for the degree Doctor of Philosophy

in

Biochemistry & Molecular Biophysics

by

Quinn Tadao Cowan

Committee in charge:

Professor Alexis Komor, Chair  
Professor Thomas Bussey  
Professor Tatum Simonson  
Professor Yitzhak Tor  
Professor Brian Zid

2024

Copyright

Quinn Tadao Cowan, 2024

All rights reserved.

The Dissertation of Quinn Tadao Cowan is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024



## **DEDICATION**

This work is dedicated to my parents (Darlene Ogoshi & Bob Cowan), my sister (Kelia Cowan), and the aunties and uncles who raised me (especially Kathy & Galen Yoshinaga, Carole Amos, & Phyllis Reynolds). I also want to dedicate my work to Thomas J. Schowalter; I wish you could be here with me to celebrate and see how far I've come from high school.

## TABLE OF CONTENTS

|                                                                                           |      |
|-------------------------------------------------------------------------------------------|------|
| DISSERTATION APPROVAL PAGE .....                                                          | iii  |
| DEDICATION.....                                                                           | iv   |
| TABLE OF CONTENTS.....                                                                    | v    |
| LIST OF FIGURES .....                                                                     | vii  |
| LIST OF TABLES .....                                                                      | ix   |
| LIST OF ABBREVIATIONS .....                                                               | x    |
| ACKNOWLEDGEMENTS .....                                                                    | xi   |
| VITA.....                                                                                 | xiii |
| ABSTRACT OF THE DISSERTATION.....                                                         | xiv  |
| Chapter 1: Introduction to Genome Editing .....                                           | 1    |
| Acknowledgements .....                                                                    | 2    |
| Chapter 2: Design, Optimization, and Characterization of Aptamer-Recruited Base Editors . | 3    |
| 2.1 Introduction and Significance .....                                                   | 3    |
| 2.2 Results .....                                                                         | 8    |
| 2.2.1 Design of orthogonal aptamer-base editor systems .....                              | 8    |
| 2.2.2 Initial aptamer-ABE designs have low, variable efficiencies .....                   | 9    |
| 2.2.3 Improving aptamer-ABE efficiency with evolved deaminases .....                      | 10   |
| 2.2.4 Optimization of aptamer-CBE deaminase and architecture .....                        | 17   |
| 2.2.5 Characterization of aptamer-BEs at genomic loci in human cells.....                 | 21   |
| Chapter 3: Multiplex Orthogonal Base Editor (MOBE) Systems.....                           | 28   |
| 3.1 Results .....                                                                         | 28   |
| 3.1.1 Combining aptamer-BEs to make MOBE systems .....                                    | 28   |
| 3.1.2 Quantification of co-occurring orthogonal edits by MOBEs .....                      | 36   |
| 3.1.3 A MOBE fluorescent reporter enriches for co-occurring edits.....                    | 41   |
| 3.1.4 MOBE compatibility with other nCas9 variants and cell types .....                   | 48   |

|                                                                                                           |                                                                                                  |     |
|-----------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|-----|
| 3.1.5                                                                                                     | Off-target analysis of MOBEs compared to their parental BEs.....                                 | 57  |
| 3.1.6                                                                                                     | MOBEs enable modeling of digenic diseases.....                                                   | 62  |
| 3.1.7                                                                                                     | MOBEs install of Schizophrenia-associated SNVs .....                                             | 68  |
| 3.1.8                                                                                                     | Comparison of MOBE systems to orthogonal Cas-derived BEs.....                                    | 69  |
| 3.2                                                                                                       | Discussion .....                                                                                 | 73  |
| 3.3                                                                                                       | Methods.....                                                                                     | 76  |
|                                                                                                           | Acknowledgements .....                                                                           | 83  |
| Chapter 4: Utilizing Base Editors to Generate Human Cell Lines Containing Single Nucleotide Variants..... |                                                                                                  | 85  |
| 4.1                                                                                                       | Introduction.....                                                                                | 85  |
| 4.1.1                                                                                                     | Strategic planning .....                                                                         | 88  |
| 4.2                                                                                                       | Basic Protocol 1: Design and Production of Plasmids for Base Editing Experiments .               | 90  |
| 4.3                                                                                                       | Basic Protocol 2: Transfecting Adherent Cells and Harvesting Genomic DNA.....                    | 103 |
| 4.4                                                                                                       | Basic Protocol 3: Genotyping Harvested Cells Using Sanger Sequencing.....                        | 108 |
| 4.5                                                                                                       | Alternate Protocol 1: Next-Generation Sequencing to Quantify Base Editing.....                   | 114 |
| 4.6                                                                                                       | Basic Protocol 4: Single Cell Isolation of Base Edited Cells Using FACS.....                     | 120 |
| 4.7                                                                                                       | Alternate Protocol 2: Single Cell Isolation of Base Edited Cells Using Dilution Plating          | 123 |
| 4.8                                                                                                       | Basic Protocol 5: Clonal Expansion to Generate Isogenic Cell Lines and Genotyping of Clones..... | 126 |
| 4.9                                                                                                       | Commentary .....                                                                                 | 131 |
| 4.9.1                                                                                                     | Background information: .....                                                                    | 131 |
| 4.9.2                                                                                                     | Critical parameters:.....                                                                        | 134 |
| 4.9.3                                                                                                     | Troubleshooting: .....                                                                           | 139 |
| 4.9.4                                                                                                     | Understanding results: .....                                                                     | 143 |
| 4.9.5                                                                                                     | Time considerations:.....                                                                        | 144 |
|                                                                                                           | Acknowledgements .....                                                                           | 145 |
| REFERENCES .....                                                                                          |                                                                                                  | 146 |

## LIST OF FIGURES

|                                                                                                                                                                                  |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 1.1: Overview of DNA base editing technologies. ....                                                                                                                      | 1   |
| Figure 2.1: Schematic overview of multiplexed orthogonal base editing. ....                                                                                                      | 7   |
| Figure 2.2: Engineering of an ABE aptamer system. ....                                                                                                                           | 13  |
| Figure 2.3: Engineering of a CBE aptamer system. ....                                                                                                                            | 19  |
| Figure 2.4: Characterization of optimized CBE and ABE aptamer systems. ....                                                                                                      | 24  |
| Figure 3.1: Editing efficiencies of multiplexed orthogonal base editor (MOBE) systems compared to multiplexing CBE and ABE. ....                                                 | 30  |
| Figure 3.2: On-target and crosstalk editing efficiencies of MOBEs and parental BEs when targeted to two protospacers at distinct chromosomes. ....                               | 32  |
| Figure 3.3: On-target and crosstalk editing efficiencies of MOBEs and parental BEs when targeted to two protospacers within the same locus. ....                                 | 35  |
| Figure 3.4: Quantification of indel rates for parental and MOBE systems. ....                                                                                                    | 36  |
| Figure 3.5: Quantification of co-occurring orthogonal edits. ....                                                                                                                | 38  |
| Figure 3.6: On-target and crosstalk editing efficiencies of MOBEs when performing multiplexed editing followed by FACS enrichment for episomal edits. ....                       | 43  |
| Figure 3.7: On-target and crosstalk editing efficiencies of MOBEs when targeted to two protospacers at distinct chromosomes followed by FACS enrichment for episomal edits. .... | 44  |
| Figure 3.8: On-target and crosstalk editing efficiencies of MOBEs when targeted to two protospacers within the same locus followed by FACS enrichment for episomal edits. ....   | 46  |
| Figure 3.9: MOBEs can be used in HeLa cells. ....                                                                                                                                | 52  |
| Figure 3.10: MOBEs can be used in SH-SY5Y cells. ....                                                                                                                            | 55  |
| Figure 3.11: Evaluation of gRNA-independent off-target DNA editing by MOBEs compared to parental systems. ....                                                                   | 58  |
| Figure 3.12: MOBEs enable the installation of orthogonal point mutations known to cause Kallmann syndrome (OLI368). ....                                                         | 66  |
| Figure 3.13: MOBEs enable the installation of orthogonal point mutations known to cause isolated anencephaly (OLI736). ....                                                      | 67  |
| Figure 3.14: MOBEs install variants associated with schizophrenia. ....                                                                                                          | 69  |
| Figure 4.1: Generating isogenic cell lines harboring a target SNV, using the MUTYH V270A variant as an example. ....                                                             | 88  |
| Figure 4.2: Overview of Basic Protocol 1. ....                                                                                                                                   | 91  |
| Figure 4.3: Representative images of successful and unsuccessful transfection experiments (Basic Protocol 2, steps 1-9). ....                                                    | 105 |
| Figure 4.4: Overview of gDNA amplification (Basic Protocol 2, steps 10-14). ....                                                                                                 | 109 |
| Figure 4.5: Representative Sanger sequencing traces from genotyping of cells in bulk and after clonal expansion (Basic Protocols 3 and 5). ....                                  | 114 |

|                                                                                                                                                                  |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 4.6: Representative images from dilution plating method (Alternate Protocol 2).....                                                                       | 125 |
| Figure 4.7: Potential pitfalls in using Base Editing to generate isogenic cell lines.....                                                                        | 131 |
| Figure 4.8: A conceptual model of base editing.....                                                                                                              | 134 |
|                                                                                                                                                                  |     |
| Supplementary Figure 2.1: Initial testing of ABE aptamer systems with the wtTadA-TadA7.10 deaminase. ....                                                        | 14  |
| Supplementary Figure 2.2: Optimization of ABE aptamer systems employing TadA8 deaminases. ....                                                                   | 16  |
| Supplementary Figure 2.3: Initial testing of CBE aptamer systems with the ancAPOBEC, evoAPOBEC1, and RrA3F deaminases. ....                                      | 20  |
| Supplementary Figure 2.4: Architecture optimization of CBE aptamer systems employing the evoAPOBEC1 deaminase. ....                                              | 21  |
| Supplementary Figure 2.5: Characterization of optimized CBE and ABE aptamer systems. ....                                                                        | 25  |
| Supplementary Figure 3.1: Genotype analysis of multiplexed base editing at the <i>HEK3</i> locus. ....                                                           | 39  |
| Supplementary Figure 3.2: Genotype analysis of multiplexed base editing at the <i>EMX1</i> locus. ....                                                           | 40  |
| Supplementary Figure 3.3: Genotype analysis of multiplexed base editing at the <i>RNF2</i> locus. ....                                                           | 41  |
| Supplementary Figure 3.4: Orthogonality scores, increases in on-target editing efficiencies, and indel rates after enrichment with the fluorescent reporter..... | 47  |
| Supplementary Figure 3.5: MOBEs are compatible with other Sp-nCas9 variants. ....                                                                                | 49  |
| Supplementary Figure 3.6: Evaluation of gRNA-dependent off-target DNA editing by MOBEs compared to parental systems.....                                         | 60  |
| Supplementary Figure 3.7: Evaluation of off-target RNA editing by MOBEs compared to parental systems.....                                                        | 62  |
| Supplementary Figure 3.8: Comparison of MOBE systems to orthogonal Cas-derived BEs. ....                                                                         | 72  |
| Supplementary Figure 3.9: Haplotype calling schematic for co-occurring edit analysis by the MOBE <sub>nto</sub> software.....                                    | 82  |
| Supplementary Figure 3.10: Example of gating scheme used for MOBE fluorescent enrichment. ....                                                                   | 83  |

## LIST OF TABLES

|                                                                                                                   |         |
|-------------------------------------------------------------------------------------------------------------------|---------|
| Table 4.1: Primers and additional sequences used to generate custom gRNAs or amplify the target locus. ....       | 93      |
| Table 4.2: Suggested base editors designed to maximize on-target editing or minimize off-target (OT) editing..... | 94      |
| Table 4.3: Next-generation sequencing barcodes or round 2 PCR primers. ....                                       | 115-116 |

## LIST OF ABBREVIATIONS

|        |                                                            |
|--------|------------------------------------------------------------|
| CRISPR | Clustered Regularly Interspersed Short Palindromic Repeats |
| Cas9   | CRISPR-associated protein 9                                |
| nCas9  | nickase Cas9                                               |
| gRNA   | guide RNA                                                  |
| PAM    | protospacer adjacent motif                                 |
| DSB    | double-strand break                                        |
| SNV    | single nucleotide variant                                  |
| VUS    | variant of uncertain significance                          |
| BE     | base editor                                                |
| CBE    | cytosine base editor                                       |
| ABE    | adenine base editor                                        |
| MOBE   | multiplexed orthogonal base editor                         |
| NGS    | next-generation sequencing                                 |
| OT     | off-target                                                 |
| nt     | nucleotide                                                 |

## ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge Dr. Alexis C. Komor for her support as a mentor and the chair of my committee. Alexis always cultivated a nurturing lab environment and encouraged me to push my own limits. She entertained my wild ideas while also drawing me in to focus on tractable goals. Alexis taught me that perfection can hinder progress and how to avoid being obsessed with the little things. Her guidance has been invaluable and none of this work would be possible without her.

I am eternally grateful to my parents (Darlene Ogoshi and Bob Cowan) for the opportunities they provided me. From a young age, my parents fostered a respect for the natural world and people of all cultures. They provided me the best education at all levels that allowed me to explore my interest in science. I would not have been able to complete a PhD without my parents. I want to thank my sister, Kelia Cowan, for being the best sister anyone could ask for. To my partner, Nichelle Camden, you have been the very best, most supporting person to share my life with. To everyone above, my family and my friends: thank you all so much I could not have done this without you; I love you.

I would also like to acknowledge all of my lab mates, the members of my committee, and members of the Chemistry and Biochemistry community (especially the members of the Biochemistry and Molecular Biophysics track) for their support. I want to thank Betsy Komives for the training and financial support through the Molecular Biophysics Training Grant (NIH T32). I also want to thank Brian Zid and Galia Debelouchina for their generous use of their research equipment. I would like to acknowledge my undergraduate advisor Dr. Irene Chiolo and my advisor during my research associate position Dr. W. Matt Michael who helped prepared me for success in graduate school.

Chapter 1, in part, is a reprint of the material as it appears in *Gene and Genome Editing*. Gu, Sifeng; Bodai, Zsolt; Cowan, Quinn T.; Komor, Alexis C., *Gene and Genome Editing*, 2021. The dissertation author was a contributing author on this paper.



Chapters 2-3, in part, are a reprint of the material as it appears in *Nature Biotechnology*. Cowan, Quinn T.; Gu, Sifeng; Gu, Wanjun; Ranzau, Brodie L.; Simonson, Tatum S.; Komor, Alexis C., *Nature Biotechnology*, 2024. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, is a reprint of the material as it appears in *Current Protocols in Molecular Biology*. Vasquez, Carlos A.; Cowan, Quinn T.; Komor, Alexis C., *Current Protocols in Molecular Biology*, 2020. The dissertation author was co-first author of this paper.

## VITA

- 2017 Bachelor of Science in Biochemistry, University of Southern California
- 2020 Master of Science in Chemistry, University of California San Diego
- 2024 Doctor of Philosophy in Biochemistry and Molecular Biophysics, University of California San Diego

## PUBLICATIONS

- Cowan, Q. T.**, Gu, S., Gu, W., Ranzau, B. L., Simonson, T. S., & Komor, A. C. Development and Characterization of Multiplexed Orthogonal Base Editing Systems. *Nature Biotechnology* **2024**, <https://doi.org/10.1038/s41587-024-02240-0>.
- Lawrence, E. S., Gu, W., Bohlender, R. J., Anza-Ramirez, C., Cole, A. M., Yu, J. J., Hu, H., Heinrich, E. C., O'Brien, K. A., Vasquez, C. A., **Cowan, Q. T.**, ... Simonson, T. S. Functional *EPAS1* / *HIF2A* missense variant is associated with hematocrit in Andean highlanders. *Science Advances* **2024**, *10*(6), eadj5661.
- Gu, S., Bodai, Z., **Cowan, Q. T.** & Komor, A. C. Base Editors: Expanding the Types of DNA Damage Products Harnessed for Genome Editing. *Gene and Genome Editing* **2021**, *1*, 100005.
- Vasquez, C. A.,\* **Cowan, Q. T.**\* & Komor, A. C. Base Editing in Human Cells to Produce Single-Nucleotide-Variant Clonal Cell Lines. *Current Protocols in Molecular Biology* **2020**, *133*, e129.
- Kim, A., Montales, K., Ruis, K., Senebandith, H., Gasparyan, H., **Cowan, Q.**, & Michael, W. M. Biochemical analysis of TOPBP1 oligomerization. *DNA Repair* **2020**, *96*.

(\* Equal contributions)

## **ABSTRACT OF THE DISSERTATION**

Development and Characterization of Multiplexed Orthogonal Base Editing Systems: Tools to Install Distinct Point Mutations in DNA

by

Quinn Tadao Cowan

Doctor of Philosophy in Biochemistry & Molecular Biophysics

University of California San Diego, 2024

Professor Alexis Komor, Chair

Base editors (BEs) enable the programmable installation of point mutations with high efficiency and precision while avoiding the use of double-strand breaks. Two major classes of BEs exist that convert C•G base pairs to T•A (CBEs) and A•T base pairs to G•C (ABEs). Although these current tools are in theory well-suited for multiplexing, in practice researchers are limited to multiplexing with a single type of BE due to the covalent fusion of the DNA modifying enzyme to the Cas9 effector causing “gRNA crosstalk”, or non-orthogonal editing. Here, we engineer and characterize aptamer-derived CBEs and ABEs that can be multiplexed with each other orthogonally. We generated four multiplexed orthogonal base

editor (MOBE) systems that enable rates of co-occurring edits of up to 7.1% without enrichment or selection strategies. We also develop a simple fluorescent enrichment strategy, which increased co-occurring edit rates up to 24.8% in HEK293T cells. We show that these systems are compatible with expanded-PAM and high-fidelity Cas9 variants, function well in multiple cell types, have equivalent or reduced off-target propensities compared to their parental systems, and can be used to model disease-relevant point mutation combinations. These MOBE systems are a valuable addition to the genome editing toolbox, and will aid significantly in efforts to study/model or correct combinations of genetic variants, such as those found in haplotypes, polygenic disorders, and driver-passenger mutation pairs.

## Chapter 1: Introduction to Genome Editing

In 1665, Robert Hooke first described the cell in the publication *Micrographia*. This direct observation via microscope would pave the way for countless discoveries in biology. Often in science, seeing is believing. Fast forward almost 300 years and Rosalind Franklin observed a different type of image, Photograph 51, that would forever reshape our view of biological systems. This X-ray diffraction pattern was a key piece of data in describing the structure and function of deoxyribose nucleic acid or DNA: the so-called molecule of life. This revolutionary discovery laid the framework for molecular biology as a whole. Since then, incredible leaps forward have been made in the study of DNA that include sequencing the human genome, a major milestone towards precision medicine. Recent progress in sequencing technologies have made the identification of genetic variants more cost-effective and accessible than ever before.

Point mutations, or single nucleotide variants (SNVs) account for over 96% of currently identified human genetic variation.<sup>1</sup> However, the majority of observed genetic variants (over 99%) currently lack a clinical interpretation.<sup>2</sup> Therefore, the ability to precisely and efficiently introduce SNVs into the genome of living cells holds great potential for the treatment of genetic diseases, as well as for modelling the impact that currently unclassified genetic variants have on cellular function and human health. The development of CRISPR-associated (Cas) nucleases as genome editing agents has led to a revolution and paved the way for the development of next generation genome editing agents.<sup>3,4</sup>

Originally functioning as adaptive immune systems in bacteria and archaea, Cas nucleases have been repurposed to perform genome editing in a wide variety of cells and organisms. Their ability to create double-stranded DNA breaks (DSBs) with high efficiencies in a programmable, RNA-guided manner has enabled their rapid adoption by the genome editing community at large. To direct Cas nucleases to a genomic locus of interest, researchers simply change the spacer sequence of the guide RNA (gRNA). This then facilitates Cas:gRNA:DNA

binding via Watson-Crick-Franklin base pairing rules between the spacer sequence of the gRNA and the protospacer sequence of the target locus. The protospacer must also be directly next to a protospacer adjacent motif (PAM), a DNA sequence required by the Cas protein. Once bound to its target DNA sequence, the Cas nuclease will introduce a DSB (in the case of traditional genome editing).

However, the introduction of a DSB is only the first step of the genome editing process. The next step involves the cellular response to this DNA damage product. In fact, all current genome editing technologies employ a two-step process to perform their function: 1) deliberately introduce a DNA lesion at a user-programmed location in the genome, followed by 2) manipulate or rely on the endogenous DNA damage repair pathways to process the lesion into a desired editing outcome. In the case of DSB repair (which has been extensively studied and reviewed<sup>5,6</sup>), multiple pathways compete with each other to process the lesion, resulting in a mixture of outcomes.

In brief, processing by the non-homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ) pathways results in insertions, deletions, translocations, and other rearrangements at the site of the break, which are useful for gene disruption.<sup>7,8</sup> Alternatively, precise DNA edits can be installed via the homology-directed repair (HDR) pathway through the use of an exogenously supplied DNA donor template. This donor template is designed to contain the desired sequence modification flanked by homology arms; the HDR machinery will use this as a template to repair the DSB, in the process incorporating the modified sequence into the genome.<sup>9</sup> HDR-mediated genome editing can be used to install any type of sequence modification of interest, including SNVs, but it is highly cell cycle dependent (and only active during G2 and S phases), and usually outcompeted by end joining processes for the repair of DSBs.<sup>10,11</sup>

Base editing is a recent technology specifically developed to combat these challenges inherent to DSB-reliant genome editing methods.<sup>12,13</sup> Base editors (BEs) facilitate the efficient

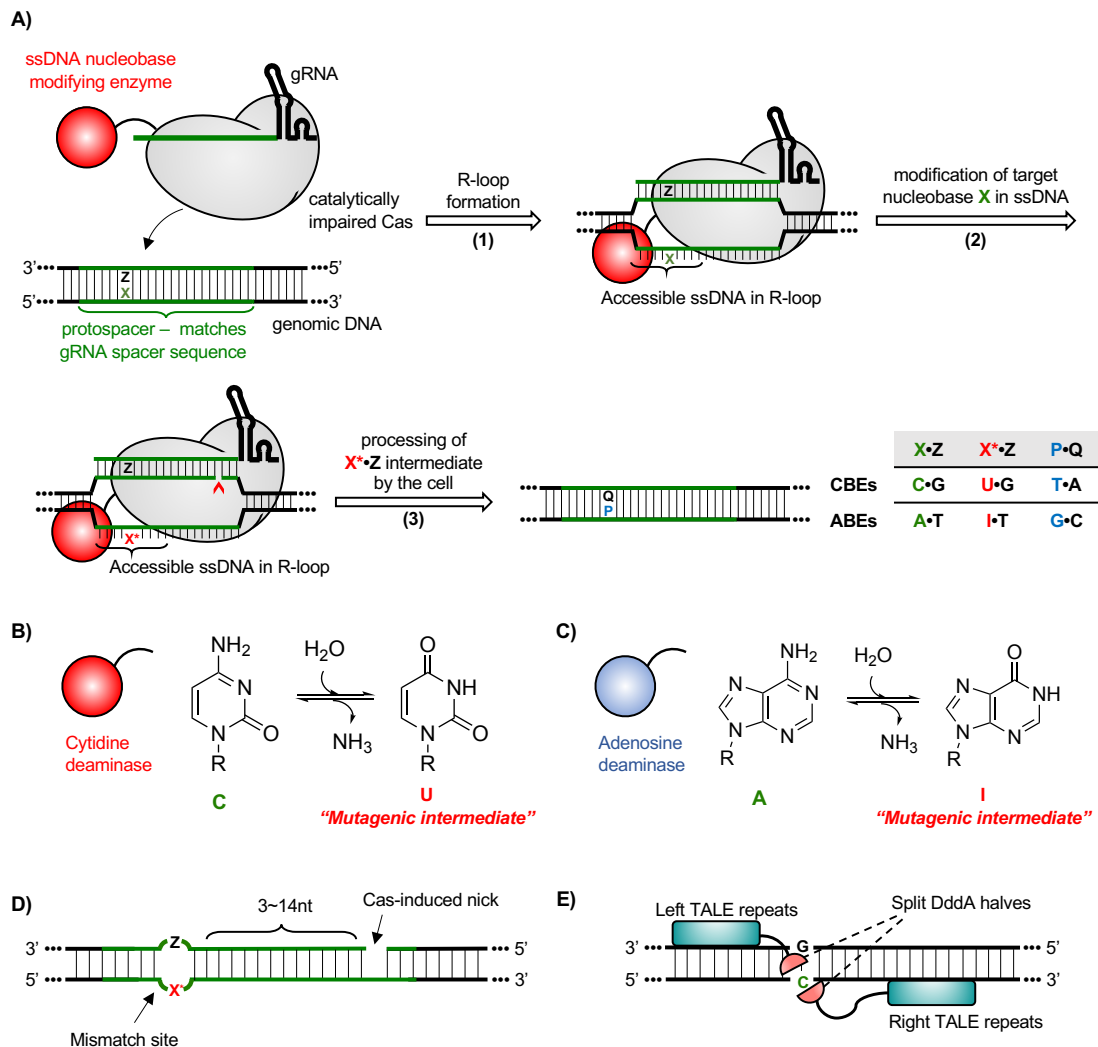
and precise introduction of SNVs using non-DSB DNA damage intermediates. DNA base editors are generally protein fusion constructs consisting of a catalytically impaired Cas nuclease, usually nickase Cas9 (nCas9), linked to a single-stranded DNA (ssDNA)-specific nucleobase modifying enzyme (Figure 1.1A). Upon localization of the Cas protein to its target DNA sequence, hybridization of the gRNA to its complementary DNA sequence displaces and exposes a small stretch of ssDNA at the PAM-distal region of the non-complementary strand. This exposure allows the ssDNA modifying enzyme to gain access to and chemically modify its target base within this ssDNA window. To date, by harnessing hydro-deamination chemistry performed by deaminase enzymes, two major classes of base editors have been developed. Cytosine base editors (CBEs) catalyze the transition mutation of C•G base pairs to T•A base pairs via the deamination of deoxycytidine to deoxyuridine as the mutagenic intermediate (Figure 1.1B). Adenine base editors (ABEs) catalyze the transition mutation of A•T base pairs to G•C base pairs via the deamination of deoxyadenosine to form deoxyinosine as the mutagenic intermediate (Figure 1.1C).

Because both deoxyuridine and deoxyinosine are noncanonical DNA nucleotides, they are recognized as DNA lesions by the cell and repaired by cellular DNA repair mechanisms distinct from those utilized in DSB-reliant methods. The Cas9-induced nick serves to bias the cellular repair mechanisms to preferentially replace this strand and use the mutagenic intermediate as a template for repair. Base editors therefore use unique types of DNA damage: a mutagenic, noncanonical DNA nucleotide on one strand, and a DNA backbone nick on the other strand, 5' upstream from the modified nucleotide (Figure 1.1D). As deoxyuridine has the base-pairing properties of deoxythymidine, this base editor was engineered to introduce C•G to T•A base pair conversions in a programmable manner.<sup>12</sup>

Unlike cytidine deaminases, there are no naturally occurring adenosine deaminases capable of using DNA as a substrate. The TadA\* (mutant TadA) enzyme used to produce the

final ABE construct (ABE7.10) was developed from seven rounds of directed evolution on the *E. coli* TadA enzyme that naturally deaminates adenosines in tRNA. Later studies reported ABE8 variants (engineered through additional rounds of directed evolution), which demonstrate enhanced editing activity and compatibility with Cas9 and Cas12a homologs.<sup>14,15</sup> ABEs produce extremely precise editing outcomes, with no A•T to non-G•C conversions and minimal indel formation.





**Figure 1.1: Overview of DNA base editing technologies.** A) DNA base editors are generally protein fusion constructs consisting of a catalytically impaired Cas nuclease linked to a ssDNA-specific nucleobase modifying enzyme. (1) Binding of a base editor to its target locus forms a DNA/RNA hybrid R-loop, which exposes the target nucleotide X in a stretch of ssDNA outside of the Cas protein. (2) The ssDNA nucleobase modifying enzyme now chemically modifies its target base X into X\*, a noncanonical nucleotide that serves as a mutagenic intermediate. (3) X\* is further processed by cellular repair or replication machinery, resulting in its conversion to N. Note that the nucleotide symbols used here are not IUPAC codes. For CBEs, X•Z, X\*•Z and N•M are C•G, U•G, and T•A, respectively. For ABEs, they are A•T, I•T, and G•C, respectively. B) The ssDNA nucleobase modifying enzymes used in CBEs are cytidine deaminases, which chemically modify deoxycytidine to deoxyuridine via hydro-deamination chemistry. C) Adenosine deaminases are used in ABEs, which chemically modify deoxyadenosine to deoxyinosine via hydro-deamination chemistry. D) The unique DNA damage intermediate harnessed for base editing, featuring a modified nucleotide(X\*)-containing mismatch and a nick located 3 to 14 nucleotides away in the 5' direction from the mismatched site on the unmodified strand.

## **Acknowledgements**

Chapter 1, in part, is a reprint of the material as it appears in Gene and Genome Editing. Gu, Sifeng; Bodai, Zsolt; Cowan, Quinn T.; Komor, Alexis C., *Gene and Genome Editing*, 2021.<sup>16</sup> The dissertation author was a contributing author on this paper.

## **Chapter 2: Design, Optimization, and Characterization of Aptamer-Recruited Base Editors**

### **2.1 Introduction and Significance**

Advances in next-generation sequencing (NGS) have made the detection of human genetic variants increasingly routine. The identification of clinically actionable and pathogenic mutations has revolutionized the field of precision medicine, but unfortunately these variants represent a small fraction of identified human genetic variants.<sup>17,18</sup> Specifically, while 787 million genetic variants have been reported in the Genome Aggregation Database,<sup>19</sup> less than 0.5% have a clinical interpretation in ClinVar.<sup>20</sup> Furthermore, 99% of identified variants are rare or population-specific, causing the prediction of their functional impact using genome-wide association studies (GWAS) or computational methods to be particularly challenging.<sup>21</sup> Additionally, the co-occurrence of variants further convolutes their interpretation using existing computational methods. In particular, different background genetic variation (such as those found in patient-derived cell lines) can lead to misclassification of risk-association or conflicting interpretations on ClinVar.<sup>22</sup> In these settings, GWAS fail to accurately determine whether individual variants are pathogenic or simply co-inherited as passenger mutations.

This presents a pressing unmet need in the field of genome editing to facilitate modeling of combinations of genetic variants within otherwise isogenic backgrounds. Functional investigation of specific variant combinations would enable researchers to deconvolute the clinical contributions of individual variants observed in polygenic disorders, combinations of variants of uncertain significance (VUS), variants inherited in haplotype blocks, or variants in modifier genes within the same biological pathway.<sup>23–25</sup> As single nucleotide variants (SNVs) account for 96% of observed human genetic variation, the development of programmable tools that can efficiently multiplex the installation of point mutations would be transformative for modelling and functionally characterizing genetic variants.

Traditional genome editing methods utilize programmable DNA double-strand breaks (DSBs) to precisely introduce SNVs in human cells. These tools, such as the wild-type CRISPR-Cas9 system,<sup>26</sup> additionally rely on the homology-directed repair (HDR) pathway to incorporate desired modifications into the genome using an exogenously supplied donor DNA template.<sup>27</sup> However, end-joining repair pathways, which are more ubiquitous and efficient than HDR, compete to process DSBs, resulting in the introduction of insertions and deletions (indels) of bases at the site of the DSB.<sup>28</sup> DSB-reliant genome editing methods therefore suffer from high rates of unwanted gene alterations (indels) and low efficiencies of the desired modification, particularly when installing point mutations.<sup>29</sup> When attempting to multiplex point mutation introduction with Cas9, the low precision of DSB-reliant technologies is exacerbated, as success rates will theoretically scale as the product of the individual editing efficiencies of each desired edit. Additionally, the incidences of translocations (when multiplexing at distinct chromosomes), large-scale deletions/inversions (when multiplexing within the same chromosome),<sup>30,31</sup> chromosomal aberrations,<sup>32</sup> and/or p53-mediated apoptosis<sup>33</sup> increase when introducing multiple DSBs. Overall, the enhanced cytotoxicity and high rates of undesired editing outcomes make the use of DSB-reliant tools impractical for multiplexing SNV introduction.<sup>34,35</sup>

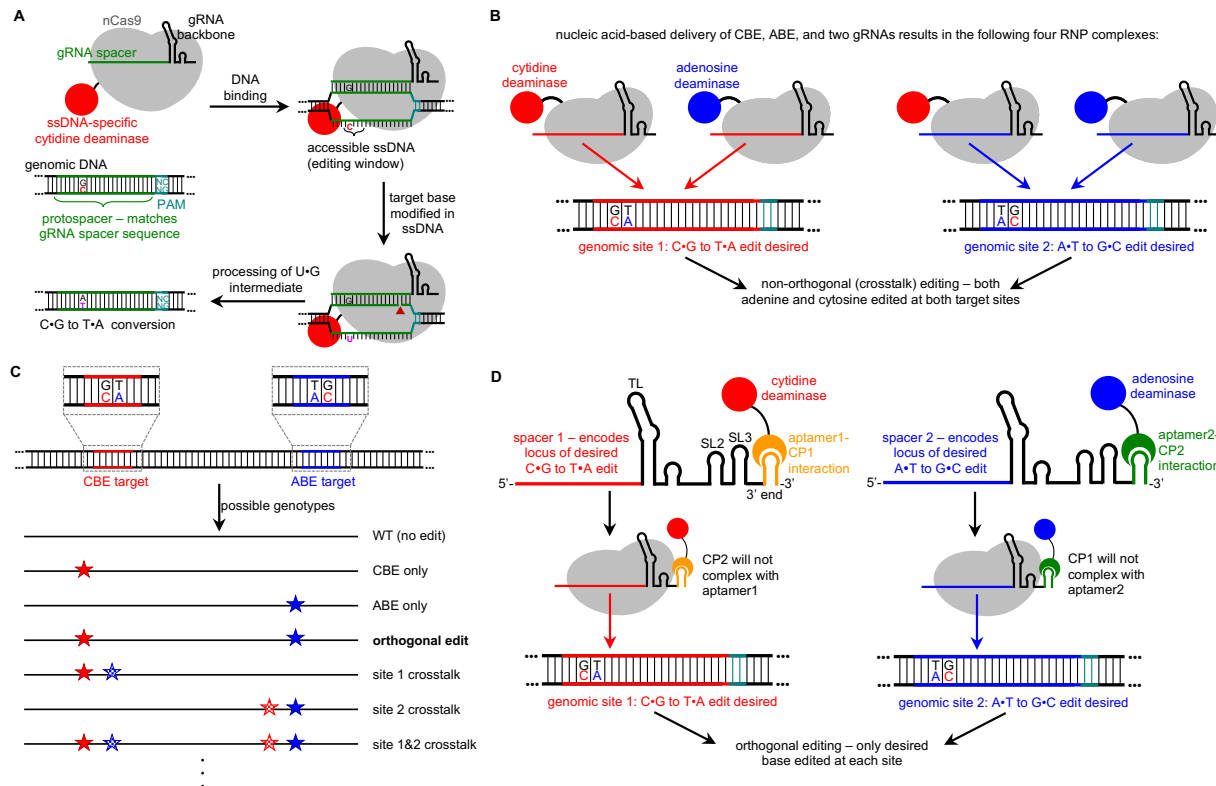
“Nontraditional” precision genome editing tools are those that avoid the introduction of DSBs, and instead use alternative DNA damage products as intermediates.<sup>12,13,28,36</sup> Specifically, base editors (BEs) are comprised of a catalytically impaired Cas9 nickase (nCas9) covalently tethered to a single-stranded DNA (ssDNA) modifying enzyme (Figure 2.1A). The Cas9 enzyme complexes with a guide RNA (gRNA) molecule, which directs the fusion protein to the target site (called the protospacer) via base-pairing rules. The protospacer must be directly next to a protospacer adjacent motif (PAM) for the Cas9:gRNA complex to bind. The Cas9:gRNA:DNA ternary complex is an R-loop, in which one DNA strand is base-paired with the gRNA, and the other is single-stranded and lacks a complement.<sup>37</sup> This in turn exposes a small (~5 nucleotide) “window” of accessible ssDNA (on the strand not bound by the gRNA) where the ssDNA

modifying enzyme directly chemically modifies target nucleotides within this window (Figure 2.1A). Two major classes of base editors have been developed that use cytosine and adenine deamination chemistries to catalyze the conversion of C•G base pairs to T•A (CBEs), and A•T base pairs to G•C (ABEs), respectively.<sup>13,38</sup> These two types of transition point mutations account for 61% of observed SNVs,<sup>39</sup> allowing for functional investigation of the majority of genetic variants using current BEs. BEs are uniquely situated to enable multiplexed genome editing as they have high on-target efficiencies (up to 90% simultaneous edits at 3 endogenous loci by a single CBE variant has been reported) and low byproduct formation (including indels and C•G to non-T•A edits by CBEs).<sup>40</sup> However, multiplexed base editing has only been demonstrated previously in mammalian cells when using only a single BE (either ABE or CBE). Multiplexing CBEs and ABEs together currently faces several challenges.

BEs that employ the *Streptococcus pyogenes* (Sp) Cas9 homolog are the most widely-used BE variants due to their high efficiencies, narrow editing windows (which reduces unwanted bystander editing), and flexible PAM requirements (there are engineered SpCas9s with NG and NR/Y PAMs; this allows for facile positioning of the target base in the center of the editing window).<sup>41,42</sup> Attempts to multiplex Sp-derived CBEs and ABEs would result in “gRNA crosstalk,” in which all SpCas9 components would complex with all Sp-gRNAs, causing both C•G to T•A and A•T to G•C editing at all targeted genomic loci (Figure 2.1B-C). This crosstalk can in theory be partially mitigated by pre-complexing each base editor protein with its appropriate gRNA in vitro, followed by direct ribonucleoprotein (RNP) delivery into cells. However, intracellular gRNA exchange, which would result in gRNA crosstalk, has been reported to be quite efficient.<sup>43,44</sup> Furthermore, the difficulty in expressing and purifying base editor protein combined with a lack of commercial sources of base editor protein make this strategy inaccessible to most laboratories. It is currently possible to orthogonally multiplex ABEs and CBEs by using BEs that employ Cas9 orthologs that utilize distinct gRNA backbones, such as SpCas9 and *Staphylococcus aureus* (Sa) Cas9. However, BEs derived from SaCas9 have

more restrictive PAM requirements (NNNRRT for the most relaxed KKH variant) and much wider editing windows (resulting in bystander editing), which severely limits their utility for disease modelling and therapeutic SNV correction.<sup>45–47</sup> Alternatively, a strategy in which the ssDNA modifying enzymes are directly recruited to their respective gRNAs (which encode the loci to be edited) would offer a modular system for multiplexed orthogonal base editing in which both BEs use the advantageous SpCas9 homolog (Figure 2.1D).

Here, we utilize RNA aptamers to engineer four multiplexed orthogonal base editor (MOBE) systems, which enable the simultaneous introduction of C•G to T•A SNVs and A•T to G•C SNVs at distinct protospacers with minimal crosstalk. These systems can be utilized with different Sp-nCas9 variants, such as expanded PAM variants (nCas9-NG and SpRY-nCas9) and high-fidelity variants (HiFi-nCas9), and are demonstrated across multiple cell types. Our MOBE systems are 19-fold more orthogonal on average (as assessed by comparing on-target editing to crosstalk editing) than when multiplexing with current CBE and ABE systems. Additionally, without any selection or enrichment strategies, our MOBE systems achieve co-occurring orthogonal editing rates of up to 7.1% when multiplexing with two protospacers within the same sequencing amplicon. We additionally developed a fluorescence-based reporter plasmid that facilitates the enrichment of cells with high MOBE activity. When using our MOBE systems with this enrichment strategy, we have achieved up to 24.8% co-occurring orthogonal edits with crosstalk rates of only 1.0%. This demonstrates the ability for base editors to create orthogonal point mutations at distinct genomic sites in human cells.



**Figure 2.1: Schematic overview of multiplexed orthogonal base editing.** A) Overview of base editing, with the cytosine base editor (CBE) shown as an example. Base editors are comprised of a ssDNA modifying enzyme (red, a cytidine deaminase in CBEs) tethered to a catalytically impaired Cas9 (nCas9, grey). Binding to genomic DNA is facilitated by base-pairing between the spacer region of the gRNA (green) and a complementary protospacer sequence (green) in the DNA. The protospacer must also be next to a PAM motif (cyan, the PAM sequence is NG for all experiments done in this study). DNA binding by the Cas9:gRNA complex forms an R-loop, which exposes a ~5 nucleotide window to the ssDNA cytidine deaminase. Any cytosines within this window are deaminated to uracils, and the opposite DNA strand is nicked by nCas9. The U•G intermediate is subsequently processed by the cell to produce an overall C•G to T•A base pair conversion. The adenine base editor (ABE) works analogously but facilitates A•T to G•C base pair conversions via inosine-containing intermediates. (B) When attempting to multiplex a CBE and an ABE, gRNA crosstalk will occur. Delivery of the BEs and gRNAs via nucleic acid-based methods (plasmid-encoded or mRNA and synthetic gRNA) results in both the CBE and ABE complexing with both gRNAs in situ. This subsequently results in both C•G to T•A and A•T to G•C editing at both protospacers. (C) Schematic representation of possible genotypes when attempting to multiplex C•G to T•A and A•T to G•C editing at two neighboring protospacers. (D) Schematic representation of an aptamer-based system for multiplexed orthogonal base editing. A single nCas9 variant is used for both editors, and the cytidine and adenosine deaminase enzymes are recruited directly to their respective gRNAs via orthogonal aptamer-coat protein interactions. As a result, only C•G to T•A editing occurs at the desired CBE target protospacer, and only A•T to G•C editing occurs at the desired ABE target protospacer.

## 2.2 Results

### 2.2.1 Design of orthogonal aptamer-base editor systems

To engineer a system that enables orthogonal, multiplexed C•G to T•A and A•T to G•C point mutation introduction at distinct genomic loci using only the SpCas9 homolog, we utilized RNA aptamer technologies (Figure 2.1D). Recruitment of the DNA modifier (the cytidine or adenosine deaminase) directly to its gRNA (encoding the genomic locus to be edited) via an aptamer-binding protein interaction provides a modular system that avoids crosstalk. Furthermore, it theoretically allows for as many simultaneous orthogonal nucleotide conversions as there are orthogonal RNA aptamer-binding protein systems (currently, there are at least four well-characterized systems, with more possible through directed evolution).<sup>48–51</sup> Finally, RNA aptamer-mediated effector recruitment has been used successfully with Cas9 systems previously; this includes orthogonal fluorescent labelling,<sup>49,50</sup> transcriptional reprogramming,<sup>48,52</sup> lncRNA recruitment,<sup>53</sup> and random mutagenesis with a hyperactive cytidine deaminase.<sup>54</sup>

Structures of the Cas9:gRNA complex show that the tetraloop (TL), stem-loop 2 (SL2), and 3'-end of the gRNA physically protrude from the complex and do not contact any Cas9 residues (Supplementary Figure 2.1A).<sup>37</sup> As such, RNA sequence additions are tolerated in these regions without affecting gRNA stability, Cas9:gRNA complex formation, or Cas9:gRNA-DNA binding.<sup>53</sup> Therefore, orthogonal recruitment of each nucleobase-specific modifying enzyme to its desired target site can be enabled by embedding RNA aptamers within the gRNA at these locations, and fusing the complimentary coat protein (CP) to either the cytidine or adenosine ssDNA modifier (Figure 2.1D). We initially focused on the most commonly used orthogonal aptamer systems: the MS2 bacteriophage system and the *Pseudomonas* phage PP7 system. The MS2 coat protein (MCP) and PP7 coat protein (PCP) are small proteins (129- and 127-amino acids, respectively) that bind as dimers to their respective RNA stem-loops<sup>55,56</sup> and do not interact with non-cognate binding proteins.<sup>48,49</sup> As cytidine deaminase-derived editing



systems have been recently developed using the MS2 aptamer,<sup>54,57</sup> we focused our engineering efforts on the development of an ABE-aptamer system.

### 2.2.2 Initial aptamer-ABE designs have low, variable efficiencies

We generated six ABE aptamer systems with PCP fused to either the N- or C-terminus of wild type (wt)TadA-TadA7.10 (the heterodimeric seventh-generation deoxyadenosine deaminase construct)<sup>13</sup> via a 93-amino acid (aa) flexible linker (Supplementary Figure 2.1B), and the PP7 aptamer embedded in the TL, SL2, or 3'-end of the Sp-gRNA (Supplementary Figure 2.1B). We tested all combinations of gRNA-aptamer and wtTadA-TadA7.10-PCP fusion by assessing editing efficiencies at 3 well-characterized genomic sites (*HEK2*, *HIRA*, and *PSMB2*). In all experiments, HEK293T cells were transfected with plasmids encoding nCas9-NG (with bicistronic mCherry expression to allow for evaluation of transfection efficiency), wtTadA-TadA7.10-PCP fusion, and gRNA-aptamer. Cells were then lysed after 72 hours, genomic loci of interest were amplified, and A•T to G•C editing efficiencies were quantified by next generation sequencing (NGS). Initial experiments showed that our ABE systems could install targeted point mutations, but with drastically lower editing efficiencies compared to the parental ABE7.10max-NG construct. Furthermore, base editing efficiencies by the aptamer complexes varied greatly by genomic site. Specifically, we observed similar A•T to G•C editing efficiencies at the *HEK2* locus by the six systems (average A•T to G•C editing of  $20.5 \pm 2.0\%$  by the six systems, Supplementary Figure 2.1B, see Methods for statistical analysis details), which represented editing efficiencies ranging from  $22.6 \pm 3.8\%$  to  $40.0 \pm 8.5\%$  of the parental ABE7.10max-NG editing. However, A•T to G•C editing efficiencies at the *HIRA* and *PSMB2* sites never surpassed  $2.4 \pm 1.0\%$  by any of the six systems (average A•T to G•C editing efficiencies by the parental ABE7.10max-NG construct at these two sites were  $44.6 \pm 18.6\%$  and  $45.2 \pm 16.9\%$ , respectively, Supplementary Figure 2.1B).

It has been shown that homodimers of later-generation mutant TadA enzymes can decrease DNA editing efficiency compared to their wtTadA-mutant TadA heterodimer counterparts.<sup>13</sup> We hypothesized that the dimeric binding mode of PCP to the PP7 aptamer could be causing *in trans* homodimerization of the mutated TadA7.10 portion of the wtTadA-TadA7.10 heterodimers. We therefore expanded our aptamer systems to include the boxB and com aptamers, which have smaller CPs ( $\lambda$ N, the coat protein of boxB, is 22-aa, and Com, the coat protein of com, is 61-aa) that bind as monomers. We embedded these aptamers into the SL2 or 3'-end of the gRNA (which are closest in 3D space to the target nucleotides) and selected two poorly edited genomic sites (*HIRA* and *RNF2*) on which to test our engineered constructs. We held the architecture of the CP-wtTadA-TadA-7.10 fusion (N-terminal CP) and linker length (93-aa) constant (Supplementary Figure 2.1C). Overall, we still observed drastically decreased A•T to G•C editing efficiencies at both sites compared to the parental ABE7.10max-NG construct (efficiencies ranged from  $0.2 \pm 0.2\%$  to  $10.1 \pm 2.1\%$  of the parental construct, Supplementary Figure 2.1C). However, the system that employed the com aptamer embedded at the 3'-end of the gRNA facilitated the highest A•T to G•C editing efficiencies at both sites ( $4.7 \pm 1.0\%$  at *HIRA*, and  $0.9 \pm 0.1\%$  at *RNF2*, Supplementary Figure 2.1C). Notably, the system with the 3'-boxB gRNA fusion, which was recently developed as an optimized aptamer-based ABE system for base editing in rice (up to 8% editing at the most-edited genomic locus), performed poorly at both sites ( $2.2 \pm 2.2\%$  at *HIRA*, and  $0.03 \pm 0.03\%$  at *RNF2*, Supplementary Figure 2.1C).<sup>58</sup> This suggests that engineered editing tools may not be transferable across organisms, and may require organism-specific re-engineering efforts to enable efficient editing.

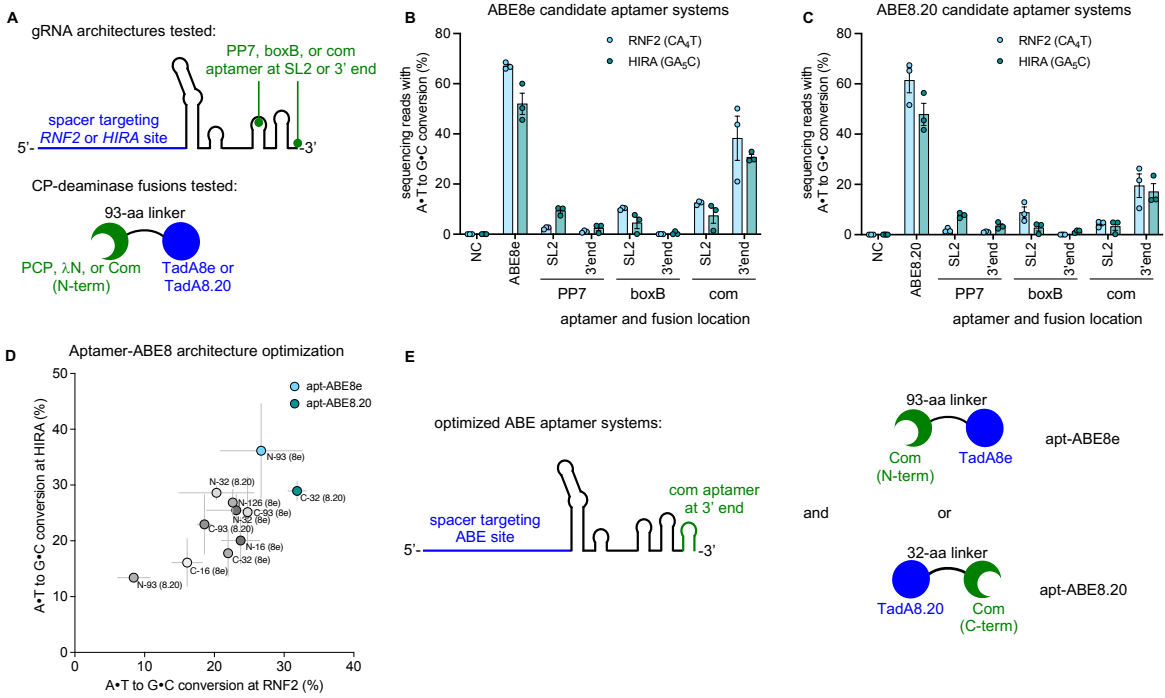
### 2.2.3 Improving aptamer-ABE efficiency with evolved deaminases

The wtTadA-TadA7.10 deoxyadenosine deaminase is known to have limited compatibility with alternate Cas enzymes and architectures beyond the parental ABE7.10 construct.<sup>59,60</sup> To combat this, several eighth-generation TadA enzymes (most notably, TadA8e

and TadA8.20) were recently independently evolved, and demonstrate increased editing efficiency, faster deamination kinetics, and enhanced compatibility with additional Cas domains.<sup>59,61</sup> We therefore postulated that these deaminase domains may also be more compatible with our aptamer systems. We generated 12 additional ABE aptamer systems with either PCP,  $\lambda$ N, or Com fused to the N-terminus of either TadA8e or TadA8.20 (monomeric deaminase) via a 93-aa flexible linker, and the corresponding aptamers embedded in the SL2 or 3'-end of the Sp-gRNA (Figure 2.2A and Supplementary Figure 2.2A). We tested all combinations of gRNA-aptamers with their cognate CP fused to TadA8 by assessing editing efficiencies at the same two poorly edited sites (*HIRA* and *RNF2*). Both evolved TadA8 deaminases displayed improved editing efficiencies when compared to their respective wtTadA-TadA7.10 counterparts with an average increase of  $9.6 \pm 2.2$ -fold for TadA8.20 and  $15.0 \pm 4.3$ -fold for TadA8e across both sites (up to  $44.2 \pm 11.1$ -fold increase was observed for the 3'-end com at *RNF2*). The systems that employed the com aptamer embedded at the 3'-end of the gRNA again facilitated the highest A•T to G•C editing efficiencies at both sites and for both TadA8 deaminases ( $30.7 \pm 1.1\%$  at *HIRA* and  $38.3 \pm 8.8\%$  at *RNF2* for TadA8e, Figure 2.2B, and  $17.1 \pm 3.2\%$  at *HIRA* and  $19.5 \pm 4.7\%$  at *RNF2* for TadA8.20, Figure 2.2C and Supplementary Figure 2.2B). As this aptamer-gRNA construct consistently facilitated the highest editing efficiencies, it was selected for further optimization with both TadA8 deaminases.

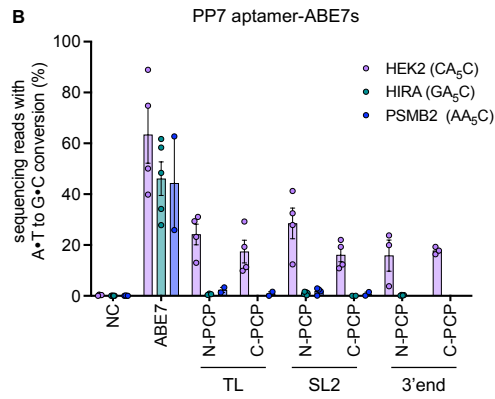
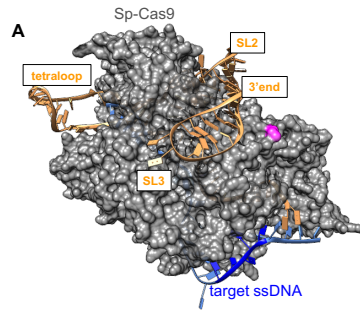
We next optimized the Com-TadA8 fusion architecture (by testing both N- and C-terminal Com fusions) and linker length (by testing flexible linkers taken from established base editors<sup>62–64</sup> with lengths ranging from 16- to 126-aa) to maximize editing efficiency (Supplementary Figure 2.2C). Out of all the aptamer-ABE systems derived from TadA8e, we observed the highest editing efficiencies when a 93-aa linker was used with an N-terminally linked Com (Com\_93-aa\_TadA8e, Figure 2.2E); editing by this construct was  $77.7 \pm 19.3\%$  of the parental ABE8e-NG (hereafter referred to as ABE8e) editing efficiency at *HIRA* and  $46.1 \pm 10.4\%$  of the parental ABE8e editing efficiency at *RNF2* (Figure 2.2D). On the other hand, out of

all the aptamer-ABE systems derived from TadA8.20, we observed the highest editing efficiencies when using a 32-aa linker with Com fused to the C-terminus of TadA8.20 (TadA8.20\_32-aa\_Com, Figure 2.2E); editing by this construct was  $88.7 \pm 5.2\%$  of the parental ABE8.20-NG (hereafter referred to as ABE8.20) editing efficiency at *HIRA* and  $63.7 \pm 2.6\%$  of the parental ABE8.20 editing efficiency at *RNF2* (Supplementary Figure 2.2D and Figure 2.2D). These two constructs, termed apt-ABE8e and apt-ABE8.20 (Figure 2.2E), were therefore both taken forward for characterization.

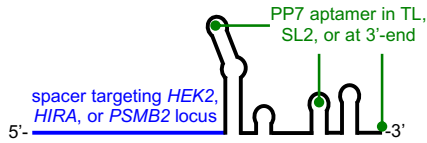


**Figure 2.2: Engineering of an ABE aptamer system.** (A) Schematics of the aptamer-embedded gRNA and deaminase-coat protein fusions tested in (B and C). (B, C, and D) Editing efficiencies at the *HIRA* and *RNF2* loci of ABE aptamer systems employing the TadA8e (B and D) or TadA8.20 (C and D) deaminase and PP7, boxB, or com aptamer-coat protein systems. HEK293T cells were transfected with plasmids encoding nCas9-NG, TadA8-CP fusion, and gRNA-aptamer (for aptamer-treated cells), or plasmids encoding the parental ABE8e-NG or ABE8.20-NG editors and unmodified gRNA (for ABE8e and ABE8.20-treated cells). Cells were then lysed after 72 hours, genomic loci of interest were amplified, and analyzed by NGS. Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA. Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the target A's indicated. Bar graphs (B and C) and dot plots (D) represent the average, and error bars represent the standard error of the mean (SEM) for n=3 biological replicates (each replicate is marked individually in [B and C]). (D) ABE aptamer architecture optimization of TadA8-Com fusions with the 3'end com-gRNA (see Supplementary Figure 2C-D for individual replicates and constructs tested). (E) Schematics of the two most efficient aptamer-embedded gRNA and deaminase-coat protein fusions that were selected as final ABE aptamer systems: apt-ABE8e and aptABE8.20.

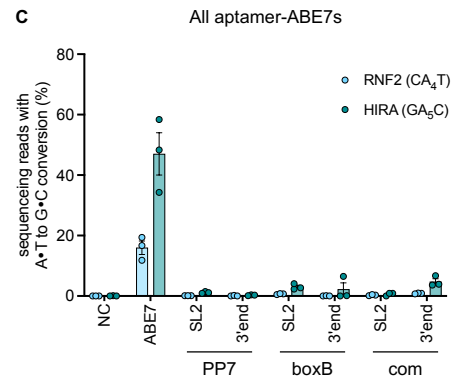
**Supplementary Figure 2.1: Initial testing of ABE aptamer systems with the wtTadA-TadA7.10 deaminase.** (A) Crystal structure (PDB ID: 5Y36) of the Cas:gRNA:DNA R-loop complex, with the target single-stranded DNA shown in blue and the various locations of the gRNA (orange) that protrude from the Cas9 protein indicated: the tetraloop, stem-loop 2 (SL2), stem-loop 3 (SL3), and 3'end. The target ssDNA window is shown in dark blue and the N-terminus of Cas9 (deaminase fusion location of most BEs) is colored pink. (B) Editing efficiencies at the *HEK2*, *HIRA*, and *PSMB2* loci of ABE aptamer systems employing the wtTadA-TadA7.10 deaminase and the PP7 aptamer-coat protein system (top), with schematics of the aptamer-containing gRNAs and deaminase-coat protein fusions shown (bottom). (C) Editing efficiencies at the *HIRA* and *RNF2* loci of ABE aptamer systems employing the wtTadA-TadA7.10 deaminase and PP7, boxB, or com aptamer-coat protein systems (top), with schematics of the aptamer-imbedded gRNAs and deaminase-coat protein fusions shown (bottom). HEK293T cells were transfected with plasmids encoding nCas9-NG, wtTadA-TadA7.10-CP fusion, and gRNA-aptamer, or plasmids encoding the parental ABE7.10max-NG editor and unmodified gRNA. Cells were then lysed after 72 hours, genomic loci of interest were amplified, and analyzed by NGS. Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the target A's indicated. Bar graphs and error bars represent the average and standard error of the mean (SEM) for n=2 or more biologically independent experiments (each replicate is marked individually). Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA.



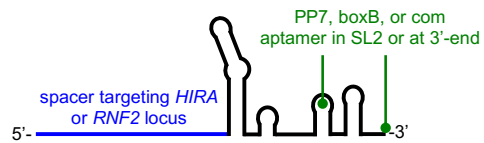
gRNA architectures tested:



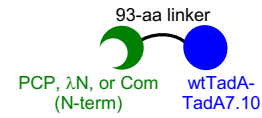
CP-deaminase fusions tested:

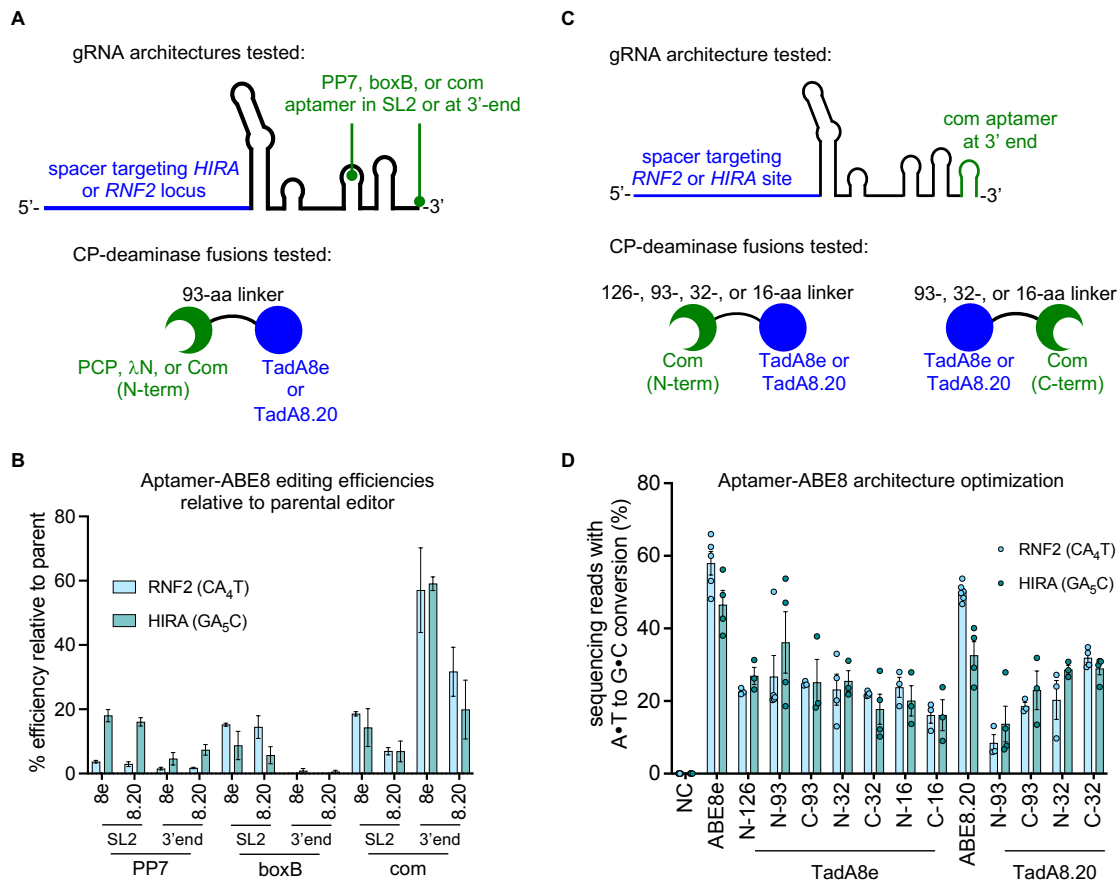


gRNA architectures tested:



CP-deaminase fusions tested:





### Supplementary Figure 2.2: Optimization of ABE aptamer systems employing TadA8

**deaminases.** (A) Schematics of the aptamer-embedded gRNAs and deaminase-coat protein fusions tested in (B). (B) Editing efficiencies at the *HIRA* and *RNF2* loci of ABE aptamer systems employing the TadA8e and TadA8.20 deaminases normalized to the parental ABE8e-NG (hereafter referred to as ABE8e) or ABE8.20-NG (hereafter referred to as ABE8.20) constructs. Cells were treated as previously described in Supplementary Figure 1. Plotted are the average percent of total DNA sequencing reads with A•T edited to G•C at the target A's indicated for the aptamer-treated cells divided by that of the cells treated with the respective parental ABE8 construct for n=3 biologically independent experiments (plotted individually in Figure 2 B-C). The values represent the ratio of the two averages and the error bars represent the propagation of uncertainty of the SEM. (C) Schematics of the aptamer-embedded gRNA and deaminase-coat protein fusions tested in (D). (D) Editing efficiencies at the *HIRA* and *RNF2* loci of ABE aptamer systems with the com aptamer embedded at the 3' end and TadA8e- or TadA8.20-Com fusion proteins. Cells were treated as previously described in Supplementary Figure 1. Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the target A's indicated. Bar graphs error bars represent the average and SEM for n=3 or more biological replicates (each replicate is marked individually). Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA.

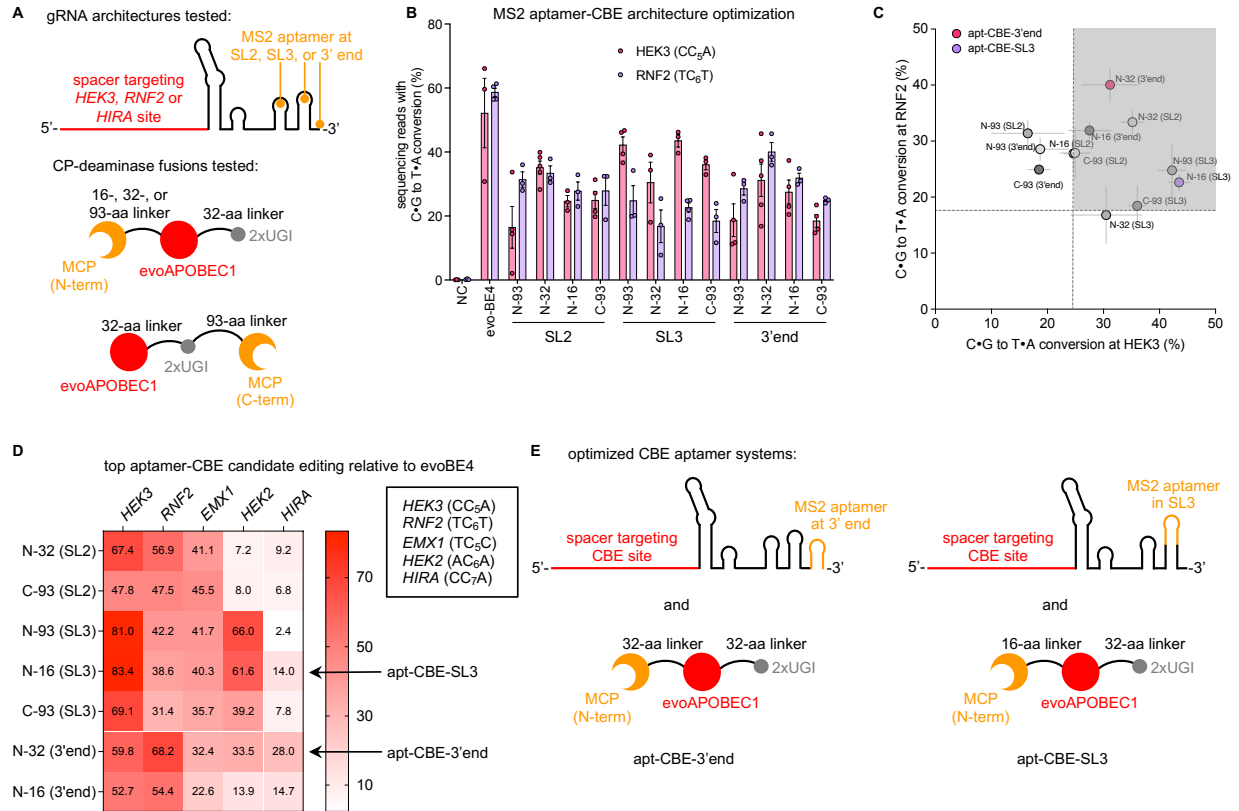


#### 2.2.4 Optimization of aptamer-CBE deaminase and architecture

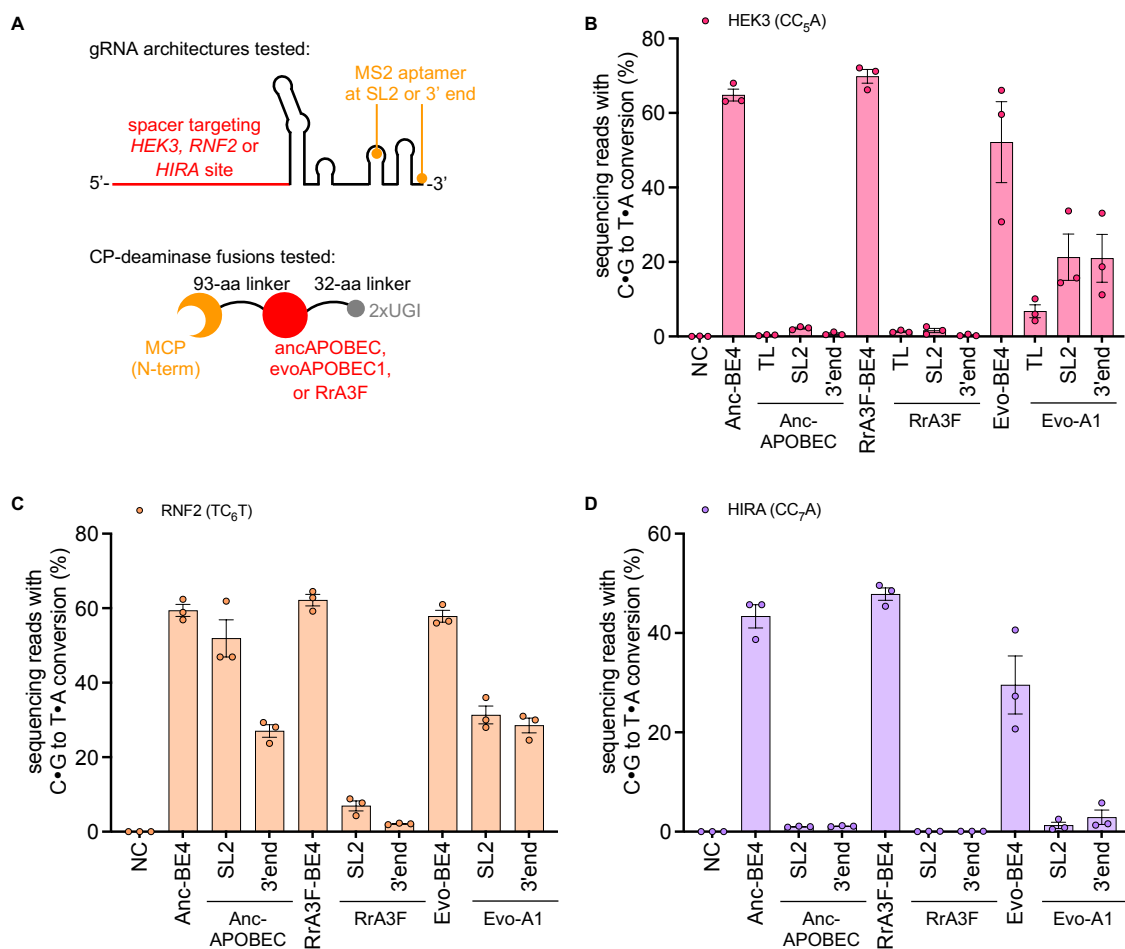
After observing the drastic impact that different deoxyadenosine deaminase variants had on the editing efficiencies of our aptamer-ABE systems, we selected three different cytidine deaminase enzymes to screen for C•G to T•A editing when used in aptamer-CBE architectures. Specifically, we chose to test ancAPOBEC (an ancestral sequence reconstruction of the rAPOBEC1 enzyme with improved editing activity),<sup>65</sup> evoAPOBEC1 (an evolved variant of the rAPOBEC1 enzyme with improved editing activity and target sequence compatibility),<sup>66</sup> and RrA3F (a cytidine deaminase identified from a BLAST search with high on-target editing efficiency and low gRNA-independent off-target editing efficiency).<sup>67</sup> We focused on the MS2-MCP aptamer system, as it has been used successfully previously with cytidine deaminases.<sup>54,57</sup> We embedded the MS2 aptamer in the SL2 or 3'-end of the Sp-gRNA and held the architecture of the MCP-APOBEC fusion (N-terminal MCP) and linker length (93-aa) constant (Supplementary Figure 2.3A). Editing precision by CBEs is enhanced when the uracil glycosylase inhibitor (UGI) peptide is included in the CBE architecture. Thus, two copies of UGI were appended to the C-terminus of the cytidine deaminase following a flexible 32-aa linker in all constructs (Supplementary Figure 2.3A). Editing efficiencies were then assessed at 3 well-characterized genomic sites (*HEK3*, *HIRA*, and *RNF2*, Supplementary Figure 2.3B-D). We found that the evoAPOBEC1-derived systems most consistently facilitated the highest C•G to T•A editing efficiencies compared to their ancAPOBEC and RrA3F counterparts, particularly at the poorly edited *HEK3* site (editing by the evoAPOBEC1 systems were on average  $30.0 \pm 8.8\%$  of the parental evoBE4-NG [hereafter referred to as evoBE4] construct editing efficiency at the *HEK3* site, compared to  $1.7 \pm 0.9\%$  for ancAPOBEC and  $1.5 \pm 0.5\%$  for RrA3F, Supplementary Figure 2.3B).

We therefore moved forward with the evoAPOBEC1 deaminase. We generated twelve total CBE aptamer systems employing the evoAPOBEC1 deaminase and tested their activity at two poorly edited sites (*HEK3* and *RNF2*). We embedded the MS2 aptamer in the SL2, SL3 (it

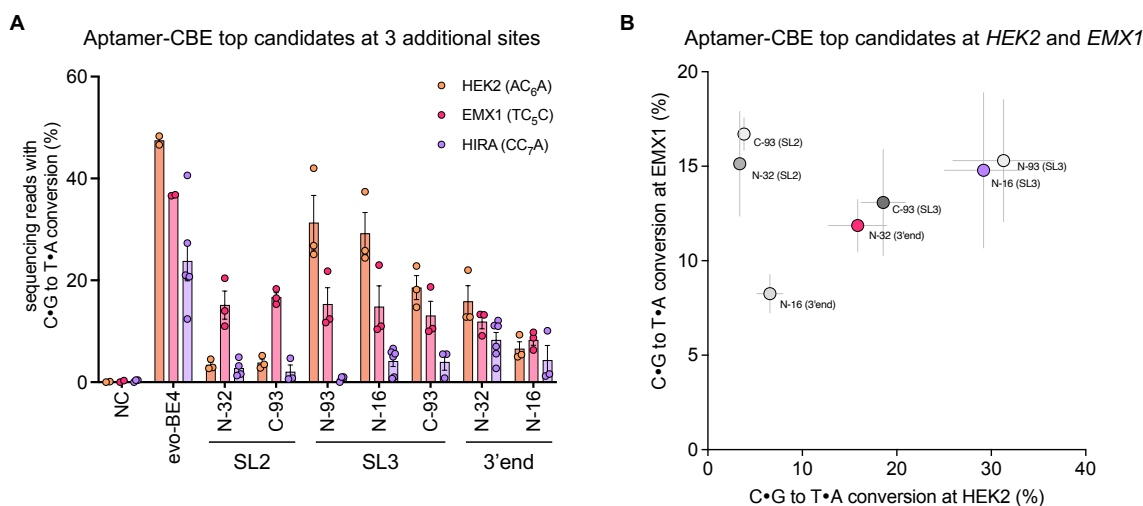
was recently reported that embedding aptamers in the SL3 portion of the gRNA can facilitate recruitment of heterologous effectors—both cytidine and adenosine deaminases concurrently—to the same site for the purpose of “dual base editing”),<sup>68</sup> or 3'-end of the Sp-gRNA (Figure 2.3A). We tested both N- and C-terminal MCP-evoAPOBEC1 fusions with 93-aa linkers, as well as N-terminal MCP-evoAPOBEC1 fusions with 32- and 16-aa linkers (Figure 2.3A). Although editing efficiencies varied across the three genomic sites, seven constructs displayed C•G to T•A editing efficiencies of >18% at the *RNF2* site and >25% at the *HEK3* site (Figure 2.3B-C, dotted lines show cutoffs). We took these seven CBE aptamer systems and evaluated their editing efficiencies at three additional genomic loci (*EMX1*, *HIRA*, and *HEK2*). Out of this extended dataset (editing efficiencies at five total sites which had target C's embedded within multiple sequence motifs and located at multiple positions within the protospacer, Figure 2.3D), we found that apt-CBE-3'end (MS2 embedded in the 3'-end of the gRNA with the MCP fused to the N-terminus of evoAPOBEC1 via a 32-aa linker, Figure 2.3E) and apt-CBE-SL3 (MS2 embedded in SL3 of the gRNA with the MCP fused to the N-terminus of evoAPOBEC1 via a 16-aa linker, Figure 2.3E) displayed the most consistently high C•G to T•A editing efficiencies. Editing by the apt-CBE-3'end system ranged from  $28.0 \pm 5.1\%$  to  $68.2 \pm 5.0\%$  (average  $44.4 \pm 8.2\%$ ) of the parental evoBE4 construct across all sites, and editing by the apt-CBE-SL3 system ranged from  $14.0 \pm 3.8\%$  to  $83.4 \pm 3.8\%$  (average  $47.6 \pm 11.7\%$ ) of the parental evoBE4-NG construct across all sites (Figure 2.3D and Supplementary Figure 2.4). We therefore chose these constructs as our final CBE aptamer systems and took them forward for characterization.



**Figure 2.3: Engineering of a CBE aptamer system.** (A) Schematics of the aptamer-embedded gRNA and deaminase-coat protein fusions tested in (B and C). (B, C) Editing efficiencies at the *HEK3* and *RNF2* loci of CBE aptamer systems employing the evoAPOBEC1 deaminase and MS2 aptamer-coat protein system. Cells were treated as previously described in Figure 2. Plotted are the percent of total DNA sequencing reads with C•G edited to T•A at the target C's indicated. Bar graphs (B) and dot plots (C) represent the average, and error bars represent the SEM for n=3 or 4 biological replicates (each replicate is marked individually in [B]). Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA. In (B), the top seven constructs are indicated with the grey box. (D) Relative editing efficiencies (compared to the parental evoBE4 editor) at the *HEK3*, *EMX1*, *RNF2*, *HEK2*, and *HIRA* loci of the top seven CBE aptamer systems employing the evoAPOBEC1 deaminase and the MS2 aptamer-coat protein system from (B) and (C). Shown are the percent of DNA sequencing reads with C•G edited to T•A at the target C's indicated relative to evoBE4-NG. Values represent the average for n=3 or 4 biological replicates. The two constructs, that consistently displayed the highest editing efficiencies across all five sites are indicated with arrows (apt-CBE-3'end and apt-CBE-SL3). (E) Schematics of the two most efficient aptamer-embedded gRNA and deaminase-coat protein fusions that were selected as final CBE aptamer systems.



**Supplementary Figure 2.3: Initial testing of CBE aptamer systems with the ancAPOBEC, evoAPOBEC1, and RrA3F deaminases.** (A) Schematics of the aptamer-embedded gRNAs and deaminase-coat protein fusions tested in (B), (C), and (D). (B, C, and D) Editing efficiencies at the *HEK3* (A), *HIRA* (B), and *RNF2* (C) loci of CBE aptamer systems employing the ancAPOBEC, evoAPOBEC1, or RrA3F deaminases and the MS2 aptamer-coat protein system. Cells were treated as previously described in Supplementary Figure 1. Plotted are the percent of total DNA sequencing reads with C•G edited to T•A at the target C's indicated. Bar graphs and error bars represent the average and SEM for n=3 biological replicates (each replicate is marked individually). Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA.



**Supplementary Figure 2.4: Architecture optimization of CBE aptamer systems employing the evoAPOBEC1 deaminase.** (A and B) Editing efficiencies at the *HEK2*, *EMX1*, and *HIRA* loci of the top seven CBE aptamer systems employing the evoAPOBEC1 deaminase and the MS2 aptamer-coat protein system. Schematics of the aptamer-embedded gRNA and deaminase-coat protein fusions are shown in Figure 3A. Cells were treated as previously described in Supplementary Figure 1. Plotted are the percent of total DNA sequencing reads with C•G edited to T•A at the target C's indicated. Bar graphs (A) and dot plots (B) represent the average, and error bars represent the SEM for n=2 or more biological replicates (each replicate is marked individually in [A]). Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA.

## 2.2.5 Characterization of aptamer-BEs at genomic loci in human cells

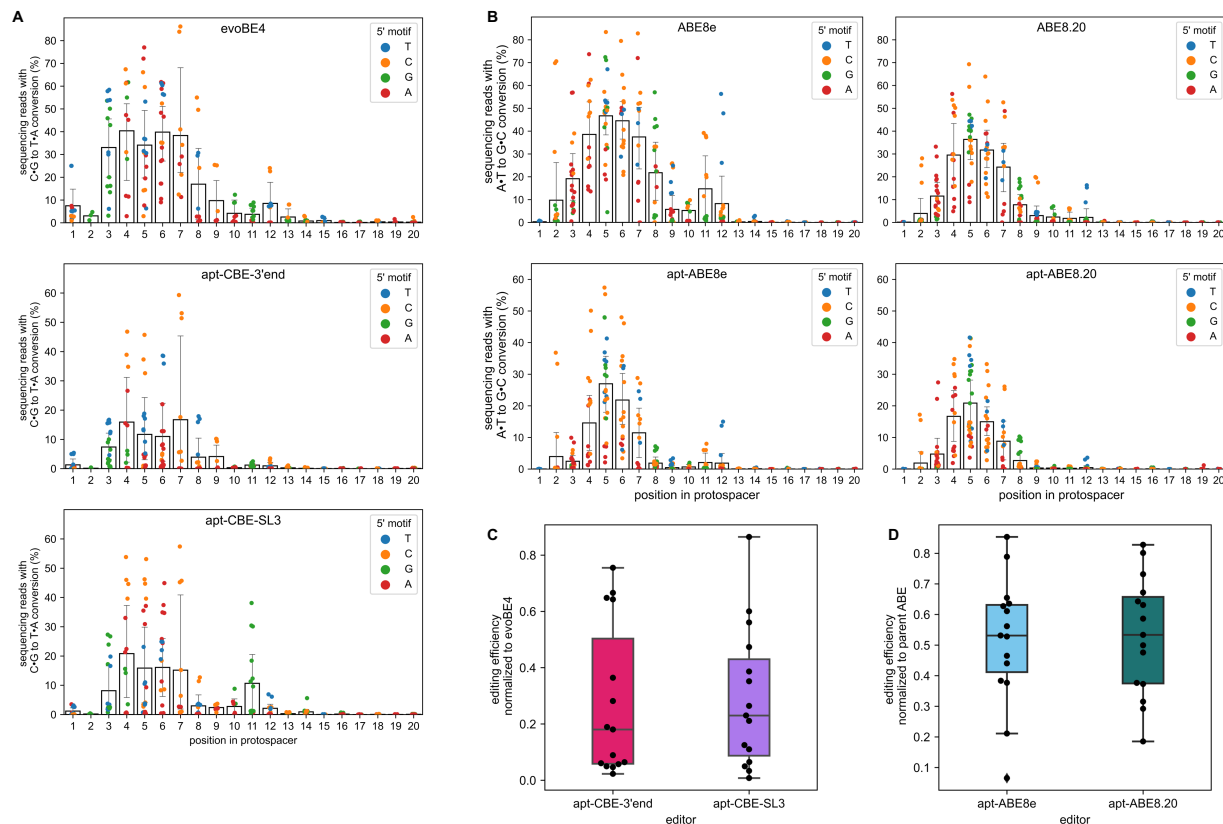
We next characterized each of the four optimized aptamer-BE systems by measuring their base editing efficiencies at 15 genomic protospacers, which collectively contained 73 C and 112 A nucleotides spanning all positions 1-20 within the protospacer (Figure 2.4A-B). We analyzed each aptamer-BE system for editing efficiency, window size, and sequence preference. To quantify the editing window, we first computed an average editing efficiency at each position, which was averaged across all measured Cs or As at that position, for each editor (shown as the black bars in Figure 2.4A-B). We define the “editing window” as the positions which have average editing efficiencies that exceed 20% of the maximal average editing efficiency observed for a given editor. All four aptamer systems displayed slightly

narrower editing windows than their parent editors when evaluated using this metric. The parental evoBE4 editor displayed an editing window of positions 3 through 9 and 12, apt-CBE-3'end displayed an editing window of 3 through 9, and apt-CBE-SL3 displayed an editing window of 3 through 7 and 11 (Supplementary Figure 2.5A and Figure 2.4A). The parental ABE8e editor displayed an editing window of positions 2 through 8 and 11, while apt-ABE8e displayed an editing window of positions 4 through 7 (Supplementary Figure 2.5A and Figure 2.4B). Finally, the parental ABE8.20 editor displayed an editing window of positions 3 through 8, while apt-ABE8.20 displayed an editing window of positions 4 through 7 (Supplementary Figure 2.5A and Figure 2.4B).

Alternatively, defining an “absolute editing window” as all positions with mean editing efficiencies above an absolute cutoff of 5%, apt-CBE-3'end exhibited a window of positions 3-7, apt-CBE-SL3 exhibited a window of positions 3-7 and 11, and the parental evoBE4 construct exhibited a window of positions 1, 3-9, and 12 (Supplementary Figure 2.5B). Using this absolute cutoff, apt-ABE8e and apt-ABE8.20 both exhibited editing windows of positions 4-7, which is much narrower than the editing windows of ABE8e (positions 2-12) and ABE8.20 (positions 3-8, Supplementary Figure 2.5B). When analyzing editing rates across the entire sequenced amplicon for each site (which included 50 to 105 base pairs on either side of the protospacer), we did not observe editing above background levels outside of the protospacers for any of the aptamer systems.

For each aptamer system, we normalized the average editing efficiency for each individual C or A to the average editing efficiency for that base by the corresponding parental editor (Supplementary Figure 2.5C-D) and found that all four aptamer systems displayed reduced editing efficiencies compared to their respective parental constructs across the 15 sites tested. Specifically, at the most-edited cytosine base, absolute editing by apt-CBE-3'end ranged from 0.2-45.4% (or 2.3-75.5% of the parental evoBE4), editing by apt-CBE-SL3 ranged from 0.2-46.0% (or 0.8-86.5% of the parental evoBE4, Figure 2.4C). At the most-edited adenine

base, editing by apt-ABE8e ranged from 1.1-45.9% (or 6.5-85.4% of the parental ABE8e, Figure 2.4D), and editing by apt-ABE8.20 ranged from 0.2-36.2% (or 18.5-82.5% of the parental ABE8.20, Figure 2.4D). While we observed that all three parental editors displayed minimal sequence context preferences (consistent with previous reports),<sup>59,61,66</sup> we found that some of the aptamer systems had gained sequence motif preferences, which were responsible for the reduced normalized editing efficiencies at certain sites. Specifically, we found that at a subset of 5' AC targets within the window (i.e., sites 2, 11, and 14), the aptamer-CBE systems displayed very low (<2%) editing despite evoBE4 editing these targets with efficiencies ranging from 8.7% to 21.2%. We found the apt-CBE-3'end construct in general prefers a 5' T or C motif (all C's with >40% activity compared to the parental system had this sequence motif), while the apt-CBE-SL3 construct had little sequence preference (Supplementary Figure 2.5C). Additionally, the apt-ABE8e construct exhibited an aversion to the 5' A motif, while the apt-ABE8.20 construct did not appear to have any sequence motif preferences (Supplementary Figure 2.5D). There was one site where both aptamer-ABE systems demonstrated very low editing; at this site, the most-edited adenine was at position 11 and only the parental ABE8e, but not ABE8.20, showed editing >2%.



**Figure 2.4: Characterization of optimized CBE and ABE aptamer systems.** (A and B) Editing efficiencies by all four optimized aptamer systems and their respective parental editors at 73 Cs and 112 As, measured from 15 different genomic protospacers in HEK293T cells. Target Cs and As are organized by their position within the protospacer (x-axis), and colored according to what the identity of the base 5' to the target C or A is (5' motif). Cells were treated as previously described in Figure 2. Plotted are the percent of total DNA sequencing reads with C•G edited to T•A (A) or A•T edited to G•C (B). Bar graphs and error bars represent the average and 95% confidence interval across all measured Cs or As at that position ( $n=3$  or 4 biological replicates for each target C or A, and each replicate is marked individually). (C and D) Editing efficiencies by all four optimized aptamer systems normalized to their respective parental editor. (C) For each CBE aptamer system, the highest edited C within each of the 15 protospacers was normalized to the parental editor by dividing the percent of total DNA sequencing reads with the target C•G edited to T•A for the aptamer system by that of the parental editor. (D) For each ABE aptamer system, the highest edited A within each of the 15 protospacers was normalized to the parental editor by dividing the percent of total DNA sequencing reads with the target A•T edited to G•C for the aptamer system by that of the parental editor. Values on the whisker plots represent the lowest observation, lower quartile, median, upper quartile, and the highest observation across the 15 different sites.



**Supplementary Figure 2.5: Characterization of optimized CBE and ABE aptamer systems.**

(A and B) Average editing efficiencies of each position in the protospacer across all 15 genomic sites. The data are either normalized to the maximal editing for each editor (A) with editing windows shown in green or shown as average absolute editing efficiencies (B) with editing windows boxed. (C and D) Editing efficiencies, organized by position within the protospacer, of all four aptamer systems normalized to their respective parental editor. Cells were treated as previously described in Supplementary Figure 1. (C) Plotted are the average percent of total DNA sequencing reads with C•G edited to T•A at each target C by each CBE aptamer system divided by that of the parental editor. (D) Plotted are the average percent of total DNA sequencing reads with A•T edited to G•C at each target A by each ABE aptamer system divided by that of the parental editor. Dot plots represent the average for n=3 or more biological replicates. Only target As and Cs where the parental editor displayed editing efficiencies >2% of are shown.

**A**

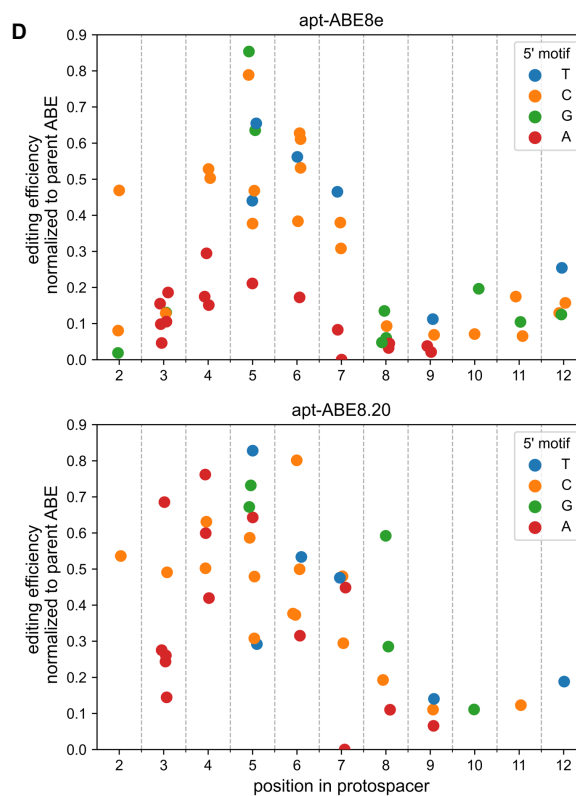
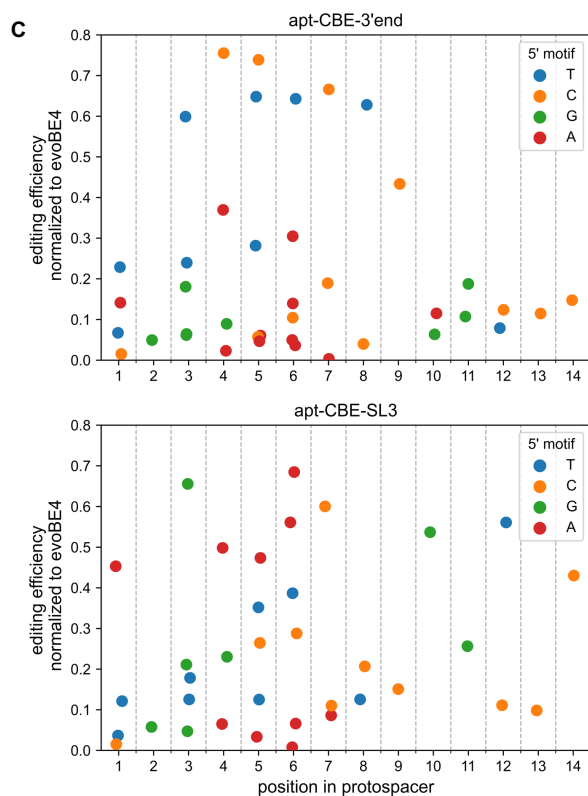
|               |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|---------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
|               | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13   | 14   | 15   | 16   | 17   | 18   | 19   | 20   |
| evoBE4        | 0.18 | 0.08 | 0.82 | 1.00 | 0.84 | 0.99 | 0.95 | 0.42 | 0.24 | 0.10 | 0.09 | 0.21 | 0.06 | 0.02 | 0.02 | 0.00 | 0.00 | 0.01 | 0.01 | 0.01 |
| apt-CBE-3'end | 0.08 | 0.01 | 0.44 | 0.95 | 0.70 | 0.66 | 1.00 | 0.23 | 0.25 | 0.02 | 0.07 | 0.06 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| apt-CBE-SL3   | 0.06 | 0.01 | 0.39 | 1.00 | 0.76 | 0.77 | 0.73 | 0.14 | 0.11 | 0.13 | 0.51 | 0.10 | 0.01 | 0.04 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
|               | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13   | 14   | 15   | 16   | 17   | 18   | 19   | 20   |
| ABE8e         | 0.01 | 0.21 | 0.41 | 0.83 | 1.00 | 0.95 | 0.80 | 0.47 | 0.12 | 0.11 | 0.32 | 0.18 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| ABE8.20       | 0.00 | 0.11 | 0.32 | 0.81 | 1.00 | 0.87 | 0.67 | 0.21 | 0.08 | 0.07 | 0.05 | 0.06 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| apt-ABE8e     | 0.00 | 0.15 | 0.09 | 0.54 | 1.00 | 0.81 | 0.42 | 0.07 | 0.02 | 0.02 | 0.08 | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| apt-ABE8.20   | 0.00 | 0.09 | 0.23 | 0.80 | 1.00 | 0.72 | 0.42 | 0.13 | 0.02 | 0.02 | 0.01 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

mean normalized editing

**B**

|               |     |     |      |      |      |      |      |      |     |     |      |     |     |     |     |     |     |     |     |     |
|---------------|-----|-----|------|------|------|------|------|------|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|               | 1   | 2   | 3    | 4    | 5    | 6    | 7    | 8    | 9   | 10  | 11   | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  |
| evoBE4        | 7.5 | 3.0 | 33.1 | 40.4 | 34.1 | 39.8 | 38.3 | 16.9 | 9.7 | 4.1 | 3.7  | 8.5 | 2.5 | 0.8 | 1.0 | 0.1 | 0.1 | 0.4 | 0.3 | 0.3 |
| apt-CBE-3'end | 1.3 | 0.2 | 7.4  | 15.9 | 11.7 | 11.0 | 16.7 | 3.9  | 4.1 | 0.3 | 1.2  | 1.0 | 0.3 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.1 |
| apt-CBE-SL3   | 1.2 | 0.2 | 8.1  | 20.8 | 15.9 | 16.1 | 15.2 | 3.0  | 2.4 | 2.8 | 10.7 | 2.1 | 0.3 | 0.9 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 0.1 |
|               | 1   | 2   | 3    | 4    | 5    | 6    | 7    | 8    | 9   | 10  | 11   | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  |
| ABE8e         | 0.3 | 9.8 | 19.2 | 38.6 | 46.7 | 44.5 | 37.4 | 21.8 | 5.7 | 5.2 | 14.7 | 8.2 | 0.2 | 0.4 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| ABE8.20       | 0.2 | 3.9 | 11.6 | 29.5 | 36.4 | 31.7 | 24.2 | 7.7  | 3.0 | 2.4 | 1.8  | 2.2 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| apt-ABE8e     | 0.0 | 3.9 | 2.5  | 14.6 | 27.0 | 21.8 | 11.5 | 1.9  | 0.4 | 0.7 | 2.1  | 1.9 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| apt-ABE8.20   | 0.0 | 1.9 | 4.7  | 16.7 | 20.9 | 15.0 | 8.8  | 2.7  | 0.3 | 0.3 | 0.2  | 0.4 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.0 |

mean absolute editing



## Acknowledgements

Chapter 2, in part, is a reprint of the material as it appears in Nature Biotechnology. Cowan, Quinn T.; Gu, Sifeng; Gu, Wanjun; Ranzau, Brodie L.; Simonson, Tatum S.; Komor, Alexis C., *Nature Biotechnology*, 2024. The dissertation author was the primary investigator and author of this paper.

## Chapter 3: Multiplex Orthogonal Base Editor (MOBE) Systems

### 3.1 Results

#### 3.1.1 Combining aptamer-BEs to make MOBE systems

Four multiplexed orthogonal base editor (MOBE) systems were then generated by combining each optimized CBE aptamer system (apt-CBE-3'end and apt-CBE-SL3) with each optimized ABE aptamer system (apt-ABE8e and apt-ABE8.20). For each MOBE system (which we label MOBE1-4, in accordance with Figure 3.1A), the two CP-deaminase constructs were joined with a tandem PT2A linker (a tandem P2A-T2A bicistronic construct, which robustly induces ribosomal skipping, ensuring that no dual deaminase fusion protein is produced). These effectors were simultaneously targeted to distinct genomic sites using two plasmids expressing their respective com or MS2 aptamer-gRNAs. Cells were transfected with plasmids encoding nCas9-NG-P2A-mCherry (the bicistronic mCherry expression allowed for evaluation of transfection efficiency but was not used to sort for transfected cells), tandem CP-deaminase fusions, and both gRNA-aptamers, lysed after three days, and editing was quantified by NGS. Editing was measured at three different protospacer combinations: *HIRA.0/HEK3.0*, *HEK2.0/RNF2.0*, and *HIRA.3/RNF2.0*, where the ABE target is listed first, and the CBE target is listed second in each combination. We additionally performed multiplexed editing with the parental, non-orthogonal ABE and CBE constructs (the evoBE4 parental CBE construct was multiplexed with either ABE8e or ABE8.20 using unmodified gRNAs targeting the same protospacers as the MOBE systems).

The MOBE systems displayed high levels of on-target activity across all target combinations; the MOBE1 system averaged 21.8% C•G to T•A conversion and 36.2% A•T to G•C conversion; the MOBE3 system averaged 17.7% C•G to T•A conversion and 32.2% A•T to G•C conversion (Figure 3.1B and Figure 3.2A-F). This is compared to their parental evoBE4/ABE8e system, which displayed an average on-target editing efficiency of 16.8% C•G

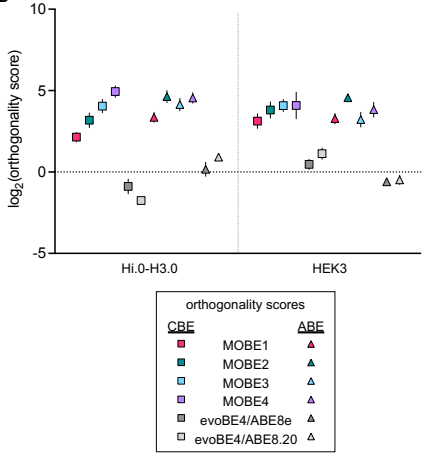
to T•A conversion across the CBE target sites and 49.0% A•T to G•C conversion across the ABE target sites. Importantly, the MOBE systems showed greatly decreased crosstalk (non-orthogonal edits) compared to the parental constructs. MOBE3 averaged 1.0% crosstalk editing across all protospacers, and MOBE1, which had the highest crosstalk editing, averaged 2.6% crosstalk editing (Figure 3.1B and Figure 3.2A-C, dotted line insets). In contrast, the evoBE4/ABE8e and evoBE4/ABE8.20 parental systems averaged 30.2% and 23.0% crosstalk editing, respectively (Figure 3.1B and Figure 3.2A-C). Specifically, when targeted to the *HIRA.0/HEK3.0* loci, MOBE3 had the highest on-target activity, with  $29.9 \pm 6.2\%$  C•G to T•A editing at the CBE target (*HEK3.0*) and  $30.9 \pm 6.9\%$  A•T to G•C editing at the ABE target (*HIRA.0*), with less than  $1.8 \pm 0.4\%$  crosstalk at either site (Figure 3.1B). This is compared to its parental evoBE4/ABE8e combination, which had  $7.6 \pm 1.6\%$  C•G to T•A editing at the CBE target (*HEK3.0*) and  $37.7 \pm 10.5\%$  A•T to G•C editing at the ABE target (*HIRA.0*), with  $14.0 \pm 3.2\%$  C•G to T•A crosstalk at the *HIRA.0* site and  $33.5 \pm 3.2\%$  A•T to G•C crosstalk at the *HEK3.0* site (Figure 3.1B).

**Figure 3.1: Editing efficiencies of multiplexed orthogonal base editor (MOBE) systems compared to multiplexing CBE and ABE.** (A) The four aptamer CBE-ABE combinations that make up each of the four MOBE systems are specified. The construct schematics of the aptamer CBE and ABE systems are listed in Figures 3 and 2, respectively. (B and C) Editing efficiencies by the MOBE1-4 systems and the evoBE4-NG/ABE8e-NG and evoBE4-NG/ABE8.20-NG combinations at the *HIRA.0/HEK3.0* (B) and *HEK3.2/HEK3.0* (C) loci. HEK293T cells were transfected with plasmids encoding nCas9-NG-P2A-mCherry, tandem CP-deaminase fusions, and both gRNA-aptamers (for the MOBE systems), or plasmids encoding the parental evoBE4-NG, ABE8e-NG or ABE8.20-NG, and both unmodified gRNAs (for the evoBE4-NG/ABE8e-NG and evoBE4-NG/ABE8.20-NG combinations). Cells were then lysed after 72 hours, genomic loci of interest were amplified, and analyzed by NGS. Plotted are the percent of total DNA sequencing reads at the most-edited base with A•T edited to G•C at the ABE target (y-axis, filled circles), C•G edited to T•A at the ABE target (y-axis, circles with X inset), C•G edited to T•A at the CBE target (x-axis, filled circles), and A•T edited to G•C at the CBE target (x-axis, circles with X inset). Dot plots and error bars represent the average and SEM for n=3 biological replicates. Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA. (D) Orthogonality scores for the MOBE1-4 systems and the evoBE4-NG/ABE8e-NG and evoBE4-NG/ABE8.20-NG combinations at the two protospacer combinations from (B) and (C). Plotted are the  $\log_2$  of the “CBE orthogonality scores” (squares), which we define for a given protospacer combination as the percent of total DNA sequencing reads with C•G edited to T•A at the CBE target divided by the percent of total DNA sequencing reads with C•G edited to T•A at the ABE target. The “ABE orthogonality scores” (triangles) are the A•T to G•C editing efficiency equivalent. Dot plots and error bars represent the average and SEM for n=3 biological replicates.

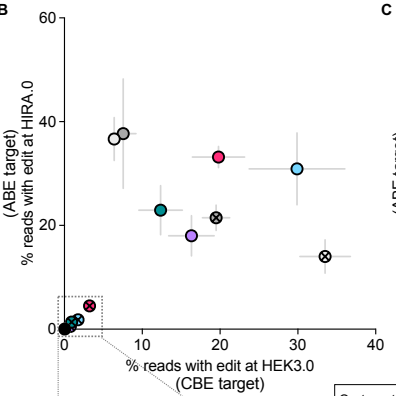
A

| MOBE system | CBE           | ABE         |
|-------------|---------------|-------------|
| MOBE1       | apt-CBE-3'end | apt-ABE8e   |
| MOBE2       | apt-CBE-3'end | apt-ABE8.20 |
| MOBE3       | apt-CBE-SL3   | apt-ABE8e   |
| MOBE4       | apt-CBE-SL3   | apt-ABE8.20 |

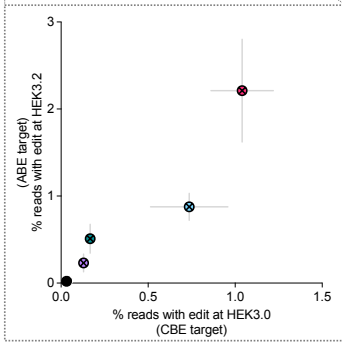
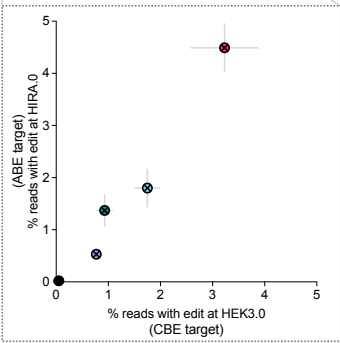
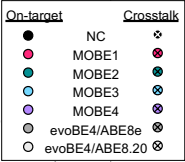
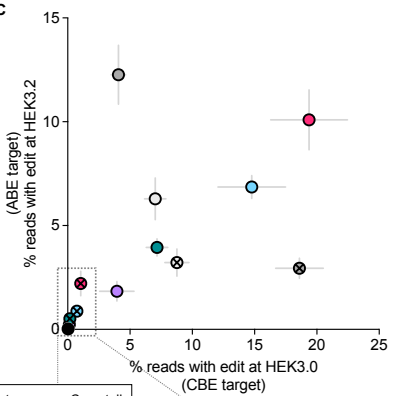
D

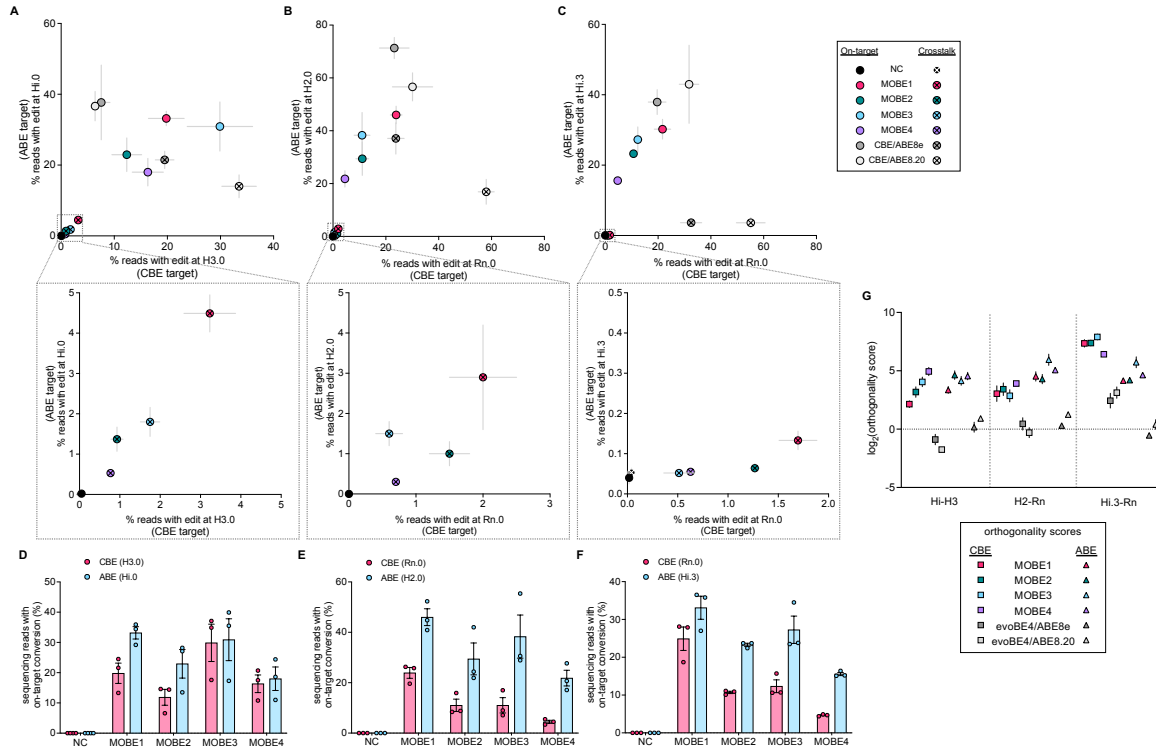


B



C





**Figure 3.2: On-target and crosstalk editing efficiencies of MOBEs and parental BEs when targeted to two protospacers at distinct chromosomes.**

(A-F) Editing efficiencies by the MOBE1-4 systems and the evoBE4/ABE8e and evoBE4/ABE8.20 combinations at the *HIRA.0/HEK3.0* (A and D), *HEK2.0/RNF2.0* (B and E), and *HIRA.3/RNF2.0* (C and F) loci. HEK293T cells were treated as previously described in Figure 3.1. (A-C) Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target (y-axis, filled circles), C•G edited to T•A at the ABE target (y-axis, circles with X inset), C•G edited to T•A at the CBE target (x-axis, filled circles), and A•T edited to G•C at the CBE target (x-axis, circles with X inset). Dot plots and error bars represent the average and SEM for n=3 biological replicates. (D-F) Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target and C•G edited to T•A at the CBE target (on-target efficiencies only). Bar graphs and error bars represent the average and SEM for n=3 biological replicates (each replicate is marked individually). Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA. (G) Orthogonality scores for the MOBE1-4 systems and the evoBE4/ABE8e and evoBE4/ABE8.20 combinations at the three protospacer combinations from (A) through (C). Plotted are the log<sub>2</sub> of the “CBE orthogonality scores”, which we define for a given protospacer combination as the percent of total DNA sequencing reads with C•G edited to T•A at the CBE target divided by the percent of total DNA sequencing reads with C•G edited to T•A at the ABE target. The “ABE orthogonality scores” are the A•T to G•C editing efficiency equivalent. Dot plots and error bars represent the average and propagation of uncertainty of the SEM for n=3 biological replicates.

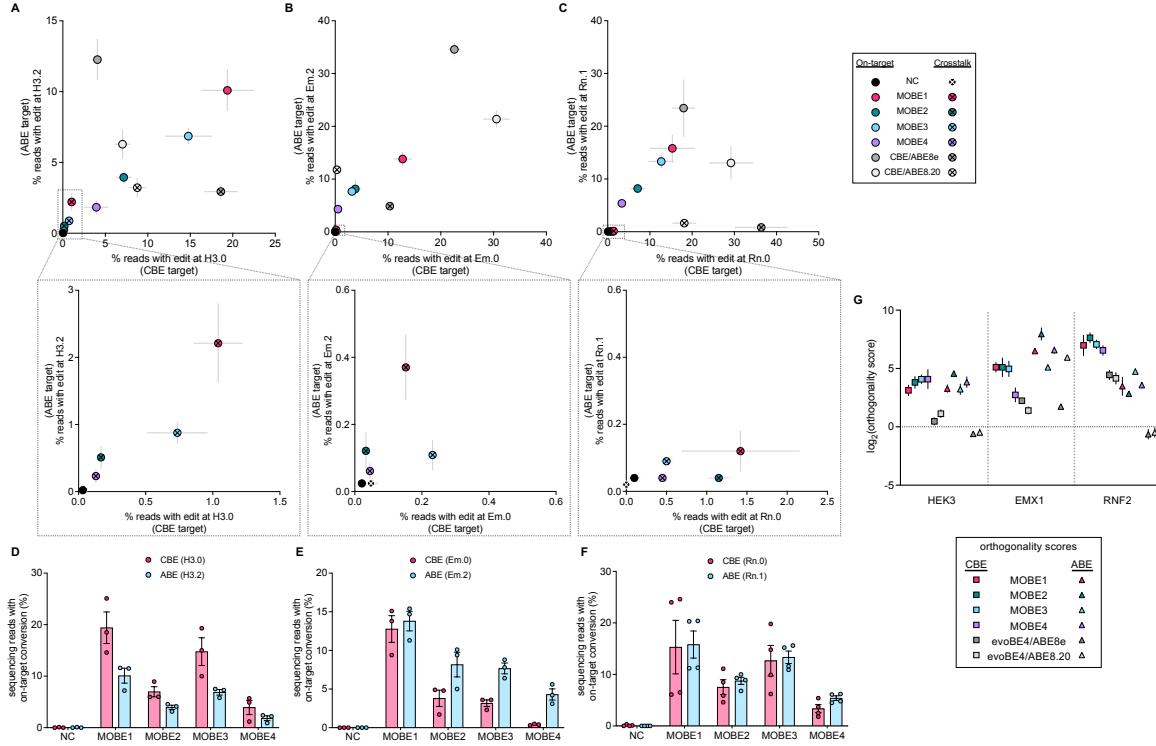


Next, we targeted two protospacers within a single genomic region to represent modelling tandem SNVs such as those inherited in haplotype blocks or observed in *cis*-regulatory elements. This approach also allows the quantification of co-occurring point mutations by targeted amplicon NGS. We selected three protospacer combinations (at the *EMX1*, *HEK3*, and *RNF2* loci) and quantified multiplexed orthogonal editing efficiencies at two protospacers within each locus by the four MOBE systems, as well as the two parental non-orthogonal systems. While we generally observed lower levels of on-target activity by MOBEs compared to their parental counterparts at these three site combinations (Figure 3.3A-F), we again detected drastically decreased crosstalk editing compared to the parental constructs. The crosstalk editing for the MOBE systems averaged  $0.45 \pm 0.11\%$  (ranging from  $0.03 \pm 0.01\%$  to  $2.2 \pm 0.6\%$ ) compared to the parental systems that averaged  $9.6 \pm 3.0\%$  crosstalk editing (ranging from  $0.32 \pm 0.01\%$  to  $36.4 \pm 6.2\%$ , Figure 3.1C and Figure 3.3A-C, dotted line insets). The highest on-target editing was typically facilitated by MOBE1 or MOBE3 (the system with the absolute highest on-target activity was site-dependent), followed by MOBE2 and then MOBE4. Specifically, at the *HEK3* site combination (PAM-out orientation, 168 bp apart), MOBE1 introduced on-target editing efficiencies of  $19.4\% \pm 3.1\%$  C•G to T•A and  $10.1 \pm 1.4\%$  A•T to G•C, with crosstalk editing of  $2.2 \pm 0.6\%$  C•G to T•A and  $1.0 \pm 0.2\%$  A•T to G•C (Figure 3.1C and Figure 3.3A). In contrast, the parental evoBE4-NG/ABE8e-NG combination introduced on-target editing efficiencies of  $4.1 \pm 0.6\%$  C•G to T•A and  $12.3 \pm 1.4\%$  A•T to G•C, with crosstalk editing of  $2.9 \pm 0.5\%$  C•G to T•A and  $18.6 \pm 1.9\%$  A•T to G•C (Figure 3.1C and Figure 3.3A).

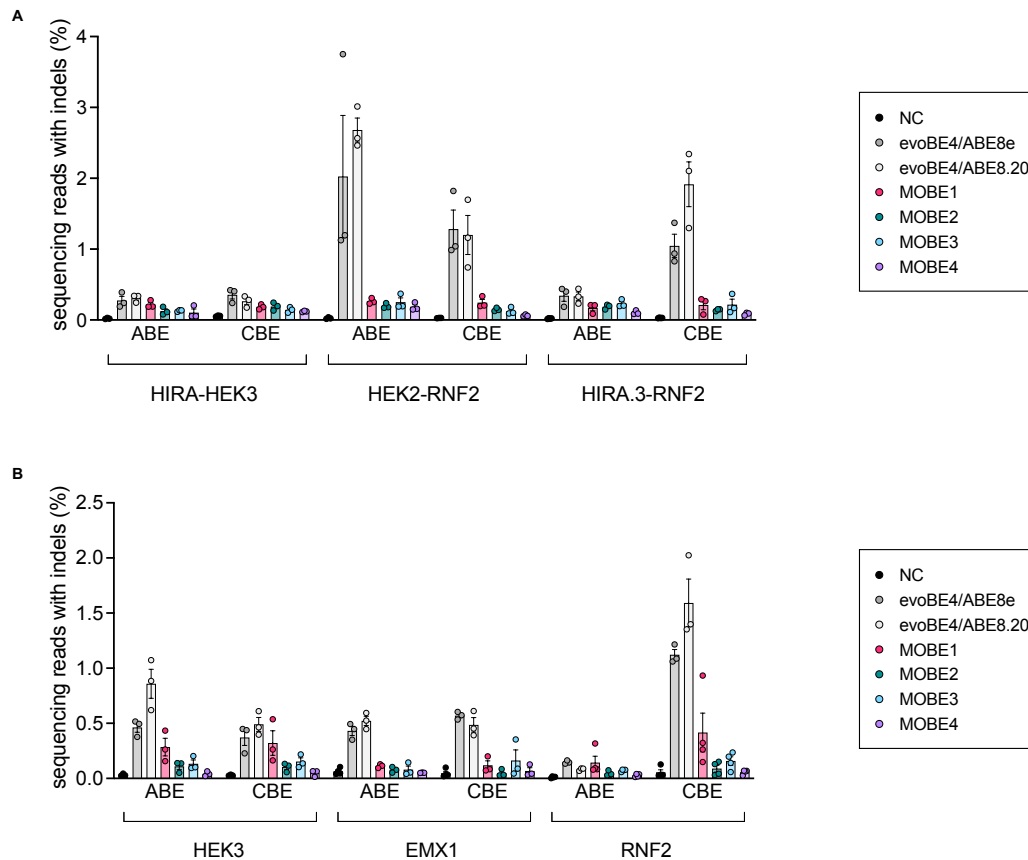
We quantified the indel rates of all four MOBE systems and their parental editors at both the CBE and ABE protospacers independently, for all six protospacer combinations. Consistent with previous reports, the parental base editors created low but detectable amounts of indels (Figure 3.4). Across the six protospacer combinations, the evoBE4/ABE8e combination introduced an average of 0.79% at the CBE site and 0.61% indels at the ABE site, while the evoBE4/ABE.20 combination introduced an average 0.99% indels at the CBE site and 0.80%

indels at the ABE site (Figure 3.4). At all sites tested, indel rates for all MOBE systems were substantially lower than the parental editors. MOBE1 displayed the highest amount of indels averaging 0.20% at the ABE targets and 0.25% at the CBE targets. Compared to its parental evoBE4/ABE8e combination, MOBE1 had a 2.5-fold lower indel rate at the ABE sites despite only a 1.6-fold lower on-target activity, and a 1.8-fold lower indel rate at the CBE sites with a 1.1-fold higher on-target activity. MOBE3 indel rates were 0.15% at the ABE targets (2.7-fold lower than ABE8e) and 0.16% at the CBE targets (3.9-fold lower than evoBE4). MOBE4 showed the lowest amount of indels, up to 7.3-fold lower than its parental editors (Figure 3.4).

We next defined the “CBE orthogonality score” of an editing system for a given protospacer combination as the ratio of the on-target C•G to T•A editing efficiency to the crosstalk C•G to T•A editing efficiency, with the “ABE orthogonality score” being the A•T to G•C editing efficiency equivalent. Thus, higher scores correspond to higher orthogonality, while scores near 1 represent equal levels of on-target and crosstalk activity. Our MOBE systems displayed a median CBE orthogonality score of 23.8, which ranged from  $4.4 \pm 0.9$  to  $239 \pm 41$  for the six total protospacer combinations tested (Figure 3.1D, 3.2G, and 3.3G). The median ABE orthogonality score of our MOBEs was 23.1, with individual ABE orthogonality scores ranging from  $7.1 \pm 1.0$  to  $249 \pm 89$ . (Figure 3.1D, 3.2G, and 3.3G). The parental evoBE4/ABE8e and evoBE4/ABE8.20 combinations displayed a median CBE orthogonality score of 2.4 (which ranged from  $0.30 \pm 0.03$  to  $21.8 \pm 6.1$ ), and a median ABE orthogonality score of 1.2 (ranging from  $0.64 \pm 0.18$  to  $61 \pm 6$ , Figure 3.1D, 3.2G, and 3.3G). When directly comparing each MOBE’s orthogonality score to its parental orthogonality score for the same protospacer combination, the four MOBE systems were found to collectively be on average 19-fold more orthogonal than their respective parental system. Finally, the MOBE3 system typically displayed the highest orthogonality scores, with median CBE and ABE orthogonality scores of 24.1 and 30.5, respectively (Figure 3.1D, 3.2G, and 3.3G).



**Figure 3.3: On-target and crosstalk editing efficiencies of MOBEs and parental BEs when targeted to two protospacers within the same locus.** (A-F) Editing efficiencies by the MOBE1-4 systems and the evoBE4/ABE8e and evoBE4/ABE8.20 combinations at the *HEK3* (A and D), *EMX1* (B and E), and *RNF2* (C and F) loci. Cells were treated as previously described in Figure 3.1. (A-C) Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target (y-axis, solid circles), C•G edited to T•A at the ABE target (y-axis, circles with X inset), C•G edited to T•A at the CBE target (x-axis, solid circles), and A•T edited to G•C at the CBE target (x-axis, circles with X inset). Dot plots and error bars represent the average and SEM for n=3 biological replicates. (D-F) Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target and C•G edited to T•A at the CBE target (on-target efficiencies only). Bar graphs and error bars represent the average and SEM for n=3 biological replicates at *EMX1* and *HEK3* or n=4 biological replicates at *RNF2* (each replicate is marked individually). Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA. (G) Orthogonality scores for the MOBE1-4 systems and the evoBE4/ABE8e and evoBE4/ABE8.20 combinations at the three protospacer combinations from (A) through (C). Plotted are the log<sub>2</sub> of the “CBE orthogonality scores”, which we define for a given protospacer combination as the percent of total DNA sequencing reads with C•G edited to T•A at the CBE target divided by the percent of total DNA sequencing reads with C•G edited to T•A at the ABE target. The “ABE orthogonality scores” are the A•T to G•C editing efficiency equivalent. Dot plots and error bars represent the average and propagation of uncertainty of the SEM for n=3 biological replicates at *EMX1* and *HEK3* or n=4 biological replicates at *RNF2*.



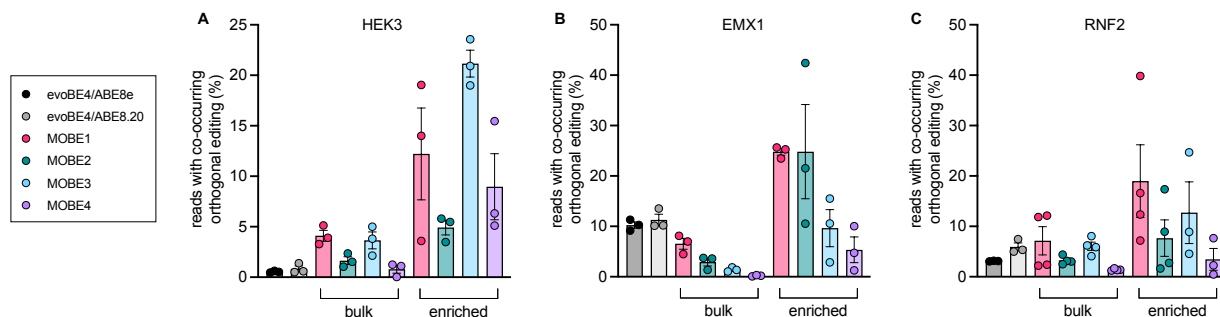
**Figure 3.4: Quantification of indel rates for parental and MOBE systems.** (A-B) Indel introduction rates by the MOBE1-4 systems and the evoBE4/ABE8e and evoBE4/ABE8.20 combinations at the *HIRA.0/HEK3.0*, *HEK2.0/RNF2.0*, *HIRA.3/RNF2.0* (A) distinct loci protospacer combinations and the *HEK3*, *EMX1*, and *RNF2* (B) single-amplicon protospacer combinations. Cells were treated as previously described in Figure 3.1. Plotted are the percent of total DNA sequencing reads with indels at each target site. Bar graphs and error bars represent the average and SEM for  $n=3$  biological replicates (for most sites) or  $n=4$  biological replicates at *RNF2* (each replicate is marked individually). Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA.

### 3.1.2 Quantification of co-occurring orthogonal edits by MOBEs

We then analyzed the genotypes of the three single-amplicon targets (Supplementary Figures 3.1-3.3). We were particularly interested in the percent of reads with simultaneous orthogonal on-target edits at both protospacers without crosstalk (which we refer to as reads with co-occurring orthogonal edits, Supplementary Figures 3.1-3.3). We found that at the *HEK3* and *RNF2* combinations, MOBE1 and MOBE3 generated the largest percent of reads with co-

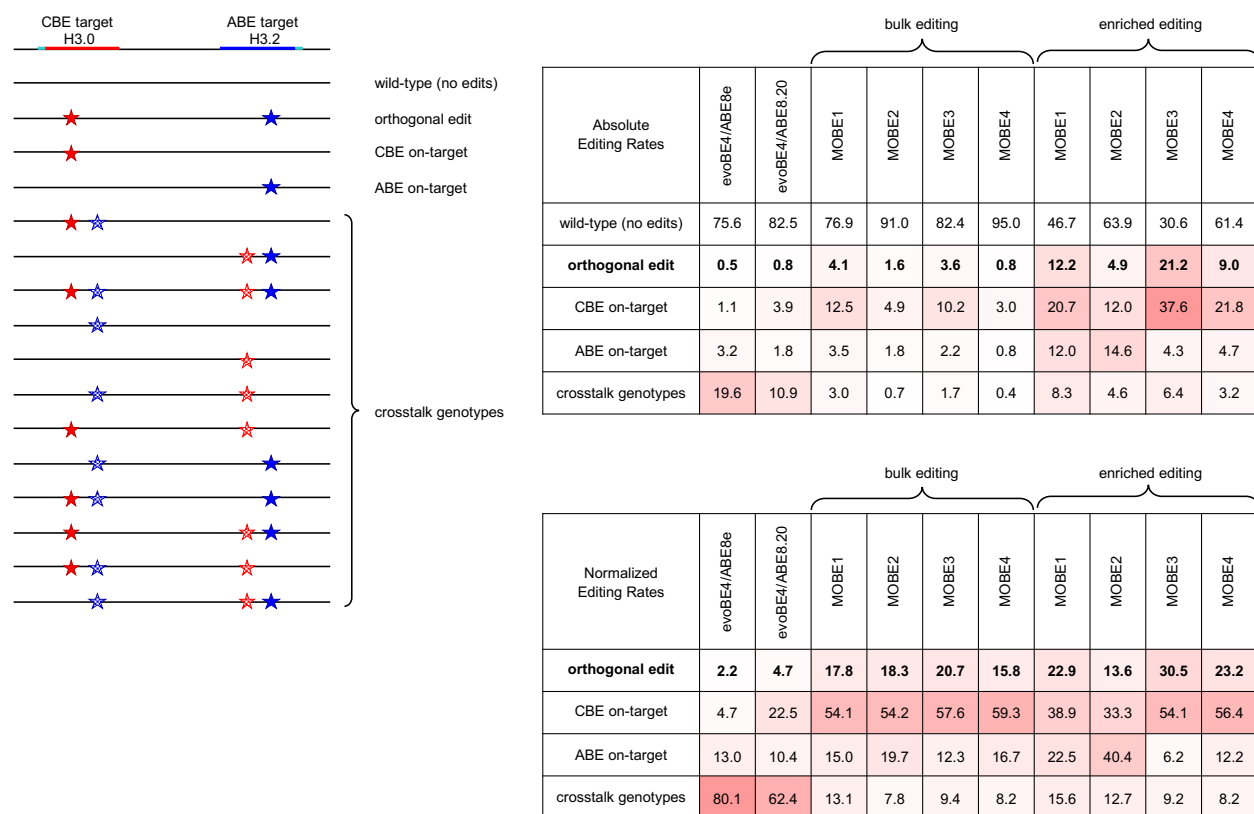
occurring orthogonal edits, while MOBE4 generated the lowest. In particular, at the *HEK3* site combination (of which both protospacers contain both a highly edited A and C base, Figure 3.3A),  $4.1 \pm 0.5\%$  of reads from the MOBE1-treated cells and  $3.6 \pm 0.8\%$  of reads from the MOBE3-treated cells contained co-occurring orthogonal edits, compared to  $0.54 \pm 0.05\%$  for their parental editor evoBE4/ABE8e (Supplementary Figure 3.1 and Figure 3.3A). At the *RNF2* site combination (of which the *RNF.1* ABE target protospacer lacks a highly edited C, making this site more naturally prone to orthogonality, Figure 3.3C),  $7.1 \pm 2.8\%$  of reads from the MOBE1-treated cells and  $6.1 \pm 0.8\%$  of reads from the MOBE3-treated cells contained co-occurring orthogonal edits, compared to  $3.1 \pm 0.0\%$  for their evoBE4/ABE8e parental editor (Supplementary Figure 3.3 and Figure 3.3C). Additionally, at all sites we observed drastic decreases in the percent of reads corresponding to genotypes with crosstalk edits compared to the parental editors, even when normalized to overall editing efficiencies (Supplementary Figures 3.1-3.3). Specifically, at the *HEK3* combination, in addition to having a 7-fold increase in the absolute level of co-occurring orthogonal edits, only  $9.4 \pm 1.4\%$  of edited reads by MOBE3 had genotypes with crosstalk editing, while  $80.1 \pm 11.0\%$  of reads from cells treated with the evoBE4/ABE8e parental system had crosstalk genotypes. Similarly, at the *RNF2* combination, in addition to having double the absolute level of co-occurring orthogonal edits compared to its parental editor, only  $6.9 \pm 1.0\%$  of edited reads by MOBE3 had genotypes with crosstalk editing, while  $83.3 \pm 2.1\%$  of reads from cells treated with the evoBE4/ABE8e parental system had crosstalk genotypes, demonstrating the enhanced precision of our MOBE systems. This high level of crosstalk editing for the evoBE4/ABE8e parental system at the *RNF2* site is responsible for the low rate of co-occurring orthogonal edits despite significantly higher rates of on-target editing (as assessed in bulk) compared to the MOBE systems (Figure 3.3C, 3.5C, and Supplementary Figure 3.3). At the *EMX1* combination, the MOBE systems reached up to  $6.5 \pm 1.1\%$  co-occurring orthogonal edits but had significantly decreased on-target editing efficiencies

compared to the parental systems. We therefore sought to develop a strategy to enrich for co-occurring orthogonal edits.



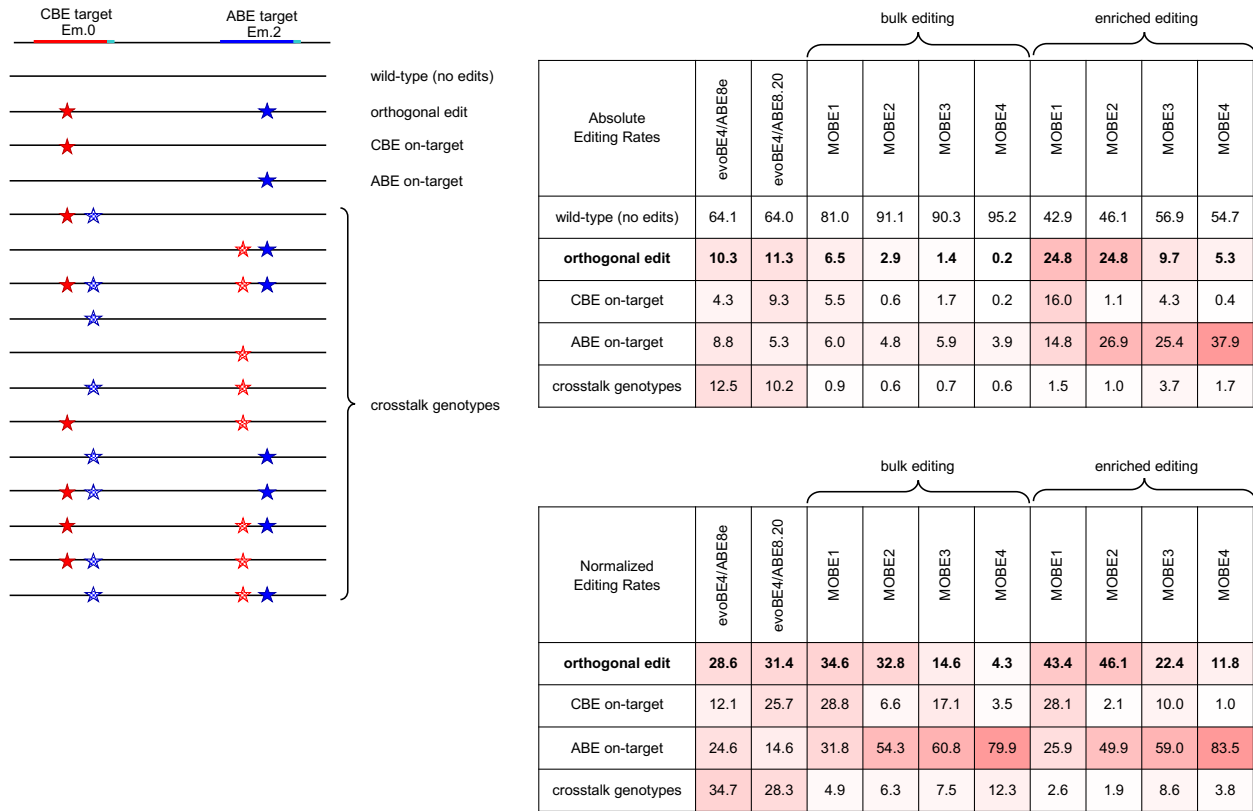
**Figure 3.5: Quantification of co-occurring orthogonal edits.**

Cells were treated as previously described in Figure 3.1 and 3.2. Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target and C•G edited to T•A at the CBE target, with no crosstalk edits when all multiplexing systems are targeted to two protospacers within the HEK3 (A), EMX1 (B), or RNF2 (C) locus. Bar graphs and error bars represent the average and SEM for n=3 biological replicates at EMX1 and HEK3 or n=4 biological replicates at RNF2 (each replicate is marked individually).



**Supplementary Figure 3.1: Genotype analysis of multiplexed base editing at the *HEK3* locus.**

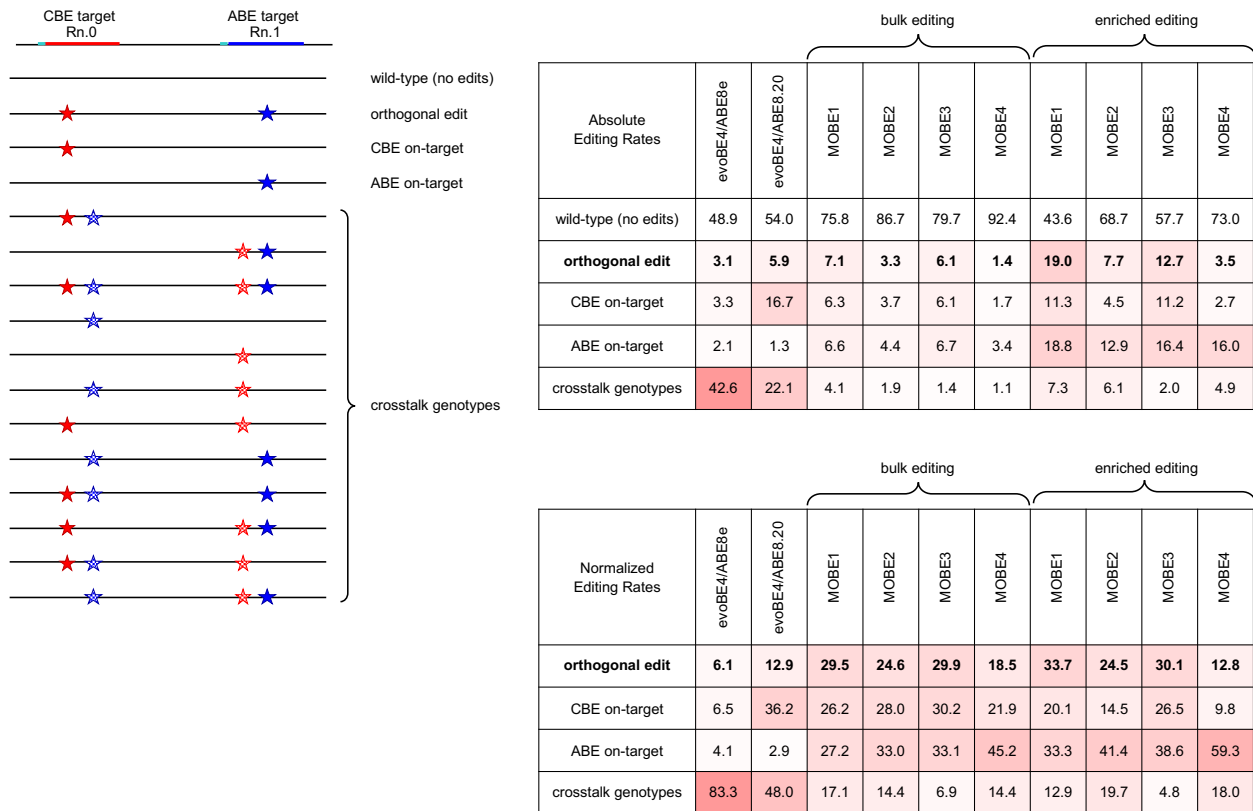
Cells were treated as previously described in Figure 3.1 (parental systems and “bulk” samples), or HEK293T cells were transfected with plasmids encoding nCas9-NG-P2A-mCherry, tandem CP-deaminase fusions, both gRNA-aptamers targeting endogenous loci, and the 2x-dead-GFP reporter (Figure 3.7A). After 96 hours, populations of GFP+/mCherry+ “enriched” cells were collected by fluorescence activated cell sorting (FACS). Genomic loci of interest were then amplified and analyzed by NGS (“enriched” samples). Plotted are the percent of total DNA sequencing reads with the genotypes illustrated on the left when all indicated multiplexing systems are targeted to two protospacers within the *HEK3* locus (absolute editing rates, top), or the listed genotypes of interest were normalized to overall editing rates (normalized editing rates, bottom). Numerical values represent the average for n=3 biological replicates.



**Supplementary Figure 3.2: Genotype analysis of multiplexed base editing at the *EMX1* locus.**

Cells were treated as previously described in Supplementary Figure 3.1. Plotted are the percent of total DNA sequencing reads with the genotypes illustrated on the left when all indicated multiplexing systems are targeted to two protospacers within the *EMX1* locus (absolute editing rates, top), or the listed genotypes of interest were normalized to overall editing rates (normalized editing rates, bottom). Numerical values represent the average for n=3 biological replicates.





**Supplementary Figure 3.3: Genotype analysis of multiplexed base editing at the *RNF2* locus.**

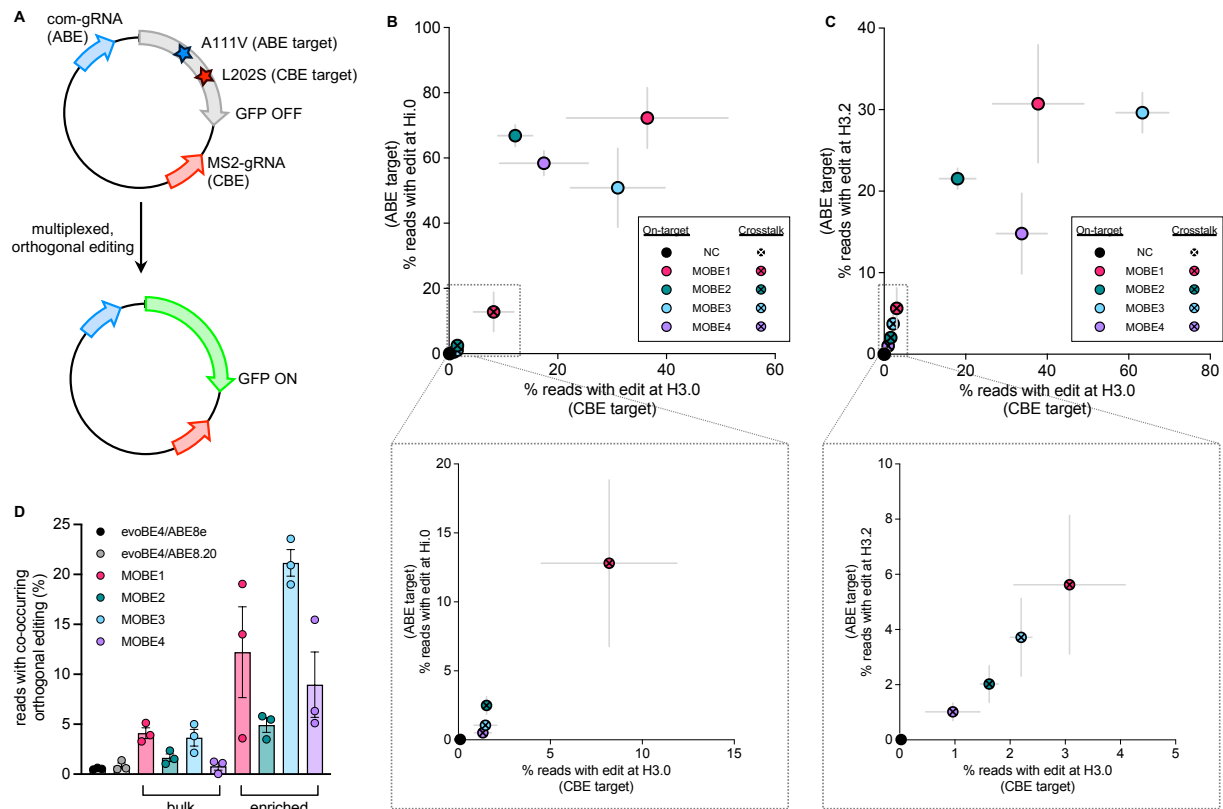
Cells were treated as previously described in Supplementary Figure 3.1. Plotted are the percent of total DNA sequencing reads with the genotypes illustrated on the left when all indicated multiplexing systems are targeted to two protospacers within the *RNF2* locus (absolute editing rates, top), or the listed genotypes of interest were normalized to overall editing rates (normalized editing rates, bottom). Numerical values represent the average for n=3 biological replicates.

### 3.1.3 A MOBE fluorescent reporter enriches for co-occurring edits

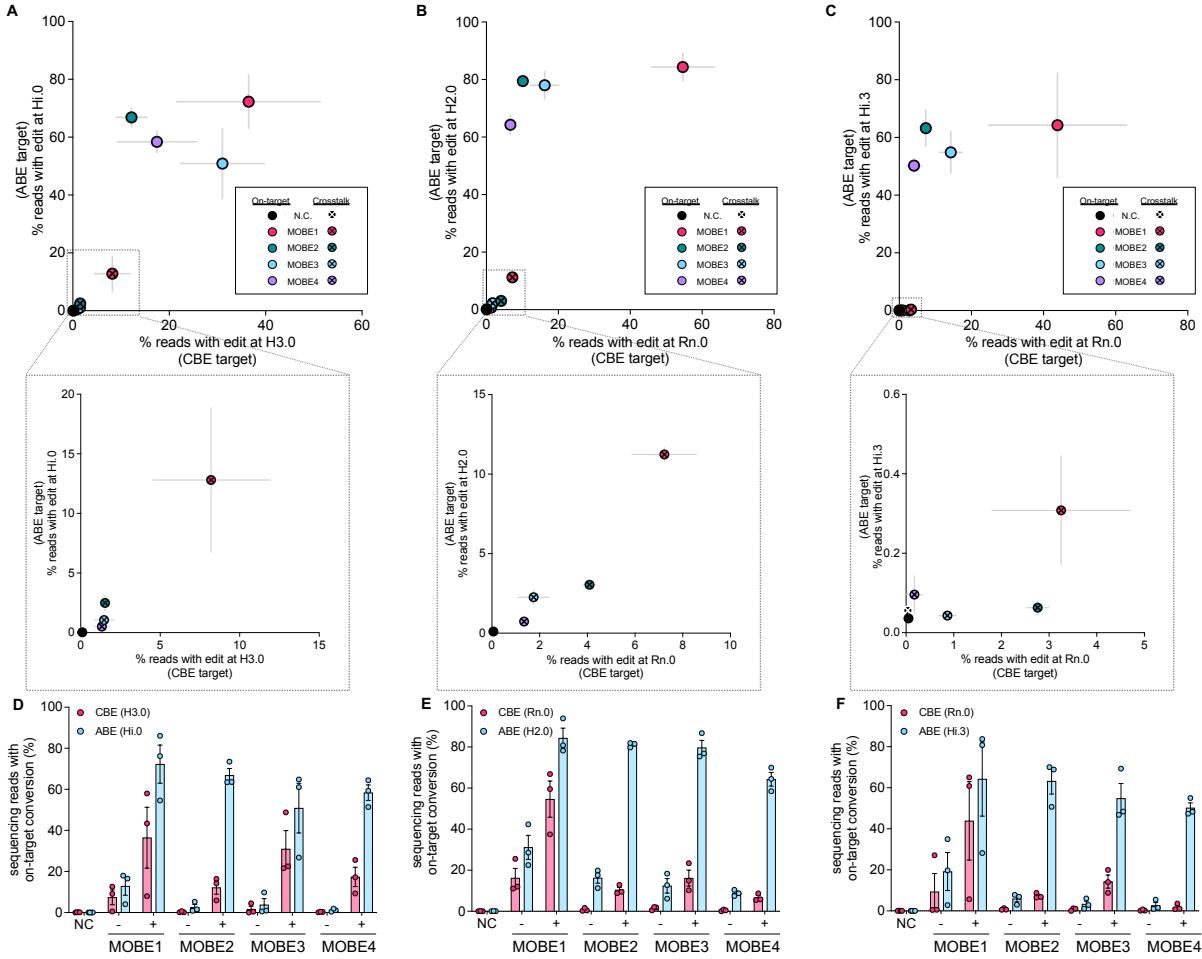
Because the MOBE systems exhibited lower on-target activity than their parental counterparts, we designed a plasmid-based enrichment system to enrich for orthogonal editing activity when co-transfected with genomic-targeting gRNAs. The reporter plasmid encodes a “2x-dead-EGFP” gene that monitors successful editing by a MOBE system. The “2x-dead-EGFP” gene harbors two missense mutations (A111V and L202S)<sup>69,70</sup> that require orthogonal A•T to G•C and C•G to T•A editing to produce functional, fluorescent GFP. MOBE components are directed to correct these mutations by two gRNA-aptamer fusions that are also expressed

from the reporter plasmid (Figure 3.6A). HEK293T cells were transfected with MOBEs, gRNAs targeted to the same six genomic site combinations used previously, and the reporter plasmid. After 96 hours, populations of both “unenriched” cells and GFP+/mCherry+ “enriched” cells were collected by fluorescence activated cell sorting (FACS) and editing efficiencies were determined by NGS. Both GFP-targeting protospacers had both a target A and C within the editing window, and crosstalk editing would result in additional amino acid changes to the GFP protein. This precluded the use of this enrichment system with the parental systems.

MOBE1 typically exhibited the highest enriched editing, averaging 40.0% on-target C•G to T•A and 55.7% on-target A•T to G•C editing across the six protospacer combinations (Figure 3.3-3.7). The highest C•G to T•A editing observed was  $63.3 \pm 6.4\%$  by MOBE3 at the *HEK3.0* protospacer (Figure 3.8A), and the highest A•T to G•C editing we observed was  $81.5 \pm 6.5\%$  by MOBE1 at the *HEK2.0* protospacer (Figure 3.3B). We observed increases in editing efficiencies ranging from  $4.9 \pm 2.6$  to  $34.0 \pm 11.5$ -fold for MOBE3 when comparing enriched cells to unenriched cells from the same sample (Supplementary Figure 3.4A), clearly demonstrating the utility of this enrichment strategy. It is important to note that we observed lower editing efficiencies in unenriched cells in these experiments compared to the previous bulk editing MOBE experiments, likely due to resource competition and an extra expansion step we employed to ensure there would be enough enriched cells for NGS analysis (see Methods).<sup>71</sup> Importantly, this increase in editing efficiency was not accompanied by a decrease in orthogonality scores; both ABE and CBE orthogonality scores for a given protospacer combination was within error of unenriched samples (Supplementary Figure 3.4B). Like the unenriched samples, indel rates were the highest for MOBE1 averaging 1.03% at the CBE site and 0.74% at the ABE site across the six protospacer combinations, while MOBE4 displayed the lowest level of indels averaging less than 0.37% at either target site (Supplementary Figure 3.4C-D).



**Figure 3.6: On-target and crosstalk editing efficiencies of MOBEs when performing multiplexed editing followed by FACS enrichment for episomal edits.** (A) Schematic diagram of the 2x-dead-GFP reporter. GFP fluorescence is only observed following successful multiplexed, orthogonal base editing within the GFP gene. GFP<sup>+</sup> cells can then be sorted to enrich for cells with highly active base editors, which enhances editing at endogenous genomic loci. (B and C) Editing efficiencies by the MOBE1-4 systems at the *HIRA.0/HEK3.0* (B) and *HEK3.2/HEK3.0* (C) loci. HEK293T cells were transfected with plasmids encoding nCas9-NG-P2A-mCherry, tandem CP-deaminase fusions, both gRNA-apartamers targeting endogenous loci, and the 2x-dead-GFP reporter. After 96 hours, populations of both “unenriched” cells and GFP<sup>+</sup>/mCherry<sup>+</sup> “enriched” cells were collected by FACS. Genomic loci of interest were then amplified and analyzed by NGS. Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target (y-axis, filled circles), C•G edited to T•A at the ABE target (y-axis, circles with X inset), C•G edited to T•A at the CBE target (x-axis, filled circles), and A•T edited to G•C at the CBE target (x-axis, circles with X inset). Dot plots and error bars represent the average and standard deviation for n=3 biological replicates. Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA. (D) Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target and C•G edited to T•A at the CBE target, with no crosstalk edits, when all multiplexing systems are targeted to two protospacers within the *HEK3* locus. The parental systems and the “bulk” samples were treated as previously described in Figure 3.1. Bar graphs and error bars represent the average and SEM for n=3 biological replicates (each replicate is marked individually).

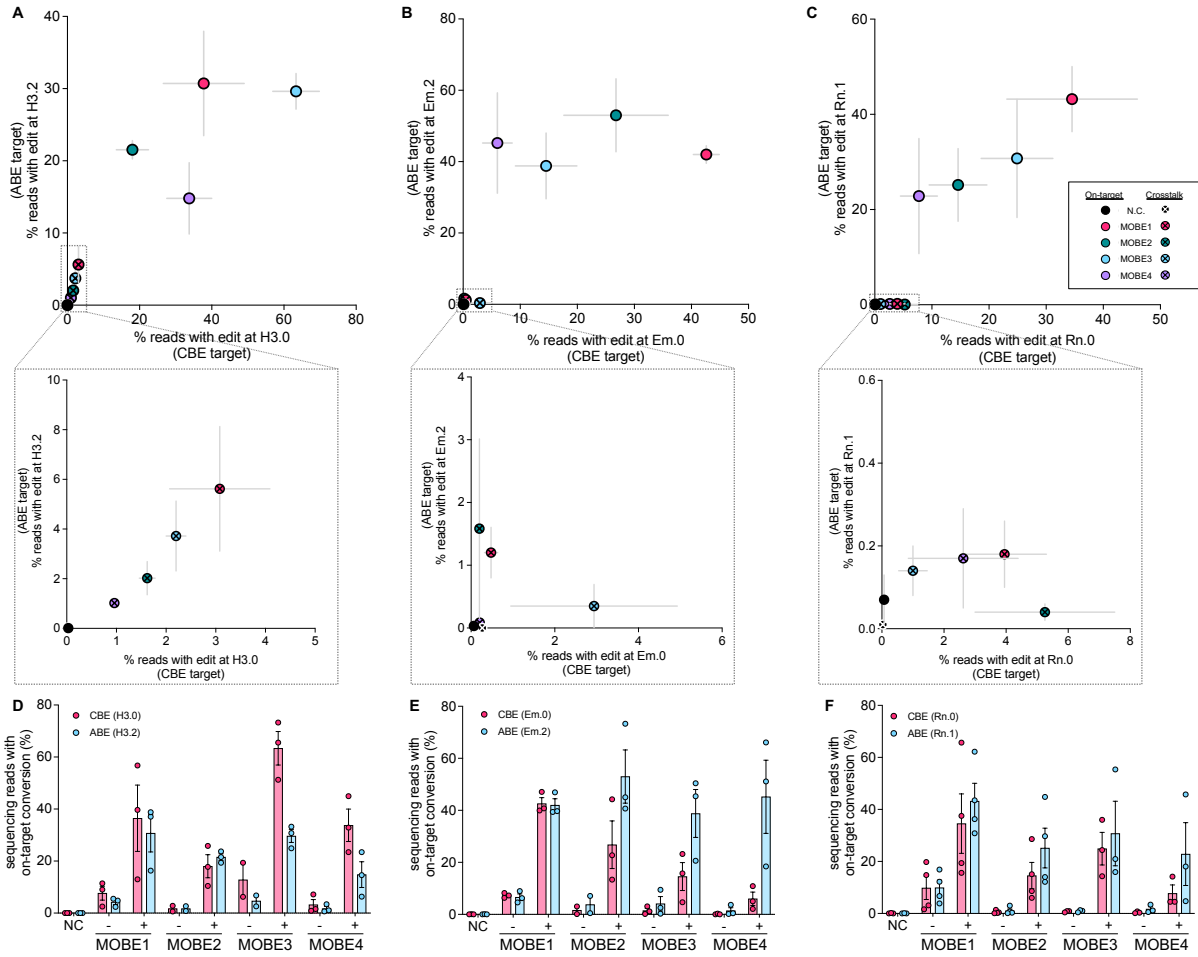


**Figure 3.7: On-target and crosstalk editing efficiencies of MOBEs when targeted to two protospacers at distinct chromosomes followed by FACS enrichment for episomal edits.**

(A-F) Editing efficiencies by the MOBE1-4 systems at the *HIRA.0/HEK3.0* (A and D), *HEK2.0/RNF2.0* (B and E), and *HIRA.3/RNF2.0* (C and F) loci. HEK293T cells were treated as previously described in Figure 3.6. (A-C) Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target (y-axis, solid circles), C•G edited to T•A at the ABE target (y-axis, circles with X inset), C•G edited to T•A at the CBE target (x-axis, solid circles), and A•T edited to G•C at the CBE target (x-axis, circles with X inset). Dot plots and error bars represent the average and SEM for n=3 biological replicates. (D-F) Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target and C•G edited to T•A at the CBE target (on-target efficiencies only). Bar graphs and error bars represent the average and propagation of uncertainty of the SEM for n=3 biological replicates (each replicate is marked individually). Negative control (NC) samples are also shown, in which cells were not transfected with any plasmid.

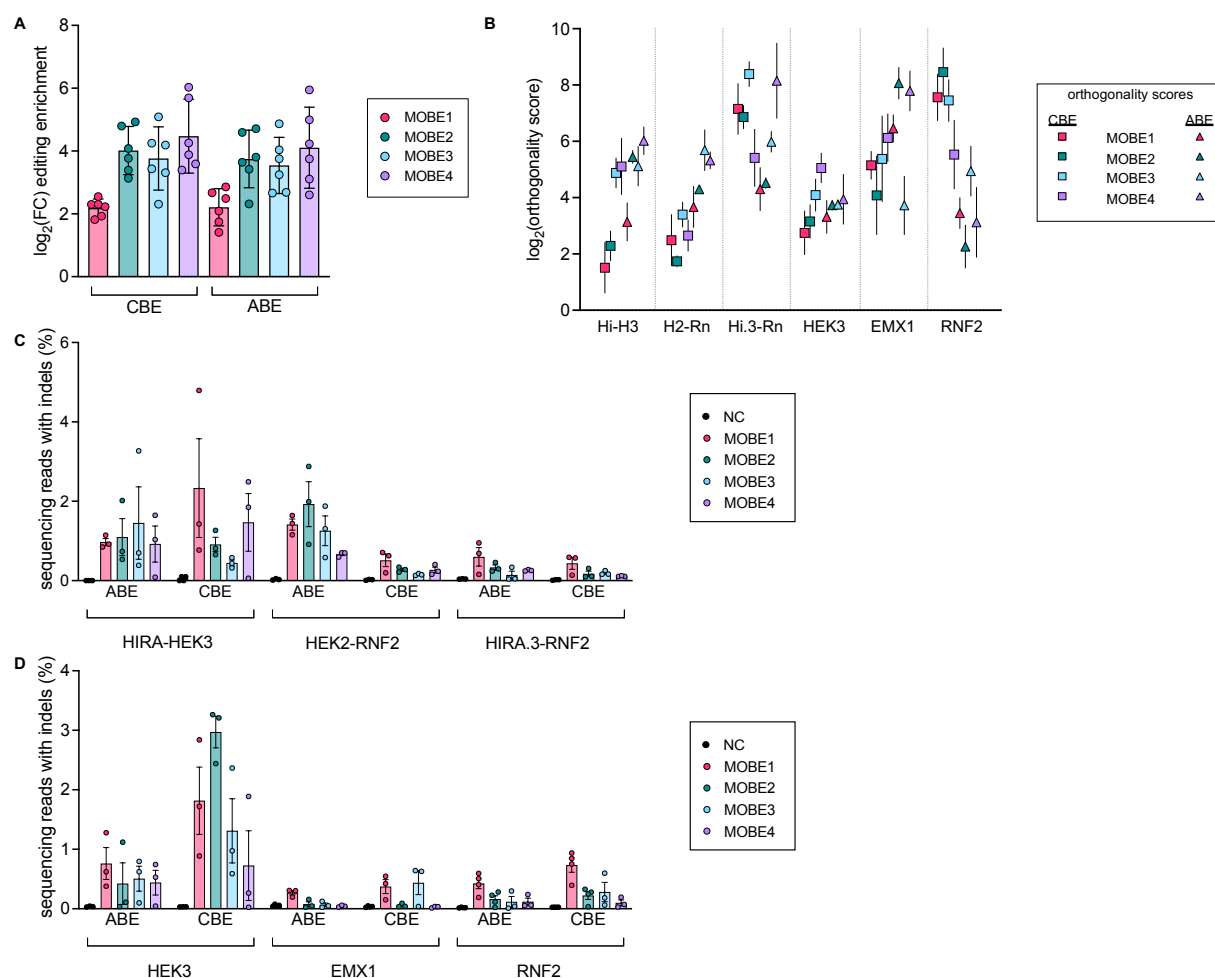
We again analyzed the genotypes of the three single-amplicon target sites and observed up to a 26-fold increase in the percent of reads with co-occurring orthogonal edits compared to

bulk samples (Supplementary Figures 3.1-3.3 and Figure 3.2). The highest incidence of co-occurring orthogonal edits was  $24.8 \pm 9.4\%$  facilitated by MOBE2 at the *EMX1* site (previously, co-occurring orthogonal editing rates were the lowest at this site, ranging from  $0.2 \pm 0.1\%$  to  $6.5 \pm 1.1\%$ , Figure 3.2). When normalized to overall editing efficiencies, MOBE2 at the *EMX1* site combination had the highest precision as well, with 46.1% of edited reads containing the desired genotype (co-occurring orthogonal edits) and only 1.9% of edited reads containing crosstalk genotypes (the remaining edited reads had either C•G to T•A editing at the CBE target only, or A•T to G•C editing at the ABE target only, Supplementary Figure 3.2). MOBE1 efficiently installed co-occurring orthogonal edits at *EMX1* ( $24.8 \pm 0.7\%$ ) and *RNF2* ( $19.0 \pm 7.2\%$ ), while MOBE3 facilitated the highest absolute rates of co-occurring edits at *HEK3* ( $21.2 \pm 1.3\%$ ), with >30% of edited reads containing the desired genotype at both sites (co-occurring orthogonal edits). All MOBE systems at all three single-amplicon targets demonstrated from 2.3- to 21.0-fold lower rates of reads with crosstalk genotypes compared to their respective parental systems. Collectively, our MOBE systems combined with this enrichment strategy facilitated up to a 39-fold increase in the absolute rate of co-occurring orthogonal edits compared to their respective parental systems (MOBE3 at the *HEK3* site). These data clearly indicate that the MOBE reporter plasmid is an effective option for enriching for cells harboring co-occurring orthogonal edits, without a decrease in precision. We imagine this strategy will be useful when utilizing MOBEs to generate cell lines or when editing poorly edited targets.



**Figure 3.8: On-target and crosstalk editing efficiencies of MOBEs when targeted to two protospacers within the same locus followed by FACS enrichment for episomal edits.**

(A-F) Editing efficiencies by the MOBE1-4 systems at the *HEK3* (A and D), *EMX1* (B and E), and *RNF2* (C and F) loci. Cells were treated as previously described in Figure 3.6. (A-C) Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target (y-axis, solid circles), C•G edited to T•A at the ABE target (y-axis, circles with X inset), C•G edited to T•A at the CBE target (x-axis, solid circles), and A•T edited to G•C at the CBE target (x-axis, circles with X inset). Dot plots and error bars represent the average and SEM for n=3 biological replicates at *EMX1* and *HEK3* or n=4 biological replicates at *RNF2*. (D-F) Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target and C•G edited to T•A at the CBE target (on-target efficiencies only). Bar graphs and error bars represent the average and SEM for n=3 biological replicates at *EMX1* and *HEK3* or n=4 biological replicates at *RNF2* (each replicate is marked individually). Negative control (NC) samples are also shown, in which cells were not transfected with any plasmid.



**Supplementary Figure 3.4: Orthogonality scores, increases in on-target editing efficiencies, and indel rates after enrichment with the fluorescent reporter.**

(A) Cells were treated as previously described in Figure 3.6. Plotted are the on-target enrichment values, which are the  $\log_2$  of the percent of total DNA sequencing reads with C•G edited to T•A at the CBE target of the “enriched” cells divided by that of the “unenriched” cells, or the A•T to G•C editing efficiency equivalent. Each dot represents the average of  $n=3$  biological replicates for a given protospacer (each MOBE has twelve individual enrich values). Bar graphs and error bars represent the average and propagation of uncertainty of the standard deviation for a given MOBE system. (B) Orthogonality scores for the MOBE1-4 systems at all six protospacer combinations tested after enrichment using the fluorescent reporter. Plotted are the  $\log_2$  of the “CBE orthogonality scores”, which we define for a given protospacer combination as the percent of total DNA sequencing reads with C•G edited to T•A at the CBE target divided by the percent of total DNA sequencing reads with C•G edited to T•A at the ABE target. The “ABE orthogonality scores” are the A•T to G•C editing efficiency equivalent. Dot plots and error bars represent the average  $\log_2$  orthogonality scores and propagation of uncertainty of the SEM for  $n=3$  biological replicates. (C-D) Plotted are the percent of total DNA sequencing reads with indels at each target site. Dot plots and error bars represent the average and propagation of uncertainty of the SEM for  $n=3$  biological replicates.

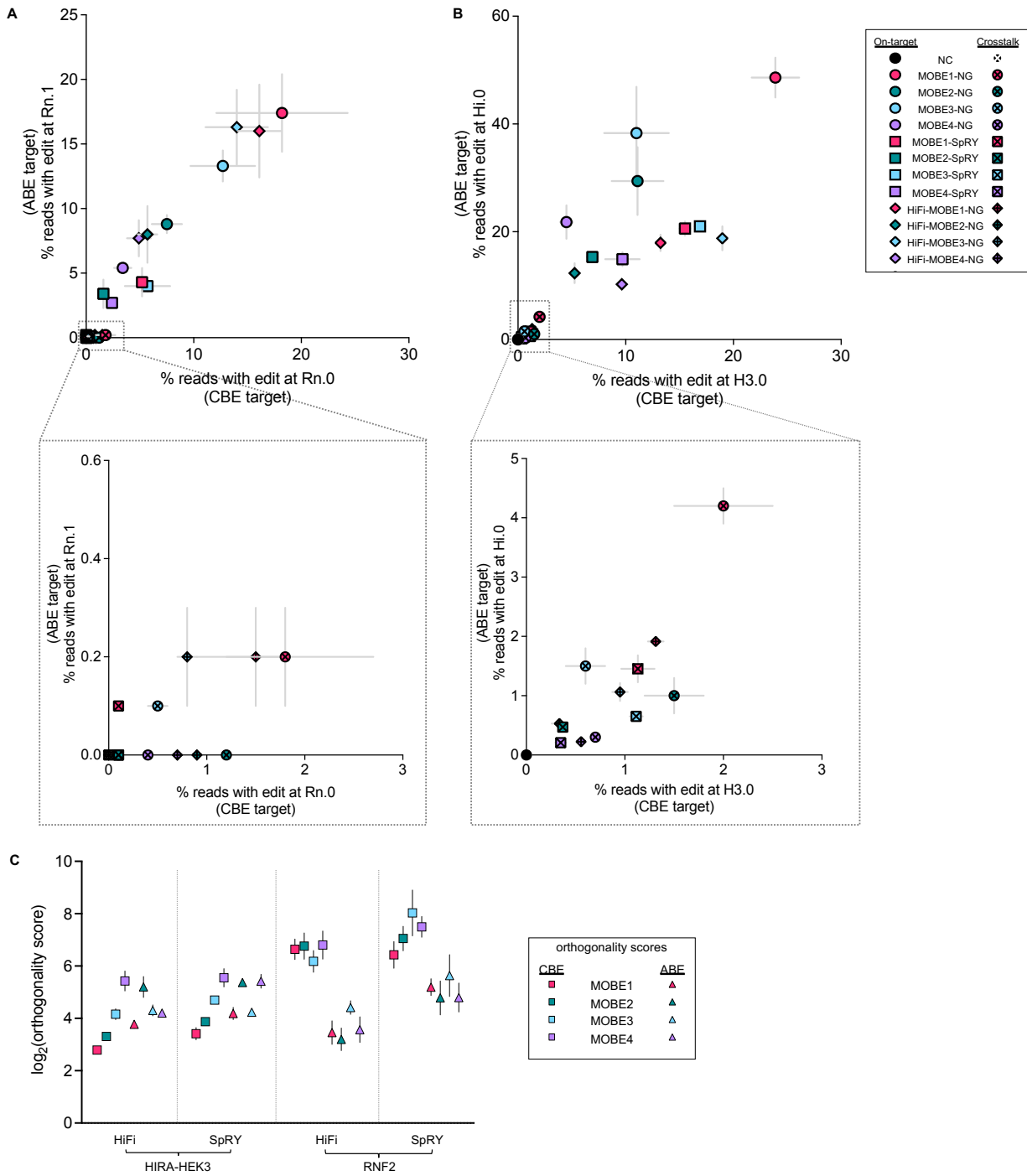
### 3.1.4 MOBE compatibility with other nCas9 variants and cell types

As the MOBE systems are modular, they can be easily modified to implement additional Sp-nCas9 variants. Specifically, the targeting scope can be expanded by using the near PAM-less SpRY nCas9 variant, which can be especially useful with the relatively narrow editing windows of the MOBE systems.<sup>42</sup> Additionally, MOBEs can be used with high-fidelity nCas9 variants such as HiFi to mitigate gRNA-dependent off-target effects.<sup>72</sup> We therefore evaluated editing efficiencies by all four MOBE systems at the *RNF2* single-amplicon protospacer combination and the *HIRA.0/HEK3.0* protospacer combination when used with nCas9-NG, nCas9-SpRY, or HiFi-nCas9 (Supplementary Figure 3.5). For the HiFi-nCas9 MOBE systems, we observed roughly equivalent editing activity compared to the corresponding nCas9-NG MOBE systems (1.1-fold higher for CBE and 1.1-fold lower for ABE). For the nCas9-SpRY MOBE systems we observed slightly reduced editing activity (1.3-fold lower for CBE and 2.1-fold lower for ABE) at the two protospacer combinations tested. Orthogonality scores followed similar trends as the nCas9-NG MOBE systems (Supplementary Figure 3.5C compared to Extended Data Figures 1G and 2G). Specifically, the nCas9-SpRY MOBE systems displayed a median CBE orthogonality score of 66 and a median ABE orthogonality score of 32, compared to 58 (median CBE orthogonality score) and 16 (median ABE orthogonality score) for HiFi-nCas9-NG, and 63 (median CBE orthogonality score) and 15 (median ABE orthogonality score) for nCas9-NG at these two site combinations (Supplementary Figure 3.5C). These experiments demonstrate the modularity of the MOBE systems for use with additional Sp-nCas9 variants.



### Supplementary Figure 3.5: MOBEs are compatible with other Sp-nCas9 variants.

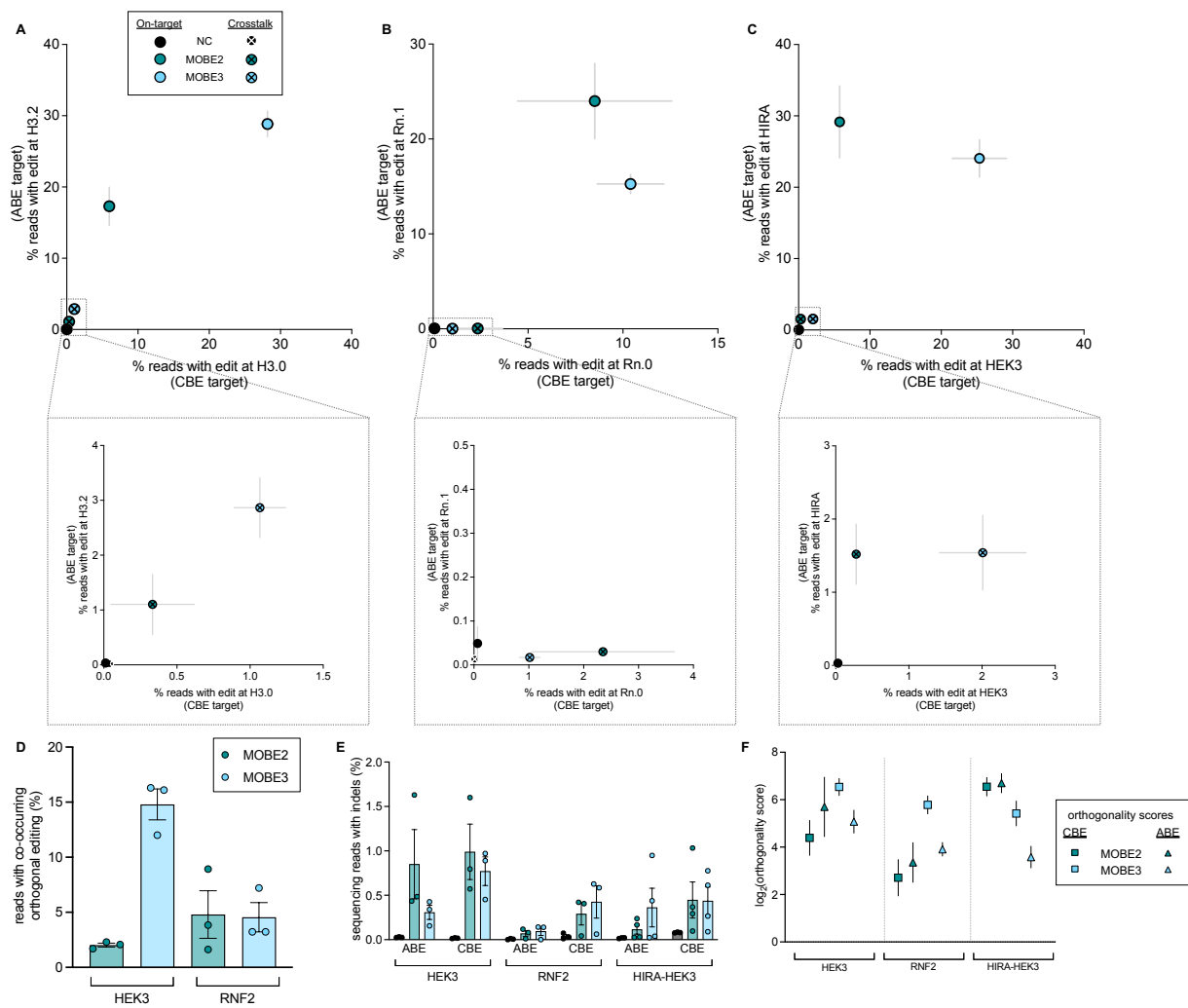
(A-B) Editing efficiencies by the MOBE1-4 systems when used with various Sp-nCas9 variants at the *HIRA.0/HEK3.0* (A) and *RNF2* single-amplicon (B) protospacer combinations. HEK293T cells were transfected with plasmids encoding tandem CP-deaminase fusions, both gRNA-aptamers, and either nCas9-NG-P2A-mCherry, SpRY-nCas9-P2A-mCherry, or HiFi-nCas9-NG-P2A-mCherry. Cells were then lysed after 72 hours, genomic loci of interest were amplified, and analyzed by NGS. Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target (y-axis, filled shapes), C•G edited to T•A at the ABE target (y-axis, shapes with X inset), C•G edited to T•A at the CBE target (x-axis, filled shapes), and A•T edited to G•C at the CBE target (x-axis, shapes with X inset). (C) Orthogonality scores for the MOBE systems at the two protospacer combinations from (A) and (B). Plotted are the  $\log_2$  of the “CBE orthogonality scores”, which we define for a given protospacer combination as the percent of total DNA sequencing reads with C•G edited to T•A at the CBE target divided by the percent of total DNA sequencing reads with C•G edited to T•A at the ABE target. The “ABE orthogonality scores” are the A•T to G•C editing efficiency equivalent. Dot plots and error bars represent the average and SEM for n=3 biological replicates. Negative control (NC) samples are also shown, in which cells were not transfected with any plasmid.



We next sought to evaluate the MOBE systems in additional mammalian cell types. We selected MOBE2 and MOBE3 (which employ different adenosine deaminase domains and different apt-CBE architectures) to evaluate in HeLa cells at the *RNF2* and *HEK3* single amplicon protospacer combinations as well as the *HIRA.0/HEK3.0* protospacer combination. HeLa cells were transfected with either of the two MOBE systems and the GFP enrichment plasmid. After 72 hours, GFP+/mCherry+ cells were collected by FACS, and editing efficiencies were determined by NGS. We observed similar trends in the relative on-target editing efficiencies of MOBE3 compared to MOBE2 and observed on-target editing efficiencies that ranged from 33% to 97% of those observed at the same genomic loci in HEK293T cells following this enrichment strategy. Specifically, the MOBE2 system averaged 6.7% C•G to T•A conversion and 23.5% A•T to G•C conversion across these three sites, and the MOBE3 system averaged 21.3% C•G to T•A conversion and 22.7% A•T to G•C conversion across these three sites (Figure 3.9A-C). These decreases in on-target editing efficiencies are likely due to the much lower transfection and reporter turn-on efficiency (which directly correlates to genomic editing efficiency) of HeLa cells compared to HEK293T cells.<sup>73</sup> Despite this fact, MOBE3 achieved up to  $14.8 \pm 1.4\%$  co-occurring orthogonal edits at the *HEK3* single-amplicon target (Figure 3.9D). Indel rates for MOBEs were also low in HeLa cells, with MOBE3 averaging 0.53% at CBE targets and 0.27% at ABE targets across the three sites tested (Figure 3.9E). Overall, orthogonality scores in HeLa cells were comparable to those in HEK293T cells as well (Figure 3.9F).

**Figure 3.9: MOBEs can be used in HeLa cells.**

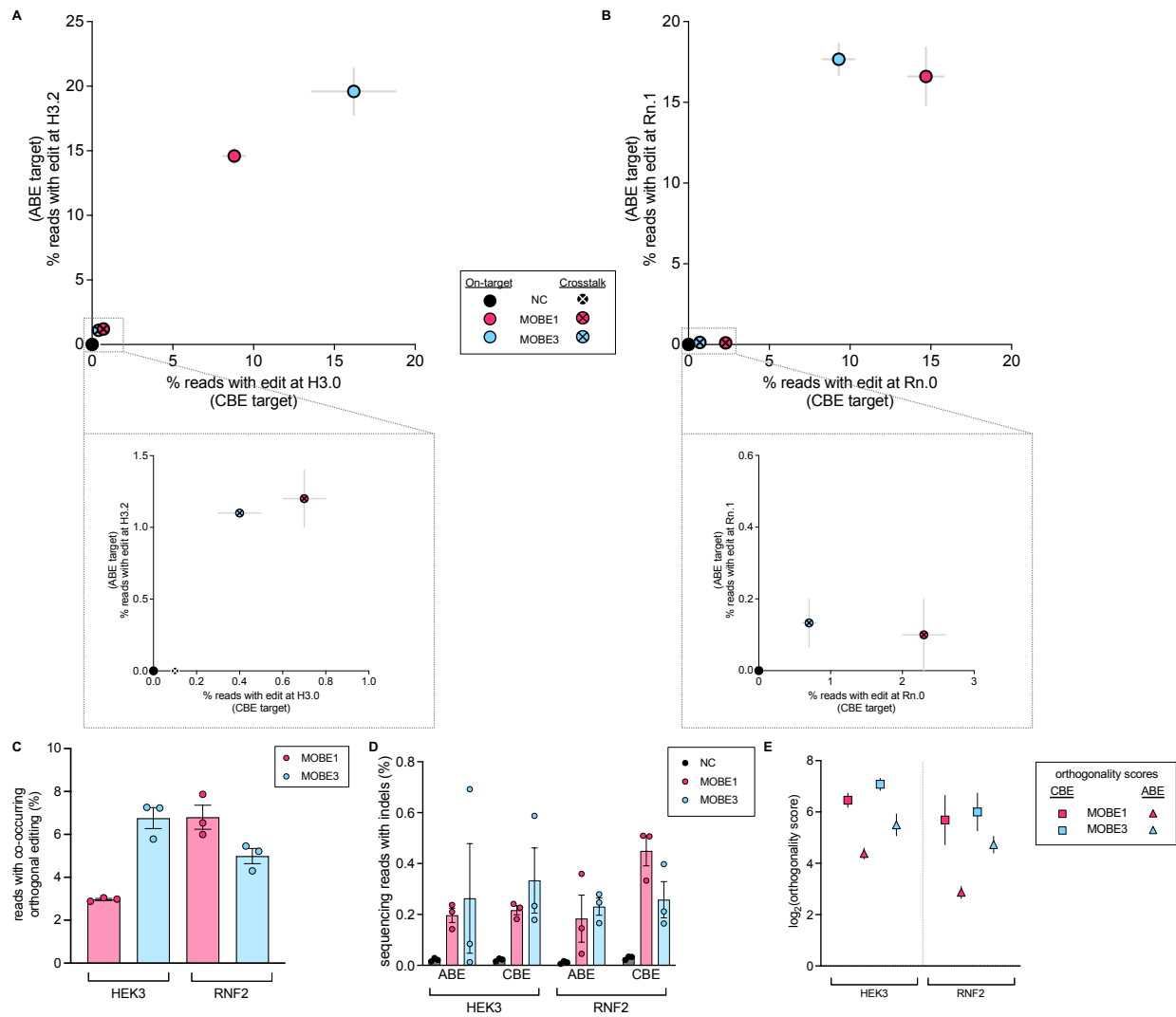
(A-C) Editing efficiencies by the MOBE2 and MOBE3 systems at the *RNF2* (A), *HEK3* (B), and *HIRA.0/HEK3.0* (C) protospacer combinations. HeLa cells were transfected with plasmids encoding nCas9-NG-P2A-mCherry, tandem CP-deaminase fusions, both gRNA-aptamers targeting endogenous loci, and the 2x-dead-GFP reporter. After 72 hours, the population of GFP+/mCherry+ “enriched” cells was collected by FACS. Genomic loci of interest were then amplified and analyzed by NGS. Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target (y-axis, solid circles), C•G edited to T•A at the ABE target (y-axis, circles with X inset), C•G edited to T•A at the CBE target (x-axis, solid circles), and A•T edited to G•C at the CBE target (x-axis, circles with X inset). (D) Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target and C•G edited to T•A at the CBE target, with no crosstalk edits when all multiplexing systems are targeted to two protospacers within the *HEK3* or *RNF2* loci. (E) Plotted are the percent of total DNA sequencing reads with indels at each target site. (F) Orthogonality scores for the MOBE systems at the two protospacer combinations from (A) through (C). Plotted are the  $\log_2$  of the “CBE orthogonality scores”, which we define for a given protospacer combination as the percent of total DNA sequencing reads with C•G edited to T•A at the CBE target divided by the percent of total DNA sequencing reads with C•G edited to T•A at the ABE target. The “ABE orthogonality scores” are the A•T to G•C editing efficiency equivalent. Dot plots and error bars represent the average and standard deviation for n=3 biological replicates. Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA.



We then selected MOBE1 and MOBE3 (the highest efficiency systems that employ different apt-CBE architectures) to evaluate in SH-SY5Y cells at the *RNF2* and *HEK3* single-amplicon protospacer combinations. SH-SY5Y cells were electroporated with either of the two MOBE systems, and after 72 hours, mCherry+ cells (marker for transfection) were collected by FACS and editing efficiencies were determined by NGS. We observed similar overall on-target editing efficiencies in SH-SY5Y cells compared to unsorted experiments in HEK293T cells. Across these two target combinations, the MOBE1 system averaged 11.7% C•G to T•A conversion and 15.6% A•T to G•C conversion, and the MOBE3 system averaged 12.8% C•G to T•A conversion and 18.6% A•T to G•C conversion (Figure 3.10A-B). SH-SY5Y cells treated with MOBE1 displayed  $6.8 \pm 0.6\%$  of reads contained co-occurring orthogonal edits at the *RNF2* target, while MOBE3-treated cells produced  $6.8 \pm 0.5\%$  of reads with co-occurring orthogonal edits at the *HEK3* target (Figure 3.10C). Indel rates remained low in SH-SY5Y cells and orthogonality scores were comparable to other cell types (Figure 3.10D-E). These data demonstrate that MOBE systems can be utilized in additional cell types beyond HEK293T cells.

**Figure 3.10: MOBEs can be used in SH-SY5Y cells.**

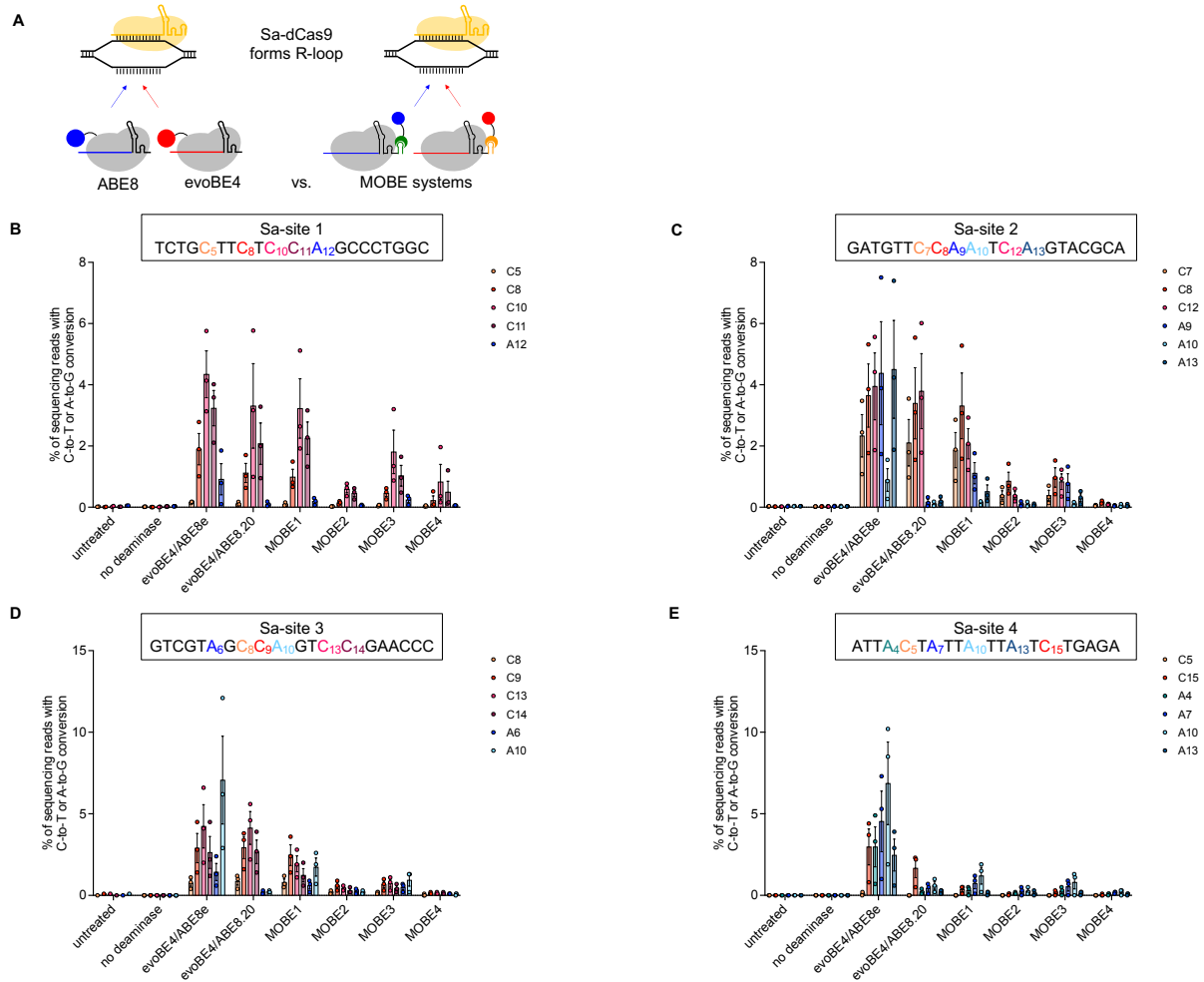
(A-B) Editing efficiencies by the MOBE1 and MOBE3 systems at the *HEK3* (A), and *RNF2* (B) single-amplicon protospacer combinations. SH-SY5Y cells were electroporated with plasmids encoding nCas9-NG-P2A-mCherry, tandem CP-deaminase fusions, and both gRNA-aptamers. After 72 hours, the population of transfected (mCherry+) cells was collected by FACS. Genomic loci of interest were then amplified and analyzed by NGS. Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target (y-axis, solid circles), C•G edited to T•A at the ABE target (y-axis, circles with X inset), C•G edited to T•A at the CBE target (x-axis, solid circles), and A•T edited to G•C at the CBE target (x-axis, circles with X inset). (C) Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the ABE target and C•G edited to T•A at the CBE target, with no crosstalk edits when all multiplexing systems are targeted to two protospacers within the *HEK3* or *RNF2* loci. (D) Plotted are the percent of total DNA sequencing reads with indels at each target site. (E) Orthogonality scores for the MOBE systems at the two protospacer combinations from (A) and (B). Plotted are the  $\log_2$  of the “CBE orthogonality scores”, which we define for a given protospacer combination as the percent of total DNA sequencing reads with C•G edited to T•A at the CBE target divided by the percent of total DNA sequencing reads with C•G edited to T•A at the ABE target. The “ABE orthogonality scores” are the A•T to G•C editing efficiency equivalent. Dot plots and error bars represent the average and standard deviation for n=3 biological replicates. Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA.





### 3.1.5 Off-target analysis of MOBEs compared to their parental BEs

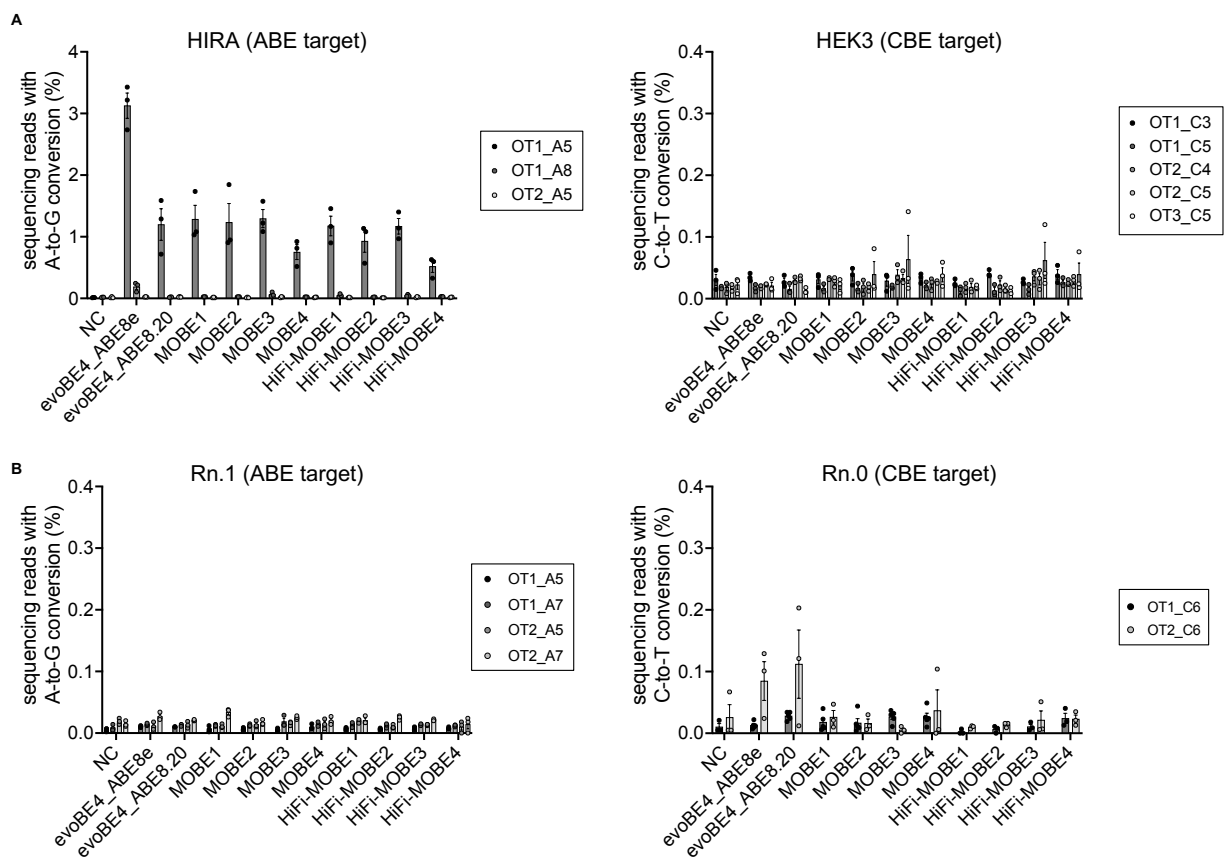
We then evaluated the off-target editing propensities of our four MOBE systems as compared to the parental, non-orthogonal combinations. We first evaluated the gRNA-independent off-target DNA editing activities of all MOBE and parental systems using an orthogonal R-loop assay (Figure 3.11A, see Methods) at four different Sa-targeted genomic loci that contain both target Cs and As within the exposed R-loop.<sup>67,74</sup> At all four sites, gRNA-independent off-target editing of both Cs and As was highest by MOBE1, followed by MOBE3, MOBE2, and MOBE4 (Figure 3.11B-E). gRNA-independent off-target editing of Cs by all four MOBE systems (which share the same cytidine deaminase domain as both parental systems) ranged from  $1.3 \pm 0.5$ -fold to  $39.6 \pm 13.4$ -fold lower than that of the parental constructs (Figure 3.11B-E). gRNA-independent off-target editing of As by MOBE1 and MOBE3 (which share the same adenosine deaminase domain as the parental ABE8e construct) ranged from  $3.8 \pm 2.5$ -fold to  $14.4 \pm 7.7$ -fold lower than that of the parental evoBE4/ABE8e combination, and gRNA-independent off-target editing of As by MOBE2 and MOBE4 (which share the same adenosine deaminase domain as the parental ABE8.20 construct) ranged from  $1.3 \pm 0.5$ -fold to  $4.2 \pm 1.3$ -fold lower than that of the parental evoBE4/ABE8.20 combination (Figure 3.11B-E).



**Figure 3.11: Evaluation of gRNA-independent off-target DNA editing by MOBEs compared to parental systems.** (A) Schematic of the orthogonal R-loop assay used to evaluate gRNA-independent off-target editing. The Sa-dCas9:gRNA complex binds to a genomic locus of interest and exposes a stretch of ssDNA via the formation of an R-loop. The deaminase components of additional BE complexes within the cell (either the MOBE systems or the parental ABE/CBE complexes) can access the exposed ssDNA and edit Cs or As within this region, which is quantified with NGS. (B-E) HEK293T cells were transfected with plasmids encoding nCas9-NG-P2A-mCherry, tandem CP-deaminase fusions, and both gRNA-aptamers (for the MOBE systems), or plasmids encoding the parental evoBE4, ABE8e or ABE8.20, and both unmodified gRNAs (for the evoBE4/ABE8e and evoBE4/ABE8.20 combinations), as well as Sa-dCas9 and Sa-gRNA targeted to the four protospacer sequences shown. Cells were then lysed after 72 hours, genomic loci of interest were amplified, and analyzed by NGS. Plotted are the percent of total DNA sequencing reads with C•G edited to T•A at the target Cs indicates, and A•T edited to G•C at the target As indicated for each of the four Sa-dCas9 protospacers indicated. Bar graphs represent the average, and error bars represent the SEM for n=3 biological replicates, with each replicate marked individually. Untreated samples are also shown, in which cells were non-transfected. Negative control (no deaminase) samples are also shown, in which cells were treated identically as the MOBE systems, but with omission of the tandem CP-deaminase fusion plasmid.

We next evaluated gRNA-dependent off-target editing of the four MOBE systems as compared to the parental systems. We chose two of our most efficiently edited protospacer combinations (the *RNF2* single amplicon protospacer combination and the *HIRA.0/HEK3.0* protospacer combination) and identified at least two putative off-target sites for each of the four protospacers. The *HEK3.0* protospacer has been previously experimentally evaluated for off-target editing;<sup>75</sup> we chose the top three off-target sites for evaluation. For the other three protospacers we used the cutting frequency determination score calculation to identify the top two putative off-target sites and evaluated editing at these two sites.<sup>76</sup> HEK293T cells were transfected with each of the four MOBEs (utilizing either nCas9-NG or HiFi-nCas9-NG) or the parental ABE/CBE combinations, lysed after three days, and the off-target genomic loci were amplified and sequenced with NGS.

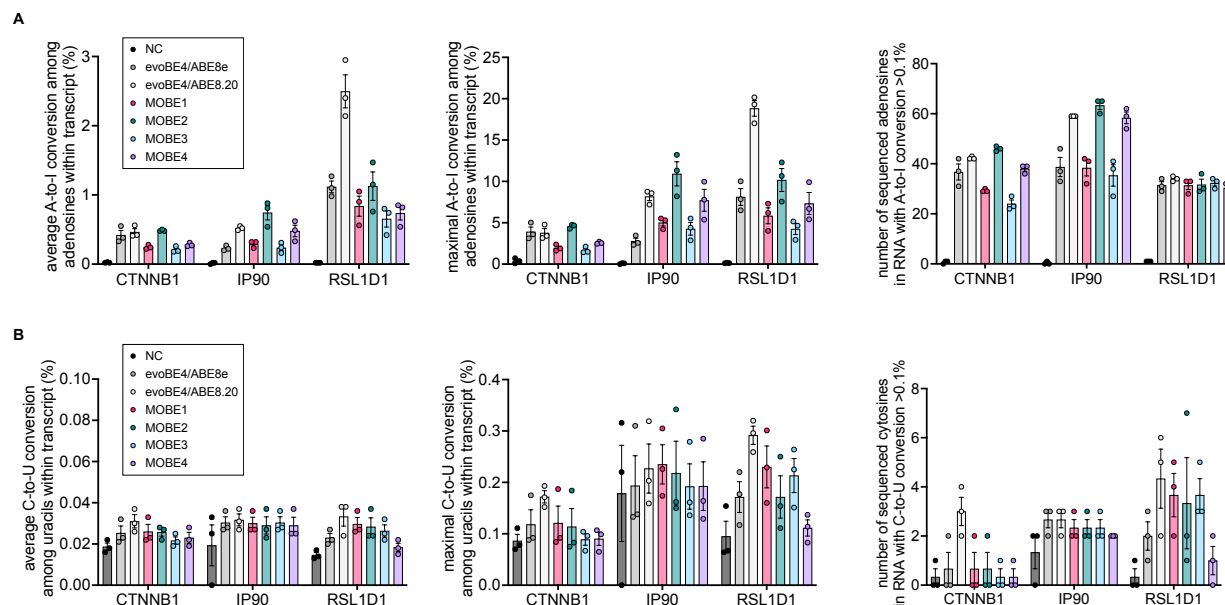
We observed editing above background levels at only one of the *HIRA* and *RNF2* off-target loci. At the *HIRA* off-target locus (an ABE target), we found that MOBE2 displayed off-target A•T to G•C editing efficiencies within error of its parental ABE8.20 system (for both the MOBE systems utilizing the nCas9-NG variant as well as the HiFi-nCas9 variant) while MOBE4 showed  $1.6 \pm 0.4$ -fold less and HiFi-MOBE4 showed  $2.3 \pm 0.7$ -fold less editing than ABE.20 (Supplementary Figure 3.6A). Meanwhile, MOBE1 and MOBE3 (which share the ABE8e deaminase) displayed off-target A•T to G•C editing efficiencies ranging from  $2.4 \pm 0.3$ -fold to  $2.7 \pm 0.3$ -fold lower than their respective parental system at the *HIRA* off-target locus (for both the MOBE systems utilizing the nCas9-NG variant as well as the HiFi-nCas9 variant, Supplementary Figure 3.6A). Off-target C•G to T•A editing efficiencies were below background levels at all sites tested except at one *RNF2* off-target site, at which both parental systems showed off-target editing levels, but all MOBEs were below background levels (Supplementary Figure 3.6B).



**Supplementary Figure 3.6: Evaluation of gRNA-dependent off-target DNA editing by MOBEs compared to parental systems.**

HEK293T cells were transfected with plasmids encoding nCas9-NG-P2A-mCherry or HiFi-nCas9-NG-mCherry, tandem CP-deaminase fusions, and both gRNA-aptamers (for the MOBE systems), or plasmids encoding the parental evoBE4-NG, ABE8e-NG or ABE8.20-NG, and both unmodified gRNAs (for the evoBE4/ABE8e and evoBE4/ABE8.20 combinations), targeted to the *HIRA.0/HEK3.0* protospacer combination (A) or the *RNF2* single amplicon protospacer combination (B). Cells were then lysed after 72 hours, genomic loci of interest were amplified, and analyzed by NGS. Plotted are the percent of total DNA sequencing reads with A•T edited to G•C at the top two putative off-target sites for the ABE protospacers, and C•G edited to T•A at the top two putative off-target sites for the CBE protospacers. Bar graphs represent the average, and error bars represent the SEM for n=3 biological replicates, with each replicate marked individually. Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA.

Finally, we evaluated unguided off-target RNA editing of the four MOBE systems as compared to the parental systems. We chose three highly expressed RNA transcripts that have previously been used to evaluate unguided off-target RNA editing by BEs.<sup>77</sup> HEK293T cells were transfected with MOBEs or the parental ABE/CBE combinations, and total RNA was extracted after 48 hours, reverse-transcribed into cDNA, and sequenced at the three transcriptomic sites of interest.<sup>77</sup> Both A to I and C to U off-target RNA editing was evaluated via three different methods: the average A to I or C to U conversion among all As or Cs within the transcript, the maximal A to I or C to U conversion among all As or Cs within the transcript, and the number of As or Cs within the transcripts with A to I or C to U conversions greater than 0.1%. By all three metrics and within all three transcripts, the gRNA-independent off-target A to I RNA editing activities of all four MOBE systems were within error or lower than their respective parental systems (Supplementary Figure 3.7A). gRNA-independent off-target C to U editing values within the *IP90* transcript were within error of negative controls, and only the evoBE4/ABE8.20 combination displayed off-target C to U editing values above negative controls within the *CTNNB1* transcript, as evaluated by all three methods (Supplementary Figure 3.7B). Within the *RSL1D1* transcript, we observed that the four MOBE systems (which share the same cytidine deaminase domain as both parental systems) displayed RNA off-target C to U editing values in between those of the two parental systems (Supplementary Figure 3.7B).



**Supplementary Figure 3.7: Evaluation of off-target RNA editing by MOBEs compared to parental systems.**

HEK293T cells were transfected with plasmids encoding nCas9-NG-P2A-mCherry, tandem CP-deaminase fusions, and both gRNA-aptamers (for the MOBE systems), or plasmids encoding the parental evoBE4-NG, ABE8e-NG or ABE8.20-NG, and both unmodified gRNAs (for the evoBE4-NG/ABE8e-NG and evoBE4-NG/ABE8.20-NG combinations). After 48 hours, the total RNA was extracted, and the mRNA was reverse-transcribed into cDNA. The three transcriptomic sites of interest (*CTNNB1*, *IP90*, and *RSL1D1*) were sequenced with NGS. Plotted are the average A to I (A) or C to U (B) conversion among all As or Cs within the transcript (left graphs), the maximal A to I (A) or C to U (B) conversion among all As or Cs within the transcript (middle graphs), and the number of As or Cs within the transcripts with A to I (A) or C to U (B) conversions greater than 0.1% (right graphs). Bar graphs represent the average, and error bars represent the SEM for  $n=3$  biological replicates, with each replicate marked individually. Negative control (NC) samples are also shown, in which cells were treated identically but transfected with a non-targeting gRNA.

### 3.1.6 MOBEs enable modeling of digenic diseases

As a proof of principle experiment to demonstrate the disease modelling potential of the MOBE systems, we installed clinically-relevant co-occurring C•G to T•A and A•T to G•C point mutation combinations in both HEK293T and SH-SY5Y cells. First, we identified two target SNV combinations from the oligogenic diseases database (OLIDA) that have been reported to cause genetic disorders.<sup>78</sup> These missense mutation combinations had suitable PAMs (NG or NA) to enable SNV positioning within the base editing window and efficient *in silico* predicted Cas9

binding (via Benchling CRISPR score) and base editing (via BE-hive and DeepBE) scores.<sup>76,79,80</sup> Further, their corresponding clinical phenotypes were related to neuronal defects, which would enable physiologically-relevant modeling in SH-SY5Y cells, which can be converted into various types of functional neurons.<sup>81,82</sup> Specifically, the OLI368 combination, *FGFR1* c.2075A>G and *DUSP6* c.545C>T (Figure 3.12A), confers Kallmann syndrome, a form of congenital hypogonadotropic hypogonadism. The OLI736 combination, *CELSR2* c.3800A>G and *INVS* c.725C>T (Figure 3.13A), confers isolated anencephaly, a neural tube defect.

We first tested a variety of protospacers with our highest efficiency MOBE1 and MOBE3 systems in HEK293T cells to assess on-target and bystander editing rates. For OLI368, we designed protospacers targeting the *FGFR1* SNV that placed the target A in positions A5 or A7 (both used NG PAMs, Figure 3.12A), and designed protospacers targeting the *DUSP6* SNV that placed the target C in positions C7 or C8 (which used NA and NG PAMs, respectively, Figure 3.12A). We tested all four protospacer combinations (which we label according to Figure 3.12A) using either the nCas9-SpRY variant (for the A5-C7 and A7-C7 combinations) or the nCas9-NG variant (for the A5-C8 and A7-C8 combinations). While the *FGFR1* A5 protospacer displayed the highest A•T to G•C conversion at the on-target position (which corresponds to a Glu692Gly mutation in the corresponding protein) of up to  $36.0 \pm 5.1\%$  with MOBE1 utilizing the nCas9-NG variant (hereafter shortened to MOBE-NG), it also facilitated  $26.5 \pm 4.0\%$  A•T to G•C conversion of a neighboring bystander adenosine (which corresponds to an Ile693Val mutation in the corresponding protein, Figure 3.12B). Fortunately, when the *FGFR1* A7 protospacer was used with any of the four MOBE systems tested (MOBE1-SpRY, MOBE3-SpRY, MOBE1-NG, or MOBE3-NG), editing of this bystander adenosine was significantly reduced (Figure 3.12B). Specifically, when using MOBE3-NG, we observed  $14.7 \pm 2.3\%$  A•T to G•C conversion with only  $0.4 \pm 0.1\%$  A•T to G•C conversion of the neighboring bystander adenosine (Figure 3.12B). When the *FGFR1* A7 protospacer was multiplexed with the *DUSP6* C8 protospacer using MOBE3-NG, we observed  $9.8 \pm 1.3\%$  C•G to T•A conversion (Figure 3.12B-C). The *DUSP6*

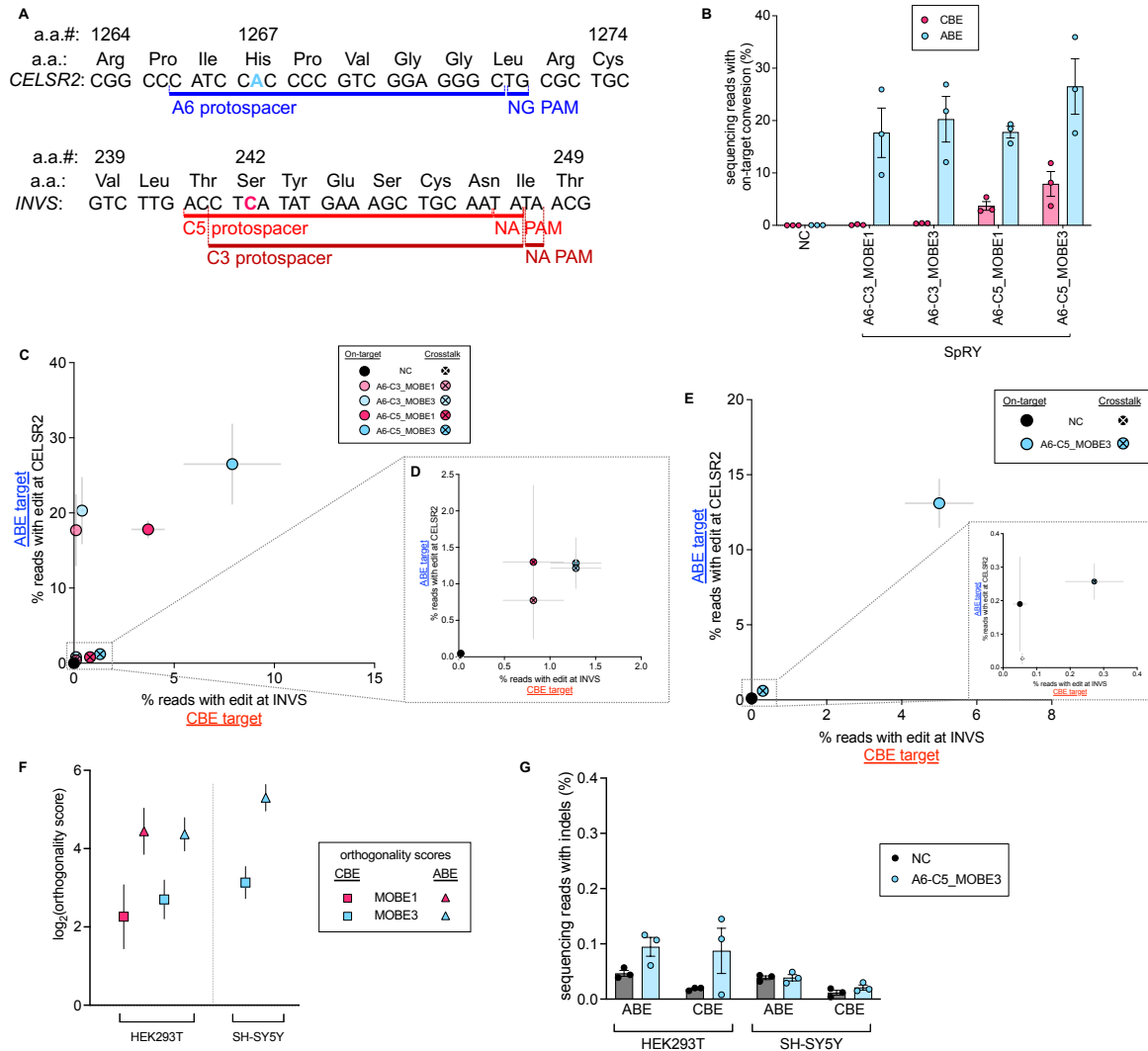
target lacked potential bystander cytidines (Figure 3.12A). We observed similar on-target editing efficiencies when multiplexing this protospacer combination with MOBE1-NG, but significantly lower crosstalk editing with the MOBE3-NG system (we observed at most only  $0.5 \pm 0.2\%$  crosstalk, Figure 3.12B-D). We also multiplexed the *FGFR1* A7 protospacer with the *DUSP6* C7 protospacer using MOBE1-SpRY and MOBE3-SpRY, but observed marginally less efficient editing (Figure 3.12B-C). We therefore used the most optimal protospacer combination (A7-C8) and MOBE3-NG with the 2x-dead-EGFP reporter to enrich for edits at the OLI368 combination and achieved  $51.4 \pm 3.3\%$  A•T to G•C conversion at the *FGFR1* (Glu692Gly) site and  $38.5 \pm 5.7\%$  C•G to T•A conversion at the *DUSP6* (Ser182Phe) site (Figure 3.12E).

For the OLI736 combination, we adopted a similar strategy and tested one *CELSR2* protospacer (A6) and two *INVS* protospacers (C3 and C5, Figure 3.13A) with MOBE1-SpRY and MOBE3-SpRY and found the optimal combination to be the A6 and C5 protospacers using MOBE3-SpRY, which enabled  $26.5 \pm 5.3\%$  A•T to G•C conversion at the *CELSR2* (His1267Arg) site and  $7.9 \pm 2.4\%$  C•G to T•A conversion at the *INVS* (Ser242Leu) site (Figure 3.13B-D). The *CELSR2* target lacked potential bystander adenosines, while the *INVS* target had one potential bystander cytidine when using the C5 protospacer, editing of which would cause a silent Thr241Thr mutation (Figure 3.13A).

We then use these optimal gRNA-MOBE combinations in the SH-SY5Y cell line, a clinically relevant cell type previously used to model these diseases.<sup>83,84</sup> SH-SY5Y cells were electroporated with the optimal gRNA combinations identified from our HEK293T cells experiments, MOBE3, and nCas9-NG or nCas9-SpRY, enriched with FACS for expression of nCas9 (via the P2A-mCherry marker) after 72 hours, and analyzed by NGS. For the OLI368 combination, the A7-C8 protospacer combination with MOBE3-NG facilitated  $16.6 \pm 1.6\%$  A•T to G•C conversion at the *FGFR1* (Glu692Gly) site and  $23.0 \pm 2.7\%$  C•G to T•A conversion at the *DUSP6* (Ser182Phe) site, with less than  $0.3 \pm 0.1\%$  crosstalk at both sites (Figure 3.12F). For the OLI736 combination, the A6-C5 protospacer combination with MOBE3-SpRY facilitated  $13.1$

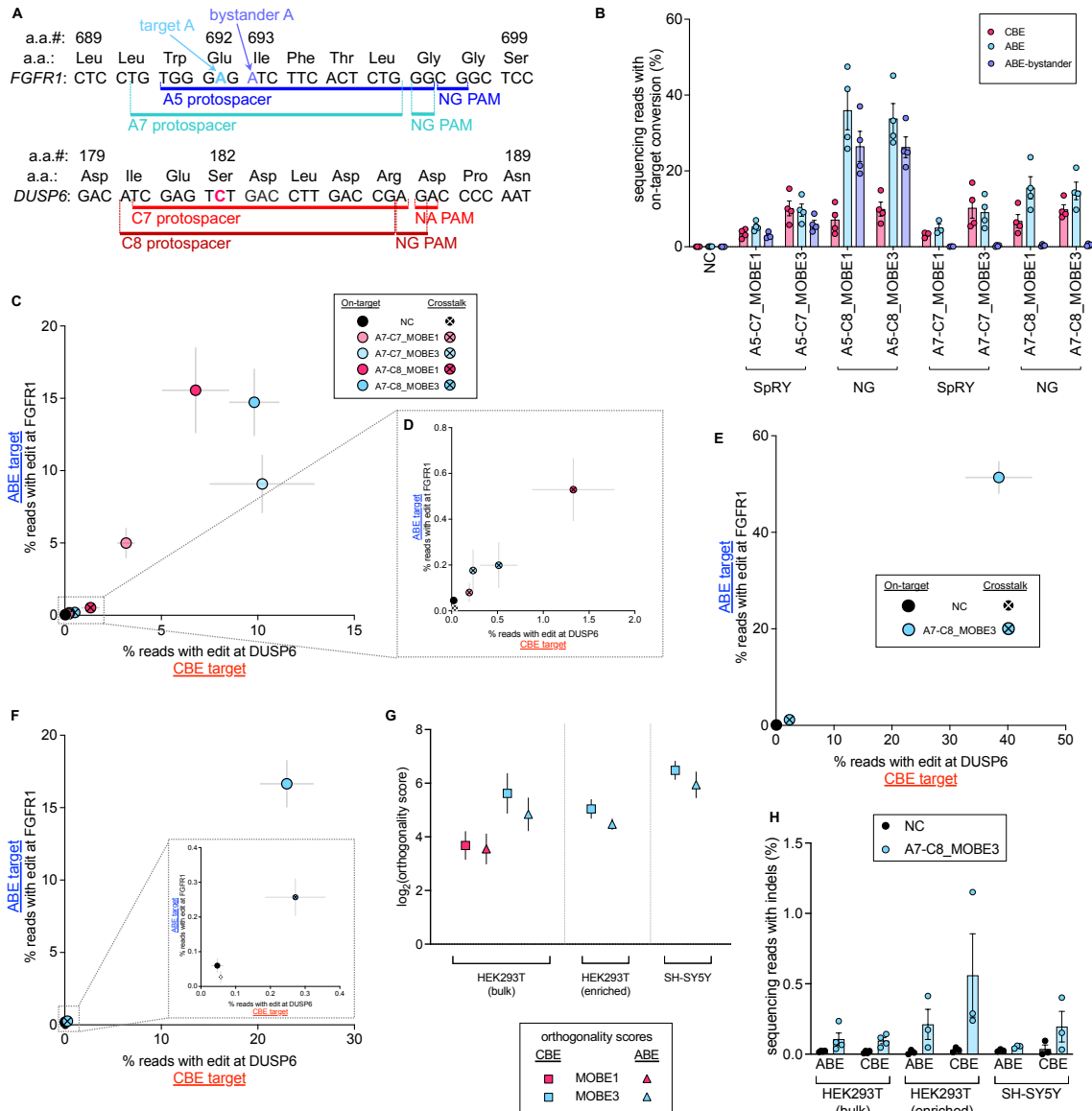


$\pm 1.6\%$  A•T to G•C conversion at the *CELSR2* (His1267Arg) site and  $5.0 \pm 0.9\%$  C•G to T•A conversion at the *INVS* (Ser242Leu) site, with at most  $0.6 \pm 0.1\%$  crosstalk (Figure 3.13E). Both OLIDA targets showed low indel rates and high orthogonality scores in HEK293T and SH-SY5Y cells (Extended Data Figures 9G-H and 10F-G). However, we will note that in the absence of single cell analyses or clonal expansion experiments, these bulk editing efficiencies do not report on efficiencies of co-occurring edits. These experiments collectively demonstrate the potential of MOBE systems as tools for modeling combinations of SNVs that cause complex human genetic diseases.



**Figure 3.12: MOBEs enable the installation of orthogonal point mutations known to cause Kallmann syndrome (OLI368).**

(A) Schematic of the DNA and protein sequences of the OLI368 combination, *FGFR1* c.2075A>G and *DUSP6* c.545C>T, with MOBE protospacer and PAMs indicated. (B) Optimization of gRNA pairs with MOBE1 and MOBE3 to avoid bystander adenine editing in HEK293T cells. (C) Bulk editing efficiencies of the *FGFR1* A7 protospacer combined with *DUSP6* C7 and C8 protospacers with inset (D) to display crosstalk in HEK293T cells. (E) Editing efficiencies of MOBE3 with the optimal A7-C8 protospacer combination after enrichment with the all-in-one fluorescent reporter plasmid in HEK293T cells. (F) Editing efficiencies of MOBE3 (A7-C8 protospacers) sorted for nCas9-P2A-mCherry expression after electroporation in SH-SY5Y cells. (G) Orthogonality scores for the A7-C8 protospacer combination in all tested cell types. (H) Indel rates of MOBE3 with A7-C8 protospacer combination in all tested cell types. All values represent the mean and SEM of n=3 or 4 independent biological replicates (shown individually).



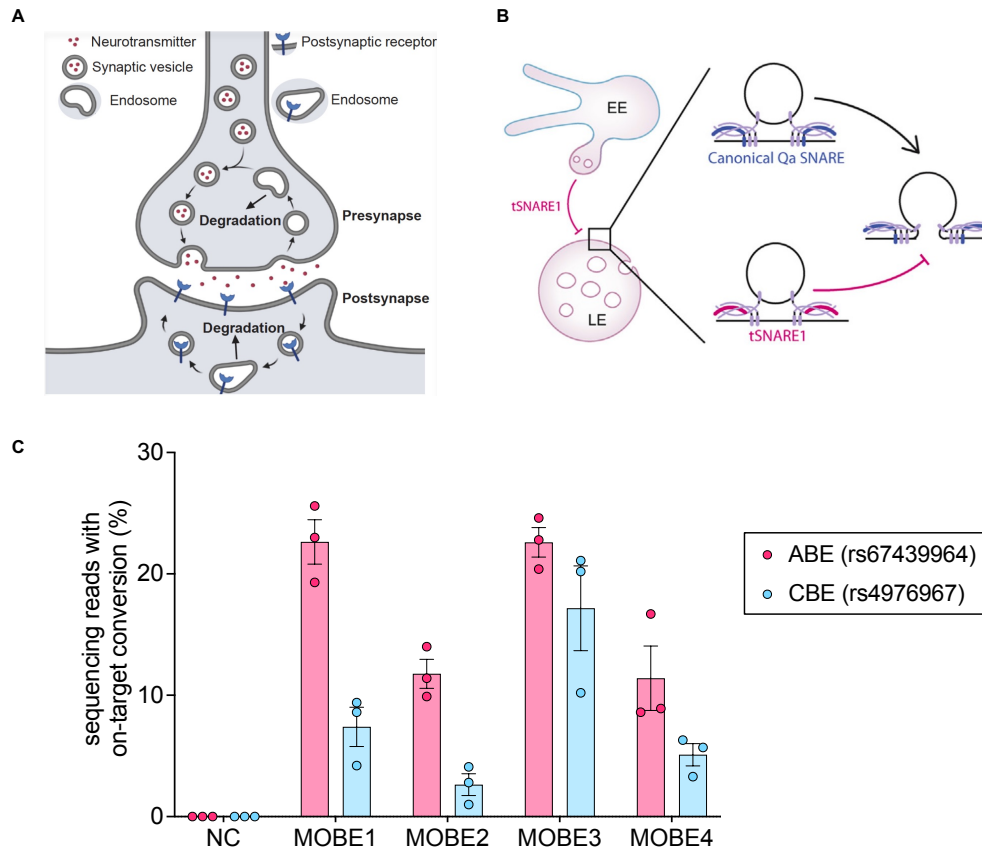
**Figure 3.13: MOBEs enable the installation of orthogonal point mutations known to cause isolated anencephaly (OLI736).**

(A) Schematic of the DNA and protein sequences of the OLI736 combination, *CELSR2* c.3800A>G and *INVS* c.725C>T, with MOBE protospacer and PAMs indicated. (B) On-target editing efficiencies of gRNA pair optimization with MOBE1 and MOBE3 in HEK293T cells. (C) Editing efficiencies of *CELSR2* A6 protospacer combined with *INVS* C3 and C5 protospacers with inset (D) to display crosstalk in HEK293T cells. (E) Editing efficiencies of MOBE3 with the optimal A6-C5 protospacer combination sorted for nCas9-P2A-mCherry expression after electroporation in SH-SY5Y cells. (F) Orthogonality scores for the A6-C5 protospacer combination in both cell types. (G) Indel rates of MOBE3 with A6-C5 protospacer combination in both cell types. All values represent the mean and SEM of n=3 independent biological replicates (shown individually).

### 3.1.7 MOBEs install of Schizophrenia-associated SNVs

Schizophrenia has an estimated 80% heritability, but the exact risk alleles are not well-characterized. Common genetic variants, which are primarily non-coding, may play a previously underappreciated role in risk for schizophrenia. Candidate genes for schizophrenia risk identified through genome- and transcriptome-wide association studies often function in endosomal trafficking.<sup>85</sup> For example, overexpression of the gene *TSNARE1* (that encodes the protein tSNARE1) has been associated with schizophrenia risk but does not have a well-defined function. Recently, a role for this protein has emerged in endosomal trafficking, in which it acts as an inhibitory SNARE protein in vesicle fusion.<sup>86</sup> Two distinct single nucleotide variants rs67439964 (A-to-G) and rs4976967 (C-to-T) have been identified that may modulate *TSNARE1* expression.

To assess the contribution of combinations of common genetic variation, multiplexed base editing represents an attractive option to model disease associated variant, and for combinations of distinct point mutations, MOBE systems would be required. We designed gRNAs to target these two variants and tested the editing efficiency in HEK293T cells. The MOBE systems install both of these point mutations simultaneously with high efficiency (Figure 3.14). Installing these variants individually or in combination will elucidate the contributions of individuals common genetic variants to schizophrenia risk. In future work, we imagine installing these variants in SH-SY5Y cells or induced pluripotent stem cells, which can be differentiated into neurons, in order to properly phenotype in a clinically relevant cell line.



**Figure 3.14: MOBEs install variants associated with schizophrenia**

A) Schematic of the endosomal recycling pathway at the synapse. B) Inhibitory role of tSNARE1 in early endosome to late endosome vesicle fusion. C) Editing efficiencies of MOBE systems when installing two different single nucleotide variants that may affect *TSNARE1* expression.

### 3.1.8 Comparison of MOBE systems to orthogonal Cas-derived BEs

Although orthogonal base editing can be achieved using orthogonal Cas proteins such as SaCas9 or Cas12, there are several key limitations with these approaches. Since Cas12 enzymes (type V CRISPR systems) contain a single nuclease domain, they cannot be made into the appropriate nickase for base editing, which dramatically decreases base editing efficiencies with these Cas enzymes. As previously mentioned, *Staphylococcus aureus* (Sa) Cas9 has much wider editing window, resulting in bystander editing, which severely limits their utility for precisely installing SNVs and modelling disease. The orthogonal Cas enzymes also have longer, more infrequent PAM sequences (NNNRRT for the most relaxed SaCas9-KKH

variant). This reduces the likelihood that the target base will fall in the center of the editing window with available PAMs, and in many cases precludes the testing of multiple gRNAs for a given target base to push potential bystander bases out of the window (as we demonstrated in Supplementary Figure 3.8A-B). Further, each additional Cas enzyme will increase the cargo needed in multiplexed experiments. We reasoned that using the SpCas9 nickase for MOBE systems would overcome all of these issues due to their high efficiencies, narrow editing windows, and flexible PAM requirements (NG and NR/Y).

We performed an initial comparison of the MOBE systems (which employ SpCas9 nickase at both ABE and CBE targets) to orthogonal Cas9-derived base editors (which employ either SpCas9 or SaCas9 at distinct targets) to demonstrate these points. In this test, we multiplexed Sp-CBE (evoBE4-NG) with Sa-ABE (either Sa-ABE8e-KKH or Sa-ABE8.20-KKH) and directly compared to each of the MOBE systems by targeting the same two protospacers at the *RNF2* locus (the ABE protospacer was adjacent to both NG and NNNRRT PAMs, Supplementary Figure 3.8A-C).

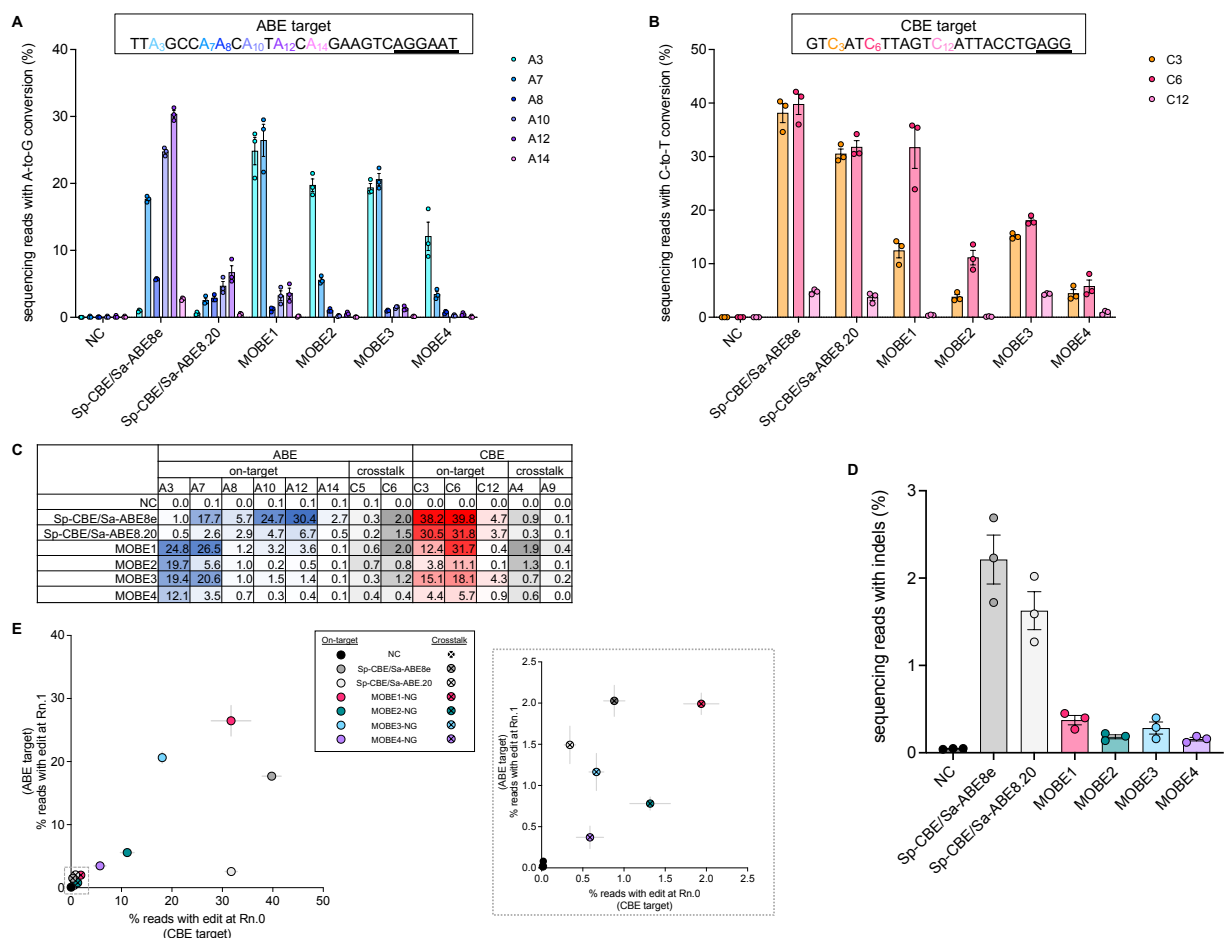
Consistent with their wider editing window, both Sa-ABE variants displayed higher editing at adenosines in positions 8-14 compared to all MOBEs (Supplementary Figure 3.8A,C). Sa-ABE8e showed highest editing at position 12 ( $30.4 \pm 0.5\%$ ) followed by positions 10 and 7, while Sa-ABE8.20 showed very low editing overall: maximum  $6.7 \pm 1.0\%$  editing at position 12 (Supplementary Figure 3.8A,C). Conversely, MOBE2 and MOBE4 which utilize apt-ABE8.20 predominantly edited position 3 (up to  $19.7 \pm 1.0\%$ ), while MOBE1 and MOBE3 (which utilize apt-ABE8e) edit positions 3 and 7 with similar efficiencies ranging from  $19.4 \pm 0.6\%$  to  $26.5 \pm 2.4\%$  (Supplementary Figure 3.8A,C). Overall, MOBE systems exhibited a narrower editing window at the ABE target, and MOBEs with apt-ABE8.20 showed up to 2.9-fold increase in maximal editing efficiency than their Sa-ABE8.20 counterparts.

At the CBE target, when multiplexing Sp-CBE/Sa-ABE, evoBE4 showed a wide window and high editing efficiencies as expected: similar editing efficiencies at positions 3 and 6

(ranging from  $30.5 \pm 0.9\%$  to  $39.8 \pm 1.9\%$ ) and up to  $4.7 \pm 0.3\%$  at position 12 (Supplementary Figure 3.8B-C). However, MOBE1 retained high editing at position 6 ( $31.7 \pm 3.9\%$ ), but had greatly reduced editing at position 3 ( $12.4 \pm 1.4\%$ ) and almost no editing at position 12 ( $0.4 \pm 0.1\%$ , Supplementary Figure 24B-C). MOBE2, which utilizes the same apt-CBE-3'end, showed a similar editing pattern, albeit with reduced overall efficiency compared to MOBE1. MOBE3, which utilizes apt-CBE-SL3, also displayed maximal editing at position 6 ( $18.1 \pm 0.5\%$ ) but with a wider editing window that agrees with our characterization data (Supplementary Figure 3.1, 3.8).

Additionally, MOBE systems generated greatly reduced indel introduction efficiencies compared to orthogonal Cas multiplexing. Indel rates averaged  $2.2 \pm 0.3\%$  when multiplexing with Sp-CBE/Sa-ABE8e and  $1.6 \pm 0.2\%$  with Sp-CBE/Sa-ABE8.20 (Supplementary Figure 3.8D). These indel rates were 6-fold to 10-fold higher than the corresponding MOBE systems, which displayed at most  $0.37 \pm 0.05\%$  indel introduction efficiencies (Supplementary Figure 3.8D). Importantly, these greatly reduced indel frequencies were not accompanied with a similar reduction in on-target editing efficiency (Supplementary Figure 3.8E). Orthogonal Cas9 multiplexing showed similar crosstalk editing to MOBE systems (Supplementary Figure 3.8E).

Although this comparison is not comprehensive, these findings agree with our comparison of MOBE systems to parental SpCas9-fused BEs, and published literature on SaCas9 BEs.<sup>45,59,61</sup> Generally, compared to multiplexing BEs with Sp-Cas9 and Sa-Cas9 orthologs, MOBE systems have a narrower editing window, similar on-target efficiencies, and generate lower indel rates. Thus, MOBE systems are a more precise tool for installation of multiple, distinct single-nucleotide variants.



**Supplementary Figure 3.8: Comparison of MOBE systems to orthogonal Cas-derived BEs.**

On-target editing efficiencies of each multiplexed editor (listed on the x-axis) for the *RNF2* locus ABE target (A) and CBE target (B). HEK293T cells were treated identically to bulk MOBE experiments as previously described in Figure 3.1. The protospacers are listed above with target bases shown in color and PAMs underlined (with the ABE target containing both NG and NNNRRT motifs to compare the MOBE systems and Sa-ABEs). (C) Average on-target and crosstalk editing efficiencies shown in table format for all A or C bases edited above background. (D) Plotted are the percent of total DNA sequencing reads with indels at each target site. (E) Plotted are the percent of total DNA sequencing reads at the most-edited base with A•T edited to G•C at the ABE target (y-axis, filled circles), C•G edited to T•A at the ABE target (y-axis, circles with X inset), C•G edited to T•A at the CBE target (x-axis, filled circles), and A•T edited to G•C at the CBE target (x-axis, circles with X inset). All plots represent the average and SEM for n=3 biological replicates (each replicate is marked individually when possible).



## 3.2 Discussion

The lack of precise genome editing tools to orthogonally multiplex SNV introduction has hindered the investigation and treatment of co-occurring variants.<sup>23–25,87</sup> BEs are ideal genome editing tools to enable multiplexed point mutation introduction, as they introduce SNVs with high efficiency and precision, and their intermediates are less toxic than DSBs. However, current BEs can only be multiplexed when using a single BE variant to perform editing at various loci. When multiplexing two or more different BE variants (such as an ABE and a CBE), gRNA crosstalk will occur, resulting in all BEs performing editing at all loci (Figure 2.1B-C). This can be prevented by using BE variants that utilize orthogonal Cas proteins (such as SpCas9, SaCas9, and Cas12 variants). However, the high efficiency, more precise editing window, and significantly less restrictive PAM requirements of the SpCas9-derived BEs cause this alternative to suffer from a variety of shortcomings (see Supplementary Figure 3.8 and Section 3.1.8 above).

In this work, we utilized RNA aptamer-coat protein technologies to develop and characterize four multiplexed orthogonal base editor (MOBE) systems that enable the introduction of combinations of C•G to T•A and A•T to G•C point mutations simultaneously throughout the genome with high efficiency and precision in multiple cell types. We additionally generated a reporter plasmid that allows for the enrichment of cells with orthogonal multiplexed edits and enabled an average 16-fold increase in editing efficiency across the four MOBEs. This simple enrichment strategy can aid researchers in generating cell-based models of SNV combinations relevant to genetic disease.

Each MOBE is comprised of a unique combination of an aptamer-based CBE system and an aptamer-based ABE system. To develop each aptamer-based BE system, we screened through a variety of constructs in which one of four different aptamers was imbedded within the TL, SL2, SL3, or 3'-end of the gRNA. These aptamer-gRNA constructs were then combined with corresponding coat protein-deaminase fusions, in which the architectures (N- and C-terminal

fusions), deaminases, and linkers were varied. We found that for the ABE aptamer systems, the use of eighth-generation evolved TadA deaminases was the single most important factor to facilitate consistently high editing activity across all tested sites, followed by the aptamer-coat protein system (the com system) and the location of the aptamer (3'-end embedded). The CBE aptamer systems were generally less stringent in their design, but again the use of a highly evolved deaminase (the evoAPOBEC1 cytidine deaminase) provided the largest boost in editing efficiency out of all the modifications tested.

Out of our four systems, MOBE1 and MOBE3 typically facilitate the highest editing efficiencies, while MOBE3 typically has the highest orthogonality scores. However, both of these metrics are site-dependent, and we thus suggest screening all four systems for a given protospacer combination. If low editing efficiencies are observed, the MOBE fluorescent reporter is recommended as orthogonality scores remain high when utilizing this strategy. Again, while MOBE1 typically facilitated the highest rates of co-occurring orthogonal edits, this was observed to be site-dependent.

We also demonstrated that the MOBE systems can be easily modified to implement additional Sp-nCas9 variants, which we explicitly demonstrated with the near PAM-less SpRY-nCas9 variant<sup>42</sup> and the high-fidelity HiFi-nCas9 variant.<sup>72</sup> The SpRY-nCas9 variant is particularly useful given the slightly narrower editing window of the MOBE systems, and enables precise positioning of target nucleotides while avoiding potential bystander edits. We also characterized undesired editing by our MOBE systems, including indel introductions, off-target DNA editing (both gRNA-dependent and -independent), and off-target RNA editing. We observed equivalent or reduced undesired editing levels for all MOBE systems compared to their parental constructs for all types of undesired edits, even when accounting for changes in on-target editing efficiencies.

In addition to the use of our fluorescent reporter, we postulate that on-target editing activity may be improved through the addition of protective pseudoknots to the 3'-end of the

aptamer-gRNAs; this strategy was shown to increase prime editing efficiency by preventing endogenous exonucleases from degrading 3' extensions on pegRNAs (which would produce RNP molecules within the cell that are still able to engage to the target site for nicking but lack the ability to install the prime edit).<sup>88</sup> This proposed mechanism of inhibition would also apply to our aptamer-BE systems, where degradation of the 3'-end aptamer (but not the Sp-gRNA backbone) could cause unproductive nCas9 activity without the ability to recruit a deaminase enzyme for base editing. Additionally, a limitation of the MOBE systems is that expression and stoichiometry have not necessarily been optimized. There is potential to improve MOBEs using drive-and-process CRISPR arrays (a multiplexed gRNA expression system) and resource aware-construct design.<sup>89,90</sup> In order to utilize MOBE systems for therapeutic correction of co-occurring SNVs, additional optimization and characterization would be necessary.

We envision the MOBE systems as powerful tools for disease modelling. They can be used to analyze epistatic effects of modifier SNVs observed in individuals with altered disease penetrance or common variants found in the afflicted population(s). MOBEs can probe SNVs in non-coding regions such as *cis*-regulatory elements that may modulate expressivity of a coding variant. They can also be applied to exploratory studies of the contributions of VUS in polygenic diseases, such as cancer, diabetes, and schizophrenia. The MOBE systems are well-positioned to model digenic diseases, which are variant combinations that are known to cause a disease (examples include Long QT syndrome and Brugada syndrome).<sup>23,25</sup> Currently, there are 615 digenic SNV combinations reported in the oligogenic diseases database (OLIDA), and 198 can be modelled by multiplexing a CBE, 18 can be modelled by multiplexing an ABE, and 91 (15%) would require a MOBE system to be modelled.<sup>78</sup> In fact, we explicitly demonstrated successful introduction of two such OLIDA variant combinations (OLI368 and OLI736) in the SH-SY5Y cell line. Further, there are 533 multigenic (3 or more co-occurring variants) SN<sup>116</sup>V combinations in OLIDA, and 220 can be modelled by multiplexing a CBE, 23 can be modelled by multiplexing an

ABE, and 148 (28%) would require a MOBE system to be modelled. While human pluripotent stem cells would be a more desirable system for disease modeling, we and others have observed low efficiencies and impure editing products by CBEs in these cell types due to their robust DNA damage response and base excision repair activities.<sup>91–94</sup> Additional work to optimize CBEs in general is required before MOBEs can be efficiently utilized in these cell types.

Finally, our MOBE systems can be leveraged for complex genome engineering with other effector modalities that use the SpCas9 ortholog. Multiplexed orthogonal base editing could be compatible with aptamer-recruited CRISPRi, CRISPRa, or targeted integration (REDIT system).<sup>48,95</sup> Possible applications include base editor screens to probe epistasis/synthetic lethal gene interactions, metabolic reprogramming, genetic logic circuits, and other multifunctional outputs using the same Cas ortholog.<sup>96,97</sup>

### 3.3 Methods

**General methods and cloning.** All primers in this study were ordered through Integrated DNA technologies (IDT). All PCR reactions were performed with Phusion DNA Green High-Fidelity Polymerase (F534L, Thermo Fisher) or Phusion U (F556L, Thermo Fisher) where appropriate. gRNA plasmids were cloned by site directed mutagenesis using a 5' tail in the primer to replace the 20nt spacer region, as described in Vasquez et al. Basic Protocol 1.<sup>98</sup> Similarly, aptamers were inserted into the *S. pyogenes* gRNA backbone using primers with 5' tails. We also created golden gate acceptor plasmids for each of the final aptamer-gRNA fusions (MS2-SL3, MS2-3'end, and Com-3'end) that will be available on Addgene. Codon optimized (GenScript) coat proteins were ordered as a gBlock gene fragment (IDT) and subcloned into CP-deaminase fusions with USER (Uracil-Specific Excision Reagent) cloning.<sup>99</sup> Linkers and deaminases were repurposed from established base editors ancBE4-max, ABE7.10-max, and target-AID (93-aa linker). Evolved base editor plasmids were obtained from Addgene: pBT281 (evo-APOBEC1-

BE4), NG-ABE8e (Addgene plasmid #122611 and 138491), ABE8.20-m (Addgene plasmid #136300), and pCMV-BE4-RrA3F (Addgene plasmid #138340).<sup>59,61,66,67</sup> The deaminase domains were subcloned into nCas9-NG backbones for parental BE controls and into CP-containing plasmids for aptamer-BE systems. The fluorescent reporter plasmids were cloned via USER cloning to insert the 2x-dead-GFP and aptamer-containing gRNAs.<sup>99</sup> We designed unique golden gate sites for each gRNA spacer (based on pDG461, Addgene plasmid #100902) and assembled the final plasmid using the established protocol.<sup>100</sup> Endotoxin-free plasmids were prepared with ZymoPURE II Midiprep Kit (D4201, Zymo Research). Sequences of all constructs are available in Supplementary Data 1.

**General human tissue culture.** HEK293T cells (CRL-3216, ATCC) and HeLa cells (CCL-2, ATCC) were cultured at 37°C with 5% CO<sub>2</sub> in DMEM + GlutaMAX (10566-024, Thermo Fisher) supplemented with 10% (v/v) fetal bovine serum (10437-028, Thermo Fisher). SH-SY5Y cells (CRL-2266, ATCC) were cultured similarly in DMEM/F12 (11320-033, Thermo Fisher).

**Transfections for base editing experiments.** HEK293T cells were seeded at 50,000 cells/well in a 48-well plate and transfected after 16 hours at ~70% confluency. DNA mixes were created in a total volume of 12.5  $\mu$ L with Opti-MEM reduced-serum medium (31985-070, Thermo Fisher), and combined with 1.5  $\mu$ L of Lipofectamine 2000 reagent in 11  $\mu$ L of Opti-MEM, according to the manufacturer's instructions. Cells were incubated for 72 hr, washed with 150  $\mu$ L PBS, and lysed for genomic DNA with freshly prepared buffer containing 10 mM Tris (pH 7.0), 0.05% SDS, and 25  $\mu$ g/mL Proteinase K (P8107S, New England Biolabs). Samples were incubated on a thermocycler at 37°C for 1 hr, 80°C for 30 min, and then held at 4°C. For individual aptamer-BE experiments (optimization and characterization), 250 ng gRNA plasmid was combined with 850 ng of nCas9-NG and 650 ng CP-deaminase (or 1000 ng parental base editor). For multiplexed orthogonal base editing experiments, 500 ng of nCas9-NG and 500 ng

of MOBE (or 500 ng of each parental base editor) were combined with 125 ng of each gRNA plasmid. For fluorescent enrichment experiments, an additional 250 ng of reporter plasmid was added. For HeLa cell experiments, 300,000 cells/well were seeded in 6-well plates and transfected after 16 hours using 10  $\mu$ L of TransIT-LT1 (Mirus) and a total of 4  $\mu$ g DNA: 480 ng of reporter plasmid, 480 ng of each gRNA, 1280 ng of nCas9-NG, and 1280 ng of MOBE. For SH-SY5Y cell experiments, the Neon transfection system (Invitrogen) was used with the 10  $\mu$ L kit (MPK1025, Invitrogen) according to the manufacturer's instructions. Briefly, SH-SY5Y cells were gently pelleted (50 x g), washed with PBS using wide bore tips, pelleted, resuspended in buffer R, mixed with 1.5  $\mu$ g of plasmid DNA (205 ng of each gRNA, 545 ng of nCas9-NG, and 545 ng of MOBE), electroporated with three 20 ms pulses of 1200V, and immediately added to pre-warmed media in a 12-well plate. For enrichment experiments, multiple wells were combined (if needed) to recover sufficient cells for next-generation sequencing.

**Off-target editing analysis.** For the orthogonal R-loop assay, Sa-dCas9 is programmed to bind to a genomic locus of interest and form an R-loop, but does not modify the DNA in any way. Concurrently, Sp-Cas9-derived BEs are co-transfected, and if the deaminase components of the BEs have gRNA-independent off-target editing activity, they will deaminate target Cs or As within the Sa-dCas9-exposed R-loop (Figure 3.11). HEK293T cells were transfected with 300 ng of *Streptococcus aureus* (Sa)-dCas9 plasmid and 200 ng Sa-gRNA plasmid to expose an R-loop at one of four previously characterized sites (Doman et al. site 5 and 6, Yu et al. site 17 and 26).<sup>67,74</sup> This was delivered along with 100 ng of each Sp-gRNA plasmid and 300 ng of MOBE and nCas9-NG plasmids or 300 ng of the parental, non-orthogonal ABE and CBE plasmids (the evoBE4 parental CBE construct was multiplexed with either ABE8e or ABE8.20 using unmodified gRNAs). After three days, the cells were lysed and targeted amplicon sequencing was performed at the Sa-dCas9 protospacer. gRNA-dependent DNA off-target edits were analyzed via targeted amplicon sequencing.<sup>101</sup> RNA off-target editing was analyzed via

targeted amplicon sequencing of cDNA as previously established.<sup>77,102</sup> Total RNA was isolated 48 hr after transfection using the Quick-RNA miniprep kit (R1054, Zymo), converted into cDNA using Superscript-III reverse transcriptase (18080093, Invitrogen), and sequenced at three highly expressed transcripts that have been previously characterized (*CTNNB*, *IP90*, and *RSL1D1*).<sup>77</sup>

**Next-generation sequencing.** Samples were prepared for targeted amplicon sequencing as described in Vasquez et al. Alternative Protocol 1.<sup>98</sup> Briefly, 1-3  $\mu$ L of genomic DNA was added in a 25  $\mu$ L PCR reaction with 0.2  $\mu$ M primers (listed in Supplementary Data 1 of the publication) for round 1 and amplified for 24-27 cycles (minimal amount to avoid PCR bias). After confirmation on a 2% agarose gel, round 2 PCR was performed to barcode samples with 8-12 cycles. Samples were pooled, purified by gel extraction, and then quantified by Qubit with the dsDNA HS assay kit (Q32854, Thermo Fisher). Sequencing was performed on an Illumina MiniSeq (2x151 paired end reads) per the manufacturer's instructions.

**Data Analysis and Statistics.** Next-generation sequencing data were collected, demultiplexed, and trimmed with Illumina Local Run Manager Generate FASTQ analysis module (v2.0) and MiniSeq control software (v2.2.1). The FASTQ files were analyzed with CRISPResso2 (version 2.0.20b) on batch mode (parameters: --base\_edit -wc -10 -w 10 -q 30) to assess genomic base editing efficiencies.<sup>103</sup> All base editing efficiencies values were reported as a percent of total DNA sequencing reads with editing (from the CRISPResso2 nucleotide percent summary output) and as averages of independent biological replicates. Indel frequencies were reported as a percent of total reads from the CRISPResso2 indel histogram output and separated by base editor (for the single-amplicon target combinations single-end CRISPResso2 analysis was performed independently on read1 and read2). Aptamer base editor characterization data (Figure 2.4 and Supplementary Figure 2.5) were analyzed on pandas (v1.4.2) and plotted on

seaborn (v0.11.2). Fold-change values compared to different types of editors were computed as the average editing efficiency of test editor divided by the average editing efficiency of control editors  $\pm$  propagation of uncertainty of the SEM. Orthogonality scores were calculated as the percent of total DNA sequencing reads with on-target conversion divided by the percent of total DNA sequencing reads with crosstalk of the same conversion (e.g. average C•G to T•A editing at the CBE target divided by average C•G to T•A editing at the ABE target)  $\pm$  propagation of uncertainty of the SEM. Fold-change enrichment values for the MOBE fluorescent reporter represent average editing efficiencies of FACS sorted GFP+ cells divided by average editing efficiencies of FACS sorted single cells regardless of color  $\pm$  propagation of uncertainty of the standard deviation.

**Co-occurring edit analysis.** Sequencing data in FASTQ format were aligned to the *Homo sapiens* Hg38 reference genome using Rsubread v2.14.<sup>104</sup> A quality control filter was applied using SAMtools v1.19.2,<sup>105</sup> and only reads with quality scores greater than or equal to 2 were considered in the analysis. Each protospacer containing the desired on-target as well as potential off-target edits was mapped onto the respective paired-end reads. The Levenshtein edit distance between the protospacers (CBE and ABE), and the sequencing reads were calculated to quantify all insertions, deletions, and substitutions in the protospacer region of the aligned reads. The genotypes at each on-target edit site as well as potential crosstalk sites were extracted as a haplotype to quantify co-occurring editing efficiency. Haplotypes called from the paired-end sequencing reads were then categorized as 1) wild-type (no-edits), 2) orthogonal edit, 3) CBE on-target edit, 4) ABE on-target edit, 5) dual edits only, 6) crosstalk only, 7) CBE or ABE only, and 8) all other genotypes (see Supplementary Figure 3.9). This software (built with R v4.2.0) is termed MOBEnto: Quantification of Haplotypes from Multiplexed Genome Editing Using Next-Generation Sequencing Data and available at <https://github.com/Komorlab/MOBEnto>.



**Fluorescence-activated Cell Sorting (FACS).** In a 48-well plate, cells were washed with 150  $\mu\text{L}$  of PBS and detached with 30  $\mu\text{L}$  of Accumax (STEMCELL Technologies) at room temperature (volumes scaled up for larger plates). After 1 min, cells were resuspended with 170  $\mu\text{L}$  of cold PBS at approximately  $2 \times 10^6$  cells per mL, passed through a 35  $\mu\text{m}$  cell strainer into a test tube, and kept on ice. Samples were run on a S3e Cell Sorter (BioRad), using cells expressing either mCherry or GFP to set scatter and fluorescence gates. To enrich for base-edited cells with the fluorescent reporter, GFP and mCherry double-positive single cells were collected per sample. An example gating scheme is shown in Supplementary Figure 3.10 (visualized with FlowJo 10.9.0) where GFP/mCherry double-positive cells were sorted for sequencing. To enrich for transfected SH-SY5Y cells, mCherry-positive single cells were collected per sample. For unenriched samples, singlets were collected regardless of color. For all samples, 5,000-50,000 cells were sorted using the S3e machine's ProSort software (v1.6, BioRad) on "Enrich" mode directly into 1.5 mL tubes containing 500  $\mu\text{L}$  of cold PBS. Cells were pelleted at 300 x g for 10 min and the supernatant was gently pipetted off. Pellets were resuspended in lysis buffer (100 mM Tris-HCl [pH 7.0], 0.05% SDS, and 25  $\mu\text{g}/\text{mL}$  Proteinase K [P8107S, New England Biolabs]) at a final concentration of 2,000 cells/ $\mu\text{L}$  to be used as template for targeted amplicon sequencing.

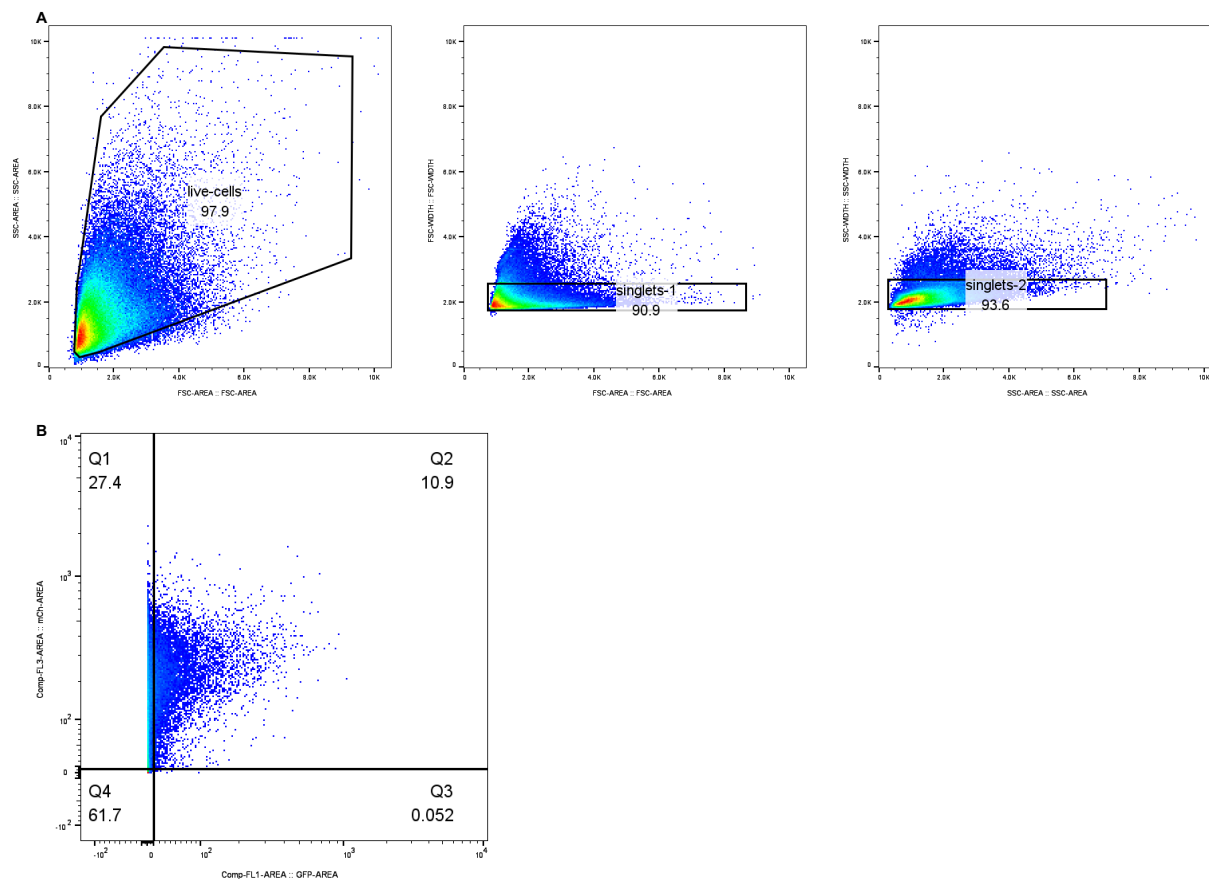
**Data availability.** Next-generation sequencing data will be available on the NCBI Sequencing Read Archive database under project number PRJNA836633. Plasmids from this study will be available at Addgene. *Homo sapiens* Hg38 was used as the reference genome for MOBEto quantification of co-occurring edits.

**Code availability.** Source code for MOBEnto (Quantification of Haplotypes from Multiplexed Genome Editing Using Next-Generation Sequencing Data) is available at <https://github.com/Komorlab/MOBEnto>.

|               |   |           |   |                     |                            |                            |                      |
|---------------|---|-----------|---|---------------------|----------------------------|----------------------------|----------------------|
| CBE on-target | Y | crosstalk | Y | dual edits only     | all other genotypes        | all other genotypes        | dual edits only      |
|               |   |           | N | all other genotypes | orthogonal edit            | CBE or ABE only both sites | CBE on-target        |
|               | N |           | Y | all other genotypes | CBE or ABE only both sites | crosstalk only             | crosstalk only       |
|               |   |           | N | dual edits only     | ABE on-target              | crosstalk only             | wild-type (no edits) |
|               |   |           |   | Y                   | N                          | Y                          | N                    |
|               |   |           |   | crosstalk           |                            |                            |                      |
|               |   |           |   | Y                   |                            | N                          |                      |
|               |   |           |   | ABE on-target       |                            |                            |                      |

**Supplementary Figure 3.9: Haplotype calling schematic for co-occurring edit analysis by the MOBEnto software.**

See Methods section for additional details.



### Supplementary Figure 3.10: Example of gating scheme used for MOBE fluorescent enrichment.

(A) Cell debris was gated out using forward scattering area against side scattering area, and doublets were gated out using width against forward and side scattering area. 50,000 cells from this population of singlets were sorted regardless of color as the unenriched sample. (B) Then, GFP (x-axis) and mCherry (y-axis) fluorescent intensities were used to sort 5,000-50,000 double-positive cells from the Q2 gate as the enriched sample. This example shows MOBE1 targeted to the *EMX1* site combination co-transfected with 250 ng of MOBE reporter plasmid. The Q4 gate represents untransfected cells and the Q1 gate represents transfected cells (via nCas9-P2A-mCherry) without reporter turn-on. See Methods section for additional details.

### Acknowledgements

Chapter 3, in part, is a reprint of the material as it appears in Nature Biotechnology. Cowan, Quinn T.; Gu, Sifeng; Gu, Wanjun; Ranzau, Brodie L.; Simonson, Tatum S.; Komor, Alexis C., *Nature Biotechnology*, 2024.<sup>106</sup> The dissertation author was the primary investigator and author of this paper.

We are grateful to all members of the Komor lab for helpful discussions. We are thankful to Galia Debelouchina for use of the Neon transfection system and Brian Zid for use of the S3e cell sorter. We are grateful to Stephanie Gupton and Melissa Plooster for the identification of the schizophrenia associated SNVs. We would also like to thank Aileena Nelson and Courtney Sawyer. This research was supported by the University of California, San Diego and the NIH through grant no. 1R35GM138317 (to A.C.K) and the Research Corporation for Science Advancement through award number 28385 (to A.C.K). Q.T.C. was supported by the Molecular Biophysics Training Grant, NIH Grant T32 GM008326. B.L.R. was supported by the Chemistry-Biology Interface Training Program, NIH Grant T32 GM112584. A.C.K. and Q.T.C. are listed as inventors in a Regents of the University of California patent application (PCT/US23/66812) on the development and application of multiplexed orthogonal base editing systems in living cells.

## Chapter 4: Utilizing Base Editors to Generate Human Cell Lines Containing Single Nucleotide Variants

### 4.1 Introduction

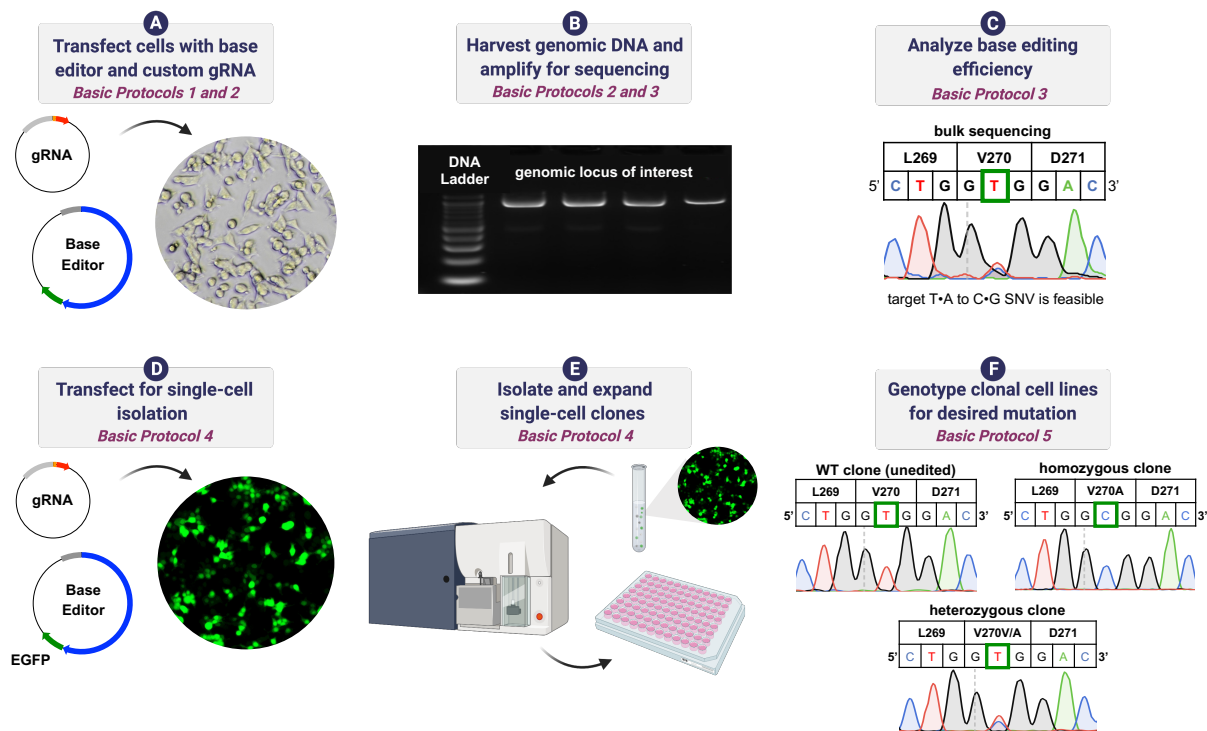
Base editing is a genome editing technique capable of precisely introducing single nucleotide variants (SNVs) into the genome of living cells.<sup>107</sup> Base editors (BEs) consist of a catalytically inactivated or impaired CRISPR-associated protein [dead Cas9 (dCas9), nickase Cas9 (nCas9), or dead Cas12a (dCas12a)] tethered to a single-stranded DNA (ssDNA)-modifying enzyme. In this system, the Cas protein will complex with a user-programmed guide RNA (gRNA), and bind to a genomic locus (termed the “protospacer”) that is complementary to the sequence of the gRNA and harbors a protospacer adjacent motif (PAM, a short DNA sequence specific to the Cas protein used, Figure 1.1).<sup>108</sup> The Cas protein locally denatures the target double-stranded DNA (dsDNA) region upon binding, to form an R-loop, exposing a small window (typically ~5 nucleotides long) of ssDNA.<sup>109</sup> Once bound, the ssDNA-modifying enzyme will catalyze the deamination of target nucleobases within this window. Subsequent DNA replication or repair of these modified base intermediates (uracil or inosine) results in permanent introduction of single-base substitutions. To date, two major classes of BEs have been described: cytosine base editors (CBEs), which change C•G base pairs to T•A,<sup>12</sup> and adenosine base editors (ABEs), which convert A•T base pairs to G•C.<sup>13</sup> Because BEs do not rely on double-stranded DNA breaks (DSBs), the introduction of unwanted gene alterations and random insertions and deletions (indels) of nucleotides at the targeted site are minimized compared to traditional genome editing methods such as ZFNs, TALENs, and the CRISPR-Cas9 system.<sup>110,111</sup> As a testament to their utility, BEs have been optimized and employed for a wide range of applications, including synthetic biology, therapeutics, and agriculture.<sup>107,112,113</sup> Of particular interest to the field of functional genomics, base-editing technologies are uniquely

positioned to facilitate studies investigating genotype-to-phenotype relationships by enabling researchers to generate isogenic cell lines with enhanced throughput.

Advances in next-generation sequencing are increasingly making the detection of SNVs routine, particularly in clinical settings.<sup>114</sup> Our ability to interpret the functional consequences of these SNVs, however, has lagged far behind. While the identification of clinically actionable and pathogenic mutations has revolutionized and improved the field of precision medicine,<sup>115,116</sup> these only represent a minority of reported human genetic variants. Indeed, while there are currently over 685 million human SNVs identified in the dbSNP database,<sup>117</sup> less than 1% have a defined clinical interpretation in ClinVar.<sup>118</sup> This issue is particularly endemic to rare genetic variants and those discovered in minoritized populations and indigenous people,<sup>119–121</sup> highlighting the need for a significant increase in studies that can functionally assess human genetic variants in laboratory settings. At the core of these studies, is the ability to introduce genetic variants in living cells with high efficiency and precision, something that can be accomplished using base editing technologies. Such approaches allow researchers to study variants of interest in their native cellular context, without confounding background genetic variation (which can occur in studies using patient-derived cell lines), and base editing facilitates the generation of these cellular models in a more time- and resource-effective manner than can be achieved with DSB-dependent editing strategies. Specifically, because SNV introduction efficiencies using DSB-mediated editing is so low (typically <1% in many cell types, compared to >20% for most targets using base editing), researchers must screen through hundreds of clones to obtain one with the desired genotype, or utilize DNA repair manipulation or donor template design “tricks” that can extend the timing of cell line generation by months or require the installation of additional, unwanted edit “scars” in the locus of interest (see the Commentary section for more details).<sup>110,122–127</sup> Using base editing, these SNV-containing isogenic cell lines can be generated in as little as ~2 weeks, with both heterozygous and homozygous genotypes generated with high efficiencies (~25% each of clones screened). This rapid and efficient

generation of isogenic cell line models could subsequently facilitate the clinical interpretation of variants with increased throughput compared to DSB-reliant methods and the identification of new therapeutic strategies in the field of precision medicine.

The protocols outlined below describe how to introduce site-specific C•G to T•A or A•T to G•C mutations in the genome of mammalian cells, with the ultimate goal of generating SNV isogenic cell lines (cell lines whose genomic sequences differ from each other at a single nucleotide). The methodology is broken down into five basic protocols, starting with designing and generating BE:gRNA combinations to introduce an SNV of interest (Basic Protocol 1). Notably, custom gRNAs can also be created using golden gate cloning and this strategy is recommended for many (>8) gRNAs. We then describe the process for validating these BE:gRNA combinations in bulk mammalian cells by Sanger or next-generation sequencing (Basic Protocol 2, Basic Protocol 3, and Alternate Protocol 1). We finish with the procedure to isolate single cells (Basic Protocol 4 and Alternate Protocol 2), then expand and genotype isogenic cell lines harboring the SNV of interest (Basic Protocol 5). We describe the whole process using four SNVs in the *MUTYH* gene [1: 45340220 (GRCh38); c.35G>A p.W12\*, c.809T>G p.V270A; c.886C>T p.L296L; and c.909G>A p.E303E] as illustrative examples. Additionally, this process (outlined in Figure 4.1) is designed to be modular, such that the researcher can adapt the presented methodology according to their specific experimental goals.



**Figure 4.1: Generating isogenic cell lines harboring a target SNV, using the MUTYH V270A variant as an example.**

(A) Transfection-quality plasmids encoding a custom-designed gRNA and base editor (BE) are transfected into mammalian cells. (B) Three days later, genomic DNA is harvested, and the genomic locus of interest is amplified via PCR and then prepared for sequencing. (C) The resulting PCR products are sequenced to assess the feasibility of generating an isogenic cell line harboring the target SNV. Shown are bulk sequencing results of the total cell population transfected with a BE:gRNA combination to introduce the V270A variant in MUTYH. The sequencing chromatogram shows base editing activity on the target base. (D) After confirmation of >10% editing efficiency, step A is repeated. The BE plasmid also encodes an EGFP fluorescent marker, allowing the use of FACS to isolate individual cells expressing a base editor. (E) FACS is utilized to sort single EGFP-positive cells into individual wells of a 96-well plate containing culture medium. (F) Single-cell-derived colonies clonally expand for 1-2 weeks. Sequencing of the resulting clonal cell lines confirms the generation of an isogenic cell lines harboring the target SNV. Data from three individual isogenic cell lines of the V270A MUTYH variant are shown, carrying wild-type, homozygous, and heterozygous genotypes. The target nucleotide is outlined in green.

#### 4.1.1 Strategic planning

The optimal BE:gRNA combination will change based on the experimental goal and sequence surrounding the target base. While selecting a base editor, gRNA spacer, and suitable PAM, there are several considerations (discussed in the Critical Parameters section)



that should be reviewed before starting. These include sequence context surrounding the target base, PAM and Cas variant selection, and acceptable levels of indels or off-target editing.

A particular challenge that many researchers encounter is “bystander editing”. This can occur if other C or A nucleobases fall within same the exposed ssDNA window as the target and are unintentionally modified; these edited bases are referred to as “bystander mutations” and can be problematic for phenotyping experiments. However, bystander editing can be minimized by using a particular deaminase variant (such as one with a shifted editing window or strict sequence preference) and careful PAM selection. The nucleotides flanking the target base—especially the 5’ nucleotide for CBEs—will also influence deaminase selection (see reference Arbab et al. 2020 for a detailed analysis of deaminase sequence preferences).<sup>79</sup> Finally, while certain experiments require no detectable indels and off-target editing (where high-fidelity Cas and deaminase variants can be used), others will benefit from using BEs with the highest on-target efficiency (see Critical Parameters). As base editing systems are modular, we recommend testing multiple BE:gRNA options initially (through Basic Protocol 3), and then selecting the combination that produces the best editing profile for isogenic cell line generation or downstream assays.

An additional matter to consider before beginning is the fact that immortalized cell lines may contain genetic variation or mutations in the target gene compared to the human reference genome; we thus recommend sequencing the locus of interest in your cell line before beginning. Within this protocol, we have included details on how to do this: see Basic Protocol 1 step 6A for primer design, Basic Protocol 2 for “Harvesting Genomic DNA”, and Basic Protocol 3 for sequencing the target site. Additionally, be sure to plan proper negative controls when generating SNV-containing isogenic cell lines (see Critical Parameters).

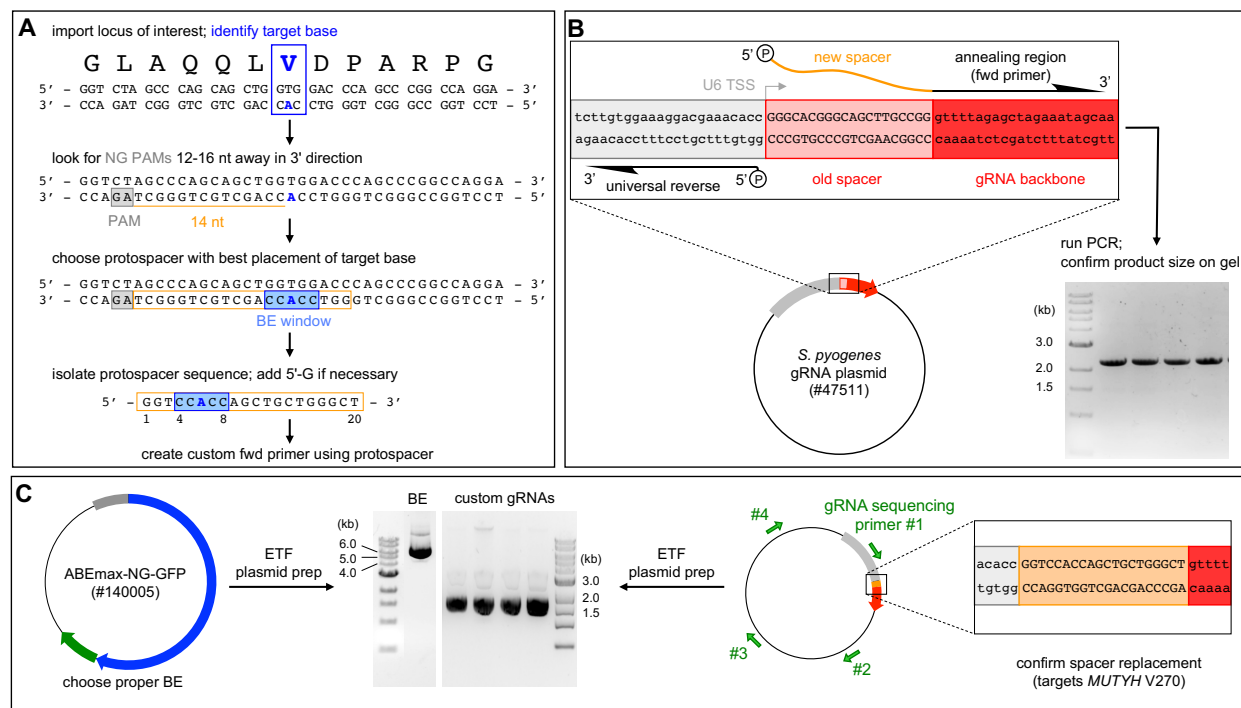
## 4.2 Basic Protocol 1: Design and Production of Plasmids for Base Editing

### Experiments

Base editing experiments rely on the preparation of two high-quality plasmids: one expresses gRNA that designates the target genomic location, and the other expresses the BE protein component (Cas9-deaminase fusion). Accurate and efficient base editing relies on careful design of the most appropriate BE:gRNA combination. Central to this, is the identification of a PAM that enables the Cas protein to bind in a manner that only positions the target C or A within the base editing window. We recommend the use of BEs that incorporate the engineered SpCas9-NG for minimal targeting restrictions with the smallest reduction in on-target activity in mammalian cells.<sup>41</sup> BEs that incorporate other engineered variants, Cas9 homologs, and Cas12a orthologs, such as SpRY, Sau/Sauri/Spy-mac (Cas9), and Lb/As/Aa (Cas12) offer additional PAM options but may have reduced on-target efficiencies at certain genomic loci or broadened editing windows, leading to bystander mutations (see the Critical Parameters section).<sup>42,45,59,61,129–131</sup> The observed editing efficiency will depend on the context of the genomic site and base editor used. We have listed recommended CBE and ABE variants in Table 2 and discuss alternatives in Critical Parameters. Additional characteristics of base editors and a decision tree for choosing the appropriate BE can be found in Anzalone et al 2020 review Figure 4 and 5 respectively.<sup>28</sup> Of note, optimized “BE-max” plasmids contain GenScript human codon optimization for maximal expression levels, bipartite NLSs on both termini for enhanced nuclear import, and an optional bicistronic enhanced green fluorescent protein (EGFP) for assessment of transfection efficiency and BE expression.<sup>65,132</sup> The gRNA expression vector should contain a U6 promoter driving expression of an *S. pyogenes* Cas9 gRNA (e.g. Addgene #47511). The initial spacer sequence is irrelevant because site-directed mutagenesis will be used in this protocol to replace the gRNA spacer using a 5'overhang on the forward primer.

This basic protocol outlines the steps to properly design gRNAs for BE experiments, replace the spacer sequence in a gRNA expression vector, prepare the appropriate plasmids for

mammalian cell transfection, and details primer design for the amplification of a target genomic locus to quantify targeted SNV introduction by Sanger or next-generation sequencing. If multiple BE:gRNA options are available, we recommend experimental validation to identify the optimal combination that maximizes target C or A editing and minimizes bystander editing. Automated programs can also be used to predict potential off-target sites (see Commentary section).



**Figure 4.2: Overview of Basic Protocol 1** (A) To design a custom gRNA (steps 1-5), first import the target locus into a viewing software of choice, being mindful of the isoform and reference genome used to annotate the SNV of interest (shown is the *MUTYH* gene surrounding residue 270, using the hg38 reference genome). Identify the target amino acid and nucleotide (blue, V270 codon, target adenine is on the template strand), and identify potential PAMs that are positioned 12-16 nt downstream, in the 3' direction, on the same DNA strand as the target nucleotide. Select the protospacer that positions the target nucleotide closest to the center of the editing window (position 6), or that positions potential bystander edits outside of the editing window. The gray "AG" PAM optimally positions the target base, while simultaneously pushing a potential bystander base to position 9, outside the canonical editing window. Order universal reverse and custom forward primers to replace the spacer sequence on a compatible gRNA expression plasmid (*S. pyogenes*, such as Addgene #47511). (B) Schematic showing the site-directed mutagenesis or around-the-horn cloning strategy for gRNA spacer replacement, and agarose gel showing confirmation of a PCR product band with the correct size (steps 8-10). (C) Preparation of endotoxin-free (ETF) midi- or max-preps of gRNA and BE plasmids (steps 19-24) with a representative agarose gel showing proper size and quality of supercoiled ETF plasmids. Additionally, in step 18, use Sanger sequencing to verify the sequence of the full gRNA plasmid.

**Materials:**

- Plasmids expressing the base editing machinery (available at Addgene): *S. pyogenes* Cas9 gRNA vector (#47511), BE4max-NG-GFP (equivalent to #125616 with ancAPOBEC or #140001 with CMV promoter) for C•G to T•A editing, and/or ABEmax-NG-GFP (#140005) for A•T to G•C editing.
- Nuclease-free water (VWR, cat. no. 82007-328)
- T4 PNK (NEB, cat. no. M0201L)
- T4 ligase buffer [ATP source] (NEB, cat. no. B0202S), or leftover from ligase kits
- Primers (Integrated DNA Technologies (IDT) or similar)
- Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, cat. no. F534L)
- 10 mM dNTPs (NEB, cat. no. N0447S)
- Reagents for agarose gel electrophoresis
- 1 kb DNA ladder (NEB cat. no. N3232L)
- QIAquick PCR purification kit (Qiagen, cat. no. 28106)
- Nanodrop spectrophotometer
- Quick ligation kit (NEB, cat. no. M2200L)
- Competent cells (Mach1, DH5a, or 10B)
- Ampicillin or carbenicillin, 100mg/mL stock solution (see Reagents and Solutions)
- Agar plates<sup>133</sup> with ampicillin/carbenicillin, final conc. 100 µg/mL
- Liquid LB or 2xYT media<sup>133</sup> with ampicillin/carbenicillin, final conc. 100 µg/mL
- 0.2 mL PCR tubes (Fisher Scientific, cat. no. 07-200-259)
- 1.5 mL nuclease-free microcentrifuge tubes (VWR, cat. no. 87003-294)
- Cryogenic storage vials, 2mL (VWR, cat. no. 10018-760)
- Plasmid mini prep kit (QIAGEN, cat. no. 28106)
- Endotoxin-free plasmid midi prep kit (Fisher Scientific, cat. no. NC0919795)

**Table 4.1: Primers and additional sequences used to generate custom gRNAs or amplify the target locus.** Underlined sequences anneal to the gRNA plasmid backbone; Bold sequences are protospacers (replace this with your custom protospacer); Italics represent 8 nt NGS barcode sequences (order multiple rd2 fwd and rev primers with difference barcodes), a unique fwd and rev barcode combination should be used for each sample.

| Usage             | #  | Primer                                      | Sequence                                                            | Notes                                             |
|-------------------|----|---------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------|
| gRNA cloning      | 1  | universal reverse                           | <u>GGTGTTCGTCCTTTCCACAAG</u>                                        | Site-directed mutagenesis (of spacer region)      |
|                   | 2  | V270 gRNA forward                           | <b><u>GGTCCACCAGCTGCTGGGCTG</u></b><br><u>TTTtagagctagaaatagca</u>  |                                                   |
|                   | 3  | E303 gRNA forward (adds 5'G)                | <b><u>GACAGGCTCTCCACAGGGCA</u></b><br><u>CGTTTTAGAGCTAGAAATAGCA</u> |                                                   |
|                   | 4  | L296 gRNA forward (adds 5'G)                | <b><u>GCCACTGTGCAGCCAGTGCCC</u></b><br><u>GTTTTAGAGCTAGAAATAGCA</u> |                                                   |
|                   | 5  | W12 (exon 1 pmSTOP) gRNA reverse            | <b><u>GTACCCACAGACGACTCAGG</u></b><br><u>GTTTTAGAGCTAGAAATAGCA</u>  |                                                   |
|                   | 6  | non-targeting gRNA spacer                   | GGTATTACTGATATTGGTGGG                                               | Spacer sequence only                              |
| Sanger sequencing | 7  | gRNA sequencing 1 (U6)                      | <u>TACGTGACGTAGAAAGTAAT</u>                                         | To verify gRNA clones                             |
|                   | 8  | gRNA sequencing 2 (ColE1)                   | <u>TTTCTCCCTTCGGGAAGCGT</u>                                         |                                                   |
|                   | 9  | gRNA sequencing 3 (backbone)                | <u>GAGATTATCAAAAAGGATCT</u>                                         |                                                   |
|                   | 10 | gRNA sequencing 4 (AmpR)                    | <u>TTGTGCAAAAAAGCGGTTAG</u>                                         |                                                   |
|                   | 11 | Sanger forward* <i>MUTYH</i> codons 193-332 | AGGAGGTGAATCAACTCTGGG C                                             | *Used to sequence V270, L296, E303                |
|                   | 12 | Sanger reverse <i>MUTYH</i> codons 193-332  | CCGAACCCTACTCAAGCCAAG A                                             |                                                   |
| NGS               | 13 | NGS rd1 adapter forward                     | ACACTCTTTCCCTACACGACGC TCTTCCGATCTNNNN                              | Add sequence to 5' end of target-specific primers |
|                   | 14 | NGS rd1 adapter reverse                     | TGGAGTTCAGACGTGTGCTCTT CCGATCT                                      |                                                   |
|                   | 15 | NGS rd1 <i>MUTYH</i> codons 264-291 forward | ACACTCTTTCCCTACACGACGC TCTTCCGATCTNNNNAGCAGCT CTGGTAGGATGTTGG       | Rd1 primers for V270 example                      |
|                   | 16 | NGS rd1 <i>MUTYH</i> codons 264-291 reverse | TGGAGTTCAGACGTGTGCTCTT CCGATCTCCCAGTAGGCTTACT CTCTGGC               |                                                   |

**Table 4.2: Suggested base editors designed to maximize on-target editing or minimize off-target (OT) editing.**

|     | High-efficiency    | Addgene #      | Reduced off-target editing | Notes                                            | Addgene #               |
|-----|--------------------|----------------|----------------------------|--------------------------------------------------|-------------------------|
| CBE | BE4max-NG-P2A-EGFP | 125616, 140001 | BE4max-NG (YE1)            | Less processive, narrowed window                 | 138159                  |
|     | RrAPOBEC3F         | 138340         | RrAPOBEC3F (F130L)         | Retains high on-target activity                  | 138341                  |
|     | PpAPOBEC1          | 138349         | PpAPOBEC1 (H122A)          | Minimal OTs, slightly reduced on-target activity | 138345, 138338          |
|     | SsAPOBEC3B         | 138343         | SsAPOBEC3B (R54Q)          | Minimal OTs, "BC" preference                     | 138344                  |
|     | AncBE4max-P2A-EGFP | 112100         | eA3A-BE3 (N57G)            | "(A)UC" preference                               | 131315                  |
| ABE | ABEmax-NG-P2A-EGFP | 140005         | evo-TadA (V106W)           | Inactivated or deleted wt-TadA                   | 125647, 138495          |
|     | NG-ABE8e           | 138491         | evo-TadA (F148A)           |                                                  | <i>Zhou et al. 2019</i> |
|     | ABE8.20-m          | 136300         | evo-TadA (V82G)            |                                                  | 131313                  |

## Protocol Steps

### Design gRNA

1. Import genomic locus of interest into sequence viewing software of choice. We frequently use Benchling (<https://www.benchling.com>) or SnapGene (<https://www.snapgene.com>).

*Be sure to import in the sense orientation using the reference genome ('build' e.g. Hg38) and isoform that matches the SNV annotation.*

2. Identify the target nucleotide (C or A) and find an NG PAM exactly 12 to 16 nucleotides downstream (i.e. towards the 3' end of the strand containing the target nucleotide, see Figure 4.2A for an example). The target can be on either strand, as long as there is a suitable PAM on the same strand. Annotate the protospacer. Alternately, free software programs such as the Benchling wizard (Komor, 2016), BE-designer,<sup>135</sup> and BE-Hive<sup>58128</sup> can be used for automated gRNA design (see Critical Parameters). As most

automated programs lack flexibility in deaminase and Cas variants, we recommend manually designing your protospacers (Figure 4.2A) and checking predicted editing efficiencies in BE-Hive.<sup>128</sup>

*If bystander edits are unavoidable, choose the protospacer with bystander nucleotides farthest from the center of the window (position 6) or one that will incorporate silent mutations (as long as it is not a splice site).*

3. If the protospacer does not start with guanine, add a 5'G to create a 21-nucleotide spacer in the gRNA. For protospacers that already start with a 5'G (such as the example V270 protospacer), move to the next step. *NOTE:* Adding a guanine drastically increases gRNA transcription from the U6 promoter, yielding high editing efficiencies, despite a possible 5' mismatch.
4. Design a custom forward primer with the sequence 5'-[N<sub>20-21</sub>]GTTTGTAGAGCTAGAAATAGCA-3', replacing the [N<sub>20-21</sub>] portion with the protospacer sequence from the previous step (e.g. Table 1, #2 for MUTYH V270). This will be used with the universal reverse primer (Table 1, #1). Analyze this primer pair for homo- and hetero-dimers using the IDT oligo analyzer (<https://www.idtdna.com/pages/tools/oligoanalyzer>). Note that amplification may be difficult using primers with a  $\Delta G < -10$  kcal/mol.

*The custom primer sequence should only include the protospacer sequence without the PAM. In base editing experiments, the spacer sequence will always be identical to the protospacer and contain the target A or C because direct nucleobase modification occurs on ssDNA of the canonical “non-target strand” that is not duplexed with gRNA. See Figure 4.2A for an example. It is also recommended to include a non-targeting gRNA as a control (see Critical Parameters and Table 1 primer #6).*

5. Order the primers from IDT or another manufacturer of choice for custom gRNA construction via site-directed mutagenesis of the spacer or “around-the-horn” cloning.

***Design primers for amplification of genomic locus***

6. A. Use a primer design tool (e.g. Primer3 or other) to generate a pair of primers that amplify a ~1 kilobase (kb) region containing the protospacer, for Sanger sequencing. See Figure 4A and Table 1, primers #11-12 for V270 example using the Benchling Primer3 design wizard.

*These primers should ideally be 18-24 base pairs (bp) long, and have 40-60% GC content, and melting temperatures of 60-65°C. To ensure that at least one of these primers is appropriately positioned to yield maximum quality sequencing data of the protospacer region of the amplicon, the target sequencing region should be at least 50 bp away from the primer used for sequencing.*

B. Alternatively, design primers for Illumina next-generation sequencing (NGS). This requires specific adapter sequences be added to both the forward and reverse primers, and a much shorter amplicon. Design a primer pair that amplifies 200-250 bp of DNA (for a 300-cycle NGS run; if a 150-cycle NGS run will be used, the amplicon should be 100-125bp long) with the same parameters as above. Then, add the adapter sequences #13 and #14 from Table 1 onto the 5'end of the forward and reverse primers, respectively. For the MUTYH V270 example, the Primer3 Wizard on Benchling was used to identify 22 nt PCR primers that amplify a 249 bp region around the protospacer and the Illumina NGS adapters were added to create primers #15 and 16 in Table 1. Order these as the round 1 PCR primers. A second round of amplification is required to provide the samples with P5/P7 tails and unique barcodes. Order round 2 PCR primers based on the Illumina barcoding system to be used during library preparation. See Alternate Protocol 1 for more information.



***Clone new spacer into gRNA expression plasmid via site-directed mutagenesis***

7. In 0.2 mL PCR tubes, 5' phosphorylate each primer (forward and reverse, from step #5 above) individually in separate 20  $\mu$ L reactions by combining the following (in the order stated):

- 15  $\mu$ L nuclease-free water
- 2  $\mu$ L primer (100  $\mu$ M stock)
- 2  $\mu$ L T4 DNA ligase buffer
- 1  $\mu$ L T4 polynucleotide kinase

*Mix well and quickly spin to collect sample. The 20  $\mu$ L of phosphorylated primer is enough for 8 PCR reactions—scale up if the primer is needed for more than 8 RNAs (i.e. the universal reverse).*

8. Place the reaction tube in a thermocycler and run the following program:

- 37°C for 20 min
- 95°C for 5 min
- Hold at 12°C

9. In a PCR tube on ice, combine the following reagents for a 50  $\mu$ L reaction (in the order stated):

- X  $\mu$ L nuclease-free water (fill to a total volume of 50  $\mu$ L)
- 10  $\mu$ L Phusion HF buffer
- 1  $\mu$ L dNTP mix
- 2.5  $\mu$ L FWD phosphorylated primer (10  $\mu$ M from previous step)
- 2.5  $\mu$ L REV phosphorylated primer (10  $\mu$ M from previous step)
- ~10 ng DNA template: gRNA expression plasmid with different spacer sequence, (Addgene # 47511, noted above)
- 0.5  $\mu$ L Phusion polymerase

10. Mix, quickly spin, and run a thermocycler program with the following cycling conditions:

|             |           |      |                        |
|-------------|-----------|------|------------------------|
| 1 cycle:    | 30 sec    | 98°C | (initial denaturation) |
| ~35 cycles: | 10 sec    | 98°C | (denaturation)         |
|             | 20 sec    | 65°C | (annealing)            |
|             | 35-45 sec | 72°C | (extension)            |
| 1 cycle:    | 5 min     | 72°C | (final extension)      |
|             | hold      | 12°C | (hold)                 |

11. Run a 1-2 uL aliquot out on an agarose gel (~1.5%) with a 1kb DNA ladder to check for a PCR product that is ~2.3 kb long (see Figure 4.2B for an example).

*If there is no PCR product, use the GC buffer, or additives such as DMSO or formamide (5%). If there are unwanted products, increasing the annealing temperature or decreasing the primer concentration can improve specificity. Gel extracting the band of interest is also possible during step 13.*

12. Add 1 µL DpnI to the PCR product, mix, and spin. Incubate at 37°C for 1 hr (if you are in a rush, 15 min should be adequate to digest any template).

13. Purify PCR products using the Qiagen PCR clean-up protocol.

*Using a vacuum manifold followed by thoroughly drying the column in a centrifuge is suggested.*

14. Quantify the concentration of the PCR product using a Nanodrop. Then, prepare a 20 uL ligation reaction to circularize the linear PCR product, which contains the new spacer sequence and 5' phosphate groups. In a PCR tube, combine:

~50 ng PCR product after clean-up (from previous step)

X µL nuclease-free water (fill to a total volume of 20 µL)

10  $\mu$ L 2X Quick Ligase buffer

1  $\mu$ L Quick Ligase

15. Mix, quickly spin, and incubate at room temperature for 5-10 min.

16. Using sterile technique, transform  $\sim$ 5  $\mu$ L of the ligated product into competent bacterial cells of choice. We recommend Alternate Protocol 1 described by Seidman and coworkers.<sup>136</sup> Plate 10-50  $\mu$ L cells on an agar plate containing 100  $\mu$ g/mL ampicillin or carbenicillin in a manner that will yield single colonies, and grow at 37°C overnight.

*The BE and gRNA plasmids we recommend confer ampicillin/carbenicillin resistance. Expect  $\sim$ 20-200 colonies, with the majority containing the desired spacer sequence replaced.*

17. Pick 1-2 colonies per clone and grow in liquid LB or 2xYT media with 100  $\mu$ g/mL ampicillin/carbenicillin until saturated.

*Using cells that double rapidly, such as Mach1, allows for short growth periods ( $\sim$ 6 hrs). Inoculating colonies in 1-3 mL of media allows for sequencing on the same day. Save the plates airtight at 4°C (using Parafilm) and mark the colonies screened.*

18. Mini-prep saturated cultures.

*We recommend saving the original culture tubes with un-lysed bacteria at 4°C to create starter cultures for correct clones.*

19. Quantify the prep using a Nanodrop, and then run an aliquot on an agarose gel (1.5%) to check the quality and size of the plasmid (Figure 4.2C for an example).

*The supercoiled gRNA plasmid runs around  $\sim$ 1.5\_kb on a linearized ladder.*

20. Send in preps for Sanger sequencing using gRNA sequencing primer #7 in Table 1 (U6), to confirm the presence of the new spacer sequence.

*Sequence the entire backbone using primers #8-10 in Table 1 (at this step or after step 26, the midi-prep) as mutations may occur during PCR. Note that the universal U6 promoter primer that many sequencing companies use anneals very close to the spacer region and may not provide high-quality sequencing data of this region of the plasmid.*

**Prepare High Quality, Endotoxin-free gRNA and BE plasmid DNA**

21. Once the base editor plasmids are obtained from Addgene, streak the surface of the bacterial stab onto an LB or 2xYT plate containing 100 µg/mL ampicillin or carbenicillin. Use sterile technique and streak for single colonies. Incubate at 37°C overnight (see <https://www.addgene.org/recipient-instructions/myplasmid/>).
22. In the morning, grow 0.5 mL starter cultures of LB or 2xYT media with 100 µg/mL ampicillin or carbenicillin for each BE and gRNA plasmid required, at 37°C for ~6 hr or until saturated. For BEs, inoculate a single colony from plate in the previous step. For gRNAs, add antibiotic-containing media to culture tube saved at step 18 or inoculate from the original plate to avoid re-transforming.
23. Dilute into 50 mL for midi-prep (or 200 mL for maxi-prep) cultures for overnight growth.
24. Once saturated, remove 0.5 mL of culture and combine with 0.5 mL 50% glycerol in a 2 mL cryogenic vial. Freeze stock at -80°C.
25. Use the remainder of the culture to prepare endotoxin-free midi- or maxi-preps based on the manufacturer's protocol.

*If solution is cloudy after the syringe filter step of the midi- or maxi-prep, centrifuge the 50 mL tube for 10 min at 4,000 rcf and proceed with supernatant. Using a vacuum manifold is suggested. Drying the wash buffer from the column is critical (spin in microcentrifuge for 2-3 min at ≥12,000 rcf). At the elution step, warming the buffer to 55°C and incubating 3-5 min increases DNA yield.*

26. Quantify the plasmid sample using a Nanodrop and run an aliquot out on an agarose gel to confirm the quality of midi- or maxi-prep (Figure 4.2C).

*BEs plasmids are high-copy while the gRNA plasmid is medium- to low-copy.*

### Golden Gate cloning: spacer insertion for gRNAs

This represents an attractive alternative to site directed mutagenesis if you are cloning many gRNAs. The following would replace steps 7-20 described in the protocol above.

#### Oligo design:

There are two type II-S recognition enzyme recognition sites (e.g. BsmBI, BsaI, BbsI) that, once digested, reveal unique overhangs in the Golden Gate (GG) backbone (in this case a gRNA expression vector). This allows the insertion of spacer sequences via annealed oligos without the need to PCR amplify. To design oligos, add the 4nt overhangs to the 5' end of your primers. In the example, the forward has a "CACC" in yellow and reverse has "AAAC" in green. This is the case for most gRNAs such as the *S. pyogenes* GG backbone (pSG141 or 941b). Make sure your spacer starts with a "G" if you are expressing from a U6 promoter.

|     | [spacer]                                         | Example for A111V-EGFP spacers |
|-----|--------------------------------------------------|--------------------------------|
| Fwd | CACC [N <sub>20</sub> -N <sub>21</sub> ]         | CACC[gctctacgcggtcttgtgt]      |
| Rev | AAAC [N <sub>20</sub> -N <sub>21</sub> _rev-com] | AAAC[actacacgacccgcgtacgac]    |

#### Phosphorylation & annealing:

- In PCR tubes, combine the following for each ABE/CBE spacer replacement:

|                                   |        |
|-----------------------------------|--------|
| Total volume:                     | 10 µL  |
| Nuclease-free H <sub>2</sub> O    | 6.5 µL |
| T4 ligase buffer                  | 1 µL   |
| 100 µM "Forward" oligo (100 pmol) | 1 µL   |
| 100 µM "Reverse" oligo (100 pmol) | 1 µL   |
| T4 PNK                            | 0.5 µL |

- Thermocycler: 30 min @ 37°C → 5 min @ 95°C → slowly ramp down to RT

- Dilute annealed oligo duplex 1:20 (add 190  $\mu\text{L}$   $\text{H}_2\text{O}$  to each)

#### Digestion & ligation:

- Combine the following in PCR tubes for either 8  $\mu\text{L}$  reactions (can scale up if needed):

|                                                      |                       |
|------------------------------------------------------|-----------------------|
| <i>Total volume:</i>                                 | 8 $\mu\text{L}$       |
| <i>Nuclease-free <math>\text{H}_2\text{O}</math></i> | 5.6 - X $\mu\text{L}$ |
| <i>T4 DNA ligase buffer</i>                          | 0.8 $\mu\text{L}$     |
| <i>Annealed oligo duplex</i>                         | 1 $\mu\text{L}$       |
| <i>20-50 ng BB vector (QTC221-3)</i>                 | X $\mu\text{L}$       |
| <i>Enzyme (BbsI or BsmBI)</i>                        | 0.4 $\mu\text{L}$     |
| <i>T4 DNA ligase</i>                                 | 0.2 $\mu\text{L}$     |

- Run a negative control without annealed oligos to see how much BB is coming through.
- Thermocycler: 5 min @ 37°C (or 42°C for BsmBI) → 5 min @ 16°C → cycle 3-6X → 10 min @ 50°C “heat soak” → 4°C hold
- Transform right away into competent cells of choice (e.g., Mach1)
  - Plate on Carb or antibiotic from GG backbone.
  - If you are cloning a lot of gRNAs, use multichannel (using <7 $\mu\text{L}$  you can get 8 per plate) and drip down near the flame.

#### Screening:

- Inoculate cultures for mini-prep:
  - Daytime (6 hr @ 37°C): 1.5 mL 2xYT media + 100  $\mu\text{g}/\text{mL}$  carbenicillin
  - Overnight (16 hr @ 37°C): 5.0 mL 2xYT media + 100  $\mu\text{g}/\text{mL}$  carbenicillin
- Mini-prep according to manufacturer’s recommendations (Zyppy ETF kit or Omega columns with homemade buffers)
- *Optionally*, set up a small-scale digest (<50 ng DNA and 0.2  $\mu\text{L}$  BbsI in 5  $\mu\text{L}$  total volume) and run it all on a gel with the uncut BB vector.

- This is to see if the BbsI cut sites were replaced by the new spacers: any that resemble uncut plasmids are likely correct (supercoiled 2.3kb so ~1.5kb on a linear ladder), while plasmids that are linearized (2.3kb) have no insert because the restriction enzyme sites are still there.
- Sanger sequence through gRNA spacer with ACK365 primer.

#### **4.3 Basic Protocol 2: Transfecting Adherent Cells and Harvesting Genomic DNA**

Before proceeding to generate isogenic cell lines, it is highly recommended to validate the BE:gRNA combination(s) from Basic Protocol 1. The most time- and resource-effective way to accomplish this is with an easy-to-transfect cell line, such as Human Embryonic Kidney 293 (HEK293) cells. While many different methods are available to deliver the BE and gRNA into cells (such as viral transduction, mRNA transfection or electroporation, and ribonucleoprotein transfection or electroporation), the protocol outlined here utilizes commercial cationic lipid reagents to transfect cells with plasmids encoding the BE and gRNA. This basic protocol details how to deliver transfection-quality plasmids into HEK293 cells and extract the genomic DNA (gDNA) for downstream sequencing.

We utilize HEK293T cells for base editing experiments. HEK293T cells contain an endogenous copy of the SV40 large T antigen. As such, plasmids containing an SV40 origin of replication will be replicated in HEK293T cells, resulting in longer-term expression of recombinant proteins, which can cause higher levels of off-target editing with genome editing agents. Because of that, the plasmids used by the authors do not contain an SV40 origin of replication.

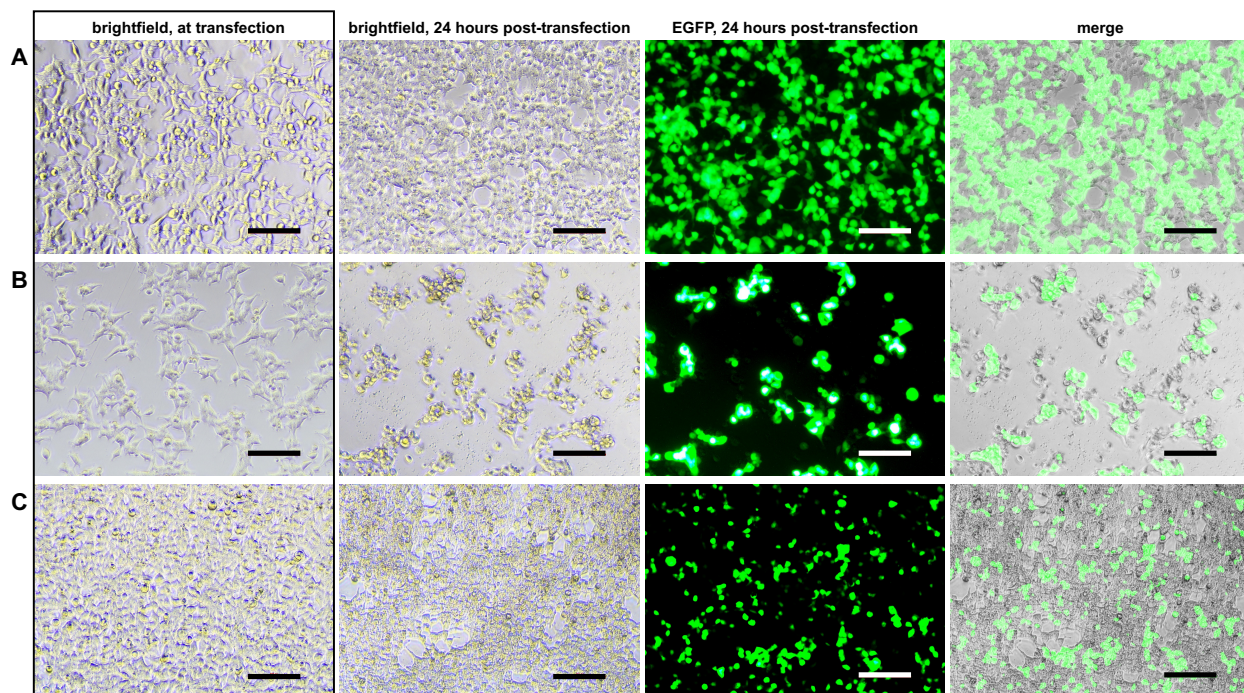
*NOTE:* All culture incubations should be performed in a humidified 37°C, 5% CO<sub>2</sub> (g) incubator. Use 10% FBS in DMEM with or without added pen/strep for all steps as indicated. All experimental work in this section must be done using sterile reagents and proper aseptic

techniques.<sup>137</sup> Failure to do so, will result in possible contamination, which will disrupt downstream workflow and cause wasting of valuable laboratory resources.

**Materials:**

- HEK293T cells (ATCC CRL-3216)
- Nuclease-free water (VWR, cat. no. 82007-328)
- Endotoxin-free base editor and gRNA plasmids from Basic Protocol 1
- 48-well plates (VWR, cat. no. 10062-898)
- Lipofectamine 2000 (Life Technologies, cat. no. 11668-019)
- Opti-MEM Reduced Serum Medium, Gibco (Life technologies cat. no. 31985070)
- Phosphate-buffered saline (PBS; Gibco Life, cat. no. 10010049)
- Dulbecco's Modified Eagle's Medium (DMEM; Thermo Fisher Scientific, cat. no. 10566024)
- Fetal Bovine Serum (FBS; Life Technologies, cat. no. 10437-028)
- 1.5 mL nuclease-free microcentrifuge tubes (VWR, cat. no. 87003-294)
- 10% (w/v) sodium dodecyl sulfate (SDS) stock solution (see Reagents and Solutions)
- 1M Tris HCL stock solution (pH 8.0; see Reagents and Solutions).
- Proteinase K enzyme (NEB, cat. no. P8107S)
- 0.2 mL PCR tubes (Fisher Scientific, cat. no. 07-200-259)
- 1.5 mL nuclease-free microcentrifuge tubes (VWR, cat. no. 87003-294)
- Fluorescence microscope
- 37°C incubator with humidified atmosphere of 5% CO<sub>2</sub> (g)





**Figure 4.3: Representative images of successful and unsuccessful transfection experiments (Basic Protocol 2, steps 1-9).** Shown are brightfield images of HEK293T cell confluency at the time of transfection (left most column, in box), as well as brightfield (middle left column), GFP fluorescence (middle right column), and merged (right most column) images of the same cells 24 hours post-transfection with gRNA and BE plasmids to generate the V270A MUTYH variant. Shown are cells that were transfected at **row (A)** – ideal confluency (70-85%), resulting in high transfection efficiency and high cell viability across the surface of the plate, **row (B)** – under confluency (<70%), resulting in high transfection efficiency but low cell viability, and **row (C)** – over confluency (>85%), resulting in low transfection efficiency but high cell viability. The latter two conditions are illustrative of poor base editing experiments. Scale bars = 166  $\mu\text{m}$ .

## Protocol Steps

### *Transfecting HEK293T Cells with BE and gRNA*

1. Plate HEK293T cells such that they will be 70-85% confluent upon transfection in medium without antibiotics (see Figure 4.3A for example).

*For a 48-well plate, 100,000 cells/well (in a well consisting of 250  $\mu\text{L}$  of 10% FBS in DMEM culture medium, or a cell solution of  $4.0 \times 10^5$  cell/mL) is usually ready to transfect 4 hours after plating. 50,000 cells/well (in 250  $\mu\text{L}$  of medium per well, or a cell solution of  $2.0 \times 10^5$  cell/mL) is usually ready to transfect 12-16 hours after plating.*

2. Warm Opti-MEM to room temperature prior to adding Lipofectamine 2000.
3. For each transfection, 200 ng of gRNA plasmid and 800 ng of BE plasmid is needed if done in a 48-well plate format. Include a negative control sample that contains a non-targeting gRNA or no gRNA (see Commentary “Inclusion of proper controls”). For different transfection formats or cell types, the total amount of plasmid and Lipofectamine will need to be adjusted (scaled up or down). Prepare two different tubes per transfection:

- a. Tube A: add 200 ng of gRNA plasmid and 800 ng of BE plasmid. Dilute the DNA with Opti-MEM to a total volume of 12.5  $\mu$ L.
- b. Tube B: add 1.5  $\mu$ L of Lipofectamine 2000 and 11  $\mu$ L of Opti-MEM.

*In some instances, you may be required to generate multiple edits in parallel. In this case, follow step 3a and 3b for each BE:gRNA combination.*

4. Mix the contents of tube A and tube B and incubate the mixture for 15 minutes at room temperature prior to transfection.
5. Carefully add the 25  $\mu$ L mixture dropwise to the cells and place back into the incubator.

*Be careful not to touch the bottom of the well with the pipet tip to avoid disturbing the cell monolayer.*

6. Monitor the cells under a fluorescence microscope after 24 hr. EGFP-positive cells should be observed evenly across the surface of the plate (see Figure 4.3A). Transfection efficiency should be 40- 70%. If not, please refer to the Troubleshooting section.
7. Optionally, after 24 hr, gently pipette off or aspirate the old medium.
8. Gently add 250  $\mu$ L of fresh pre-warmed culture medium (10% FBS in DMEM, optionally with 1% Pen/Strep to avoid contamination).

*Lipofectamine can be toxic to the cells. If cells appear healthy, no media change is required or changing after 48 hr is also sufficient. If high cell toxicity is*

*observed, as seen in Figure 4.3B, change to fresh DMEM medium prior to 24 hr. Replacing just 200µL of media is also an option if you do not want to lose any cells.*

9. Return to incubator for two additional days after media is replaced (3 days total after transfection).

### **Harvesting Genomic DNA**

10. Prepare cell lysis solution prior to taking cells from step 9 out of the incubator. 100 µL of cell lysis solution will be required per well when done in a 48-well plate format. As such, prepare a master mix solution containing the following: 10 mM Tris (pH 7.5), 0.05% SDS, 25 µg/mL proteinase K.

*For example, a 1-mL master mix cell lysis solution will contain the following: 983.75 µL of nuclease-free water, 10 µL of 1M Tris HCL stock solution (pH 8.0), 5 µL of 10% SDS stock solution, and 1.25 µL of Proteinase K enzyme. The Tris and SDS may be prepared in advance, but the Proteinase K must be added immediately prior to use.*

11. Pipette or aspirate 250 µL of the old medium and gently rinse the cells with 150 µL of PBS. Carefully pipette or aspirate off the PBS.
12. Add 100 µL of cell lysis solution prepared in step 10 to each well and wait for 1-2 minutes.
13. Collect the cell lysate into PCR tubes.

*Slightly tilt the 48-well plate and use a scraping motion to push all the lysed cells to the bottom corner. Slowly collect 100 µL of cell lysate on the first suction; you may need to wait a few seconds for the viscous lysate to be drawn up the pipette. Do not pipette the cell lysate up and down, especially vigorously, as it will shear the genomic DNA and result in air bubbles. Be advised—the viscosity of genomic*

*DNA may create “stringy” filaments coming from the pipette tips so be careful not to cross-contaminate samples of lysate.*

14. Incubate the cell lysis reaction in the thermocycler:

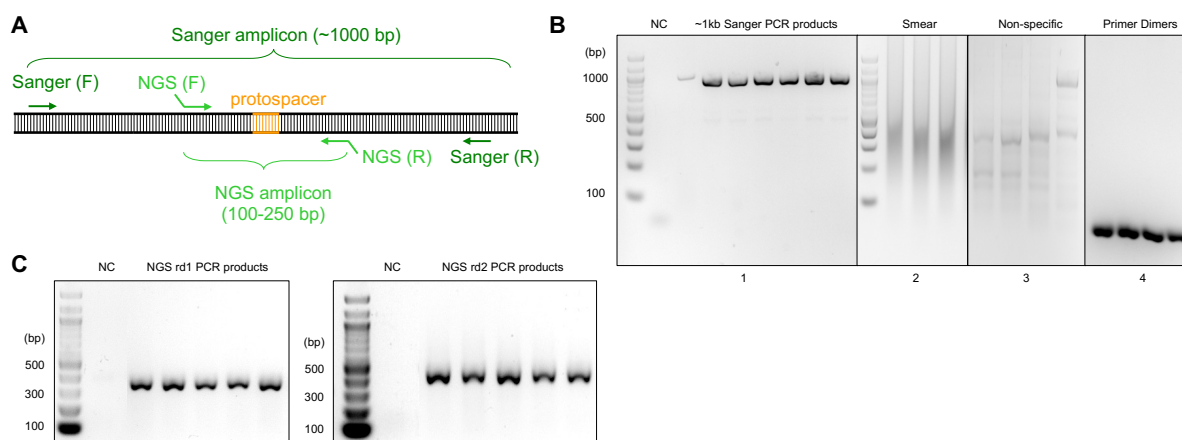
- 37°C for 1 hr 30 min
- 80°C for 30 min
- Hold at 4°C

*Keep cell lysate at 4°C and avoid freeze/thaw cycles. Amplification works best on fresh gDNA, which can be used directly or diluted (1:2-1:10) depending on the cell density before harvesting. Cell lysates will be more viscous when higher amounts of gDNA are present.*

#### **4.4 Basic Protocol 3: Genotyping Harvested Cells Using Sanger Sequencing**

After transfection with the BE:gRNA combination(s), base editing efficiency needs to be assessed to determine the feasibility of generating an isogenic cell line harboring the target SNV. Generally, an editing efficiency >10% is sufficient to proceed (Figures 2C; 6A). However, base editing efficiencies of >30% greatly increase the chances of generating all genotypes.

Sanger sequencing is a cost-effective alternative to next-generation sequencing (Alternate Protocol 1) and is usually sufficient to determine the feasibility of generating cell lines with the SNV of interest. Please refer to the “Using EditR to Quantify Base Editing Efficiencies” section to understand how to determine editing efficiency using Sanger Sequencing (directly after step 4 of Basic Protocol 3).



**Figure 4.4: Overview of gDNA amplification (Basic Protocol 2, steps 10-14).** (A) When designing primers for Sanger sequencing, ensure there is at least 50 bp between the protospacer and the beginning of the sequencing read. Positioning the protospacer in the center of the read will give the best chance for high quality data within the protospacer region. NGS amplicons should amplify a gDNA region based on the number of cycles in the NGS run (i.e. for a 300-cycle run, the amplicon should be ~250 bp, and for a 150-cycle run, the amplicon should be ~100 bp). (B) For MUTYH V270 Sanger sequencing, primers #11 and 12 (Table 1) were used for amplification of a ~1 kb amplicon. Panel 1 displays a 2% agarose gel showing a clean negative control (NC, lane 1) and ideal gDNA amplification products (lanes 3-8). Additional 2% gels show common issues we have experienced. Specifically, panel 2 shows a PCR product following the use of too much gDNA template (as evidence by smears), panel 3 shows a PCR product with off-target bands, and panel 4 shows a PCR outcome that predominantly consists of primer dimer bands. (C) Representative 2% agarose gels of round 1 and 2 PCR products for NGS using a ~280 bp amplicon as an example. The negative control samples contain water (round 1) or the previous negative control PCR product (round 2) as a template instead of gDNA. Round 1 primer adapter sequences will add 66 bp to the length of the amplicon, creating a ~350 bp product in this example. Round 2 primer sequences add an additional 74 bp to the length of the amplicon, resulting in a ~420 bp product (total of 140bp added to the amplicon by PCR).

#### Materials:

- QIAquick PCR purification kit (Qiagen, cat. no. 28106)
- Cell lysate from Basic Protocol 2
- Primers
- dNTP mix, 10mM each (NEB, cat. no. N0447L)
- Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, cat. no. F534L)
- DMSO

- 0.2 mL PCR tubes (Fisher Scientific, cat. no. 07-200-259)
- Thermocycler
- Reagents for agarose gel electrophoresis
- 100 bp DNA ladder (NEB, cat. no. N3231L)
- Nanodrop spectrophotometer
- Access to Sanger sequencing facility (e.g. GENEWIZ/Azenta)
- Data analysis platform to analyze the sequencing results (e.g. Benchling)

## Protocol Steps

### ***PCR of genomic DNA***

1. For each sample, combine the following reagents (in the order stated) for a 50  $\mu$ L PCR reaction (can be scaled down to 25  $\mu$ L to save reagents) to amplify the genomic locus of interest:

- X  $\mu$ L Nuclease-free water (fill to a total volume of 50  $\mu$ L)
- 10  $\mu$ L GC Buffer
- 1  $\mu$ L dNTP mix (10 mM each)
- 0.5  $\mu$ L 10 $\mu$ M forward primer (in this example, #11 **Table 1**)
- 0.5  $\mu$ L 10 $\mu$ M reverse primer (in this example, #12 **Table 1**)
- 1.5  $\mu$ L 100% DMSO (final concentration 3%)
- 0.5  $\mu$ L of Phusion DNA polymerase
- 1.0  $\mu$ L cell lysate (gDNA) from step 14 of Basic Protocol 2

*It is imperative to include a negative control sample every time amplification from gDNA is performed. In this sample, 0.5  $\mu$ L of water is added instead of the cell lysis solution (gDNA) to control for reagent contamination. We suggest making a master mix, then adding the cell lysate to individual aliquots.*

2. Mix, quickly spin, and run a thermocycler program with the following cycling conditions:

|                   |        |        |                        |
|-------------------|--------|--------|------------------------|
| 1 cycle:          | 60 sec | 98°C   | (initial denaturation) |
| 25 to 28 cycles*: | 16 sec | 98°C   | (denaturation)         |
|                   | 20 sec | 65°C** | (annealing)            |
|                   | 16 sec | 72°C   | (extension)            |
| 1 cycle:          | 5 min  | 72°C   | (final extension)      |
| 1 cycle:          | ∞      | 12°C   | (hold)                 |

\*See “Quantification of base editing efficiency in bulk cells” in the *Critical Parameters* section for more information; use the fewest number of cycles during gDNA amplification that yields an on-target band to avoid PCR bias.

\*\*The annealing temperature is specific to the primers chosen. In this protocol, it is recommended that researchers use an automated primer design tool (Step 6A of Basic Protocol 1). Using the Benchling Primer3 design wizard, the default optimal annealing temperature is 65°C. If the primer pair’s annealing temperature is not 65°C, please adjust accordingly.

3. Run a 2-5 µL aliquot of each PCR reaction on a 2% agarose gel using a 100 bp ladder for size comparison (see Figure 4.4B for an example).

*If you do not get the correct length of PCR product, please refer to the Troubleshooting section.*

4. After confirming the presence of a correct length PCR product, prepare the samples for sequencing. Genewiz/Azenta (<https://www.genewiz.com/en>) offers two Sanger sequencing options, but this may depend on the sequencing vendor of choice:
  - a. Sending purified PCR product: Purify the product with QIAquick PCR purification kit following the manufacturer’s instructions. Measure the DNA concentration after PCR clean-up of each sample using a Nanodrop. Prepare the appropriate amount of DNA

along with the sequencing primer (we recommend using either the forward or reverse primer from the PCR reaction that is positioned at least 50-75 bp away from the target base of interest) in the same tube according to service provider's instructions.

- b. Sending unpurified PCR product: unpurified PCR samples sent for Sanger sequencing will undergo an enzymatic PCR purification protocol. Provide the appropriate amount of the successful PCR reaction in one tube according to sequencing company guidelines. Provide the appropriate amount of sequencing primer (we recommend using either the forward or reverse primer from the PCR reaction, whichever one is at least 25-bp away from the target base of interest) in a separate tube. We recommend submitting a picture of the labelled gel as well to the sequencing company as this can help with quantification and troubleshooting, if necessary.

*Sending unpurified PCR product (step 4b.) usually results in better sequencing coverage quality and thus, an easier to interpret chromatogram, but costs extra.*

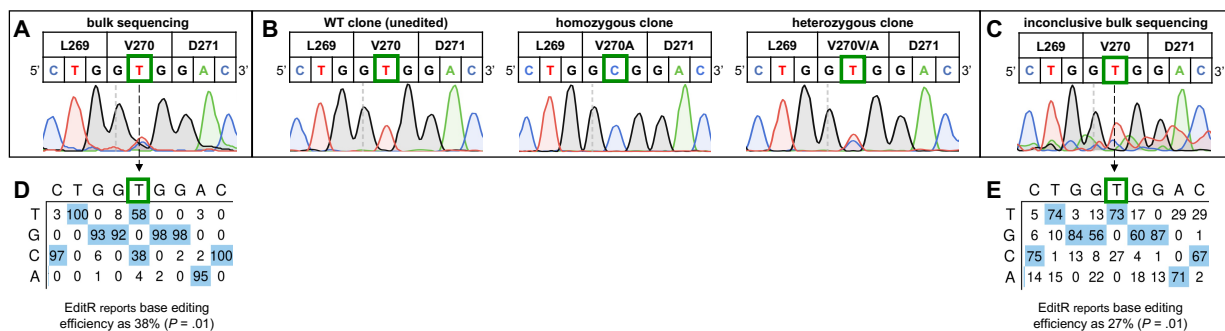
### **Using EditR to Quantify Base Editing Efficiencies**

Quantifying base editing efficiency of the target nucleotide is needed either (a) to determine the feasibility of generating a cell line with a given SNV as seen in Basic Protocol 3, or (b) to confirm the introduction of either a homozygous edit or heterozygous edit into mammalian cells as seen in Basic Protocol 5. In both instances, we recommend measuring the base editing efficiency of the total harvested cells in bulk using a simple and publicly available program, termed "EditR." ([https://moriaritylab.shinyapps.io/editr\\_v10/](https://moriaritylab.shinyapps.io/editr_v10/)).<sup>138</sup> EditR is a free online tool which requires an .ab1 Sanger sequencing file of the potentially edited region (~300-700bp) and the gRNA protospacer sequence (~20bp) to predict where a base edit occurred. Once the .ab1 file and DNA protospacer sequence are correctly uploaded, EditR generates a plot displaying editing efficiencies at each base within the protospacer (Figure 4.5D-E).



*NOTE:* If we state, “an editing efficiency of ~10%”, this means that ~10% of the cells of the total harvested cell population have a successful edit at the target base. This is will be indicated by overlapping chromatogram peaks at the targeted nucleotide. In this hypothetical example, it would show ~90% unedited base and ~10% edited base in the sequencing chromatogram. Using MUTYH V270 as an example, Figure 4.5A and 4.5D shows a chromatogram where the base edit efficiency is reported as 38% ( $P = .01$ ) by EditR. It is important to note that a poor-quality Sanger sequencing chromatograph (such as that shown in Figure 4.5C) may produce an EditR output file that falsely shows editing at the target base (see Figure 4.5E). It is, therefore, important to confirm that non-A/C bases in the protospacer display less than ~7% editing. The following steps have been adapted from the online protocol found at [https://moriaritylab.shinyapps.io/editr\\_v10/](https://moriaritylab.shinyapps.io/editr_v10/):

5. Upload your .ab1 file of the sequenced region.
6. Enter the gRNA sequence protospacer sequence.
  - If your gRNA is antisense to the .ab1 file, check the “Guide sequence is reverse complement” box.
7. Click the “Predicted Editing” tab.
8. Examine the gRNA protospacer chromatogram and underlying tile plot to determine if base editing occurred. Observing >10% editing efficiency at the target base (as indicated by a double peak at the targeted nucleotide showing <90% unedited base and >10% edited base in the sequencing chromatogram) is sufficient to proceed (Figure 4.5A).
  - All colored tiles represent base calls that are deemed significant, i.e. if there are multiple colored tiles under a single base call, base editing likely occurred.
9. If you wish to download a report of the operations performed on your data, click the "Download Report" tab on the top of the page.



**Figure 4.5: Representative Sanger sequencing traces from genotyping of cells in bulk and after clonal expansion (Basic Protocols 3 and 5).** A) Sanger sequencing trace of the bulk population of total cells after transfecting with a BE:gRNA combination to introduce the MUTYH V270A mutation. (B) Following re-transfection and clonal expansion, single cell-derived colonies potentially harboring the MUTYH V270A variant were then sequenced. Sanger sequencing chromatograms show the generation of three individual isogenic cell lines, in this instance, with genotypes of wild-type (left), homozygous (middle), and heterozygous (right). (C) A low-quality chromatogram with a high degree of background noise obtained after the bulk population of total cells were sequenced via Sanger sequencing. (D) Analysis of the .ab1 file of the bulk population of total cells, as described in panel A. EditR quantifies a base editing efficiency of 38% ( $P = .01$ ) The observed base editing efficiency at the target nucleotide indicates the feasibility of generating the targeted V270A variants, since the editing efficiency is  $>10\%$ . (E) Analysis of the .ab1 file of the bulk population of total cells with a degree of background noise, as described in panel C. EditR quantifies a base editing efficiency of 27% ( $P = .01$ ), which may mislead the researcher into incorrectly concluding a successful initial base editing experiment. Please refer to Basic Protocol 3 for information on EditR.

#### 4.5 Alternate Protocol 1: Next-Generation Sequencing to Quantify Base Editing

Some base editing experiments require rigorous quantification to accurately determine absolute editing efficiencies or to deconvolute editing patterns in bulk. Next-generation sequencing enables the researcher to quantify individual edited alleles, which can be helpful to establish the frequency of bystander mutations in bulk before isolating single cells. NGS is also desirable when genotyping edited cell lines at target sites with increased copy number (due to gene duplication) or if robust quantification of indels is needed. In these cases, the harvested genomic DNA samples from Basic Protocol 2 or Basic Protocol 5 should be amplified and prepared for Illumina NGS. Targeted amplicon sequencing for genome editing experiments has been previously described in depth by Veeranagouda and coworkers, as well as by Yang and

coworkers.<sup>139,140</sup> We recommend quantifying base editing using a 300-cycle, paired-end NGS run with a 200-250 bp amplicon. However, the amplicon length can be easily altered for other types of sequencing runs. Primers for the initial amplification (round 1) should be designed with the proper adapter sequence and distance from the target site, as described in Basic Protocol 1, step 6B and Figure 4.4A. Then, sample barcodes (for de-multiplexing) and Illumina-specific P5/P7 tail sequences will be added during the round 2 PCR. Once the data is acquired, we suggest performing data analysis with the free CRISPResso2 software using the batch mode and base editor output.<sup>103</sup> See the Critical Parameters section for additional considerations.

**Table 4.3: Next-generation sequencing barcodes or round 2 PCR primers.**

| #  | Name  | Primer sequence                                               | Barcode  |
|----|-------|---------------------------------------------------------------|----------|
| 1  | Fwd-A | AATGATACGGCGACCACCGAGATCTACACATTACTC<br>GACACTCTTTCCCTACACGAC | ATTACTCG |
| 2  | Fwd-B | AATGATACGGCGACCACCGAGATCTACACTCCGGA<br>GAACACTCTTTCCCTACACGAC | TCCGGAGA |
| 3  | Fwd-C | AATGATACGGCGACCACCGAGATCTACACCGCTCAT<br>TACACTCTTTCCCTACACGAC | CGCTCATT |
| 4  | Fwd-D | AATGATACGGCGACCACCGAGATCTACACGAGATTC<br>CACACTCTTTCCCTACACGAC | GAGATTCC |
| 5  | Fwd-E | AATGATACGGCGACCACCGAGATCTACACATTCAGA<br>AACACTCTTTCCCTACACGAC | ATTCAGAA |
| 6  | Fwd-F | AATGATACGGCGACCACCGAGATCTACACGAATTCG<br>TACACTCTTTCCCTACACGAC | GAATTCGT |
| 7  | Fwd-G | AATGATACGGCGACCACCGAGATCTACACCTGAAGC<br>TACACTCTTTCCCTACACGAC | CTGAAGCT |
| 8  | Fwd-H | AATGATACGGCGACCACCGAGATCTACACTAATGCG<br>CACACTCTTTCCCTACACGAC | TAATGCGC |
| 9  | Fwd-I | AATGATACGGCGACCACCGAGATCTACACCGGCTAT<br>GACACTCTTTCCCTACACGAC | CGGCTATG |
| 10 | Fwd-J | AATGATACGGCGACCACCGAGATCTACACTCCGCG<br>AAACACTCTTTCCCTACACGAC | TCCGCGAA |
| 11 | Fwd-K | AATGATACGGCGACCACCGAGATCTACACTCTCGCG<br>CACACTCTTTCCCTACACGAC | TCTCGCGC |
| 12 | Fwd-L | AATGATACGGCGACCACCGAGATCTACACAGCGATA<br>GACACTCTTTCCCTACACGAC | AGCGATAG |

**Table 4.3: Next-generation sequencing barcodes or round 2 PCR primers (Continued).**

| #  | Name   | Primer sequence                                                   | Barcode  |
|----|--------|-------------------------------------------------------------------|----------|
| 13 | Rev-1  | CAAGCAGAAGACGGCATAACGAGATATCGTGATGTGA<br>CTGGAGTTCAGACGTGTGCTCTTC | ATCACGAT |
| 14 | Rev-2  | CAAGCAGAAGACGGCATAACGAGATATACATCGGTGA<br>CTGGAGTTCAGACGTGTGCTCTTC | CGATGTAT |
| 15 | Rev-3  | CAAGCAGAAGACGGCATAACGAGATATGCCTAAGTGA<br>CTGGAGTTCAGACGTGTGCTCTTC | TTAGGCAT |
| 16 | Rev-4  | CAAGCAGAAGACGGCATAACGAGATATTGGTCAGTGA<br>CTGGAGTTCAGACGTGTGCTCTTC | TGACCAAT |
| 17 | Rev-5  | CAAGCAGAAGACGGCATAACGAGATATCACTGTGTGA<br>CTGGAGTTCAGACGTGTGCTCTTC | ACAGTGAT |
| 18 | Rev-6  | CAAGCAGAAGACGGCATAACGAGATATATTGGCGTGA<br>CTGGAGTTCAGACGTGTGCTCTTC | GCCAATAT |
| 19 | Rev-7  | CAAGCAGAAGACGGCATAACGAGATATGATCTGGTGA<br>CTGGAGTTCAGACGTGTGCTCTTC | CAGATCAT |
| 20 | Rev-8  | CAAGCAGAAGACGGCATAACGAGATATTCAAGTGTGA<br>CTGGAGTTCAGACGTGTGCTCTTC | ACTTGAAT |
| 21 | Rev-9  | CAAGCAGAAGACGGCATAACGAGATATCTGATCGTGA<br>CTGGAGTTCAGACGTGTGCTCTTC | GATCAGAT |
| 27 | Rev-15 | CAAGCAGAAGACGGCATAACGAGATTCTGACATGTGA<br>CTGGAGTTCAGACGTGTGCTCTTC | ATGTCAGA |
| 28 | Rev-16 | CAAGCAGAAGACGGCATAACGAGATCGGGACGGGTG<br>ACTGGAGTTCAGACGTGTGCTCTTC | CCGTCCCG |
| 29 | Rev-18 | CAAGCAGAAGACGGCATAACGAGATGTGCGGACGTG<br>ACTGGAGTTCAGACGTGTGCTCTTC | GTCCGCAC |
| 30 | Rev-19 | CAAGCAGAAGACGGCATAACGAGATCGTTTCACGTGA<br>CTGGAGTTCAGACGTGTGCTCTTC | GTGAAACG |
| 31 | Rev-20 | CAAGCAGAAGACGGCATAACGAGATAAGGCCACGTG<br>ACTGGAGTTCAGACGTGTGCTCTTC | GTGGCCTT |
| 32 | Rev-21 | CAAGCAGAAGACGGCATAACGAGATTCCGAAACGTG<br>ACTGGAGTTCAGACGTGTGCTCTTC | GTTTCGGA |
| 33 | Rev-22 | CAAGCAGAAGACGGCATAACGAGATTACGTACGGTG<br>ACTGGAGTTCAGACGTGTGCTCTTC | CGTACGTA |
| 34 | Rev-23 | CAAGCAGAAGACGGCATAACGAGATATCCACTCGTGA<br>CTGGAGTTCAGACGTGTGCTCTTC | GAGTGGAT |
| 35 | Rev-25 | CAAGCAGAAGACGGCATAACGAGATAAAGGAATGTG<br>ACTGGAGTTCAGACGTGTGCTCTTC | ATTCCTTT |
| 36 | Rev-27 | CAAGCAGAAGACGGCATAACGAGATATATCAGTGTGA<br>CTGGAGTTCAGACGTGTGCTCTTC | ACTGATAT |

**Materials:**

- Cell lysate from Basic Protocol 2
- Primers for round 1 PCR (IDT or any other DNA provider)

- Primers for round 2 PCR (IDT or any other DNA provider) from **Table 3**, or TG Nextera® XT Index Kit v2 (Illumina, cat. no. TG-131-2001)
- dNTP mix, 10mM each (NEB, cat. no. N0447L)
- Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, cat. no. F534L)
- DMSO
- Nuclease-free water (VWR, cat. no. 82007-328)
- 0.2 mL PCR tubes (Fisher Scientific, cat. no. 07-200-259)
- Thermocycler
- Reagents for agarose gel electrophoresis
- 100 bp DNA ladder (NEB, cat. no. N3231L)
- QIAquick PCR purification kit (Qiagen, cat. no. 28106)
- Qubit fluorometer with dsDNA HS assay kit (Q32854), or qPCR with NEBNext Library Quant Kit for Illumina (E7630L)

## Protocol Steps

### *Next-Generation Sequencing*

1. To amplify the genomic locus of interest, create a 25 µL round 1 (rd1) PCR reaction for each gDNA sample (BE:gRNA combination) and a negative control containing water. To save reagents, scale down to a total volume of 12.5 µL. For each sample, combine the following reagents (in the order stated):
  - X µL Nuclease-free water (fill to a total volume of 25 µL)
  - 5 µL GC Buffer
  - 0.5 µL 10 mM dNTPs
  - 0.5 µL 10µM forward NGS rd1 primer (for the MUTYH V270 example, primer #15 **Table 1**)
  - 0.5 µL 10µM reverse NGS rd1 primer (for the MUTYH V270 example, primer #16 **Table 1**)

- 0.75  $\mu$ L 100% DMSO
- 0.25  $\mu$ L of Phusion DNA polymerase
- 1.0-0.5  $\mu$ L cell lysate (gDNA)

*It is imperative to include a negative control sample every time amplification from gDNA is performed. In this sample, 0.5  $\mu$ L of water is added instead of the cell lysis solution, to control for gDNA contamination. We suggest making a master mix (adding polymerase last), then adding the cell lysate to individual aliquots.*

2. Mix, quickly spin, and run a thermocycler program with the following cycling conditions:

|                   |          |      |                        |
|-------------------|----------|------|------------------------|
| 1 cycle:          | 60 sec   | 98°C | (initial denaturation) |
| 24 to 28 cycles*: | 16 sec   | 98°C | (denaturation)         |
|                   | 20 sec   | 65°C | (annealing)            |
|                   | 16 sec   | 72°C | (extension)            |
| 1 cycle:          | 5 min    | 72°C | (final extension)      |
| 1 cycle:          | $\infty$ | 12°C | (hold)                 |

*\*To avoid PCR bias, use the minimum number of cycles that provides robust amplification. This may require optimization for each target locus (see also “Quantification of base editing efficiency in bulk cells” in the Critical Parameters section).*

3. Run a 2  $\mu$ L aliquot of each PCR reaction on a 2% agarose gel using a 100 bp ladder (see Figure 4.4C for an example).

*NGS round 1 adapter sequences add 66 bp to the length of your amplicon. If you do not get the correct length of PCR product bands, please refer to the*

*Troubleshooting section. The most important parameters to vary are amount of gDNA and primer concentration.*

4. Round 2 (rd2) PCR adds a unique barcode designated by an 8 nt sequence in each of the primers: A1 in this example (Fwd-A/Rev-1). For each rd1 sample, create a 25  $\mu$ L rd2 PCR reaction (could scale down to 12.5  $\mu$ L to save reagents) by combining the following reagents (in the order stated):

- X  $\mu$ L Nuclease-free water (fill to a total volume of 25  $\mu$ L)
- 5  $\mu$ L HF Buffer
- 0.5  $\mu$ L 10 mM dNTPs
- 0.5  $\mu$ L 10 $\mu$ M forward NGS rd2 primer (custom barcode; #1 **Table 3**, A for example)
- 0.5  $\mu$ L 10 $\mu$ M reverse NGS rd2 primer (custom barcode; #13 **Table 3**, 1 for example)
- 0.25  $\mu$ L of Phusion DNA polymerase
- 1.0  $\mu$ L rd1 PCR product

*The negative control sample in this PCR should contain 0.5  $\mu$ L of the round 1 negative control PCR product. Barcodes can be custom-generated and input into the sequencer, such as those provided in Table 3, or ordered in a kit through Illumina (TG Nextera® XT Index Kit v2).*

5. Mix, quickly spin, and run a thermocycler program with the following cycling conditions:

|                  |          |      |                        |
|------------------|----------|------|------------------------|
| 1 cycle:         | 60 sec   | 98°C | (initial denaturation) |
| 8 to 12 cycles*: | 16 sec   | 98°C | (denaturation)         |
|                  | 20 sec   | 65°C | (annealing)            |
|                  | 16 sec   | 72°C | (extension)            |
| 1 cycle:         | 5 min    | 72°C | (final extension)      |
| 1 cycle:         | $\infty$ | 12°C | (hold)                 |

*\*Use the minimum number of cycles that provides robust amplification.*

6. Run a 2 uL aliquot of each PCR reaction on a 2% agarose gel using a 100 bp ladder (see Figure 4.4C for an example).

*NGS round 2 adapter sequences add 74 bp to the length of your rd1 PCR product (or a total of 140 bp to your original amplicon).*

7. Pool rd2 PCR products from all samples together (e.g. no gRNA control and each BE:gRNA combination at the locus of interest). Perform a gel extraction on the pooled samples to eliminate lower molecular weight products that would decrease the quality of the NGS data.

*Perform a second PCR clean-up on the elution to remove any agarose or guanidine (peak absorbance at 230nm) that remains after gel purification. A Nanodrop can be used to determine purity and roughly quantify concentration. Then, dilute into the quantification range of Qubit for the next step.*

8. Accurately quantify the DNA concentration of the libraries following the Qubit dsDNA HS assay kit protocol. Alternatively, the NEBNext Library Quant Kit for Illumina protocol describes quantification by qPCR.
9. Set up the NGS run according to Illumina protocols or submit to a sequencing core facility.

*We recommend analyzing data with CRISPResso2, discussed in the Critical Parameters section.*

#### **4.6 Basic Protocol 4: Single Cell Isolation of Base Edited Cells Using FACS**

Following confirmation of >10% editing efficiency, the next step is to isolate single cells and clonally expand, to obtain isogenic cell lines harboring the SNV of interest. In this article, we describe two different methods for this: one that utilizes fluorescence-activated cell sorting (FACS) (this protocol), and one that utilizes dilution plating (Alternate Protocol 2). Using FACS followed by clonal expansion is preferred over dilution plating for two key reasons. Firstly, FACS



allows for the discrimination between single cells and multiple cells, which eliminates the chances of obtaining doublet cell-derived colonies (a frequent observation seen in dilution plating, Figure 4.6E). Secondly, FACS allows for the discrimination between untransfected and transfected cells when using plasmids with fluorescent markers, which is particularly important when working with cell lines with low transfection efficiencies. Dilution plating, however, is more cost-effective, gentler on the cells, and does not require specialized instrumentation. Dilution plating, therefore, may be preferred over FACS when working with sensitive cells or if the researcher does not have access to FACS instrumentation. This Basic Protocol describes the process for transfecting and isolating single cells for clonal expansion.

*NOTE:* We used ABEmax-NG-P2A-GFP (Addgene #140005) to generate the V270A MUTYH variant used in this example. In this plasmid, although EGFP and ABE are both transcribed in a single mRNA transcript, they are translated into unique and separate proteins via the P2A linker which induces ribosomal skipping. Therefore, FACS can be used to sort individual EGFP-positive cells, which selects cells that are actively expressing the BE. There are also methods to select for cells with high BE activity (rather than simply expression); see the Critical Parameters section for more details on this.

***Materials:***

- HEK293T cells (ATCC CRL-3216)
- Nuclease-free water (VWR, cat. no. 82007-328)
- Endotoxin-free base editor plasmids and cloned gRNA plasmids from **Basic Protocol 1**
- 48-well plates (VWR, cat. no. 10062-898)
- 96-well plates (VWR, cat. no. 10062-900)
- Lipofectamine 20000 (Thermo Fisher Scientific, cat. no. 11668-019)
- Opti-MEM Reduced Serum Medium, Gibco (Life technologies cat. no. 31985070)
- Phosphate-buffered saline (PBS; Gibco, cat. no. 10010049)
- Dulbecco's Modified Eagle's Medium (DMEM; CoreBio, cat. no. 10566-024)

- Fetal Bovine Serum (FBS; Thermo Fisher Scientific, cat. no. 10437-028)
- Pen/Strep (Gibco, cat. no. 15070063)
- 1.5 mL nuclease-free microcentrifuge tubes (VWR, cat. no. 87003-294)
- 5 mL round bottom polystyrene test tube, with cell strainer snap cap (Thermo Fisher Scientific, cat. no. 352235)
- Sterile 15 mL conical tubes
- Accumax (Accutase, cat. no. AM105)
- Propidium Iodide (PI; Thermo Fisher Scientific, cat. no. P3566)
- Fluorescence microscope
- Centrifuge
- 37°C incubator with humidified atmosphere of 5% CO<sub>2</sub> (g)

## Protocol Steps

### ***Transfect for Single Cell Isolation***

1. Transfect the HEK293T cells with the BE:gRNA combination(s) that show desired editing activity in bulk, as described in Basic Protocol 2 steps 1 to 8. Incubate the cells for three days post transfection.

*It is important to include a proper negative control sample (see Critical Parameters), such as a non-targeting gRNA or no gRNA.*

2. Prepare 96-well plates which contain 100 µL of culture medium in each well (30-50% FBS, 1% Pen/Strep). Generally, 2 plates per well in step 1 is sufficient to obtain cells containing the SNVs of interest, as referred to in the Understanding Results section. Place the plates into the tissue-culture incubator before sorting cells to equilibrate.

*Preparing extra collection plates is recommended for lower rates of desired editing.*

## Preparation of the Cells for FACS

3. Three days post-transfection, gently pipette or aspirate 250  $\mu$ L of the old medium from the transfected cells and gently rinse cells with 150  $\mu$ L of PBS. Carefully remove the PBS.
4. Add 30  $\mu$ L of Accumax to each well and incubate at 37° for 2-3 min, or until cells are dislodged from the bottom of the well.

*The use of trypsin is not advised to prepare cells for sorting as it is too harsh and will result in fewer cells surviving the clonal expansion process.*

7. Resuspend each sample with 170  $\mu$ L PBS.

*Volumes of Accumax and PBS given are for fully confluent wells but should be scaled to the confluency of cells at the time of sample prep and the cell sorter used. For example, approximately 1-2 million cells per mL final volume is recommended for the BioRad S3e sorter.*

8. Filter cells through a sterile 35  $\mu$ m cell strainer in a 5 mL polystyrene round-bottom tube. Place on ice until sort.
9. Using a cell sorter capable of sorting single cells into individual wells (FACS Aria II system or equivalent), and under sterile conditions, sort single EGFP-positive cells into individual wells of the 96-well plates containing 100  $\mu$ L of culture media containing DMEM with 30-50% FBS, 1% Pen/Strep.
10. Place plates into tissue-culture incubator immediately after sorting.

#### **4.7 Alternate Protocol 2: Single Cell Isolation of Base Edited Cells Using Dilution**

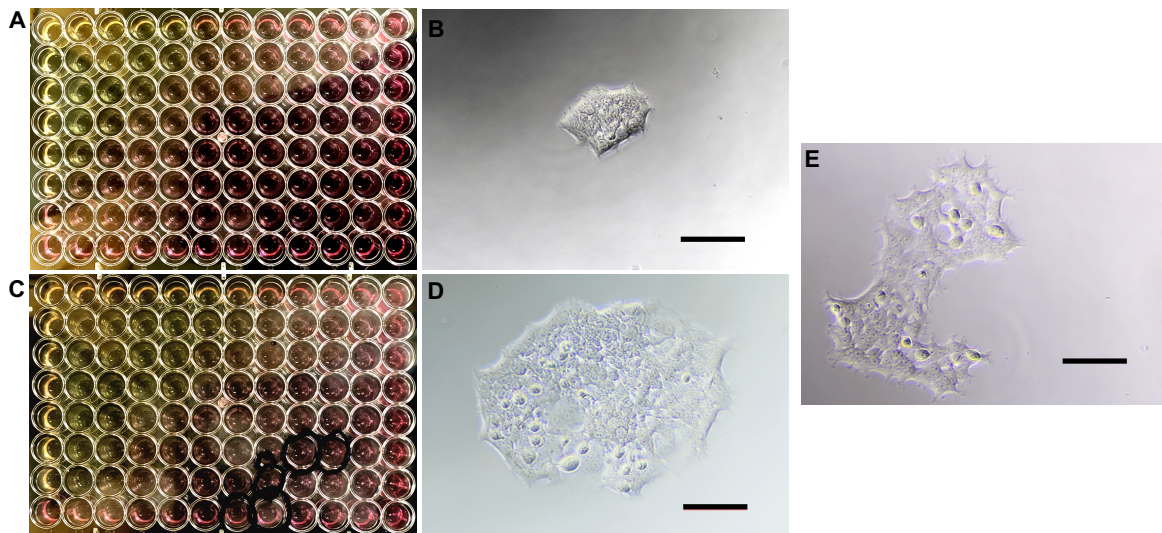
##### **Plating**

Dilution plating offers a cost-effective alternative when researchers do not have access to a sorting machine or when they are working with plasmids which do not possess fluorescent markers. Furthermore, dilution plating might be preferred when working with sensitive cell lines that are less likely to survive the sorting conditions. However, there are two major limitations to

dilution plating. Firstly, the chances of obtaining doublet cell-derived colonies increases. We have found, however, that using a 35µm sterile cell strainer prior to dilution plating reduces the chances of obtaining doublet-derived colonies. Secondly, dilution plating does not allow for the discrimination between untransfected and transfected cells, something that is possible when using FACS. Therefore, the rates of successfully obtaining a cell line harboring the SNV of interest are reduced (Please refer to the Understanding Results section). Therefore, screening more single cell-derived colonies per transfection may be necessary to obtain the cell line of interest. This can be accomplished by using more than two plates per transfection. This protocol uses an adaptation of methods previously described.<sup>141</sup>

***Materials:***

- 96-well plates (VWR, cat. no. 10062-900)
- Dulbecco's Modified Eagle's Medium (DMEM; CoreBio, cat. no. 10566-024)
- Fetal Bovine Serum (FBS; Thermo Fisher Scientific, cat. no. 10437-028)
- Pen/Strep (Gibco, cat. no. 15070063)
- 35µm sterile cell strainer snap cap (Life Sciences, cat. no. 352235)
- 8-channel 200µL micropipettor
- Reagent dispensing reservoir/tray (VWR, cat. no. 89094-662 )
- TrypLE Express (Thermo Fisher Scientific, cat. no. 12605028)
- Transfected HEK293T cells from Basic Protocol 2
- 37°C incubator with humidified atmosphere of 5% CO<sub>2</sub> (g)



**Figure 4.6: Representative images from dilution plating method (Alternate Protocol 2).** (A) Image of a 96-well plate taken four days post-dilution plating. The top left wells have a higher concentration of cells as indicated by the low pH level (yellow-colored culture medium). (B) 20X magnification image of a well from (A) with a single cell-derived colony, taken four days post-dilution plating. (C) Image of the 96-well plate from (A), taken seven days post-dilution plating. Wells which are harboring single cell-derived clones are marked. (D) 20X magnification image of the same colony from (B), taken seven days post-dilution plating. (E) 20X magnification image of a doublet cell-derived colony, taken seven days post-dilution plating. Scale bars = 166  $\mu\text{m}$ .

## Protocol Steps

### *Single Cell Isolation Using Dilution Plating*

1. Add 100  $\mu\text{L}$  of pre-warmed DMEM medium (10% FBS, 1% Pen/Strep) to all the wells in a 96-well plate, except to well A1.

*In general, we recommend doing four plates per SNV when the editing efficiency from Basic Protocol 3 is >20 %. More plates may be needed per SNV when editing efficiency is <20%.*

2. After the cells from Basic Protocol 2 step 6 have been transfected and incubated for three days, pipette off 250  $\mu\text{L}$  of the old medium from the cells and gently rinse cells with 150  $\mu\text{L}$  of PBS. Carefully aspirate off the PBS.

3. Add 50  $\mu\text{L}$  of TrypLE to each well and incubate at 37°C for 5 to 10 min, or until cells are completely dislodged from the bottom of the well.
4. Resuspend the transfected cells with culture medium.
5. Add 200  $\mu\text{L}$  of the cell suspension from step 4 to well A1 through a 35  $\mu\text{m}$  sterile cell strainer.

*The use of the strainer is optional, but when used, we have observed a higher frequency of isolating single cell-derived colonies as opposed to doublet cell-derived colonies.*

4. Using a single channel pipette, make 1:2 dilutions by transferring 100  $\mu\text{L}$  of the cell suspension from A1 down the first column (B1 to H1) using the same tip. Discard 100  $\mu\text{L}$  of cells from the last well.

*Mix gently before each transfer.*

5. Add an additional 100  $\mu\text{L}$  of medium to each well in column 1.
6. Using an 8-channel pipette, make 1:2 dilutions by transferring 100  $\mu\text{L}$  of the cell suspension across each column of the plate starting from column 1 and ending at column 12. Discarding 100  $\mu\text{L}$  of cell suspension from the last column is optional.
7. Place plates into tissue-culture incubator undisturbed after dilutions are made. Observe cells 4-7 days later (Figure 4.6).

#### **4.8 Basic Protocol 5: Clonal Expansion to Generate Isogenic Cell Lines and Genotyping of Clones**

Regardless of which method the researcher uses to isolate single cell-derived colonies (FACS Basic Protocol 4 or dilution plating Alternate Protocol 2), colony formation of HEK293T cells should be apparent after 4-7 days and they should be ready to subculture 10-14 days after the isolation protocol. Figures 7B and 7D provide examples of single cell-derived colonies. This Basic Protocol describes the process for clonally expanding the isolated cells from Basic

Protocol 4 into isogenic cell lines, genotyping the resultant lines, and subculturing the appropriate colonies into larger flasks for storage or downstream experiments. For the SNV of interest, aim to obtain at least three different isogenic cell lines, one harboring a wild-type genotype, one harboring a heterozygous genotype, and one harboring a homozygous genotype. All these cell lines can be used for comparison and use in downstream experiments (Figure 4.5B).

**Materials:**

- 48-well plates (VWR, cat. no. 10062-898)
- Cells from Basic Protocol 4, step 10 or Alternate Protocol 2, step 7
- T-75 flask (Victorio, cat. no. 10062-860-KLM)
- Phosphate-buffered saline (PBS; Gibco, cat. no. 10010049)
- Dulbecco's Modified Eagle's Medium (DMEM; CoreBio, cat. no. 10566-024)
- Fetal Bovine Serum (FBS; Thermo Fisher Scientific, cat. no. 10437-028)
- Pen/Strep (Gibco, cat. no. 15070063)
- 37°C incubator with humidified atmosphere of 5% CO<sub>2</sub> (g)
- TrypLE Express (Thermo Fisher Scientific, cat. no. 12605028)
- Microscope
- 37°C incubator with humidified atmosphere of 5% CO<sub>2</sub> (g)

**Protocol Steps**

***Subculturing and Genotyping Single Cells***

1. Carefully inspect each well of the plates from either Basic Protocol 4 or Alternate Protocol 2 under the microscope and circle the wells that are harboring single cell-derived clones (usually, it takes around 3-4 days to notice distinguishable colony formation, but it can also be longer depending on clonal expansion rate, see Time

Considerations). Be wary of potential doublet cell-derived colonies (see Figure 4.6 for examples).

*We have found that each 96-well plate will yield around 8 colonies per plate when prepared according to the FACS method, and around 11 single cell-derived colonies per plate when prepared according to the dilution method.*

2. Once the colony covers at least 30% of the well's surface area, pipette off 100  $\mu$ L of the old medium and gently rinse the cells with 50  $\mu$ L of PBS.

*Since clonal expansion rate may differ from clone to clone, we recommend checking the colonies daily after they have been identified in step 1 of Basic Protocol 5. This step usually takes 1-2 weeks but can vary. Please see Time Considerations.*

3. Add 30  $\mu$ L of TrypLE Express. Wait 3-5 minutes, or until cells are completely dislodged from the bottom of the well.
4. While cells are being detached, add 175  $\mu$ L of pre-warmed (37°C ) culture medium (10% FBS, 1% Pen/Strep) to each well of a separate 48-well plate. Allocate two wells for each single cell-derived colony.

*Half of the cells will be clonally expanded while the other half will be used for genotyping.*

5. Resuspend the trypsinized cells with 120  $\mu$ L of pre-warmed DMEM medium (10% FBS, 1% Pen/Strep).
6. Passage two individual 75  $\mu$ L aliquots of the clonal cell suspension into each well from step 4.
7. Allow the cells to reach 80-90% confluency before proceeding to step 8. This will usually take 3 days, but it is recommended to check the cells under the microscope every day until they reach the appropriate confluency.



8. Two simultaneous steps must be followed. Thus, proper labelling of wells for identification of which clone it corresponds to is crucially important:
  - a. Once both wells have reached the appropriate confluency, take one of the wells, and harvest the genomic DNA and genotype, as previously described in Basic Protocols 2 and 3 or Alternate Protocol 1.
  - b. For the other well containing the same clone, continue passaging the cells into larger wells or flasks before storage and until the sequencing data from step 8a is obtained. We recommend moving from a 48-well plate (250  $\mu$ L of DMEM medium with 10% FBS, 1% Pen/Strep), to a 6-well plate (1 mL of DMEM medium with 10% FBS, 1% Pen/Strep), then to a T25 flask (5 mL of DMEM medium with 10% FBS, 1% Pen/Strep), and finally to a T75 flask (10 mL of DMEM medium with 10% FBS, 1% Pen/Strep). Passage the cells when they are at 85-90% confluent, therefore, it is recommended to monitor the cells under the microscope every day. We recommend keeping at least three clones, three cell lines containing wild-type, three clones containing heterozygous, and three clones containing homozygous genotypes (Figure 4.5B). Please refer to the Understanding Results section.
9. Continue to clonally expand the sequenced-validated cell lines by passaging the cells into larger wells or flasks as indicated in step 8b. We additionally recommend preparing a fourth cell line per each genotype that is a mixture of each of the clones with that specific genotype.
10. Cryopreserve cell lines.

*There are many protocols that show researchers how to cryopreserve mammalian cell lines.*<sup>142,143</sup> We also recommend using <https://www.abcam.com/protocols/cryopreservation-of-mammalian-cell-lines-video-protocol> as a resource.

## **REAGENTS AND SOLUTIONS**

### **Ampicillin or carbenicillin 1000X stock solution (100 mg/mL; 5 mL)**

Dissolve 500 mg of powdered ampicillin or carbenicillin sodium salt (such as GoldBio, cat. no. A-301 or C-103) in 5 mL of deionized water to create a 100 mg/mL stock solution. Store at -20°C. Note: carbenicillin is more stable but more expensive.

### **Agar plates and liquid media with ampicillin or carbenicillin, final concentration 100 µg/mL**

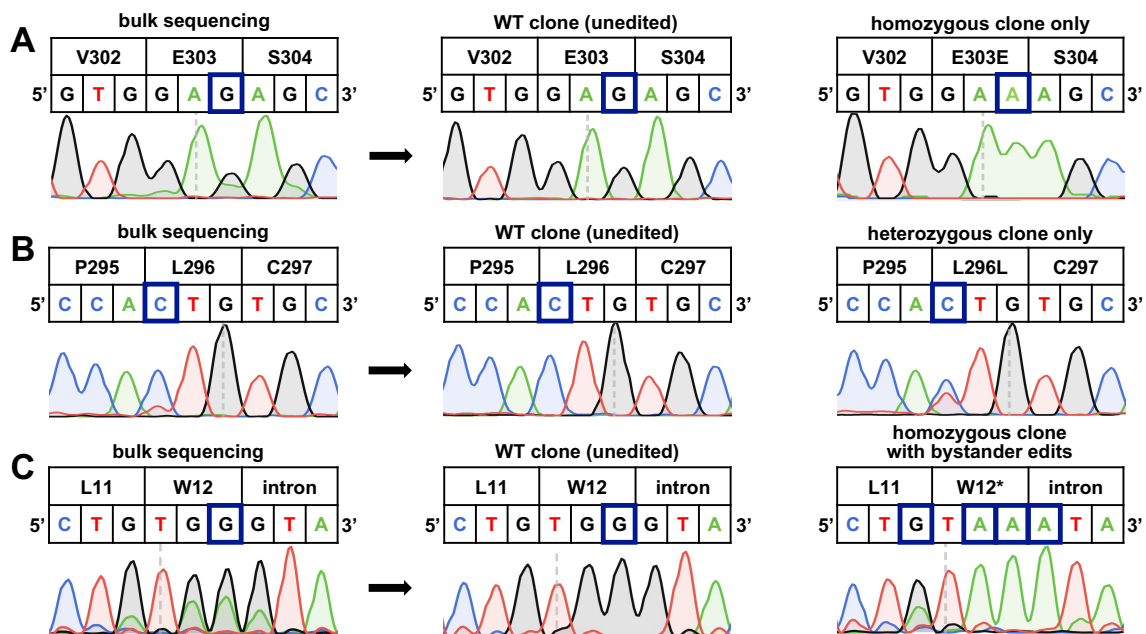
Prepare agar plates and LB or 2xYT liquid media as previously described.<sup>133</sup> Dilute the ampicillin/carbenicillin stock solution 1000X to the final concentration of 100 µg/mL before pouring plates or inoculating liquid cultures. For example, add 500 µL of antibiotic in 500 mL of agar-containing media to make ~20 plates with 25 mL each.

### **10% (w/v) SDS stock solution (100 mL)**

Dissolve 10 g of SDS (Thermo Fisher Scientific, cat. no. 15525017) in 80 mL of deionized water. Bring final volume to 100 mL with deionized water.

### **Tris HCL (1M) stock solution (pH 8.0; 1 Liter)**

Dissolve 121.14 g Tris (Thermo Fisher Scientific, cat. no. PI17926) in 800 mL of deionized water. Adjust pH to 8 with 1 M NaOH. Bring final volume to 1 liter with deionized water. Autoclave and store at room temperature.



**Figure 4.7: Potential pitfalls in using Base Editing to generate isogenic cell lines.** (A) Attempted introduction of the MUTYH E303E mutation resulted in only WT and homozygous cell lines, with no heterozygous clones observed. Experiment can be repeated with a less active BE variant, or cells with lower GFP signal isolated. (B) Attempted introduction of the MUTYH L296L mutation resulted in only a WT and heterozygous cell lines, with no homozygous clones observed. Check gene and mutation for lethality, and repeat experiment using base editor activity selection scheme. (C) Attempted introduction of the MUTYH W12\* mutation resulted in bystander edits of non-target bases in the protospacer region. Because a premature stop codon is being introduced, bystander edits are acceptable.

## 4.9 Commentary

### 4.9.1 Background information:

Base editing is a technique capable of precisely and efficiently introducing point mutations in mammalian cells by relying on nucleobase chemistry rather than on the introduction of double-stranded DNA breaks (DSBs).<sup>110,111</sup> The system utilizes the targeting ability of the Cas9:gRNA complex but avoids the pitfalls of DSB-mediated genome editing, as discussed below.

All targeted nuclease systems, such as ZFNs, TALENs, and Cas enzymes rely on phosphodiester backbone cleavage and subsequent DNA repair either to create random insertions and deletions (indels) through the non-homologous end-joining (NHEJ) pathway,<sup>144</sup> or for the incorporation of a donor DNA sequence containing the mutation of interest via homology-directed repair (HDR).<sup>145</sup> While NHEJ is active in all phases of the cell cycle, HDR normally requires a sister chromatid as a template and is, therefore, only active during the late S and G2 phases, making HDR-mediated genome editing only possible in actively dividing cells. Thus, attempting to create targeted point mutations using traditional DSB-mediated genome editing inevitably leads to low levels of accurate SNV introduction (typically, less than 1%, unless cell cycle or DNA repair perturbations are employed, or silent “blocking mutations” are incorporated).<sup>10,125,126</sup> Further, DSBs can be toxic to cells and cause p53-dependent apoptosis{Citation} or large genomic rearrangements,<sup>8</sup> which is exacerbated during multiplexed genome editing (i.e. simultaneous targeting of distinct loci).<sup>146,147</sup> This results in decreased viability, the requirement of screening hundreds of clones, the introduction of genomic modifications other than the modification of interest, and the potential enrichment for cells with p53 mutations (the latter two of which can cause artifacts in phenotypic characterization). Recently, deleterious on-target effects, such as large deletions or loss-of-heterozygosity, were shown to escape standard genotyping methods in up to 40% of edited clones (where hemizygous ones appeared homozygous).<sup>148</sup>

A common method to facilitate the enrichment and selection of genome edited cell lines rely on knock-in strategies of selectable markers through homologous recombination, such as antibiotic resistance or fluorescent markers.<sup>149–152</sup> However, removing these integrated selection markers without permanently damaging or “scarring” the host genome is non-trivial and can sometimes require multiple clonal expansion events.<sup>153,154</sup> Another alternative method to DSB-dependent editing strategies is the use of adeno-associated-viral (AAV)-mediated insertion of genomic cassettes into human somatic cells.<sup>155,156</sup> However, the packaging limit for AAVs is

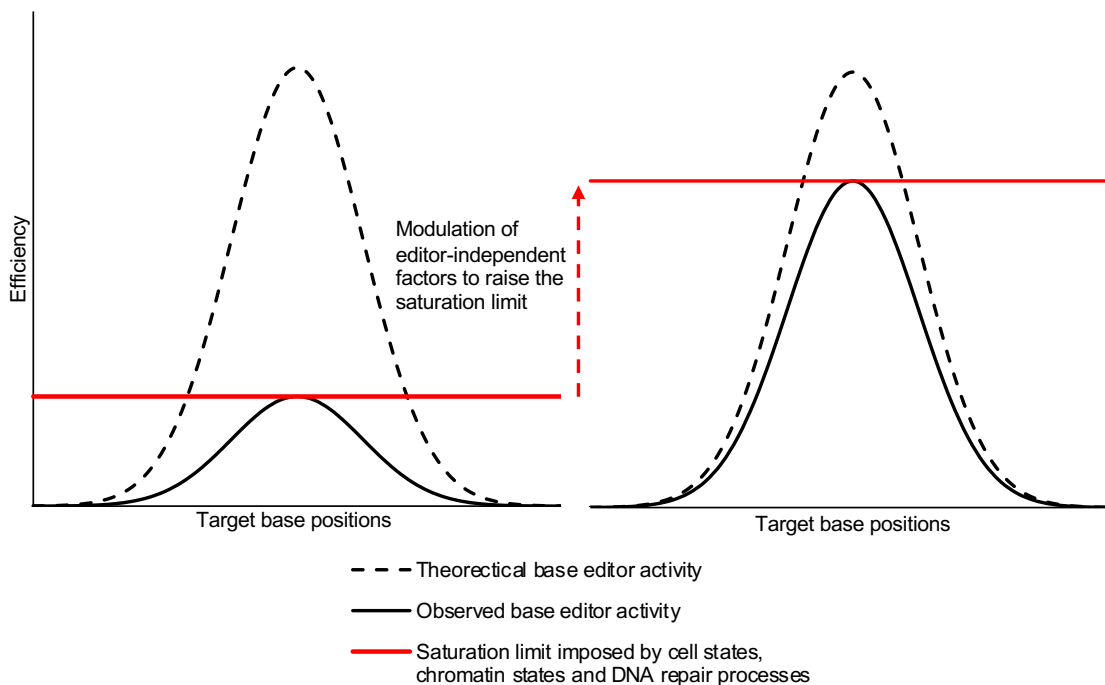
4.7kb (smaller than some genes of interest), and as this involves ectopic expression of the gene of interest, it is less physiologically-relevant than editing of the endogenous locus.<sup>157,158</sup>

Base editors avoid DSBs by instead relying on the direct chemical modification of target nucleobases. When the edited base is used as a template for DNA replication or repair, a transition mutation is faithfully introduced with high efficiency (up to 80% in many cases). This addition to the genome engineering toolkit appeared in 2016 with the advent of CBEs (which introduce C•G to T•A point mutations using uracil intermediates), followed by ABEs (which introduce A•T to G•C point mutations using inosine intermediates) the following year. Because of their dependence on more ubiquitous and reliable DNA repair pathways, base editors have the added advantage of working in non-dividing cells.

However, like all current genome editing technologies, base editing has limitations. Most importantly, base editors cannot be used to introduce transversion mutations, insertions, or deletions. Second, base editors have a strict requirement for a PAM to be positioned 12-16 nucleotides from the target base, making certain targets inaccessible. However, ongoing efforts to engineer Cas enzymes with PAMs complimentary to Cas9-NG have yielded successful base editors.<sup>129–131,159–162</sup> Additionally, it is important to bear in mind that even with a suitable PAM, not all genomic sites are amenable to base editing. As with other forms of genome editing, the sequence or chromatin context of the target site may limit the Cas9:gRNA complex's ability to binding DNA. Some base editors additionally contain deaminase enzymes with an intrinsic sequence preference. Finally, CBEs can produce C-to-non-T edits, likely due to high efficiencies of excision of the uracil intermediate by endogenous DNA repair enzymes. A better understanding of the cellular mechanisms responsible for high BE efficiency and product purity (i.e. target conversion only into the desired base rather than unwanted bases) is needed to improve the system (Figure 4.8).

Recently described “prime” editors, a novel CRISPR-Cas9 gene editing technique, utilize a reverse transcriptase and a template sequence as a 3'-extension to the gRNA to

precisely edit DNA.<sup>163</sup> This tool is not limited to transition mutations, avoids any bystander nucleotides, and has more flexible PAM positioning. However, prime editing is less efficient than base editing when a target A or C can be optimally positioned within the protospacer and requires more extensive gRNA optimization per target.<sup>163</sup> Base editors are also more compact and produce fewer indels,<sup>163</sup> which can be important factors with *in vivo* applications. It is important to consider the goal of your study while designing an experiment to select the most appropriate editing technique and set of tools.



**Figure 4.8: A conceptual model of base editing.** The editing efficiency reaches the maximum at the optimal target base position. At certain poorly edited sites, the observed base editing efficiencies are low even a highly active base editor is used. We reason that the saturation limits of those sites imposed by editor-independent factors, such as cell states, chromatin structures, and DNA repair processes, are low. Raising the limits by modulation of these editor-independent factors can potentially increase the editing efficiencies.

#### 4.9.2 Critical parameters:

##### *gRNA design considerations*

Programs including the Benchling wizard (<https://benchling.com/pub/liu-base-editor>; limited to CBEs), BE-designer (<http://www.rgenome.net/be-designer/>), DeepBE (<https://deepcrispr.info/DeepBE/>) and BE-Hive (<https://www.crisprbehive.design/>) can be used for automated gRNA design or scoring.<sup>38,80,128,135</sup> BE-Hive incorporates ABEs and alternate PAMs, making it the most comprehensive software. This machine learning algorithm enables predictions of both editing outcome and efficiency for a given BE:gRNA combination. Other programs are available for designing and scoring gRNAs for traditional Cas9 genome editing, and have been previously reviewed.<sup>164</sup> These may be helpful for predicting potential off-targets or gauging how effectively Cas9 will bind to a given target site.

The human U6 promoter drastically increases expression of the gRNA if guanine is the first nucleotide transcribed. Any decrease in Cas9 binding due to a mismatch at the first position of the protospacer that this may cause is more than compensated for by the increased expression levels. The mouse U6 promoter exhibits high expression with A or G as the first nucleotide, and can be used as an alternative.<sup>165</sup> Additionally, the *S. pyogenes* gRNA backbone contains an early transcription termination signal (UUUU) that may decrease expression levels, and can be mutated to increase cellular gRNA levels.<sup>166</sup> While we have found that low base editing efficiency is usually caused by other factors, this may be an issue in certain cell types.

### ***Bystander editing***

Bystander edits occur when C's or A's beyond the desired target are also present in the editing window and become mutated concurrently with the target base. Bystander editing can be avoided by designing a protospacer that pushes potential bystander bases out of the editing window, as shown with the gRNA example targeting V270 with an adenine in position 9 (Figure 4.2A). Bystander edits may also be mitigated by using deaminase enzymes that are modified to be less processive (such as the YE1, YE2, and YEE CBEs)<sup>160,167</sup> or sequence-specific (such as the eA3A CBE),<sup>168</sup> or by alternative BE architectures, such as rigid linkers.<sup>169</sup> In some cases,

bystander edits are acceptable, such as in gene knockout experiments (Figure 4.7C) or if the bystander creates a silent mutation or intronic mutation (as long as it is not a splice site). When attempting to knock out genes via premature stop codon introduction or splice site disruption, disruptions targeted to the first and last few exons can be inefficient due to translation reinitiating and/or alternative splicing.<sup>170</sup> We suggest base editing splice donor sites in the middle of the gene body to achieve the highest level of protein loss.<sup>147</sup>

### ***Base editor selection considerations***

There is a large selection of base editor constructs from which to choose, and the most suitable choice depends on your experimental goal. The two editors that we recommend (BE4max-NG-P2A-EGFP and ABEmax-NG-P2A-EGFP) are good choices with which to start, and modifications can be made according to specific experimental requirements. In most cases, nCas9 (D10A) should be utilized, which will direct DNA repair machinery to use the modified base as a template, but dCas9 can be employed if indels must be avoided at all costs (we note that indel formation with CBEs is target site-dependent, and ABEs generally do not introduce indels even when using nCas9). If an NG PAM is not available, we suggest using SaCas9-KKH (which recognizes an NNNRRT PAM, but will result in a widened window which can cause bystander editing<sup>160</sup> or the SpRY variant<sup>162</sup> instead of the SpCas9-NG.

A consideration is the ability of the deaminase enzyme portion of the BE to deaminate other free ssDNA or RNA in the cell,<sup>171–175</sup> causing off-target editing. Engineering efforts have rationally designed deaminase variants that substantially decrease off-target editing,<sup>67,74,176,177</sup> but many times at the expense of sequence specificity, causing decreased on-target activity at certain sites. Specifically, rAPOBEC1-YE1 and the next-generation CBEs listed in Table 2 are examples of CBEs with greatly reduced off-target DNA editing, and ABEs containing the mutations V106W, F148A, and V82G show almost no off-target RNA editing (ABEs inherently do not display off-target DNA editing).<sup>173</sup> If small amounts of off-target editing are acceptable,



utilizing the most efficient base editors will aid in isogenic cell line generation. ABE8s significantly increase editing efficiencies but also have a wider editing window, increasing the chance of bystander mutations.

### ***Inclusion of proper controls***

The process of clonal expansion represents an enormous genetic bottleneck and puts the cells under extreme selective pressure. This can result in genetic, epigenetic, and/or phenotypic variation of single cell-derived lines despite accurate CRISPR-Cas9 genome editing.<sup>141,178</sup> Additionally, some CBEs can increase 2-fold the inherent mutation rate during isogenic cell line generation.<sup>179</sup> It is, therefore, of the utmost importance to include proper controls to confirm that observed differences in phenotypes or protein activity are due to the mutation of interest. When generating isogenic lines, transfections should contain a negative control sample that lacks a gRNA or has a non-targeting gRNA sequence that is absent from the human genome (Table 1, #6). For each set of transfections to generate isogenic cell lines (Basic Protocol 4), single cell clones should be isolated from this negative control sample alongside the on-target gRNA samples. Using these clones as the ‘wild-type’ cell lines for phenotyping and SNV characterization can control for the process of clonal expansion and potential effects due to DNA damage introduction by the BEs. Additionally, for each genotype of interest (wild-type, heterozygous, and homozygous for each SNV of interest), it is crucial to generate at least three lines derived from different clones. We also recommend generating an additional “control” line for each genotype that is a mixture of each of the individual clones with that specific genotype. Whole-genome sequencing or whole-exome sequencing of the resultant cell lines is highly encouraged to determine where other genomic modifications might have occurred during the process of clonal expansion.

### ***Quantification of base editing efficiency in bulk cells***

After base editing has had 3-5 days to occur, select the appropriate endpoint analysis for your experiment. For bulk Sanger sequencing, EditR can be used to reduce background signal and generate editing percentages based on a p-value with a detection limit for base editing of ~7%.<sup>138</sup> However, the reliability of this method is highly dependent on the quality of the Sanger sequencing read. Low quality Sanger sequencing reads may mislead the researcher into incorrectly concluding that their base editing efficiency is high enough to proceed to cell line generation (see Figure 4.5C and 4.5E for an example). Genomic DNA samples can also be prepared for Illumina NGS to robustly quantitate editing efficiencies. Targeted amplicon sequencing is the most common method with which to do this, and has been previously described.<sup>13,139,140</sup> We recommend the use of CRISPResso2 as an open-access software to quantify base editing efficiencies from FASTQ files.<sup>103</sup> However, it is important to note that over-amplification during either round of PCR can create PCR bias, which will result in inaccurate quantification; care should be taken to use the fewest possible PCR cycles during gDNA amplification and barcoding.

### ***Cell line considerations***

HEK293T cells are easily transfected and robust enough to tolerate clonal expansion. This provides researchers with a relatively 'well-behaved' cell line to optimize experimental conditions and assess the feasibility of generating a cell line with a given SNV before moving forward with other mammalian cell lines. The ultimate cell line to use will be entirely dependent on downstream experimental goals and should be thoroughly investigated before embarking on isogenic cell line generation. Furthermore, immortalized cell lines, such as HEK293T cells, often harbor gene duplications, chromosomal rearrangements, and mutations that allow them to effectively propagate in tissue culture. The exact genomic modifications may even vary from laboratory to laboratory for a given cell line. Additionally, these modifications can differ greatly from the reference genome. As such, it is prudent to first sequence the target locus of interest

for any mutations or variation before designing gRNA sequences. Additionally, if the data is available for that specific cell line, check the ploidy at each locus to determine the copy number of the gene of interest.<sup>180</sup>

#### **4.9.3 Troubleshooting:**

##### ***Low or no cell viability following transfection.***

A common problem is low cell viability following transfection. This may be solved by using less cationic lipid reagent, but usually at the expense of a lower transfection efficiency. However, the most common cause of low cell viability is transfecting under-confluent cells (see Figure 4.3B). Determining the confluency of cells is highly subjective; we have included pictures of cells with the optimal density for transfection, in Figure 4.3A.

##### ***Low cell viability following transfection, with all surviving cells concentrated in one area***

Another common problem after transfection is low cell viability, with “islands” of cells or cells only on the periphery of the wells. Additionally, cells in the middle of the islands are not transfected. This is usually due to an uneven distribution of cells when plating. This can be solved by using a cell strainer during plating and/or gently tapping the plate repeatedly against the hood surface after plating to halt circular movement of media (eddies) inside the well, which can push the cells to the edges of the wells.

##### ***Low editing efficiency after transfection.***

Sometimes observing an editing efficiency of <10% is due to low transfection efficiency. This can be due to transfecting over-confluent cells (see Figure 4.3C), which results in high cell viability but low transfection and editing efficiencies. If over-confluent cells are transfected and bulk sequencing shows some editing efficiency, using FACS to enrich for transfected cells prior

to re-assessing editing efficiency in bulk is advised. However, it is generally recommended for the researcher to re-transfect the cells when the confluency is ideal, as in Figure 4.3A.

If the transfection efficiency was >80%, and low or no base editing efficiency was observed, there could be an issue with the gRNA, or the target sequence may be incompatible with the base editor that was used. To assess if there is an issue with the gRNA, we suggest repeating the same experiment, but using wild-type Cas9 instead of the base editor, and then monitor the target site for indel formation. This will reveal if the Cas9:gRNA complex can access the target site. If low or no indels are formed, we recommend trying a different Cas enzyme such as the SpRY Cas9 and a different spacer sequence. If the gRNA results in high levels of indel introduction, the problem may be with the sequence context of the target, in which case we suggest using a different deaminase (see Table 2) and quantifying editing efficiency in bulk again. If even after optimization, editing efficiency is less than 10%, generating SNV-containing isogenic cell lines may still be feasible, but will require selection strategies, and dilution plating should not be used. See the Troubleshooting section on *“Certain genotypes are not observed after clonal expansion.”*

***Accidentally disrupted or washed away a vast majority of the cell monolayer.***

There are two instances in the protocol where indelicate techniques can result in loss of cells and be detrimental to the experiment if actions are not taken to correct for them. During harvesting of genomic DNA for genotyping (Basic Protocol 2), one might lose cells while washing. In this case, even a small number of cells remaining can be sufficient for PCR amplification of the targeted site, albeit a higher volume of cell lysate will have to be used (please refer to the Troubleshooting genomic PCR section). In this case, however, zero base editing efficiency may be due to only the presence of untransfected cells remaining. During sub-culturing of single cell-derived clones (Basic Protocol 5) this can happen as well. If some cells

remain adhered, add the appropriate amount of prewarmed culture media and place back into the incubator.

***Multiple PCR bands observed during gDNA amplification.***

If your primers are specific and still result in off-target amplification, increasing the annealing temperature or decreasing primer concentration can improve specificity. Furthermore, a high primer concentration can result in primer dimers and no band of the desired size. We have observed that using a lower concentration of primer than the manufacturer's recommendation (up to 10-fold less) usually generates cleaner on-target product bands. See Figure 4.4B for examples.

***No PCR bands observed during gDNA amplification.***

Increasing DMSO concentration or trying other additives may help with the genomic DNA PCR. Additionally, the amount of genomic DNA added to the PCR mixture can have drastic effects on PCR success. The recommended volume suggested in Basic Protocol 2 is based on 90-100% confluency pre-harvesting, but sometimes that may be too much (in certain cases, this results in smears on the gel). We recommend first trying 10-fold less genomic DNA template in the PCR reaction, by diluting the genomic DNA mixture in water. If fewer cells were harvested, increasing the amount of cell lysate up to 10-fold may be required. Decreasing primer concentration may improve the specificity and strength of amplification, especially with NGS. Additionally, gDNA can also be harvested using a kit, which can help with difficult gDNA PCRs.

***No colony expansion is observed after sorting for single cells.***

Sorting via FACS can be stressful for the cells, but there are some modifications that can help improve the success rate of clonal expansion. Reducing the time between harvesting and

sorting can improve the outcome of colony formation after sorting. Using a higher concentration of FBS in the collection media can also promote single cell survival and proliferation. Another suggestion is adding cell-conditioned media as a supplement to the collection media to facilitate cell growth. This can be done by collecting the supernatant from cells that have been passaged 24 hours prior, spinning it down, and adding it to the collection media through a 0.45 µm filter. Dilution plating is less harsh on the cells and can have higher success rates of clonal expansion, but as there is no selection for transfected cells, the percentage of cell lines with the SNV of interest will be lower. One option is to use FACS to enrich for transfected cells, allow them to recover, and then use dilution plating to isolate and clonally expand single cells.

### ***Certain genotypes are not observed after clonal expansion***

If no homozygous clones are obtained, it is possible that the homozygous edit with the target SNV is lethal, especially if the gene of interest is essential. We recommend using tools such as ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) to assess the pathogenicity of the target SNV in humans (if the SNV has only been observed in the heterozygous form in humans, the homozygous genotype may be lethal), and DepMap (<https://depmap.org/portal/>) to determine gene essentiality in the specific cell type. If the culprit is that initial base editing efficiencies are low, then additional selection strategies (beyond just selecting for cells with high base editor expression levels) can be employed. These selection strategies make use of reporter systems that involve the modulation of the presence or color of a fluorescent protein by base editing activity concurrently with editing of a genomic locus (e.g. ACE reporter, BE-FLARE, BIG-TREE, GO reporter and others).<sup>181–185</sup> In these cases, the reporter plasmid and a gRNA targeting the fluorescent protein are co-transfected with the BE and target gRNA. Individual cells with active base editor can be isolated by FACS, to enrich for cells with active base editor.

Conversely, if only homozygous clones are obtained, less active base editor variants can be used to reduce activity (see Table 2). Or, different populations of cells can be isolated with

FACS, such as cells with intermediate base editor expression or activity levels. Additionally, recent work has demonstrated that gRNA mismatches can titrate the activity of other genome editing agents. While this strategy has not been explicitly tested with base editors, it may act to reduce base editing activity as well.<sup>186</sup>

***The single cell-derived clones have a different morphology than the examples provided, or ones shown in literature.***

Although generally consistent, the morphology of single cell-derived clones can vary from clone to clone and from cell line to cell line. For this reason, we recommend aiming to generate at least three clones for each SNV in its respective cell line to confirm the reproducibility of your results. The resulting cell lines can be subjected to whole-genome or whole-exome genotyping to confirm that only the point mutations of interest have been introduced.

#### **4.9.4 Understanding results:**

Having the capability to generate both homozygous and heterozygous clones harboring SNVs highlights the importance of using base editing technologies; generating matched wild-type, heterozygous, and homozygous knock-in clones with traditional genome editing methods is typically quite inefficient or impossible without the use of “blocking mutations”<sup>125</sup> or multiple clonal expansion steps.<sup>187</sup> Using this protocol, we have found that base editors typically introduce their respective SNVs with >10-fold higher efficiency and >100-fold higher precision than traditional genome editing methods without the use of potentially undesired “blocking mutations”, as quantified by NGS, and without having to physically manipulate cells into phase-enriched populations (physical fractionization) or having to add exogenous chemicals to block cells into specific phases of the cell cycle (chemical blockade). This is in direct contrast to

traditional genome editing methods, where typically, accurate SNV introduction is less than 1%.<sup>110,122,124–127</sup>

When using FACS to isolate and clonally expand single cells, we have obtained homozygous edits with a success rate of approximately 25%, and heterozygous edits with a success rate of approximately 22%. This is measured from an average 8 clones that we typically obtain per 96-well plate. Additionally, in these cases, initial base editing efficiencies (when measured in bulk, as in Basic Protocol 3) were estimated at around 23%. When using the dilution method, we have obtained homozygous edits with a success rate of approximately 5%, and heterozygous edits with a success rate of approximately 14%. This is measured from an average of 11 clones that we typically obtain per 96-well plate. Additionally, in these cases, initial base editing efficiencies (when measured in bulk, as in Basic Protocol 3) were estimated at around 38%.

#### **4.9.5 Time considerations:**

Generating the gRNA plasmids will take 1-2 days, followed by another 1-2 days to sequence the plasmid and verify accurate spacer replacement. Preparing high quality, endotoxin-free plasmids for transfections will take 1 day. Time considerations for the initial assessment of base editing efficiencies (feasibility of generating an isogenic cell line) are as follows: 3-5 days to transfect the BE:gRNA combination(s) and allow for editing, 1 day to harvest the gDNA and PCR-amplify the locus of interest, followed by 1-2 days to determine the percentage desired SNVs introduction by Sanger sequencing. Once > 10% editing is confirmed in bulk, transfections of the same BE:gRNA combination(s) and preparation for single cell isolation requires 3 days. Regardless of which isolation method the researcher uses (FACS or dilution plating), expansion of single cell clones can span from 1-4 weeks depending on the growth rate of the cell line and the phenotypic effects of the introduced SNV. For HEK293T



cells, single cell-derived colonies are generally seen 4-7 days after single cell isolation and are ready to be sub-cultured and/or harvested 1 to 2 weeks thereafter. Subsequent sequencing of final clones will take 1-2 days; clonal expansion of the desired cell lines for cryo-storage will take 1-3 weeks depending on the clonal cell line's proliferation rate. Downstream experimental work (researcher-specific assays) can be simultaneously performed during clonal expansion for storage.

### **Acknowledgements**

Chapter 4, in part, a reprint of the material as it appears in Current Protocols in Molecular Biology. Vasquez, Carlos A.; Cowan, Quinn T.; Komor, Alexis C., *Current Protocols in Molecular Biology*, 2020. The dissertation author was co-first author of this paper. Chapter 4, in part (Figure 4.8), is a reprint of the material as it appears in Gene and Genome Editing. Gu, Sifeng; Bodai, Zsolt; Cowan, Quinn T.; Komor, Alexis C., *Gene and Genome Editing*, 2021. The dissertation author was a contributing author on this paper.

## REFERENCES

- (1) The 1000 Genomes Project Consortium. A Global Reference for Human Genetic Variation. *Nature* **2015**, 526 (7571), 68–74. <https://doi.org/10.1038/nature15393>.
- (2) Landrum, M. J.; Lee, J. M.; Benson, M.; Brown, G.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Hoover, J.; Jang, W.; Katz, K.; Ovetsky, M.; Riley, G.; Sethi, A.; Tully, R.; Villamarin-Salomon, R.; Rubinstein, W.; Maglott, D. R. ClinVar: Public Archive of Interpretations of Clinically Relevant Variants. *Nucleic Acids Res.* **2016**, 44 (D1), D862–D868. <https://doi.org/10.1093/nar/gkv1222>.
- (3) Pickar-Oliver, A.; Gersbach, C. A. The next Generation of CRISPR–Cas Technologies and Applications. *Nat. Rev. Mol. Cell Biol.* **2019**, 20 (8), 490–507. <https://doi.org/10.1038/s41580-019-0131-5>.
- (4) Jiang, F.; Doudna, J. A. CRISPR–Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **2017**, 46 (1), 505–529. <https://doi.org/10.1146/annurev-biophys-062215-010822>.
- (5) Scully, R.; Panday, A.; Elango, R.; Willis, N. A. DNA Double-Strand Break Repair-Pathway Choice in Somatic Mammalian Cells. *Nat. Rev. Mol. Cell Biol.* **2019**, 20 (11), 698–714. <https://doi.org/10.1038/s41580-019-0152-0>.
- (6) Altieri, F.; Grillo, C.; Maceroni, M.; Chichiarelli, S. DNA Damage and Repair: From Molecular Mechanisms to Health Implications. *Antioxid. Redox Signal.* **2008**, 10 (5), 891–938. <https://doi.org/10.1089/ars.2007.1830>.
- (7) Chang, H. H. Y.; Pannunzio, N. R.; Adachi, N.; Lieber, M. R. Non-Homologous DNA End Joining and Alternative Pathways to Double-Strand Break Repair. *Nat. Rev. Mol. Cell Biol.* **2017**, 18 (8), 495–506. <https://doi.org/10.1038/nrm.2017.48>.
- (8) Kosicki, M.; Tomberg, K.; Bradley, A. Repair of Double-Strand Breaks Induced by CRISPR–Cas9 Leads to Large Deletions and Complex Rearrangements. *Nat. Biotechnol.* **2018**, 36 (8), 765–771. <https://doi.org/10.1038/nbt.4192>.
- (9) Jasin, M.; Rothstein, R. Repair of Strand Breaks by Homologous Recombination. *Cold Spring Harb. Perspect. Biol.* **2013**, 5 (11), a012740–a012740. <https://doi.org/10.1101/cshperspect.a012740>.
- (10) Yeh, C. D.; Richardson, C. D.; Corn, J. E. Advances in Genome Editing through Control of DNA Repair Pathways. *Nat. Cell Biol.* **2019**, 21 (12), 1468–1478. <https://doi.org/10.1038/s41556-019-0425-z>.
- (11) 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* **2014**, 3, e04766. <https://doi.org/10.7554/eLife.04766>.
- (12) 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* **2016**, 533 (7603), 420–424. <https://doi.org/10.1038/nature17946>.

- (13) 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* **2017**, *551* (7681), 464–471. <https://doi.org/10.1038/nature24644>.
- (14) Richter, M. F.; Zhao, K. T.; Eton, E.; Lapinaite, A.; Newby, G. A.; Thuronyi, B. W.; Wilson, C.; Koblan, L. W.; Zeng, J.; Bauer, D. E.; Doudna, J. A.; Liu, D. R. Phage-Assisted Evolution of an Adenine Base Editor with Improved Cas Domain Compatibility and Activity. *Nat. Biotechnol.* **2020**, *38* (7), 883–891. <https://doi.org/10.1038/s41587-020-0453-z>.
- (15) Gaudelli, N. M.; Lam, D. K.; Rees, H. A.; Solá-Esteves, N. M.; Barrera, L. A.; Born, D. A.; Edwards, A.; Gehrke, J. M.; Lee, S.-J.; Liquori, A. J.; Murray, R.; Packer, M. S.; Rinaldi, C.; Slaymaker, I. M.; Yen, J.; Young, L. E.; Ciaramella, G. Directed Evolution of Adenine Base Editors with Increased Activity and Therapeutic Application. *Nat. Biotechnol.* **2020**, *38* (7), 892–900. <https://doi.org/10.1038/s41587-020-0491-6>.
- (16) Gu, S.; Bodai, Z.; Cowan, Q. T.; Komor, A. C. Base Editors: Expanding the Types of DNA Damage Products Harnessed for Genome Editing. *Gene Genome Ed.* **2021**. <https://doi.org/10.1016/j.ggedit.2021.100005>.
- (17) Green, R. C.; Berg, J. S.; Grody, W. W.; Kalia, S. S.; Korf, B. R.; Martin, C. L.; McGuire, A. L.; Nussbaum, R. L.; O'Daniel, J. M.; Ormond, K. E.; Rehm, H. L.; Watson, M. S.; Williams, M. S.; Biesecker, L. G. ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing. *Genet. Med.* **2013**, *15* (7), 565–574. <https://doi.org/10.1038/gim.2013.73>.
- (18) Rehm, H. L.; Berg, J. S.; Brooks, L. D.; Bustamante, C. D.; Evans, J. P.; Landrum, M. J.; Ledbetter, D. H.; Maglott, D. R.; Martin, C. L.; Nussbaum, R. L.; Plon, S. E.; Ramos, E. M.; Sherry, S. T.; Watson, M. S.; ClinGen. ClinGen—the Clinical Genome Resource. *N. Engl. J. Med.* **2015**, *372* (23), 2235–2242. <https://doi.org/10.1056/NEJMSr1406261>.
- (19) Karczewski, K. J.; Francioli, L. C.; Tiao, G.; Cummings, B. B.; Alföldi, J.; Wang, Q.; Collins, R. L.; Laricchia, K. M.; Ganna, A.; Birnbaum, D. P.; Gauthier, L. D.; Brand, H.; Solomonson, M.; Watts, N. A.; Rhodes, D.; Singer-Berk, M.; England, E. M.; Seaby, E. G.; Kosmicki, J. A.; Walters, R. K.; Tashman, K.; Farjoun, Y.; Banks, E.; Poterba, T.; Wang, A.; Seed, C.; Whiffin, N.; Chong, J. X.; Samocha, K. E.; Pierce-Hoffman, E.; Zappala, Z.; O'Donnell-Luria, A. H.; Minikel, E. V.; Weisburd, B.; Lek, M.; Ware, J. S.; Vittal, C.; Armean, I. M.; Bergelson, L.; Cibulskis, K.; Connolly, K. M.; Covarrubias, M.; Donnelly, S.; Ferriera, S.; Gabriel, S.; Gentry, J.; Gupta, N.; Jeandet, T.; Kaplan, D.; Llanwarne, C.; Munshi, R.; Novod, S.; Petrillo, N.; Roazen, D.; Ruano-Rubio, V.; Saltzman, A.; Schleicher, M.; Soto, J.; Tibbetts, K.; Tolonen, C.; Wade, G.; Talkowski, M. E.; Aguilar Salinas, C. A.; Ahmad, T.; Albert, C. M.; Ardissino, D.; Atzmon, G.; Barnard, J.; Beaucher, L.; Benjamin, E. J.; Boehnke, M.; Bonnycastle, L. L.; Bottinger, E. P.; Bowden, D. W.; Bown, M. J.; Chambers, J. C.; Chan, J. C.; Chasman, D.; Cho, J.; Chung, M. K.; Cohen, B.; Correa, A.; Dabelea, D.; Daly, M. J.; Darbar, D.; Duggirala, R.; Dupuis, J.; Ellinor, P. T.; Elosua, R.; Erdmann, J.; Esko, T.; Färkkilä, M.; Florez, J.; Franke, A.; Getz, G.; Glaser, B.; Glatt, S. J.; Goldstein, D.; Gonzalez, C.; Groop, L.; Haiman, C.; Hanis, C.; Harms, M.; Hiltunen, M.; Holli, M. M.; Hultman, C. M.; Kallela, M.; Kaprio, J.; Kathiresan, S.; Kim, B.-J.; Kim, Y. J.; Kirov, G.; Kooner, J.; Koskinen, S.; Krumholz, H. M.; Kugathasan, S.; Kwak, S. H.; Laakso, M.; Lehtimäki, T.; Loos, R. J. F.; Lubitz, S. A.; Ma, R. C. W.; MacArthur, D. G.; Marrugat, J.; Mattila, K. M.; McCarroll, S.; McCarthy, M. I.; McGovern, D.; McPherson, R.; Meigs, J. B.; Melander, O.; Metspalu, A.; Neale, B. M.; Nilsson, P. M.; O'Donovan, M. C.; Ongur, D.;

- Orozco, L.; Owen, M. J.; Palmer, C. N. A.; Palotie, A.; Park, K. S.; Pato, C.; Pulver, A. E.; Rahman, N.; Remes, A. M.; Rioux, J. D.; Ripatti, S.; Roden, D. M.; Saleheen, D.; Salomaa, V.; Samani, N. J.; Scharf, J.; Schunkert, H.; Shoemaker, M. B.; Sklar, P.; Soininen, H.; Sokol, H.; Spector, T.; Sullivan, P. F.; Suvisaari, J.; Tai, E. S.; Teo, Y. Y.; Tiinamäijä, T.; Tsuang, M.; Turner, D.; Tusie-Luna, T.; Vartiainen, E.; Vawter, M. P.; Ware, J. S.; Watkins, H.; Weersma, R. K.; Wessman, M.; Wilson, J. G.; Xavier, R. J.; Neale, B. M.; Daly, M. J.; MacArthur, D. G.; Genome Aggregation Database Consortium. The Mutational Constraint Spectrum Quantified from Variation in 141,456 Humans. *Nature* **2020**, *581* (7809), 434–443. <https://doi.org/10.1038/s41586-020-2308-7>.
- (20) Landrum, M. J.; Lee, J. M.; Benson, M.; Brown, G. R.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; Karapetyan, K.; Katz, K.; Liu, C.; Maddipatla, Z.; Malheiro, A.; McDaniel, K.; Ovetsky, M.; Riley, G.; Zhou, G.; Holmes, J. B.; Kattman, B. L.; Maglott, D. R. ClinVar: Improving Access to Variant Interpretations and Supporting Evidence. *Nucleic Acids Res.* **2018**, *46* (D1), D1062–D1067. <https://doi.org/10.1093/nar/gkx1153>.
- (21) Starita, L. M.; Ahituv, N.; Dunham, M. J.; Kitzman, J. O.; Roth, F. P.; Seelig, G.; Shendure, J.; Fowler, D. M. Variant Interpretation: Functional Assays to the Rescue. *Am. J. Hum. Genet.* **2017**, *101* (3), 315–325. <https://doi.org/10.1016/j.ajhg.2017.07.014>.
- (22) Shah, N.; Hou, Y.-C. C.; Yu, H.-C.; Sainger, R.; Caskey, C. T.; Venter, J. C.; Telenti, A. Identification of Misclassified ClinVar Variants via Disease Population Prevalence. *Am. J. Hum. Genet.* **2018**, *102* (4), 609–619. <https://doi.org/10.1016/j.ajhg.2018.02.019>.
- (23) Cooper, D. N.; Krawczak, M.; Polychronakos, C.; Tyler-Smith, C.; Kehrer-Sawatzki, H. Where Genotype Is Not Predictive of Phenotype: Towards an Understanding of the Molecular Basis of Reduced Penetrance in Human Inherited Disease. *Hum. Genet.* **2013**, *132* (10), 1077–1130. <https://doi.org/10.1007/s00439-013-1331-2>.
- (24) Rahit, K. M. T. H.; Tarailo-Graovac, M. Genetic Modifiers and Rare Mendelian Disease. *Genes* **2020**, *11* (3), 239. <https://doi.org/10.3390/genes11030239>.
- (25) Mukherjee, S.; Cogan, J. D.; Newman, J. H.; Phillips, J. A.; Hamid, R.; Meiler, J.; Capra, J. A. Identifying Digenic Disease Genes via Machine Learning in the Undiagnosed Diseases Network. *Am. J. Hum. Genet.* **2021**, *108* (10), 1946–1963. <https://doi.org/10.1016/j.ajhg.2021.08.010>.
- (26) Jinek, M.; East, A.; Cheng, A.; Lin, S.; Ma, E.; Doudna, J. RNA-Programmed Genome Editing in Human Cells. *eLife* **2013**, *2*, e00471. <https://doi.org/10.7554/eLife.00471>.
- (27) Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M. RNA-Guided Human Genome Engineering via Cas9. *Science* **2013**, *339* (6121), 823–826. <https://doi.org/10.1126/science.1232033>.
- (28) Anzalone, A. V.; Koblan, L. W.; Liu, D. R. Genome Editing with CRISPR–Cas Nucleases, Base Editors, Transposases and Prime Editors. *Nat. Biotechnol.* **2020**, *38* (7), 824–844. <https://doi.org/10.1038/s41587-020-0561-9>.
- (29) Findlay, G. M.; Daza, R. M.; Martin, B.; Zhang, M. D.; Leith, A. P.; Gasperini, M.; Janizek, J. D.; Huang, X.; Starita, L. M.; Shendure, J. Accurate Classification of BRCA1 Variants

- with Saturation Genome Editing. *Nature* **2018**, 562 (7726), 217–222. <https://doi.org/10.1038/s41586-018-0461-z>.
- (30) Kosicki, M.; Tomberg, K.; Bradley, A. Repair of Double-Strand Breaks Induced by CRISPR–Cas9 Leads to Large Deletions and Complex Rearrangements. *Nat. Biotechnol.* **2018**, 36 (8). <https://doi.org/10.1038/nbt.4192>.
  - (31) Turchiano, G.; Andrieux, G.; Klermund, J.; Blattner, G.; Pennucci, V.; el Gaz, M.; Monaco, G.; Poddar, S.; Mussolino, C.; Cornu, T. I.; Boerries, M.; Cathomen, T. Quantitative Evaluation of Chromosomal Rearrangements in Gene-Edited Human Stem Cells by CAST-Seq. *Cell Stem Cell* **2021**, 28 (6), 1136–1147.e5. <https://doi.org/10.1016/j.stem.2021.02.002>.
  - (32) Leibowitz, M. L.; Papathanasiou, S.; Doerfler, P. A.; Blaine, L. J.; Sun, L.; Yao, Y.; Zhang, C.-Z.; Weiss, M. J.; Pellman, D. Chromothripsis as an On-Target Consequence of CRISPR–Cas9 Genome Editing. *Nat. Genet.* **2021**, 53 (6), 895–905. <https://doi.org/10.1038/s41588-021-00838-7>.
  - (33) Haapaniemi, E.; Botla, S.; Persson, J.; Schmierer, B.; Taipale, J. CRISPR–Cas9 Genome Editing Induces a P53-Mediated DNA Damage Response. *Nat. Med.* **2018**, 24 (7), 927–930. <https://doi.org/10.1038/s41591-018-0049-z>.
  - (34) Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **2013**, 339 (6121), 819–823. <https://doi.org/10.1126/science.1231143>.
  - (35) Ihry, R. J.; Worringer, K. A.; Salick, M. R.; Frias, E.; Ho, D.; Theriault, K.; Kommineni, S.; Chen, J.; Sondey, M.; Ye, C.; Randhawa, R.; Kulkarni, T.; Yang, Z.; McAllister, G.; Russ, C.; Reece-Hoyes, J.; Forrester, W.; Hoffman, G. R.; Dolmetsch, R.; Kaykas, A. P53 Inhibits CRISPR–Cas9 Engineering in Human Pluripotent Stem Cells. *Nat. Med.* **2018**, 24 (7), 939–946. <https://doi.org/10.1038/s41591-018-0050-6>.
  - (36) Anzalone, A. V.; Randolph, P. B.; Davis, J. R.; Sousa, A. A.; Koblan, L. W.; Levy, J. M.; Chen, P. J.; Wilson, C.; Newby, G. A.; Raguram, A.; Liu, D. R. Search-and-Replace Genome Editing without Double-Strand Breaks or Donor DNA. *Nature* **2019**, 576 (7785), 149–157. <https://doi.org/10.1038/s41586-019-1711-4>.
  - (37) Jiang, F.; Taylor, D. W.; Chen, J. S.; Kornfeld, J. E.; Zhou, K.; Thompson, A. J.; Nogales, E.; Doudna, J. A. Structures of a CRISPR–Cas9 R-Loop Complex Primed for DNA Cleavage. *Science* **2016**, 351 (6275), 867–871. <https://doi.org/10.1126/science.aad8282>.
  - (38) 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* **2016**, 533 (7603), 420–424. <https://doi.org/10.1038/nature17946>.
  - (39) Sherry, S. T.; Ward, M. H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E. M.; Sirotkin, K. dbSNP: The NCBI Database of Genetic Variation. *Nucleic Acids Res.* **2001**, 29 (1), 308–311. <https://doi.org/10.1093/nar/29.1.308>.
  - (40) Webber, B. R.; Lonetree, C.; Kluesner, M. G.; Johnson, M. J.; Pomeroy, E. J.; Diers, M. D.; Lahr, W. S.; Draper, G. M.; Slipek, N. J.; Smeester, B. A.; Lovendahl, K. N.; McElroy, A. N.;

- Gordon, W. R.; Osborn, M. J.; Moriarity, B. S. Highly Efficient Multiplex Human T Cell Engineering without Double-Strand Breaks Using Cas9 Base Editors. *Nat. Commun.* **2019**, *10* (1), 5222. <https://doi.org/10.1038/s41467-019-13007-6>.
- (41) Nishimasu, H.; Shi, X.; Ishiguro, S.; Gao, L.; Hirano, S.; Okazaki, S.; Noda, T.; Abudayyeh, O. O.; Gootenberg, J. S.; Mori, H.; Oura, S.; Holmes, B.; Tanaka, M.; Seki, M.; Hirano, H.; Aburatani, H.; Ishitani, R.; Ikawa, M.; Yachie, N.; Zhang, F.; Nureki, O. Engineered CRISPR-Cas9 Nuclease with Expanded Targeting Space. *Science* **2018**, *361* (6408), 1259–1262. <https://doi.org/10.1126/science.aas9129>.
- (42) Walton, R. T.; Christie, K. A.; Whittaker, M. N.; Kleinstiver, B. P. Unconstrained Genome Targeting with Near-PAMless Engineered CRISPR-Cas9 Variants. *Science* **2020**, *368* (6488), 290–296. <https://doi.org/10.1126/science.aba8853>.
- (43) Ting, P. Y.; Parker, A. E.; Lee, J. S.; Trussell, C.; Sharif, O.; Luna, F.; Federe, G.; Barnes, S. W.; Walker, J. R.; Vance, J.; Gao, M.-Y.; Klock, H. E.; Clarkson, S.; Russ, C.; Miraglia, L. J.; Cooke, M. P.; Boitano, A. E.; McNamara, P.; Lamb, J.; Schmedt, C.; Snead, J. L. Guide Swap Enables Genome-Scale Pooled CRISPR–Cas9 Screening in Human Primary Cells. *Nat. Methods* **2018**, *15* (11), 941–946. <https://doi.org/10.1038/s41592-018-0149-1>.
- (44) Glaser, V.; Flugel, C.; Kath, J.; Du, W.; Drosdek, V.; Franke, C.; Stein, M.; Pruß, A.; Schmuck-Henneresse, M.; Volk, H.-D.; Reinke, P.; Wagner, D. L. Combining Different CRISPR Nucleases for Simultaneous Knock-in and Base Editing Prevents Translocations in Multiplex-Edited CAR T Cells. *Genome Biol.* **2023**, *24* (1), 89. <https://doi.org/10.1186/s13059-023-02928-7>.
- (45) Kim, Y. B.; Komor, A. C.; Levy, J. M.; Packer, M. S.; Zhao, K. T.; Liu, D. R. Increasing the Genome-Targeting Scope and Precision of Base Editing with Engineered Cas9-Cytidine Deaminase Fusions. *Nat. Biotechnol.* **2017**, *35* (4), 371–376. <https://doi.org/10.1038/nbt.3803>.
- (46) Porto, E. M.; Komor, A. C.; Slaymaker, I. M.; Yeo, G. W. Base Editing: Advances and Therapeutic Opportunities. *Nat. Rev. Drug Discov.* **2020**, *19* (12), 839–859. <https://doi.org/10.1038/s41573-020-0084-6>.
- (47) Liu, Z.; Lu, Z.; Yang, G.; Huang, S.; Li, G.; Feng, S.; Liu, Y.; Li, J.; Yu, W.; Zhang, Y.; Chen, J.; Sun, Q.; Huang, X. Efficient Generation of Mouse Models of Human Diseases via ABE- and BE-Mediated Base Editing. *Nat. Commun.* **2018**, *9* (1), 2338. <https://doi.org/10.1038/s41467-018-04768-7>.
- (48) Zalatan, J. G.; Lee, M. E.; Almeida, R.; Gilbert, L. A.; Whitehead, E. H.; La Russa, M.; Tsai, J. C.; Weissman, J. S.; Dueber, J. E.; Qi, L. S.; Lim, W. A. Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell* **2015**, *160* (1–2), 339–350. <https://doi.org/10.1016/j.cell.2014.11.052>.
- (49) Ma, H.; Tu, L.-C.; Naseri, A.; Huisman, M.; Zhang, S.; Grunwald, D.; Pederson, T. Multiplexed Labeling of Genomic Loci with dCas9 and Engineered sgRNAs Using CRISPRainbow. *Nat. Biotechnol.* **2016**, *34* (5), 528–530. <https://doi.org/10.1038/nbt.3526>.
- (50) Wang, S.; Su, J.-H.; Zhang, F.; Zhuang, X. An RNA-Aptamer-Based Two-Color CRISPR Labeling System. *Sci. Rep.* **2016**, *6* (1), 26857. <https://doi.org/10.1038/srep26857>.

- (51) Fukunaga, K.; Yokobayashi, Y. Directed Evolution of Orthogonal RNA–RBP Pairs through Library-vs-Library *in Vitro* Selection. *Nucleic Acids Res.* **2021**, gkab527. <https://doi.org/10.1093/nar/gkab527>.
- (52) Konermann, S.; Brigham, M. D.; Trevino, A. E.; Joung, J.; Abudayyeh, O. O.; Barcena, C.; Hsu, P. D.; Habib, N.; Gootenberg, J. S.; Nishimasu, H.; Nureki, O.; Zhang, F. Genome-Scale Transcriptional Activation by an Engineered CRISPR-Cas9 Complex. *Nature* **2015**, 517 (7536), 583–588. <https://doi.org/10.1038/nature14136>.
- (53) Shechner, D. M.; Hacisuleyman, E.; Younger, S. T.; Rinn, J. L. Multiplexable, Locus-Specific Targeting of Long RNAs with CRISPR-Display. *Nat. Methods* **2015**, 12 (7), 664–670. <https://doi.org/10.1038/nmeth.3433>.
- (54) Hess, G. T.; Frésard, L.; Han, K.; Lee, C. H.; Li, A.; Cimprich, K. A.; Montgomery, S. B.; Bassik, M. C. Directed Evolution Using dCas9-Targeted Somatic Hypermutation in Mammalian Cells. *Nat. Methods* **2016**, 13 (12), 1036–1042. <https://doi.org/10.1038/nmeth.4038>.
- (55) Vålegård, K.; Murray, J. B.; Stonehouse, N. J.; van den Worm, S.; Stockley, P. G.; Liljas, L. The Three-Dimensional Structures of Two Complexes between Recombinant MS2 Capsids and RNA Operator Fragments Reveal Sequence-Specific Protein-RNA Interactions. *J. Mol. Biol.* **1997**, 270 (5), 724–738. <https://doi.org/10.1006/jmbi.1997.1144>.
- (56) Tars, K.; Fridborg, K.; Bundule, M.; Liljas, L. The Three-Dimensional Structure of Bacteriophage PP7 from *Pseudomonas Aeruginosa* at 3.7-Å Resolution. *Virology* **2000**, 272 (2), 331–337. <https://doi.org/10.1006/viro.2000.0373>.
- (57) Collantes, J. C.; Tan, V. M.; Xu, H.; Ruiz-Urigüen, M.; Alasadi, A.; Guo, J.; Tao, H.; Su, C.; Tyc, K. M.; Selmi, T.; Lambourne, J. J.; Harbottle, J. A.; Stombaugh, J.; Xing, J.; Wiggins, C. M.; Jin, S. Development and Characterization of a Modular CRISPR and RNA Aptamer Mediated Base Editing System. *CRISPR J.* **2021**, 4 (1), 58–68. <https://doi.org/10.1089/crispr.2020.0035>.
- (58) Li, C.; Zong, Y.; Jin, S.; Zhu, H.; Lin, D.; Li, S.; Qiu, J.-L.; Wang, Y.; Gao, C. SWISS: Multiplexed Orthogonal Genome Editing in Plants with a Cas9 Nickase and Engineered CRISPR RNA Scaffolds. *Genome Biol.* **2020**, 21 (1), 141. <https://doi.org/10.1186/s13059-020-02051-x>.
- (59) Richter, M. F.; Zhao, K. T.; Eton, E.; Lapinaite, A.; Newby, G. A.; Thuronyi, B. W.; Wilson, C.; Koblan, L. W.; Zeng, J.; Bauer, D. E.; Doudna, J. A.; Liu, D. R. Phage-Assisted Evolution of an Adenine Base Editor with Improved Cas Domain Compatibility and Activity. *Nat. Biotechnol.* **2020**, 38 (7), 883–891. <https://doi.org/10.1038/s41587-020-0453-z>.
- (60) Nguyen Tran, M. T.; Mohd Khalid, M. K. N.; Wang, Q.; Walker, J. K. R.; Lidgerwood, G. E.; Dilworth, K. L.; Lisowski, L.; Pébay, A.; Hewitt, A. W. Engineering Domain-Inlaid SaCas9 Adenine Base Editors with Reduced RNA off-Targets and Increased on-Target DNA Editing. *Nat. Commun.* **2020**, 11 (1), 4871. <https://doi.org/10.1038/s41467-020-18715-y>.
- (61) Gaudelli, N. M.; Lam, D. K.; Rees, H. A.; Solá-Estevés, N. M.; Barrera, L. A.; Born, D. A.; Edwards, A.; Gehrke, J. M.; Lee, S.-J.; Liquori, A. J.; Murray, R.; Packer, M. S.; Rinaldi, C.; Slaymaker, I. M.; Yen, J.; Young, L. E.; Ciaramella, G. Directed Evolution of Adenine Base

Editors with Increased Activity and Therapeutic Application. *Nat. Biotechnol.* **2020**, 38 (7), 892–900. <https://doi.org/10.1038/s41587-020-0491-6>.

- (62) Komor, A. C.; Zhao, K. T.; Packer, M. S.; Gaudelli, N. M.; Waterbury, A. L.; Koblan, L. W.; Kim, Y. B.; Badran, A. H.; Liu, D. R. 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.* **2017**, 3 (8). <https://doi.org/10.1126/sciadv.aao4774>.
- (63) Nishida, K.; Arazoe, T.; Yachie, N.; Banno, S.; Kakimoto, M.; Tabata, M.; Mochizuki, M.; Miyabe, A.; Araki, M.; Hara, K. Y.; Shimatani, Z.; Kondo, A. Targeted Nucleotide Editing Using Hybrid Prokaryotic and Vertebrate Adaptive Immune Systems. *Science* **2016**, 353 (6305), aaf8729. <https://doi.org/10.1126/science.aaf8729>.
- (64) Zhang, X.; Chen, L.; Zhu, B.; Wang, L.; Chen, C.; Hong, M.; Huang, Y.; Li, H.; Han, H.; Cai, B.; Yu, W.; Yin, S.; Yang, L.; Yang, Z.; Liu, M.; Zhang, Y.; Mao, Z.; Wu, Y.; Liu, M.; Li, D. Increasing the Efficiency and Targeting Range of Cytidine Base Editors through Fusion of a Single-Stranded DNA-Binding Protein Domain. *Nat. Cell Biol.* **2020**, 22 (6), 740–750. <https://doi.org/10.1038/s41556-020-0518-8>.
- (65) Koblan, L. W.; Doman, J. L.; Wilson, C.; Levy, J. M.; Tay, T.; Newby, G. A.; Maianti, J. P.; Raguram, A.; Liu, D. R. Improving Cytidine and Adenine Base Editors by Expression Optimization and Ancestral Reconstruction. *Nat. Biotechnol.* **2018**, 36 (9), 843–848. <https://doi.org/10.1038/nbt.4172>.
- (66) Thuronyi, B. W.; Koblan, L. W.; Levy, J. M.; Yeh, W.-H.; Zheng, C.; Newby, G. A.; Wilson, C.; Bhaumik, M.; Shubina-Oleinik, O.; Holt, J. R.; Liu, D. R. Continuous Evolution of Base Editors with Expanded Target Compatibility and Improved Activity. *Nat. Biotechnol.* **2019**, 37 (9), 1070–1079. <https://doi.org/10.1038/s41587-019-0193-0>.
- (67) Yu, Y.; Leete, T. C.; Born, D. A.; Young, L.; Barrera, L. A.; Lee, S.-J.; Rees, H. A.; Ciaramella, G.; Gaudelli, N. M. Cytosine Base Editors with Minimized Unguided DNA and RNA Off-Target Events and High on-Target Activity. *Nat. Commun.* **2020**, 11 (1), 2052. <https://doi.org/10.1038/s41467-020-15887-5>.
- (68) Wang, Y.; Zhou, L.; Tao, R.; Liu, N.; Long, J.; Qin, F.; Tang, W.; Yang, Y.; Chen, Q.; Yao, S. sgBE: A Structure-Guided Design of sgRNA Architecture Specifies Base Editing Window and Enables Simultaneous Conversion of Cytosine and Adenosine. *Genome Biol.* **2020**, 21 (1), 222. <https://doi.org/10.1186/s13059-020-02137-6>.
- (69) Fu, J. L.; Kanno, T.; Liang, S.-C.; Matzke, A. J. M.; Matzke, M. GFP Loss-of-Function Mutations in *Arabidopsis Thaliana*. *G3 GenesGenomesGenetics* **2015**, 5 (9), 1849–1855. <https://doi.org/10.1534/g3.115.019604>.
- (70) Martin, A. St.; Salamango, D. J.; Serebrenik, A. A.; Shaban, N. M.; Brown, W. L.; Harris, R. S. A Panel of eGFP Reporters for Single Base Editing by APOBEC-Cas9 Editosome Complexes. *Sci. Rep.* **2019**, 9 (1), 497. <https://doi.org/10.1038/s41598-018-36739-9>.
- (71) Di Blasi, R.; Marbiah, M. M.; Siciliano, V.; Polizzi, K.; Ceroni, F. A Call for Caution in Analysing Mammalian Co-Transfection Experiments and Implications of Resource Competition in Data Misinterpretation. *Nat. Commun.* **2021**, 12 (1), 2545. <https://doi.org/10.1038/s41467-021-22795-9>.



- (72) Vakulskas, C. A.; Dever, D. P.; Rettig, G. R.; Turk, R.; Jacobi, A. M.; Collingwood, M. A.; Bode, N. M.; McNeill, M. S.; Yan, S.; Camarena, J.; Lee, C. M.; Park, S. H.; Wiebking, V.; Bak, R. O.; Gomez-Ospina, N.; Pavel-Dinu, M.; Sun, W.; Bao, G.; Porteus, M. H.; Behlke, M. A. A High-Fidelity Cas9 Mutant Delivered as a Ribonucleoprotein Complex Enables Efficient Gene Editing in Human Hematopoietic Stem and Progenitor Cells. *Nat. Med.* **2018**, *24* (8), 1216–1224. <https://doi.org/10.1038/s41591-018-0137-0>.
- (73) Standage-Beier, K.; Tekel, S. J.; Brookhouser, N.; Schwarz, G.; Nguyen, T.; Wang, X.; Brafman, D. A. A Transient Reporter for Editing Enrichment (TREE) in Human Cells. *Nucleic Acids Res.* **2019**, *47* (19), e120–e120. <https://doi.org/10.1093/nar/gkz713>.
- (74) Doman, J. L.; Raguram, A.; Newby, G. A.; Liu, D. R. Evaluation and Minimization of Cas9-Independent off-Target DNA Editing by Cytosine Base Editors. *Nat. Biotechnol.* **2020**, *38* (5), 620–628. <https://doi.org/10.1038/s41587-020-0414-6>.
- (75) Tsai, S. Q.; Zheng, Z.; Nguyen, N. T.; Liebers, M.; Topkar, V. V.; Thapar, V.; Wyvekens, N.; Khayter, C.; Iafrate, A. J.; Le, L. P.; Aryee, M. J.; Joung, J. K. GUIDE-Seq Enables Genome-Wide Profiling of off-Target Cleavage by CRISPR-Cas Nucleases. *Nat. Biotechnol.* **2015**, *33* (2), 187–197. <https://doi.org/10.1038/nbt.3117>.
- (76) Doench, J. G.; Fusi, N.; Sullender, M.; Hegde, M.; Vaimberg, E. W.; Donovan, K. F.; Smith, I.; Tothova, Z.; Wilen, C.; Orchard, R.; Virgin, H. W.; Listgarten, J.; Root, D. E. Optimized sgRNA Design to Maximize Activity and Minimize Off-Target Effects of CRISPR-Cas9. *Nat. Biotechnol.* **2016**, *34* (2), 184–191. <https://doi.org/10.1038/nbt.3437>.
- (77) Rees, H. A.; Wilson, C.; Doman, J. L.; Liu, D. R. Analysis and Minimization of Cellular RNA Editing by DNA Adenine Base Editors. *Sci. Adv.* **2019**, *5* (5), eaax5717. <https://doi.org/10.1126/sciadv.aax5717>.
- (78) Nachtegaal, C.; Gravel, B.; Dillen, A.; Smits, G.; Nowé, A.; Papadimitriou, S.; Lenaerts, T. Scaling up Oligogenic Diseases Research with OLIDA: The Oligogenic Diseases Database. *Database* **2022**, 2022, baac023. <https://doi.org/10.1093/database/baac023>.
- (79) Arbab, M.; Shen, M. W.; Mok, B.; Wilson, C.; Matuszek, Ż.; Cassa, C. A.; Liu, D. R. Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. *Cell* **2020**, *182* (2), 463–480.e30. <https://doi.org/10.1016/j.cell.2020.05.037>.
- (80) Kim, N.; Choi, S.; Kim, S.; Song, M.; Seo, J. H.; Min, S.; Park, J.; Cho, S.-R.; Kim, H. H. Deep Learning Models to Predict the Editing Efficiencies and Outcomes of Diverse Base Editors. *Nat. Biotechnol.* **2024**, *42* (3), 484–497. <https://doi.org/10.1038/s41587-023-01792-x>.
- (81) Kovalevich, J.; Langford, D. Considerations for the Use of SH-SY5Y Neuroblastoma Cells in Neurobiology. In *Neuronal Cell Culture*; Amini, S., White, M. K., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2013; Vol. 1078, pp 9–21. [https://doi.org/10.1007/978-1-62703-640-5\\_2](https://doi.org/10.1007/978-1-62703-640-5_2).
- (82) Shipley, M. M.; Mangold, C. A.; Szpara, M. L. Differentiation of the SH-SY5Y Human Neuroblastoma Cell Line. *J. Vis. Exp.* **2016**, No. 108, 53193. <https://doi.org/10.3791/53193>.

- (83) Garaffo, G.; Conte, D.; Provero, P.; Tomaiuolo, D.; Luo, Z.; Pincioli, P.; Peano, C.; D'Atri, I.; Gitton, Y.; Etzion, T.; Gothilf, Y.; Gays, D.; Santoro, M. M.; Merlo, G. R. The Dlx5 and Foxg1 Transcription Factors, Linked via miRNA-9 and -200, Are Required for the Development of the Olfactory and GnRH System. *Mol. Cell. Neurosci.* **2015**, *68*, 103–119. <https://doi.org/10.1016/j.mcn.2015.04.007>.
- (84) Yan, Y.; Wang, G.; Luo, X.; Zhang, P.; Peng, S.; Cheng, X.; Wang, M.; Yang, X. Endoplasmic Reticulum Stress-Related Calcium Imbalance Plays an Important Role on Zinc Oxide Nanoparticles-Induced Failure of Neural Tube Closure during Embryogenesis. *Environ. Int.* **2021**, *152*, 106495. <https://doi.org/10.1016/j.envint.2021.106495>.
- (85) Plooster, M.; Brennwald, P.; Gupton, S. L. Endosomal Trafficking in Schizophrenia. *Curr. Opin. Neurobiol.* **2022**, *74*, 102539. <https://doi.org/10.1016/j.conb.2022.102539>.
- (86) Plooster, M.; Rossi, G.; Farrell, M. S.; McAfee, J. C.; Bell, J. L.; Ye, M.; Diering, G. H.; Won, H.; Gupton, S. L.; Brennwald, P. Schizophrenia-Linked Protein tSNARE1 Regulates Endosomal Trafficking in Cortical Neurons. *J. Neurosci.* **2021**, *41* (45), 9466–9481. <https://doi.org/10.1523/JNEUROSCI.0556-21.2021>.
- (87) Riordan, J. D.; Nadeau, J. H. From Peas to Disease: Modifier Genes, Network Resilience, and the Genetics of Health. *Am. J. Hum. Genet.* **2017**, *101* (2), 177–191. <https://doi.org/10.1016/j.ajhg.2017.06.004>.
- (88) Nelson, J. W.; Randolph, P. B.; Shen, S. P.; Everette, K. A.; Chen, P. J.; Anzalone, A. V.; An, M.; Newby, G. A.; Chen, J. C.; Hsu, A.; Liu, D. R. Engineered pegRNAs Improve Prime Editing Efficiency. *Nat. Biotechnol.* **2022**, *40* (3), 402–410. <https://doi.org/10.1038/s41587-021-01039-7>.
- (89) Yuan, Q.; Gao, X. Multiplex Base- and Prime-Editing with Drive-and-Process CRISPR Arrays. *Nat. Commun.* **2022**, *13* (1), 2771. <https://doi.org/10.1038/s41467-022-30514-1>.
- (90) Di Blasi, R.; Pisani, M.; Tedeschi, F.; Marbiah, M. M.; Polizzi, K.; Furini, S.; Siciliano, V.; Ceroni, F. Resource-Aware Construct Design in Mammalian Cells. *Nat. Commun.* **2023**, *14* (1), 3576. <https://doi.org/10.1038/s41467-023-39252-4>.
- (91) Maynard, S.; Swistowska, A. M.; Lee, J. W.; Liu, Y.; Liu, S.-T.; Da Cruz, A. B.; Rao, M.; De Souza-Pinto, N. C.; Zeng, X.; Bohr, V. A. Human Embryonic Stem Cells Have Enhanced Repair of Multiple Forms of DNA Damage. *Stem Cells* **2008**, *26* (9), 2266–2274. <https://doi.org/10.1634/stemcells.2007-1041>.
- (92) Wang, X.; Ding, C.; Yu, W.; Wang, Y.; He, S.; Yang, B.; Xiong, Y.-C.; Wei, J.; Li, J.; Liang, J.; Lu, Z.; Zhu, W.; Wu, J.; Zhou, Z.; Huang, X.; Liu, Z.; Yang, L.; Chen, J. Cas12a Base Editors Induce Efficient and Specific Editing with Low DNA Damage Response. *Cell Rep.* **2020**, *31* (9), 107723. <https://doi.org/10.1016/j.celrep.2020.107723>.
- (93) Park, J.-C.; Jang, H.-K.; Kim, J.; Han, J. H.; Jung, Y.; Kim, K.; Bae, S.; Cha, H.-J. High Expression of Uracil DNA Glycosylase Determines C to T Substitution in Human Pluripotent Stem Cells. *Mol. Ther. - Nucleic Acids* **2022**, *27*, 175–183. <https://doi.org/10.1016/j.omtn.2021.11.023>.

- (94) Park, J.-C.; Kim, Y.-J.; Kang, C. Y.; Bae, S.; Cha, H.-J. *Enhancing Precise Genome Editing in Human Pluripotent Stem Cells through Dual Inhibition of DNA Damage Response and Repair Pathways*; preprint; In Review, 2023. <https://doi.org/10.21203/rs.3.rs-3028152/v1>.
- (95) Wang, C.; Cheng, J. K. W.; Zhang, Q.; Hughes, N. W.; Xia, Q.; Winslow, M. M.; Cong, L. Microbial Single-Strand Annealing Proteins Enable CRISPR Gene-Editing Tools with Improved Knock-in Efficiencies and Reduced off-Target Effects. *Nucleic Acids Res.* **2021**, 49 (6), e36–e36. <https://doi.org/10.1093/nar/gkaa1264>.
- (96) Najm, F. J.; Strand, C.; Donovan, K. F.; Hegde, M.; Sanson, K. R.; Vaimberg, E. W.; Sullender, M. E.; Hartenian, E.; Kalani, Z.; Fusi, N.; Listgarten, J.; Younger, S. T.; Bernstein, B. E.; Root, D. E.; Doench, J. G. Orthologous CRISPR–Cas9 Enzymes for Combinatorial Genetic Screens. *Nat. Biotechnol.* **2018**, 36 (2), 179–189. <https://doi.org/10.1038/nbt.4048>.
- (97) McCarty, N. S.; Graham, A. E.; Studená, L.; Ledesma-Amaro, R. Multiplexed CRISPR Technologies for Gene Editing and Transcriptional Regulation. *Nat. Commun.* **2020**, 11 (1), 1281. <https://doi.org/10.1038/s41467-020-15053-x>.
- (98) Vasquez, C. A.; Cowan, Q. T.; Komor, A. C. Base Editing in Human Cells to Produce Single-Nucleotide-Variant Clonal Cell Lines. *Curr. Protoc. Mol. Biol.* **2020**, 133 (1). <https://doi.org/10.1002/cpmb.129>.
- (99) Salomonsen, B.; Mortensen, U. H.; Halkier, B. A. USER-Derived Cloning Methods and Their Primer Design. In *DNA Cloning and Assembly Methods*; Valla, S., Lale, R., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2014; Vol. 1116, pp 59–72. [https://doi.org/10.1007/978-1-62703-764-8\\_5](https://doi.org/10.1007/978-1-62703-764-8_5).
- (100) Adikusuma, F.; Pfitzner, C.; Thomas, P. Q. Versatile Single-Step-Assembly CRISPR/Cas9 Vectors for Dual gRNA Expression. *PLOS ONE* **2017**, 12 (12), e0187236. <https://doi.org/10.1371/journal.pone.0187236>.
- (101) 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. *Nat. Commun.* **2017**, 8 (1), 15790. <https://doi.org/10.1038/ncomms15790>.
- (102) Rallapalli, K. L.; Ranzau, B. L.; Ganapathy, K. R.; Paesani, F.; Komor, A. C. Combined Theoretical, Bioinformatic, and Biochemical Analyses of RNA Editing by Adenine Base Editors. *CRISPR J.* **2022**, 5 (2), 294–310. <https://doi.org/10.1089/crispr.2021.0131>.
- (103) Clement, K.; Rees, H.; Canver, M. C.; Gehrke, J. M.; Farouni, R.; Hsu, J. Y.; Cole, M. A.; Liu, D. R.; Joung, J. K.; Bauer, D. E.; Pinello, L. CRISPResso2 Provides Accurate and Rapid Genome Editing Sequence Analysis. *Nat. Biotechnol.* **2019**, 37 (3), 224–226. <https://doi.org/10.1038/s41587-019-0032-3>.
- (104) Liao, Y.; Smyth, G. K.; Shi, W. The R Package Rsubread Is Easier, Faster, Cheaper and Better for Alignment and Quantification of RNA Sequencing Reads. *Nucleic Acids Res.* **2019**, 47 (8), e47–e47. <https://doi.org/10.1093/nar/gkz114>.

- (105) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, 25 (16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- (106) Cowan, Q. T.; Gu, S.; Gu, W.; Ranzau, B. L.; Simonson, T. S.; Komor, A. C. Development of Multiplexed Orthogonal Base Editor (MOBE) Systems. *Nat. Biotechnol.* **2024**. <https://doi.org/10.1038/s41587-024-02240-0>.
- (107) Rees, H. A.; Liu, D. R. Base Editing: Precision Chemistry on the Genome and Transcriptome of Living Cells. *Nat. Rev. Genet.* **2018**, 19 (12), 770–788. <https://doi.org/10.1038/s41576-018-0059-1>.
- (108) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, 337 (6096), 816–821. <https://doi.org/10.1126/science.1225829>.
- (109) Jiang, F.; Taylor, D. W.; Chen, J. S.; Kornfeld, J. E.; Zhou, K.; Thompson, A. J.; Nogales, E.; Doudna, J. A. Structures of a CRISPR-Cas9 R-Loop Complex Primed for DNA Cleavage. *Science* **2016**, 351 (6275), 867–871. <https://doi.org/10.1126/science.aad8282>.
- (110) Ran, F. A.; Hsu, P. D.; Wright, J.; Agarwala, V.; Scott, D. A.; Zhang, F. Genome Engineering Using the CRISPR-Cas9 System. *Nat. Protoc.* **2013**, 8 (11), 2281–2308. <https://doi.org/10.1038/nprot.2013.143>.
- (111) Hess, G. T.; Tycko, J.; Yao, D.; Bassik, M. C. Methods and Applications of CRISPR-Mediated Base Editing in Eukaryotic Genomes. *Mol. Cell* **2017**, 68 (1), 26–43. <https://doi.org/10.1016/j.molcel.2017.09.029>.
- (112) Evanoff, M.; Komor, A. C. Base Editors: Modular Tools for the Introduction of Point Mutations in Living Cells. *Emerg. Top. Life Sci.* **2019**, 3 (5), 483–491. <https://doi.org/10.1042/ETLS20190088>.
- (113) Molla, K. A.; Yang, Y. CRISPR/Cas-Mediated Base Editing: Technical Considerations and Practical Applications. *Trends Biotechnol.* **2019**, 37 (10), 1121–1142. <https://doi.org/10.1016/j.tibtech.2019.03.008>.
- (114) Lappalainen, T.; Scott, A. J.; Brandt, M.; Hall, I. M. Genomic Analysis in the Age of Human Genome Sequencing. *Cell* **2019**, 177 (1), 70–84. <https://doi.org/10.1016/j.cell.2019.02.032>.
- (115) Ahmed, Z.; Zeeshan, S.; Mendhe, D.; Dong, X. Human Gene and Disease Associations for Clinical-Genomics and Precision Medicine Research. *Clin. Transl. Med.* **2020**, 10 (1), 297–318. <https://doi.org/10.1002/ctm2.28>.
- (116) Malone, E. R.; Oliva, M.; Sabatini, P. J. B.; Stockley, T. L.; Siu, L. L. Molecular Profiling for Precision Cancer Therapies. *Genome Med.* **2020**, 12. <https://doi.org/10.1186/s13073-019-0703-1>.

- (117) Sherry, S. T.; Ward, M.-H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E. M.; Sirotkin, K. dbSNP: The NCBI Database of Genetic Variation. *Nucleic Acids Res.* **2001**, *29* (1), 308–311.
- (118) Landrum, M. J.; Lee, J. M.; Benson, M.; Brown, G. R.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; Karapetyan, K.; Katz, K.; Liu, C.; Maddipatla, Z.; Malheiro, A.; McDaniel, K.; Ovetsky, M.; Riley, G.; Zhou, G.; Holmes, J. B.; Kattman, B. L.; Maglott, D. R. ClinVar: Improving Access to Variant Interpretations and Supporting Evidence. *Nucleic Acids Res.* **2018**, *46* (D1), D1062–D1067. <https://doi.org/10.1093/nar/gkx1153>.
- (119) 1000 Genomes Project Consortium; Auton, A.; Brooks, L. D.; Durbin, R. M.; Garrison, E. P.; Kang, H. M.; Korbel, J. O.; Marchini, J. L.; McCarthy, S.; McVean, G. A.; Abecasis, G. R. A Global Reference for Human Genetic Variation. *Nature* **2015**, *526* (7571), 68–74. <https://doi.org/10.1038/nature15393>.
- (120) Fox, K.; Rallapalli, K. L.; Komor, A. C. Rewriting Human History and Empowering Indigenous Communities with Genome Editing Tools. *Genes* **2020**, *11* (1), 88. <https://doi.org/10.3390/genes11010088>.
- (121) Popejoy, A. B.; Fullerton, S. M. Genomics Is Failing on Diversity. *Nat. News* **2016**, *538* (7624), 161. <https://doi.org/10.1038/538161a>.
- (122) Coggins, N. B.; Stultz, J.; O’Geen, H.; Carvajal-Carmona, L. G.; Segal, D. J. Methods for Scarless, Selection-Free Generation of Human Cells and Allele-Specific Functional Analysis of Disease-Associated SNPs and Variants of Uncertain Significance. *Sci. Rep.* **2017**, *7* (1), 15044. <https://doi.org/10.1038/s41598-017-15407-4>.
- (123) Kwart, D.; Paquet, D.; Teo, S.; Tessier-Lavigne, M. Precise and Efficient Scarless Genome Editing in Stem Cells Using CORRECT. *Nat. Protoc.* **2017**, *12* (2), 329–354. <https://doi.org/10.1038/nprot.2016.171>.
- (124) Ousterout, D. G.; Kabadi, A. M.; Thakore, P. I.; Majoros, W. H.; Reddy, T. E.; Gersbach, C. A. Multiplex CRISPR/Cas9-Based Genome Editing for Correction of Dystrophin Mutations That Cause Duchenne Muscular Dystrophy. *Nat. Commun.* **2015**, *6*, 6244. <https://doi.org/10.1038/ncomms7244>.
- (125) Paquet, D.; Kwart, D.; Chen, A.; Sproul, A.; Jacob, S.; Teo, S.; Olsen, K. M.; Gregg, A.; Noggle, S.; Tessier-Lavigne, M. Efficient Introduction of Specific Homozygous and Heterozygous Mutations Using CRISPR/Cas9. *Nature* **2016**, *533* (7601), 125–129. <https://doi.org/10.1038/nature17664>.
- (126) Riesenberger, S.; Chintalapati, M.; Macak, D.; Kanis, P.; Maricic, T.; Pääbo, S. Simultaneous Precise Editing of Multiple Genes in Human Cells. *Nucleic Acids Res.* **2019**, *47* (19), e116. <https://doi.org/10.1093/nar/gkz669>.
- (127) Yeh, C. D.; Richardson, C. D.; Corn, J. E. Advances in Genome Editing through Control of DNA Repair Pathways. *Nat. Cell Biol.* **2019**, *21* (12), 1468–1478. <https://doi.org/10.1038/s41556-019-0425-z>.

- (128) Arbab, M.; Shen, M. W.; Mok, B.; Wilson, C.; Matuszek, Ż.; Cassa, C. A.; Liu, D. R. Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. *Cell* **2020**, *182* (2), 463–480.e30. <https://doi.org/10.1016/j.cell.2020.05.037>.
- (129) Hu, Z.; Wang, S.; Zhang, C.; Gao, N.; Li, M.; Wang, D.; Wang, D.; Liu, D.; Liu, H.; Ong, S.-G.; Wang, H.; Wang, Y. A Compact Cas9 Ortholog from *Staphylococcus Auricularis* (SauriCas9) Expands the DNA Targeting Scope. *PLoS Biol.* **2020**, *18* (3), e3000686. <https://doi.org/10.1371/journal.pbio.3000686>.
- (130) Li, X.; Wang, Y.; Liu, Y.; Yang, B.; Wang, X.; Wei, J.; Lu, Z.; Zhang, Y.; Wu, J.; Huang, X.; Yang, L.; Chen, J. Base Editing with a Cpf1-Cytidine Deaminase Fusion. *Nat. Biotechnol.* **2018**, *36* (4), 324–327. <https://doi.org/10.1038/nbt.4102>.
- (131) Liu, Z.; Shan, H.; Chen, S.; Chen, M.; Song, Y.; Lai, L.; Li, Z. Efficient Base Editing with Expanded Targeting Scope Using an Engineered Spy-Mac Cas9 Variant. *Cell Discov.* **2019**, *5*, 58. <https://doi.org/10.1038/s41421-019-0128-4>.
- (132) Zafra, M. P.; Schatoff, E. M.; Katti, A.; Foronda, M.; Breinig, M.; Schweitzer, A. Y.; Simon, A.; Han, T.; Goswami, S.; Montgomery, E.; Thibado, J.; Kasthuber, E. R.; Sánchez-Rivera, F. J.; Shi, J.; Vakoc, C. R.; Lowe, S. W.; Tschaharganeh, D. F.; Dow, L. E. Optimized Base Editors Enable Efficient Editing in Cells, Organoids and Mice. *Nat. Biotechnol.* **2018**, *36* (9), 888–893. <https://doi.org/10.1038/nbt.4194>.
- (133) Elbing, K. L.; Brent, R. Recipes and Tools for Culture of *Escherichia Coli*. *Curr. Protoc. Mol. Biol.* **2019**, *125* (1), e83. <https://doi.org/10.1002/cpmb.83>.
- (134) *Guidelines for base editing in mammalian cells* · Benchling. <https://benchling.com/pub/liu-base-editor> (accessed 2020-06-19).
- (135) Hwang, G.-H.; Park, J.; Lim, K.; Kim, S.; Yu, J.; Yu, E.; Kim, S.-T.; Eils, R.; Kim, J.-S.; Bae, S. Web-Based Design and Analysis Tools for CRISPR Base Editing. *BMC Bioinformatics* **2018**, *19* (1), 542. <https://doi.org/10.1186/s12859-018-2585-4>.
- (136) Seidman, C. E.; Struhl, K.; Sheen, J.; Jessen, T. Introduction of Plasmid DNA into Cells. *Curr. Protoc. Mol. Biol.* **2001**, *37* (1), mb0108s37. <https://doi.org/10.1002/0471142727.mb0108s37>.
- (137) Phelan, M. C. Techniques for Mammalian Cell Tissue Culture. *Curr. Protoc. Protein Sci.* **2006**, *Appendix 3*, Appendix 3C. <https://doi.org/10.1002/0471140864.psa03cs46>.
- (138) Kluesner, M. G.; Nedveck, D. A.; Lahr, W. S.; Garbe, J. R.; Abrahante, J. E.; Webber, B. R.; Moriarity, B. S. EditR: A Method to Quantify Base Editing from Sanger Sequencing. *CRISPR J.* **2018**, *1* (3), 239–250. <https://doi.org/10.1089/crispr.2018.0014>.
- (139) Veeranagouda, Y.; Debono-Lagneaux, D.; Fournet, H.; Thill, G.; Didier, M. CRISPR-Cas9-Edited Site Sequencing (CRES-Seq): An Efficient and High-Throughput Method for the Selection of CRISPR-Cas9-Edited Clones. *Curr. Protoc. Mol. Biol.* **2018**, *121* (1), 31.14.1–31.14.11. <https://doi.org/10.1002/cpmb.53>.

- (140) Yang, L.; Yang, J. L.; Byrne, S.; Pan, J.; Church, G. M. CRISPR/Cas9-Directed Genome Editing of Cultured Cells. *Curr. Protoc. Mol. Biol.* **2014**, 107 (1), 31.1.1-31.1.17. <https://doi.org/10.1002/0471142727.mb3101s107>.
- (141) Giuliano, C. J.; Lin, A.; Girish, V.; Sheltzer, J. M. Generating Single Cell-Derived Knockout Clones in Mammalian Cells with CRISPR/Cas9. *Curr. Protoc. Mol. Biol.* **2019**, 128 (1), e100. <https://doi.org/10.1002/cpmb.100>.
- (142) Stacey, G. N.; Masters, J. R. Cryopreservation and Banking of Mammalian Cell Lines. *Nat. Protoc.* **2008**, 3 (12), 1981–1989. <https://doi.org/10.1038/nprot.2008.190>.
- (143) Yokoyama, W. M.; Thompson, M. L.; Ehrhardt, R. O. Cryopreservation and Thawing of Cells. *Curr. Protoc. Immunol.* **2012**, 99 (1), A.3G.1-A.3G.5. <https://doi.org/10.1002/0471142735.ima03gs99>.
- (144) Hsu, P. D.; Lander, E. S.; Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **2014**, 157 (6), 1262–1278. <https://doi.org/10.1016/j.cell.2014.05.010>.
- (145) Liu, M.; Rehman, S.; Tang, X.; Gu, K.; Fan, Q.; Chen, D.; Ma, W. Methodologies for Improving HDR Efficiency. *Front. Genet.* **2019**, 9. <https://doi.org/10.3389/fgene.2018.00691>.
- (146) Shin, H. Y.; Wang, C.; Lee, H. K.; Yoo, K. H.; Zeng, X.; Kuhns, T.; Yang, C. M.; Mohr, T.; Liu, C.; Hennighausen, L. CRISPR/Cas9 Targeting Events Cause Complex Deletions and Insertions at 17 Sites in the Mouse Genome. *Nat. Commun.* **2017**, 8, 15464. <https://doi.org/10.1038/ncomms15464>.
- (147) Webber, B. R.; Lonetree, C.-L.; Kluesner, M. G.; Johnson, M. J.; Pomeroy, E. J.; Diers, M. D.; Lahr, W. S.; Draper, G. M.; Slipek, N. J.; Smeester, B. A.; Lovendahl, K. N.; McElroy, A. N.; Gordon, W. R.; Osborn, M. J.; Moriarity, B. S. Highly Efficient Multiplex Human T Cell Engineering without Double-Strand Breaks Using Cas9 Base Editors. *Nat. Commun.* **2019**, 10 (1), 5222. <https://doi.org/10.1038/s41467-019-13007-6>.
- (148) Weisheit, I.; Kroeger, J. A.; Malik, R.; Klimmt, J.; Crusius, D.; Dannert, A.; Dichgans, M.; Paquet, D. Detection of Deleterious On-Target Effects after HDR-Mediated CRISPR Editing. *Cell Rep.* **2020**, 31 (8), 107689. <https://doi.org/10.1016/j.celrep.2020.107689>.
- (149) Martin, R. M.; Ikeda, K.; Cromer, M. K.; Uchida, N.; Nishimura, T.; Romano, R.; Tong, A. J.; Lemgart, V. T.; Camarena, J.; Pavel-Dinu, M.; Sindhu, C.; Wiebking, V.; Vaidyanathan, S.; Dever, D. P.; Bak, R. O.; Laustsen, A.; Lesch, B. J.; Jakobsen, M. R.; Sebastiano, V.; Nakauchi, H.; Porteus, M. H. Highly Efficient and Marker-Free Genome Editing of Human Pluripotent Stem Cells by CRISPR-Cas9 RNP and AAV6 Donor-Mediated Homologous Recombination. *Cell Stem Cell* **2019**, 24 (5), 821-828.e5. <https://doi.org/10.1016/j.stem.2019.04.001>.
- (150) Norris, A. D.; Kim, H.-M.; Colaiácovo, M. P.; Calarco, J. A. Efficient Genome Editing in *Caenorhabditis Elegans* with a Toolkit of Dual-Marker Selection Cassettes. *Genetics* **2015**, 201 (2), 449–458. <https://doi.org/10.1534/genetics.115.180679>.

- (151) Ochiai, H.; Miyamoto, T.; Kanai, A.; Hosoba, K.; Sakuma, T.; Kudo, Y.; Asami, K.; Ogawa, A.; Watanabe, A.; Kajii, T.; Yamamoto, T.; Matsuura, S. TALEN-Mediated Single-Base-Pair Editing Identification of an Intergenic Mutation Upstream of BUB1B as Causative of PCS (MVA) Syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (4), 1461–1466. <https://doi.org/10.1073/pnas.1317008111>.
- (152) Sur, I. K.; Hallikas, O.; Vähärautio, A.; Yan, J.; Turunen, M.; Enge, M.; Taipale, M.; Karhu, A.; Aaltonen, L. A.; Taipale, J. Mice Lacking a Myc Enhancer That Includes Human SNP Rs6983267 Are Resistant to Intestinal Tumors. *Science* **2012**, *338* (6112), 1360–1363. <https://doi.org/10.1126/science.1228606>.
- (153) H, G.; Yr, Z.; K, R. *Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting*. *Cell*. [https://doi.org/10.1016/0092-8674\(93\)90644-6](https://doi.org/10.1016/0092-8674(93)90644-6).
- (154) Watkins, J. A.; Irshad, S.; Grigoriadis, A.; Tutt, A. N. Genomic Scars as Biomarkers of Homologous Recombination Deficiency and Drug Response in Breast and Ovarian Cancers. *Breast Cancer Res. BCR* **2014**, *16* (3), 211. <https://doi.org/10.1186/bcr3670>.
- (155) Di Nicolantonio, F.; Arena, S.; Gallicchio, M.; Zecchin, D.; Martini, M.; Flonta, S. E.; Stella, G. M.; Lamba, S.; Cancelliere, C.; Russo, M.; Geuna, M.; Appendino, G.; Fantozzi, R.; Medico, E.; Bardelli, A. Replacement of Normal with Mutant Alleles in the Genome of Normal Human Cells Unveils Mutation-Specific Drug Responses. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (52), 20864–20869. <https://doi.org/10.1073/pnas.0808757105>.
- (156) S, S.; R, P.; F, B.; C, R.; La, D.; Kw, K.; B, V.; N, P. *A panel of isogenic human cancer cells suggests a therapeutic approach for cancers with inactivated p53*. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.0813333106>.
- (157) Gray, S. J.; Blake, B. L.; Criswell, H. E.; Nicolson, S. C.; Samulski, R. J.; McCown, T. J. Directed Evolution of a Novel Adeno-Associated Virus (AAV) Vector That Crosses the Seizure-Compromised Blood–Brain Barrier (BBB). *Mol. Ther.* **2010**, *18* (3), 570–578. <https://doi.org/10.1038/mt.2009.292>.
- (158) Kohlbrenner, E.; Aslanidi, G.; Nash, K.; Shklyayev, S.; Campbell-Thompson, M.; Byrne, B. J.; Snyder, R. O.; Muzyczka, N.; Warrington, K. H.; Zolotukhin, S. Successful Production of Pseudotyped rAAV Vectors Using a Modified Baculovirus Expression System. *Mol. Ther.* **2005**, *12* (6), 1217–1225. <https://doi.org/10.1016/j.ymthe.2005.08.018>.
- (159) Gaudelli, N. M.; Lam, D. K.; Rees, H. A.; Solá-Esteves, N. M.; Barrera, L. A.; Born, D. A.; Edwards, A.; Gehrke, J. M.; Lee, S.-J.; Liquori, A. J.; Murray, R.; Packer, M. S.; Rinaldi, C.; Slaymaker, I. M.; Yen, J.; Young, L. E.; Ciaramella, G. Directed Evolution of Adenine Base Editors with Increased Activity and Therapeutic Application. *Nat. Biotechnol.* **2020**. <https://doi.org/10.1038/s41587-020-0491-6>.
- (160) Kim, Y. B.; Komor, A. C.; Levy, J. M.; Packer, M. S.; Zhao, K. T.; Liu, D. R. Increasing the Genome-Targeting Scope and Precision of Base Editing with Engineered Cas9-Cytidine Deaminase Fusions. *Nat. Biotechnol.* **2017**, *35* (4), 371–376. <https://doi.org/10.1038/nbt.3803>.



- (161) Richter, M. F.; Zhao, K. T.; Eton, E.; Lapinaite, A.; Newby, G. A.; Thuronyi, B. W.; Wilson, C.; Koblan, L. W.; Zeng, J.; Bauer, D. E.; Doudna, J. A.; Liu, D. R. Phage-Assisted Evolution of an Adenine Base Editor with Improved Cas Domain Compatibility and Activity. *Nat. Biotechnol.* **2020**. <https://doi.org/10.1038/s41587-020-0453-z>.
- (162) Walton, R. T.; Christie, K. A.; Whittaker, M. N.; Kleinstiver, B. P. Unconstrained Genome Targeting with Near-PAMless Engineered CRISPR-Cas9 Variants. *Science* **2020**, 368 (6488), 290–296. <https://doi.org/10.1126/science.aba8853>.
- (163) Anzalone, A. V.; Randolph, P. B.; Davis, J. R.; Sousa, A. A.; Koblan, L. W.; Levy, J. M.; Chen, P. J.; Wilson, C.; Newby, G. A.; Raguram, A.; Liu, D. R. Search-and-Replace Genome Editing without Double-Strand Breaks or Donor DNA. *Nature* **2019**, 576 (7785), 149–157. <https://doi.org/10.1038/s41586-019-1711-4>.
- (164) Tycko, J.; Myer, V. E.; Hsu, P. D. Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Mol. Cell* **2016**, 63 (3), 355–370. <https://doi.org/10.1016/j.molcel.2016.07.004>.
- (165) Wang, D.; Zhang, C.; Wang, B.; Li, B.; Wang, Q.; Liu, D.; Wang, H.; Zhou, Y.; Shi, L.; Lan, F.; Wang, Y. Optimized CRISPR Guide RNA Design for Two High-Fidelity Cas9 Variants by Deep Learning. *Nat. Commun.* **2019**, 10 (1), 4284. <https://doi.org/10.1038/s41467-019-12281-8>.
- (166) Dang, Y.; Jia, G.; Choi, J.; Ma, H.; Anaya, E.; Ye, C.; Shankar, P.; Wu, H. Optimizing sgRNA Structure to Improve CRISPR-Cas9 Knockout Efficiency. *Genome Biol.* **2015**, 16 (1), 280. <https://doi.org/10.1186/s13059-015-0846-3>.
- (167) Liu, Z.; Chen, S.; Shan, H.; Jia, Y.; Chen, M.; Song, Y.; Lai, L.; Li, Z. Efficient Base Editing with High Precision in Rabbits Using YFE-BE4max. *Cell Death Dis.* **2020**, 11 (1), 1–11. <https://doi.org/10.1038/s41419-020-2244-3>.
- (168) Gehrke, J. M.; Cervantes, O.; Clement, M. K.; Wu, Y.; Zeng, J.; Bauer, D. E.; Pinello, L.; Joung, J. K. An APOBEC3A-Cas9 Base Editor with Minimized Bystander and off-Target Activities. *Nat. Biotechnol.* **2018**, 36 (10), 977–982. <https://doi.org/10.1038/nbt.4199>.
- (169) Tan, J.; Zhang, F.; Karcher, D.; Bock, R. Engineering of High-Precision Base Editors for Site-Specific Single Nucleotide Replacement. *Nat. Commun.* **2019**, 10 (1), 439. <https://doi.org/10.1038/s41467-018-08034-8>.
- (170) Smits, A. H.; Ziebell, F.; Joberty, G.; Zinn, N.; Mueller, W. F.; Clauder-Münster, S.; Eberhard, D.; Fälth Savitski, M.; Grandi, P.; Jakob, P.; Michon, A.-M.; Sun, H.; Tessmer, K.; Bürckstümmer, T.; Bantscheff, M.; Steinmetz, L. M.; Drewes, G.; Huber, W. Biological Plasticity Rescues Target Activity in CRISPR Knock Outs. *Nat. Methods* **2019**, 16 (11), 1087–1093. <https://doi.org/10.1038/s41592-019-0614-5>.
- (171) Grünewald, J.; Zhou, R.; Garcia, S. P.; Iyer, S.; Lareau, C. A.; Aryee, M. J.; Joung, J. K. Transcriptome-Wide off-Target RNA Editing Induced by CRISPR-Guided DNA Base Editors. *Nature* **2019**, 569 (7756), 433–437. <https://doi.org/10.1038/s41586-019-1161-z>.
- (172) Jin, S.; Zong, Y.; Gao, Q.; Zhu, Z.; Wang, Y.; Qin, P.; Liang, C.; Wang, D.; Qiu, J.-L.; Zhang, F.; Gao, C. Cytosine, but Not Adenine, Base Editors Induce Genome-Wide off-

- Target Mutations in Rice. *Science* **2019**, 364 (6437), 292–295. <https://doi.org/10.1126/science.aaw7166>.
- (173) Zhou, C.; Sun, Y.; Yan, R.; Liu, Y.; Zuo, E.; Gu, C.; Han, L.; Wei, Y.; Hu, X.; Zeng, R.; Li, Y.; Zhou, H.; Guo, F.; Yang, H. Off-Target RNA Mutation Induced by DNA Base Editing and Its Elimination by Mutagenesis. *Nature* **2019**, 571 (7764), 275–278. <https://doi.org/10.1038/s41586-019-1314-0>.
- (174) Zuo, E.; Sun, Y.; Wei, W.; Yuan, T.; Ying, W.; Sun, H.; Yuan, L.; Steinmetz, L. M.; Li, Y.; Yang, H. Cytosine Base Editor Generates Substantial Off-Target Single-Nucleotide Variants in Mouse Embryos. *Science* **2019**, 364 (6437), 289–292. <https://doi.org/10.1126/science.aav9973>.
- (175) Lee, H. K.; Smith, H. E.; Liu, C.; Willi, M.; Hennighausen, L. Cytosine Base Editor 4 but Not Adenine Base Editor Generates Off-Target Mutations in Mouse Embryos. *Commun. Biol.* **2020**, 3. <https://doi.org/10.1038/s42003-019-0745-3>.
- (176) Grünewald, J.; Zhou, R.; Iyer, S.; Lareau, C. A.; Garcia, S. P.; Aryee, M. J.; Joung, J. K. CRISPR DNA Base Editors with Reduced RNA Off-Target and Self-Editing Activities. *Nat. Biotechnol.* **2019**, 37 (9), 1041–1048. <https://doi.org/10.1038/s41587-019-0236-6>.
- (177) Zuo, E.; Sun, Y.; Yuan, T.; He, B.; Zhou, C.; Ying, W.; Liu, J.; Wei, W.; Zeng, R.; Li, Y.; Yang, H. A Rationally Engineered Cytosine Base Editor Retains High On-Target Activity While Reducing Both DNA and RNA off-Target Effects. *Nat. Methods* **2020**, 17 (6), 600–604. <https://doi.org/10.1038/s41592-020-0832-x>.
- (178) Grav, L. M.; Sergeeva, D.; Lee, J. S.; Marin de Mas, I.; Lewis, N. E.; Andersen, M. R.; Nielsen, L. K.; Lee, G. M.; Kildegaard, H. F. Minimizing Clonal Variation during Mammalian Cell Line Engineering for Improved Systems Biology Data Generation. *ACS Synth. Biol.* **2018**, 7 (9), 2148–2159. <https://doi.org/10.1021/acssynbio.8b00140>.
- (179) Yu, Y.; Leete, T. C.; Born, D. A.; Young, L.; Barrera, L. A.; Lee, S.-J.; Rees, H. A.; Ciaramella, G.; Gaudelli, N. M. Cytosine Base Editors with Minimized Unguided DNA and RNA Off-Target Events and High on-Target Activity. *Nat. Commun.* **2020**, 11 (1), 2052. <https://doi.org/10.1038/s41467-020-15887-5>.
- (180) Lin, Y.-C.; Boone, M.; Meuris, L.; Lemmens, I.; Van Roy, N.; Soete, A.; Reumers, J.; Moisse, M.; Plaisance, S.; Drmanac, R.; Chen, J.; Speleman, F.; Lambrechts, D.; Van de Peer, Y.; Tavernier, J.; Callewaert, N. Genome Dynamics of the Human Embryonic Kidney 293 Lineage in Response to Cell Biology Manipulations. *Nat. Commun.* **2014**, 5. <https://doi.org/10.1038/ncomms5767>.
- (181) Brookhouser, N.; Tekel, S. J.; Standage-Beier, K.; Nguyen, T.; Schwarz, G.; Wang, X.; Brafman, D. A. BIG-TREE: Base-Edited Isogenic hPSC Line Generation Using a Transient Reporter for Editing Enrichment. *Stem Cell Rep.* **2020**, 14 (2), 184–191. <https://doi.org/10.1016/j.stemcr.2019.12.013>.
- (182) Coelho, M. A.; Li, S.; Pane, L. S.; Firth, M.; Ciotta, G.; Wrigley, J. D.; Cuomo, M. E.; Maresca, M.; Taylor, B. J. M. BE-FLARE: A Fluorescent Reporter of Base Editing Activity Reveals Editing Characteristics of APOBEC3A and APOBEC3B. *BMC Biol.* **2018**, 16. <https://doi.org/10.1186/s12915-018-0617-1>.

- (183) Katti, A.; Foronda, M.; Zimmerman, J.; Diaz, B.; Zafra, M. P.; Goswami, S.; Dow, L. E. GO: A Functional Reporter System to Identify and Enrich Base Editing Activity. *Nucleic Acids Res.* **2020**, *48* (6), 2841–2852. <https://doi.org/10.1093/nar/gkaa124>.
- (184) St. Martin, A.; Salamango, D.; Serebrenik, A.; Shaban, N.; Brown, W. L.; Donati, F.; Munagala, U.; Conticello, S. G.; Harris, R. S. A Fluorescent Reporter for Quantification and Enrichment of DNA Editing by APOBEC–Cas9 or Cleavage by Cas9 in Living Cells. *Nucleic Acids Res.* **2018**, *46* (14), e84. <https://doi.org/10.1093/nar/gky332>.
- (185) Ranzau, B. L.; Rallapalli, K. L.; Evanoff, M.; Paesani, F.; Komor, A. C. The Wild-Type tRNA Adenosine Deaminase Enzyme TadA Is Capable of Sequence-Specific DNA Base Editing. *ChemBioChem* **2023**, *24* (16), e202200788. <https://doi.org/10.1002/cbic.202200788>.
- (186) Jost, M.; Santos, D. A.; Saunders, R. A.; Horlbeck, M. A.; Hawkins, J. S.; Scaria, S. M.; Norman, T. M.; Hussmann, J. A.; Liem, C. R.; Gross, C. A.; Weissman, J. S. Titrating Gene Expression Using Libraries of Systematically Attenuated CRISPR Guide RNAs. *Nat. Biotechnol.* **2020**, *38* (3), 355–364. <https://doi.org/10.1038/s41587-019-0387-5>.
- (187) Kim, S.-I.; Matsumoto, T.; Kagawa, H.; Nakamura, M.; Hirohata, R.; Ueno, A.; Ohishi, M.; Sakuma, T.; Soga, T.; Yamamoto, T.; Woltjen, K. Microhomology-Assisted Scarless Genome Editing in Human iPSCs. *Nat. Commun.* **2018**, *9*. <https://doi.org/10.1038/s41467-018-03044-y>.